

Washington University in St. Louis
Washington University Open Scholarship

All Theses and Dissertations (ETDs)

5-24-2009

Genetic and Biochemical Properties of Arabidopsis RNA Polymerases IV and V

Jeremy Haag

Washington University in St. Louis

Follow this and additional works at: <https://openscholarship.wustl.edu/etd>

Recommended Citation

Haag, Jeremy, "Genetic and Biochemical Properties of Arabidopsis RNA Polymerases IV and V" (2009). *All Theses and Dissertations (ETDs)*. 886.

<https://openscholarship.wustl.edu/etd/886>

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Plant Biology

Dissertation Examination Committee:

Craig S. Pikaard, Ph.D., Chair

Peter Burgers, Ph.D.

Douglas L. Chalker, Ph.D.

Sarah R. Elgin, Ph.D.

Joseph Jez, Ph.D.

John Majors, Ph.D.

GENETIC AND BIOCHEMICAL PROPERTIES OF *ARABIDOPSIS* RNA

POLYMERASES IV AND V

by

Jeremy Richard Haag

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2009

Saint Louis, Missouri

Copyright by
Jeremy Richard Haag
2009

ABSTRACT OF THE DISSERTATION

Genetic and Biochemical Properties of *Arabidopsis* RNA Polymerases IV and V

by

Jeremy Richard Haag

Doctor of Philosophy in Plant Biology

Washington University in St. Louis, 2009

Professor Craig S. Pikaard, Ph.D., Chairman

RNA Polymerases IV and V (Pol IV and Pol V) are plant-specific enzyme complexes with subunit homology to RNA Polymerase II (Pol II). The largest subunits in Pol IV and Pol V, NRPD1 and NRPE1 respectively, share a second largest subunit, NRPD2/NRPE2. The evolutionarily conserved Metal A and Metal B binding sites are required for Pol IV and V *in vivo* function fitting the prediction that these are functional polymerases. The Defective Chloroplast and Leaves-like (DeCL) domain at the C-terminus of both NRPD1 and NRPE1 is also required for complementation but other domains in the NRPE1 CTD are largely dispensable. Biochemical analysis reveals Pol IV to be a DNA-dependent RNA Polymerase capable of producing RNA from a tripartite template that mimics an open transcription bubble. The Metal A binding site is required for Pol IV *in vitro* transcription while the enzyme is resistant to alpha-amanitin, a potent Pol II inhibitor. Pol IV has also been found to physically interact with RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) *in vivo* providing an explanation for how Pol IV RNA products are channeled specifically to RDR2 for the

production of double-stranded RNA and eventual dicing. Biochemical analysis has also revealed that RDR2 is capable of transcribing both single-stranded RNA and DNA *in vitro*, consistent with previously analyzed RNA-dependent RNA polymerases from plants and other organisms.

ACKNOWLEDGEMENTS

It has been a long road to reach this point in my graduate career and I could not have done it without those who traveled by my side and those I met along the way. First and foremost, I must thank my wife, April, for her unwavering love, support, understanding and encouragement. Without her I don't know how I would have reached this day. Next, I must thank my family and friends for helping maintain my sanity and reminding me there is a life outside the lab. My parents, James and Annette, and brother, Christopher, might not have always understood what exactly I was doing but recognized the sense of fulfillment it provided and patiently listened about setbacks and triumphs alike.

My lab mates became a second family of sorts and I thank them for the many valued discussions, collaborations, lunches at Fitz's and canoe trips shared. Keith Earley was an early lab mentor and set an example that I still strive to meet. Tom Ream and I were partners in crime when it came to tackling the genetics and biochemistry of Pol IV and Pol V. Our six years of close teamwork provided both professional enrichment and personal enjoyment. Together with Olga Pontes and Pedro Costa-Nunes, the four of us formed a team that will be difficult to replicate.

I cannot thank Craig Pikaard enough for taking me into his lab and setting the example for what a mentor should be. Craig never lost faith in my techniques or me and always offered an encouraging word. His support in and out of the lab has been greatly appreciated. I must also thank my thesis committee (including Eric Richards) for their thoughtful discussions and advice.

Lastly, I may never have pursued graduate school had it not been for Craig Coates, my undergraduate research advisor at Texas A&M and Rodolfo Aramayo who hired me as a research assistant following graduation. They introduced me to the biological sciences and encouraged me to pursue a higher degree. The time spent in their labs set a solid foundation and prepared me well for graduate school.

This thesis is dedicated to my grandfather, Earl Haag, Jr. He worked as a chemist and has not let retirement stop him from seeking to understand the world around him. His inquisitive nature is an inspiration.

TABLE OF CONTENTS

Abstract	
Acknowledgments	
Chapter 1: Introduction	1
i. RNA Polymerase II: A model for RNA Polymerase IV and V	2
a. Subunit nomenclature, composition and structure	2
i. Catalytic subunits	4
ii. Assembly subunits	5
iii. Auxiliary subunits	6
b. Sequence conservation and divergence among the Pol II, IV and V largest and second-largest subunits	9
i. Catalytic core	9
ii. C-terminal domain features	14
c. Regulation via the Pol II largest subunit C-terminal domain	16
d. DNA-dependent RNA polymerase activity	20
ii. Roles of RNA polymerases IV and V in gene silencing	23
a. RNA-directed DNA methylation	24
i. siRNA biogenesis	25
ii. AGO-RISC assembly	27
iii. Pol V transcription, AGO4-RISC and DNA methylation	29
b. Paramutation	32
c. Flowering and development	33
d. Abiotic and biotic stress-inducible responses	35
e. Short- and Long-distance spread of silencing	38
iii. Scope of Thesis	40
iv. References cited	44
Chapter 2: Gateway-compatible vectors for plant functional genomics and proteomics	56
Chapter 3: Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation	72
Chapter 4: Metal A and Metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing.	106
Chapter 5: DNA-dependent RNA polymerase IV and RNA-dependent RNA polymerase 2 are physically coupled to produce siRNA precursors.	131
Chapter 6: Functional analysis of NRPD1 and NRPE1 C-terminal domains required for RNA-directed DNA methylation.	160

Chapter 7: Conclusions and future directions	219
Prologue	220
i. Biochemical elucidation of the RNA-directed DNA methylation pathway	220
a. Introduction	220
b. RNA Polymerase IV	221
c. Characterization of the Pol IV-RDR2 relationship	224
d. RNA Polymerase V	230
e. Steps towards <i>in vitro</i> reconstitution of the RdDM pathway	232
ii. Roles of the NRPD1 and NRPE1 C-Terminal domains	233
a. Introduction	233
b. Defective Chloroplast and Leaves-like Domain	234
c. Platform for protein-protein interactions	237
d. Target of post-translational modifications	245
e. Applications for dominant suppression of RdDM	247
iii. Structure-Function Analysis	250
a. Introduction	250
b. Determination of Pol IV and Pol V structures	250
c. Discovery of Pol IV-nucleic acid contacts	253
d. Elucidation of the eukaryotic DdRP subunit assembly pathway	254
v. References cited	255
Appendix A: Roles of RNA polymerase IV in gene silencing.	260
Appendix B: The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center.	270
Appendix C: Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II	290
Appendix D: Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes.	350
Appendix E: RNA polymerase V transcription guides ARGONAUTE4 to chromatin.	373
Appendix F: Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis.	380
Appendix G: RNA polymerase I: a multifunctional molecular machine.	399
Appendix H: <i>Curriculum vitae</i>	403

LIST OF TABLES AND FIGURES

Chapter 2

Table 1:	Gateway compatible plant destination vectors.	59
Figure 1:	A summary of available Gateway-compatible vectors for use in plants.	60
Figure 2:	Overview of Gateway cloning for generation of fusion proteins.	61
Figure 3:	<i>In vitro</i> evaluation of AcV5, HA, FLAG and cMyc epitope detection in commonly studied plants.	64
Figure 4:	pEarleyGate plant transformation vectors.	66
Figure 5:	Immunoblot detection of epitope-tagged recombinant proteins expressed from pEarleyGate-derived T-DNAs in <i>Arabidopsis thaliana</i> .	67
Figure 6:	Affinity purification of FLAG, HA, or cMyc-tagged HDA6 expressed in <i>Arabidopsis thaliana</i> transgenic plants.	67
Figure 7:	Use of pEarleyGate vectors for protein localization experiments.	68

Chapter 3

Figure 1:	Evidence for RNA Pol IV in plants.	75
Figure 2:	NRPD2 does not cofractionate with Pol II or with DNA-dependent RNA polymerase activity.	77
Figure 3:	Heterochromatin is disrupted in <i>nrpd2</i> mutants.	78
Figure 4:	NRPD1 and NRPD2 are required for 5S gene and <i>AtSN1</i> cytosine methylation and siRNA accumulation.	80
Tables S1 and S2:	GenBank accessions for the DNA-dependent RNA polymerase largest subunits analyzed in Figure 1.	86
Figure S1:	Multiple alignment of RPD1 with DNA-dependent RNA polymerase largest subunits of <i>A. thaliana</i> (At), <i>S. cerevisiae</i> (Sc), and <i>E. coli</i> (Ec).	89
Figure S2:	Multiple alignment of RPD2 with DNA-dependent RNA	96

polymerase second-largest subunits of *A. thaliana* (At),
S. cerevisiae (Sc), and *E. coli* (Ec).

Figure S3: Comparison of conserved domains A-H in RPD1a and DNA-dependent RNA polymerase largest subunits in <i>A. thaliana</i> (At), <i>S. cerevisiae</i> (Sc), and <i>E. coli</i> (Ec).	100
Figure S4: Comparison of conserved domains A-I in RPD2 and DNA-dependent RNA polymerase second-largest subunits in <i>A. thaliana</i> (At), <i>S. cerevisiae</i> (Sc), and <i>E. coli</i> (Ec).	101
Figure S5: Determination of the full-length mRNA sequence for RPD2a by RT-PCR, 5' RACE, and primer extension.	102
Table S3: Cytological changes in <i>rpd2</i> mutants.	104

Chapter 4

Figure 1: Catalytic residues that comprise the Metal A and Metal B binding sites of DNA-dependent RNA polymerases are conserved in the NRPD1, NRPE1/NRPD1b and NRPD2 subunits.	110
Figure 2: Pol IV and Pol V active site amino acids are required for rescue of small RNA production but not Pol IV and Pol V subunit assembly.	111
Figure 3: Pol IV and Pol V active site amino acids are required for the RNA-directed methylation of 5S rRNA gene repeats.	112
Figure 4: DNA methylation and transcriptional silencing of <i>AtSN1</i> retrotransposons requires the Pol IV and Pol V active sites.	113
Figure 5: NRPD1 and NRPE1/NRPD1b proteins mutated at their active sites fail to display characteristic Pol IV and Pol V punctate localization patterns in Arabidopsis nuclei.	114
Table S1: DNA oligonucleotides used in this study.	118
Table S2: Positions of amino acids that are invariant among Arabidopsis Pol I, II and III and yeast Pol II but have diverged in Arabidopsis Pol IV and Pol V largest and second-largest subunits.	119
Figure S1: Multiple alignment of <i>A. thaliana</i> RNAP largest subunits and the yeast Pol II largest subunit.	122
Figure S2: Flowering time control is dependent upon the Pol IV and Pol V	130

active sites.

Chapter 5

Figure 1: Pol IV and RDR2 interact <i>in vivo</i> .	153
Figure 2: Pol IV and RDR2 interaction is independent of Pol IV transcripts.	154
Figure 3: Pol IV displays DNA-dependent RNA polymerase activity.	155
Figure 4: RDR2 transcribes single-stranded RNA and DNA.	156
Figure S1: <i>RDR2</i> HA-tagged genomic transgene rescues <i>rdr2-1</i> mutant.	157
Figure S2: Pol IV and Pol V are predicted to be alpha-amanitin insensitive.	158

Chapter 6

Figure 1: The NRPE1 DeCL-liked domain is required for <i>nrpe1</i> <i>in vivo</i> complementation.	187
Figure 2: The NRPD1 DeCL-like domain is required for <i>nrpd1</i> <i>in vivo</i> complementation.	188
Figure 3: The NRPE1 repetitive elements and majority of WG motifs are not required for <i>nrpe1</i> complementation.	189
Figure 4: The NRPE1 WG motifs are important but not required for <i>nrpe1</i> <i>in vivo</i> complementation.	190
Figure 5: Over-expression of the NRPE1 CTD dominantly suppresses the RdDM pathway.	191
Table S1: Primers used in this study.	192
Figure S1: Comparison of NRPD1 and NRPE1 C-terminal domain architectures among diverse plant species.	197
Figure S2: Predicted NRPE1 protein sequences among diverse plant species with key domain features denoted to the right-hand side.	198
Figure S3: Predicted NRPD1 protein sequences among diverse plant species with key domain features denoted to the right-hand side.	207

Figure S4: Flowering time experiment with Arabidopsis plants grown under short-day conditions (8 hrs light/16 hrs dark) and randomly rotated every 4 to 6 days.	215
Figure S5: Visible phenotypes observed among wild type Arabidopsis plants transformed with pEarleyGate202-NRPD1 aa 1337-1453 (Line #258, T2 generation).	216
Figure S6: AGO4 <i>in vitro</i> interaction with the NRPE1 CTD.	217
Figure S7: Failure to verify reported NRPE1-AGO4 interaction <i>in vivo</i> .	218
<u>Chapter 7: Conclusions and future directions</u>	
Figure 1: Test for intrinsic RNA cleavage activity.	225
Figure 2: Western blot analysis of Pol IV-RDR2 interaction specificity by co-IP.	226
Figure 3: Model of the proposed Pol IV-RDR2 interface via the NRPD4/7 subcomplex.	227
Figure 4: Two proposed models for RDR2 polymerase initiation using Pol IV transcript as a template.	229
Figure 5: ClustalW2 protein sequence alignments of small DeCL domain containing proteins in Arabidopsis.	236
Figure 6: Conserved DeCL domain sequence block in NRPD1 and NRPE1 proteins from diverse plant species with AtDCL, DOMINO1 and At3g46630 reference sequences.	238
Table 1: Mediator subunits found in LC-MS/MS analysis of FLAG-NRPE1 aa 1851-1977 (NRPE1 QS).	241
Table 2: Selection of candidate proteins identified by LC-MS/MS that may interact with the NRPE1 CTD.	242
Table 3: Predicted nuclear-localized proteins that interact with AtDCL by Y2H.	244
Table 4: Predicted and experimentally observed NRPE1 amino acids that are phosphorylated or ubiquitinated.	246
Figure 7: Luciferase reporter screen to detect defects in the RdDM pathway.	247

Figure 8: Chop-PCR experiment to assay DNA methylation at the <i>AtSNI</i> -locus	250
---	-----

Appendix A:

Glossary	262
Figure 1: Catalytic subunits of DNA-dependent RNA polymerase.	263
Box 1: Pol IV subunit nomenclature.	264
Figure 2: A variety of proteins participate in Pol IVa-dependent silencing pathways.	266
Figure 3: Possible modes of Pol IVa function.	267

Appendix B:

Figure 1: Loss of siRNAs and cytosine methylation at repeated DNA sequences in mutants of the nuclear siRNA pathway.	273
Figure 2: Nuclear localization of siRNAs.	275
Figure 3: Immunolocalization of nuclear siRNA pathway proteins.	277
Figure 4: siRNAs colocalize with NRPD1b, RDR2, DCL3, and AGO4.	278
Figure 5: Pol IV colocalizes with endogenous repeat loci.	279
Figure 6: Effects of mutations on the localization of proteins involved in Nuclear siRNA biogenesis.	280
Figure 7: A spatial and temporal model for nuclear siRNA biogenesis.	281
Figure S1: Antibody specificity controls.	286
Figure S2: NRPD1a and NRPD1b immunolocalization signals are not lost in Dnase I-Treated nuclei.	287
Figure S3: Immunolocalization of DRD1, NRPD1b, and NRPD1a in <i>drd1-6</i> mutant nuclei.	287
Table S1: Supporting data for Figure 2A: siRNA probe hybridization patterns and frequencies.	288

Table S2: Supporting data for Figure 3A: protein localization and effect of RNase.	288
Table S3: Supporting data for Figure 3C: pairwise detection of nuclear siRNA pathway proteins.	288
Table S4: Supporting data for Figure 4: protein-siRNA colocalization.	289
Table S5: Supporting data for Figure 5: localization of proteins relative to NORs and 5S gene loci.	289
Table S6: Supporting data for Figure 6: protein localization in various nuclear siRNA pathway mutants.	289

Appendix C:

Figure 1: Relationship of <i>Arabidopsis</i> Pol II, IV, and V subunits to <i>E. coli</i> , Archaeal, and yeast RNA Pol II subunits.	294
Figure 2: Verification of Pol V subunit associations.	295
Figure 3: Pol V utilizes a distinct RPB3 variant, NRPE3b, as well as an NRPE3a variant corresponding to the Pol II NRPB3 subunit.	296
Figure 4: CoIP tests of Pol V, IV and II subunit associations.	297
Figure 5: <i>nrpe5</i> mutants are defective in RNA-directed DNA methylation and retrotransposon silencing.	298
Figure 6: Comparison of RNA polymerase subunits in Pol II, IV, and V.	300
Table S1: Genes whose known or predicted sequences were used in peptide coverage maps and/or protein alignments.	308
Table S2: Subunits of <i>Arabidopsis</i> Pol V identified by LC-MS/MS analysis of immunoprecipitated FLAG-NRPE5.	309
Table S3: List of primer sequences.	310
Figure S1: Peptide coverage maps of DNA-directed RNA polymerase subunits detected by LC MS/MS in affinity purified Pol V (NRPE1-FLAG).	311
Figure S2: Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS analysis of affinity purified Pol IV (NRPD1-FLAG).	317

Figure S3: Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS analysis of affinity purified Pol II (NRPB2-FLAG).	323
Figure S4: RPB4 family alignment.	329
Figure S5: RPB5 family alignment.	330
Figure S6: RPB6 family alignment.	331
Figure S7: RPB7 family alignment.	332
Figure S8: RPB8 family alignment.	333
Figure S9: RPB9 family alignment.	333
Figure S10: RPB10 family alignment.	333
Figure S11: RPB11 alignment.	333
Figure S12: RPB12 family alignment.	334
Figure S13: Expression patterns of the RPB5 family.	334
Figure S14: Analysis of <i>nripd5-1</i> and <i>nripe11-1</i> T-DNA insertion mutants.	335
Figure S15: Flowering time of individual plants from wild-type (ecotype Col-0) and <i>nripe5-1</i> populations.	336
Figure S16: Alignment of RPB5 family variants in diverse plants with non-plant RPB5s.	337
Figure S17: The N-terminal extension of NRPE5 is required for the protein's stability and function.	341
Figure S18: Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS in affinity purified FLAG-NRPE5 samples.	343

Appendix D:

Figure 1: Detection of intergenic Pol V-dependent transcripts.	354
Figure 2: Characterization of Pol V-dependent transcripts.	355
Figure 3: Evidence that Pol V synthesizes IGN transcripts.	357

Figure 4: RNA polymerase activity of Pol V is necessary for silencing adjacent transposons and repetitive elements.	359
Figure 5: Pol V-dependent transcription is necessary for heterochromatin formation.	361
Figure 6: Pol V-dependent transcription requires the chromatin remodeler DRD1, but not siRNA production or DNA methylation.	362
Figure 7: Possible modes of Action for Pol V in RNA-directed transcriptional silencing	363
Table S1: Oligonucleotides used in this study.	368
Figure S1: The chromosomal contexts of <i>IGN7</i> and <i>IGN15</i> loci at which Pol V-dependent transcripts have been identified (Fig 1F).	369
Figure S2: 5' ends of Pol V-dependent transcripts identified by 5'RACE.	370
Figure S3: The chromosomal contexts of the <i>AtSN1</i> and <i>solo LTR</i> loci tested in our study.	371
Figure S4: Quantitative PCR of control reactions in which no antibody was included in the chromatin immunoprecipitation (ChIP) experiments shown in Fig. 4F and Figs. 5A-C.	372

Appendix E:

Figure 1: Pol V, AGO4 and DMS3 work nonredundantly in heterochromatin formation.	376
Figure 2: AGO4 is not required for Pol V transcription.	376
Figure 3: Pol V transcription is necessary for AGO4-chromatin interactions.	377
Figure 4: AGO4 physically interacts with Pol V transcripts.	377
Figure 5: The SMC hinge-domain protein DMS3 is required for Pol V transcription and detectable Pol V-chromatin interactions.	378
Figure 6: A model for Pol V and siRNA-dependent heterochromatin formation.	378

Appendix F:

Figure 1: Sex-biased transmission of disrupted alleles for second-largest subunits of RNA polymerases I, II, and III (<i>NRPA2</i> , <i>NRPB2</i> , and <i>NRPC2</i> respectively).	385
Table 1: Genotypes of progeny of Pol I, II, and III hemizygotes.	386
Table 2: Male-specific transmission of Pol I, II, and III mutant alleles.	386
Figure 2: Failed seed development in siliques of <i>nrpa2-1</i> , <i>nrpb2-1</i> , <i>nrpb2-2</i> , <i>nrpc2-1</i> , and <i>nrpc2-2</i> hemizygotes.	387
Figure 3: Developmental arrest of mutant female gametophytes in flowers just prior to antithesis was visualized by confocal fluorescence microscopy.	388
Table 3: Female gametophyte development in polymerase mutants.	389
Figure 4: Development and early tube elongation of pollen are unaffected by defects in RNA polymerases.	390
Figure 5: Reduced paternal transmission of <i>nrpa2-1</i> , <i>nrpa2-2</i> , <i>nrpb2-1</i> , <i>nrpb2-2</i> , <i>nrpc2-1</i> , and <i>nrpc2-2</i> alleles relative to wild-type alleles in self-fertilized hemizygotes is due to decreased, competitive fertilization of ovules farthest from the stigma.	391
Table S1: Male-specific transmission of <i>RPB12a</i> mutant alleles.	396
Table S2: Transgene rescue allows maternal transmission of mutant alleles.	397
Figure S1: Developmentally arrested mutant female gametophytes within pistils just prior to antithesis, visualized by confocal fluorescent microscopy.	398

Appendix G:

Figure 1: RNA polymerase I.	401
-----------------------------	-----

CHAPTER ONE

INTRODUCTION

RNA POLYMERASE II: A MODEL FOR RNA POLYMERASES IV AND V

RNA Polymerases IV and V (Pol IV and Pol V) evolved from well-studied RNA Polymerase II (Pol II), and a thorough exploration of Pol IV and Pol V necessitates comparison with their evolutionary precursor. The DNA-dependent RNA polymerase (DdRP) enzyme carries out transcription of genetic information from DNA to RNA, by catalyzing the formation of phosphodiester bonds using dsDNA as a template. All DdRPs likely evolved from a common ancestral enzyme. Bacteria and archaea contain a single multisubunit RNA polymerase that is responsible for transcribing rRNA, mRNA and tRNA. The bacterial RNA polymerase is composed of five subunits, whereas the archaeal RNA polymerase is more complex with twelve subunits. Archaeal RNA polymerase is most likely the progenitor of the eukaryotic RNA Polymerases I, II and III given their similar, more complicated subunit compositions and structures. Eukaryotic RNA polymerases have a division of duties within the nucleus. Pol I transcribes 45S rRNA, Pol II transcribes mRNA, most micro RNA precursors and snRNA, and Pol III transcribes 5S rRNA and tRNA. The plant-specific Pol IV and Pol V evolved from Pol II and have specialized functions in RNA-mediated gene silencing.

Subunit nomenclature, composition and structure

We will focus here on the subunit composition and structure of yeast RNA Polymerase II (Pol II) and compare and contrast what is known about the subunit compositions of Arabidopsis Pol II, IV and V. Yeast Pol II is composed of twelve

subunits conserved among all eukaryotes. Pol II subunits are named by the prefix “Rpb” which is short for “RNA polymerase” with the letter “b” designating it is a Pol II subunit. Pol I subunits use the letter “a” and Pol III subunits use the letter “c”. The subunits are numbered 1 to 12 in order of molecular mass from largest to smallest. Thus the largest subunit of Pol II is Rpb1, the second-largest subunit of Pol II is Rpb2 and so on to the smallest subunit, Rpb12. This naming convention has been kept in Arabidopsis but reflects subunit homology to individual yeast subunits rather than the molecular mass, as the numbering would actually be different between yeast and Arabidopsis subunits. Due to conflicts with previously named genes in the Arabidopsis genome, the letter “N” has been added before the subunit name reflecting the nuclear localization of Pol I, II, III, IV and V. Extending the letter designation system, Pol IV uses the letter “d” and Pol V the letter “e”. Thus the Arabidopsis Pol II subunit homologous to yeast Rpb1 is named NRPB1 and the Pol IV subunit homologous to yeast Rpb7 is named NRPD7. Finally, some subunits are shared by two or more RNA polymerases. In this situation, the subunit can go by alternate names reflecting the RNA polymerase context. An example of this is the shared second-largest subunit of Pol IV and Pol V that is named both NRPD2 and NRPE2.

While most genomes contain a single gene encoding each RNA polymerase subunit, plant genomes have undergone many duplication events giving rise to multi-gene subunit families (2000). This holds true for the RNA polymerase subunits as there are multiple genes encoding Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12-like subunits in Arabidopsis shared by Pol I, II and III in yeast and mammals. In addition, the Pol II-specific Rpb3, Rpb4 and Rpb7 subunits are also in multi-gene families. Thus, nine of the

twelve homologous Pol II subunits have undergone gene expansion events in Arabidopsis making it difficult to predict whether only one or multiple genes in each gene family contribute functional subunits to Pol II. Given Pol IV and Pol V evolution from Pol II, such a prediction becomes even more difficult as subunit variants may have become specialized components of individual RNA polymerases.

Work performed by the Pikaard lab utilizing immunopurified Arabidopsis Pol II, Pol IV and Pol V samples subjected to tryptic digest and analyzed by liquid chromatography coupled tandem mass-spectrometry (LC-MS/MS), in addition to co-immunoprecipitation experiments, defined the complete subunit compositions of these three RNA polymerases (Ream et al., 2009). Identification of subunits in a partial Pol V complex purified from cauliflower (Huang et al., 2009), a forward genetics screen (He et al., 2009a) and a reverse genetics candidate approach (Lahmy et al., 2009) has supported this work. These findings demonstrate that Pol IV and Pol V are specialized forms of Pol II (Ream et al., 2009). The subunits can be categorized by their roles in the RNA polymerase complex (Werner, 2007) and will be briefly discussed in this context.

Catalytic subunits

The yeast largest and second-largest RNA polymerase subunits, Rpb1 and Rpb2, are homologous to the bacterial β' and β subunits, respectively. Yeast RNA Polymerases I, II and III each use unique largest and second-largest subunits. The largest and second-largest subunits interact to form the active center of the enzyme in their shared interior and make the majority of contacts with the DNA template and RNA product (Gnatt et al., 2001). The Metal A and Metal B sites in the largest and second-largest subunits,

respectively, coordinate two magnesium ions forming the catalytic active site of the enzyme required for transcription, backtracking and cleavage activities (Sosunov et al., 2003; Sosunov et al., 2005).

Arabidopsis Pol II, Pol IV and Pol V each use unique largest subunits named NRPB1, NRPD1 and NRPE1, respectively (Herr et al., 2005; Pontes et al., 2006; Pontier et al., 2005; Ream et al., 2009). The second-largest subunit of Pol II, NRPB2, is unique, whereas Pol IV and Pol V use a common second-largest subunit encoded by the same gene, NRPD2/NRPE2 (Onodera et al., 2005; Pontes et al., 2006; Pontier et al., 2005; Ream et al., 2009). Interestingly, NRPE1 but not NRPD1 *in vivo* protein levels are drastically reduced in *nrpe2* mutants and NRPE2 protein levels are reduced in *nrpe1* but not *nrpd1* mutants (Pontier et al., 2005). It has been suggested that either protein stability is compromised and/or the majority of NRPD2/NRPE2 protein is associated with Pol V with a smaller fraction associated with Pol IV. Follow-up studies to address these interpretations have yet to be performed but are consistent with observed lower concentration levels of *S. pombe* Rpb1, Rpb2 and Rpb3 subunits relative to the smaller Pol II subunits (Kimura et al., 2001).

Assembly subunits

The yeast Rpb3, Rpb10, Rpb11 and Rpb12 subunits are referred to as assembly subunits and help stabilize the RNA polymerase. Yeast RNA Polymerases I, II and III share the Rpb10 and Rpb12 subunits. Rpb3 and Rpb11 are functional equivalents of the bacterial α subunit. Bacterial RNA polymerase assembly begins with dimerization of two α subunits followed by β subunit binding to form a $\alpha_2\beta$ assembly intermediate that is

finally completed with binding of the β' subunit to form $\alpha_2\beta\beta'$ (Ishihama, 1981). Evidence suggests that eukaryotic RNA polymerase II assembly begins in a similar fashion with the Rpb2-Rpb3-Rpb11 subunits, which are equivalent to the bacterial RNA polymerase assembly intermediate $\alpha_2\beta$ (Kimura et al., 1997; Kimura et al., 2001). The Rpb3, Rpb10, Rpb11 and Rpb12 subunits form a compact subassembly in the yeast Pol II crystal structure (Cramer et al., 2001).

The Arabidopsis Rpb3 and Rpb11 homologs were demonstrated to interact *in vitro* and *in vivo* in an early study (Ulmasov et al., 1996). Arabidopsis Pol II, IV and V share the same Rpb10, Rpb11 and Rpb12 gene-encoded subunits, which are named NRPB10/NRPD10/NRPE10, NRPB11/NRPD11/NRPE11, and NRPB12/NRPD12/NRPE12, respectively (Huang et al., 2009; Ream et al., 2009). Two Arabidopsis genes encode the yeast Rpb3 homolog. Pol II, IV and V share one of the variants, NRPB3/NRPD3/NRPE3a, whereas the other variant, NRPE3b, exclusively associates with Pol V (Huang et al., 2009; Ream et al., 2009).

Auxiliary subunits

The auxiliary subunits in yeast Pol II are Rpb4, Rpb5, Rpb6, Rpb7, Rpb8 and Rpb9. Yeast RNA Polymerases I, II and III share the Rpb5, Rpb6 and Rpb8 subunits. The auxiliary subunits help stabilize interactions between the RNA polymerase, nucleic acids and exogenous transcription factors (Werner, 2007). Rpb5 and Rpb9 are positioned near the DNA entry point of Pol II, whereas Rpb4, Rpb6, Rpb7 and Rpb8 are located in the region of the RNA exit pore (Armache et al., 2005; Cramer et al., 2001). As with the

catalytic and assembly subunits discussed above, important comparisons can be made between the yeast auxiliary subunits and those of Arabidopsis Pol II, IV and V.

The yeast Rpb5 subunit is positioned near the DNA entry point into the RNA polymerase (Cramer et al., 2001) and the Rpb5 N-terminal “jaw domain” interacts with the downstream DNA (Gnatt et al., 2001). The yeast Rpb5 ortholog exists as a five-member gene family in Arabidopsis with at least two of the variant proteins expressed at the protein level (Larkin et al., 1999). The Rpb5 variant most similar to that used by Pol I, II and III in yeast associates with Arabidopsis Pol I, II and III purified complexes (Larkin et al., 1999; Saez-Vasquez and Pikaard, 1997). Later LC-MS/MS analysis and co-immunoprecipitation experiments demonstrated that Pol IV also shares this same Rpb5 variant (Lahmy et al., 2009; Ream et al., 2009). Thus, the Arabidopsis Rpb5 variant most similar to yeast Rpb5 and shared by Pol I, II, III and IV is called NRPA5/NRPB5/NRPC5/NRPD5. The second Arabidopsis Rpb5 variant is a Pol V-specific subunit, NRPE5, which has a unique N-terminal sequence extension of unknown significance (Huang et al., 2009; Lahmy et al., 2009; Ream et al., 2009). The remaining Arabidopsis Rpb5 variants have yet to be fully characterized but may have some functional redundancy as *nrpe5* mutants display less severe mutant phenotypes than *nrpe1* mutants (Huang et al., 2009; Lahmy et al., 2009; Ream et al., 2009). The differential use of Rpb5 variants in Arabidopsis has been hypothesized to play a role in template specificity and/or association with recruitment factors (Ream et al., 2009).

The Rpb4 and Rpb7 subunits form a Pol II-dissociable subcomplex dispensable for Pol II promoter binding (Edwards et al., 1991; Larkin and Guilfoyle, 1998; Orlicky et al., 2001) positioned near the RNA exit pore and adjacent to the Rpb1 CTD linker

(Armache et al., 2005). The yeast Rpb4 subunit is dispensable *in vivo* when yeast strains are grown under optimal conditions (Choder and Young, 1993), though under some stresses like low or high temperatures or starvation, Pol II loses its ability to transcribe most, if not all, genes (Choder and Young, 1993; Farago et al., 2003; Maillet et al., 1999; Miyao et al., 2001; Sheffer et al., 1999). Furthermore, the temperature-sensitive phenotype of Rpb4 deletion strains can be rescued by over-expression of Rpb7 (Maillet et al., 1999; Sheffer et al., 1999; Tan et al., 2000). Rpb7 has a functional RNA binding domain (Djupedal et al., 2005; Kato et al., 2005; Mitsuzawa et al., 2003; Ujvari and Luse, 2006) and is required, along with Rpb2, for siRNA-dependent heterochromatin formation in *S. pombe* (Djupedal et al., 2005; Kato et al., 2005). The Rpb4/7 subcomplex has roles during initiation and RNA 3' end processing (Mitsuzawa et al., 2003; Orlicky et al., 2001), and interacts with the RNA product co-transcriptionally in the nucleus. The subcomplex is able to dissociate from Pol II and chaperone mRNA to the cytoplasm to stimulate mRNA decay (Goler-Baron et al., 2008; Lotan et al., 2005; Lotan et al., 2007; Selitrennik et al., 2006).

The Arabidopsis genome encodes two Rpb4 variants and three Rpb7 variants. The NRPB4 variant is unique to Pol II, whereas Pol IV and Pol V share the second variant, NRPD4/NRPE4 (He et al., 2009a; Ream et al., 2009). The three Arabidopsis Rpb7 variants are all functionally distinct as Pol II uses NRPB7, Pol IV uses NRPD7 and Pol V uses NRPE7 (Ream et al., 2009). NRPD4/NRPE4 localization within the nucleus demonstrates it does not always co-localize with Pol IV and Pol V suggesting that the NRPD4/7 and NRPE4/7 sub complexes may also be able to dissociate and play some role in chaperoning Pol IV and Pol V transcripts (He et al., 2009a).

The RNA exit pore spatially separates the yeast subunits Rpb6 and Rpb8, whereas the Rpb9 subunit makes contact with the downstream DNA as it enters Pol II (Cramer et al., 2001). Rpb9 is involved in transcription start site selection (Furter-Graves et al., 1994; Hull et al., 1995) and is required for the transcript cleavage function of TFIIIS (Awrey et al., 1997). The Arabidopsis genome encodes two variants for each of the Rpb6, Rpb8 and Rpb9 subunits. LC-MS/MS and co-immunoprecipitation experiments demonstrate that Pol II, Pol IV and Pol V share both variants for each subunit with no observable preference (Ream et al., 2009).

It should be noted that yeast RNA Polymerases I and III are composed of equivalent Pol II-like 12 subunit cores with an additional two and five subunits, respectively. The additional subunits are likely due to Pol I and Pol III annexing exogenous proteins for dedicated polymerase functions (Werner, 2007). The Pol I-specific Rpa49 and Rpa34 subunits heterodimerize and promote elongation much as the Pol II-associated TFIIIF (Kuhn et al., 2007). The Pol III-specific Rpc82/34/31 subcomplex directs binding of Pol III to the TFIIIB-DNA complex (Wang and Roeder, 1997; Werner et al., 1992). The Pol III-specific Rpc53/37 subcomplex participates in Pol III termination and with Rpc11 promotes re-initiation for additional rounds of transcription (Landrieux et al., 2006).

Sequence conservation and divergence among the Pol II, IV and V largest and second-largest subunits

Catalytic core

Multisubunit DNA-dependent RNA polymerases (DdRP) are evolutionarily related having a high degree of sequence conservation among the largest and second-largest subunits in prokaryotes, viruses, archaea and eukaryotes (Allison et al., 1985; Bergsland and Haselkorn, 1991; Patel and Pickup, 1989; Puhler et al., 1989; Schneider et al., 1987; Sweetser et al., 1987). The largest subunit is characterized by the presence of eight conserved domains, named domains A-H (Allison et al., 1985; Jokerst et al., 1989), whereas the second-largest subunit has nine conserved domains, named domains A-I (Sweetser et al., 1987). Yeast Pol II analyses demonstrated that *S. cerevisiae* Rpb1 and Rpb2 mutations leading to conditional phenotypes were predominantly mapped to invariant amino acids within the conserved domains suggesting these amino acids were important for function (Martin et al., 1990; Scafe et al., 1990). Structural analyses of *E. coli* RNA polymerase and *S. cerevisiae* Pol II have substantiated these early interpretations revealing that the conserved domains are clustered around the interior polymerase active center (Cramer et al., 2001; Zhang et al., 1999), whereas amino acids that map to the exterior surfaces have little to no homology between prokaryotes and eukaryotes due to differences in subunit and regulatory machinery interactions (Cramer et al., 2001).

RNA polymerases IV and V were originally identified during annotation of the Arabidopsis genome (2000). Subunits for nuclear DNA-dependent RNA polymerases I, II and III were identified in addition to two additional atypical largest and two additional atypical second-largest RNA polymerase subunits. The atypical subunits appeared Pol II-like but had clearly diverged. Since all genomes only encode a single gene for each largest and second-largest subunit, it was unclear whether these additional subunits made

up a functional plant-specific RNA polymerase of simple or complex composition (2000).

Sequence analysis of the two atypical largest subunits, originally named NRPD1a and NRPD1b but now known as NRPD1 and NRPE1, respectively, shows they contain conserved domains A-H (Pikaard et al., 2008), though a region between domains F and G is deleted (Luo and Hall, 2007). The atypical second-largest subunit NRPD2, previously known as NRPD2a, contains conserved domains A-I (Pikaard et al., 2008). The remaining atypical second-largest subunit, NRPD2b, is encoded by a pseudogene with a premature stop codon in the first exon and thus is not expressed (Pontier et al., 2005). NRPD1 and NRPE1 have an estimated amino acid substitution rate 20 times greater than Arabidopsis NRPB1, whereas NRPD2 has a substitution rate 10 times greater than Arabidopsis NRPB2 (Luo and Hall, 2007).

The idea of a conserved catalytic mechanism among multisubunit DdRPs is supported by the conserved sequences and tertiary structures in regions of the largest and second-largest subunits that comprise the active center. Interestingly, the Pol IV and V largest and second-largest subunits have remained relatively well conserved in sequences that are predicted to lie at the periphery and exterior surfaces using homology to Pol II, while the greatest proportion of divergence has occurred in the vicinity of the active center including sequences around the Metal A site, trigger loop, bridge helix, cleft and funnel domains of NRPD1 and NRPE1 and the hybrid binding region of NRPD2 (Haag et al., 2009) (Chapter 4). This has led many to question if Pol IV and Pol V are functional RNA polymerases and if they use an alternative template for transcription.

One notable example of divergence is that the region between conserved domains F and G in NRPD1 and NRPE1 has been completely deleted (Luo and Hall, 2007) with NRPD1 proteins having a unique conserved sequence block that replaces the G domain (Erhard et al., 2009). Neither NRPD1 nor NRPE1 proteins have any detectable conservation with the trigger loop encoded by the conserved G domain and appear to completely lack the flexible tip of the trigger loop and the bridge helix (Haag et al., 2009; Landick, 2009). The trigger loop is a mobile structural element conserved in both prokaryotic and eukaryotic RNA polymerases that forms hydrogen bonds with the NTP substrate and is important for transcription elongation, control and fidelity (Bar-Nahum et al., 2005; Kaplan et al., 2008; Kireeva et al., 2008; Touloukhonov et al., 2007; Wang et al., 2006). The trigger loop is a target of alpha-amanitin binding (Brueckner and Cramer, 2008; Bushnell et al., 2002) causing potent inhibition of Pol II transcription (Jacob et al., 1970; Kedinger et al., 1970; Lindell et al., 1970), and to a lesser extent Pol III transcription (Weil and Blatti, 1975). The bridge helix plays a role in RNA polymerase translocation helping to hold the RNA-DNA hybrid helix tightly (Gnatt et al., 2001) and appears to have concerted movements with the trigger loop during elongation based on structural analysis (Brueckner and Cramer, 2008). Without these structural elements the processivity and fidelity of Pol IV and Pol V transcription are called into question unless compensatory changes have been made through the course of evolution.

Arguably the most important feature to analyze is the RNA polymerase active site composed of the Metal A and Metal B sites that each bind a magnesium ion and are required for transcription. The magnesium ions guide free nucleoside triphosphates (NTP) into the active site for RNA synthesis, stabilize the transition state of the growing

RNA chain and participate in transcript cleavage events during polymerase backtracking, a process which helps prevent polymerase arrest at pause sites (Cramer, 2006; Sosunov et al., 2003). Three invariant aspartate amino acids compose the Metal A site of DdRP largest subunits and permanently bind a magnesium ion (metal A), which binds the RNA 3' end (Cramer et al., 2001). Among archaeal and eukaryotic Pol I, II and III largest subunits, the Metal A site is embedded within a YNADDFDGDEMN conserved sequence motif. NRPD1 and NRPE1 sequences conserve the three invariant aspartates in keeping with their evolution from Pol II but have divergent sequences in the larger context of the Metal A site. NRPD1 proteins only conserve the DFDGD motif, whereas NRPE1 proteins conserve the ADFDGD motif (Haag et al., 2009)(Chapter 4). The Metal B site of DdRP second-largest subunits coordinates a mobile magnesium ion (metal B) that binds the NTP triphosphate moiety (Westover et al., 2004). The Metal B site is composed of an invariant glutamate and aspartate amino acid pair that are part of the larger G(Y/F)NQEDS sequence motif conserved among NRPD2/NRPE2, Pol II and prokaryotes (Haag et al., 2009) (Chapter 4).

Taken alone, the Pol IV and Pol V conserved Metal A and Metal B sites support the hypothesis that these plant-specific RNA polymerases are transcriptionally competent. Mutation of any one of the invariant amino acids composing the Metal A and Metal B sites is enough to disrupt binding of the magnesium ions and abrogate transcription in prokaryotes (Zaychikov et al., 1996), archaea (Werner and Weinzierl, 2002) and eukaryotes (Dieci et al., 1995). Thus, given the increased divergence rate of the Pol IV and Pol V largest and second-largest subunits, it is suggested that there is a selective pressure to conserve the invariant Metal A and Metal B sites. To test if these

sites were required for Pol IV and Pol V function, the Metal A sites of NRPD1 and NRPE1 as well as the Metal B site of NRPD2 were each mutated to alanines and analyzed for *in vivo* complementation of the respective mutants (Haag et al., 2009) (Chapter 4). Results concluded that Pol IV and Pol V require the Metal A and Metal B sites for *in vivo* complementation of defects in siRNA production, DNA methylation and retrotransposon transcript suppression. In support of this, an EMS mutagenesis screen identified a NRPE1 D451N mutant, *nrpe1-3*, that corresponds to a missense mutation in the second aspartate of the NRPE1 Metal A site (Lahmy et al., 2009), providing additional evidence for Pol IV and Pol V being functional RNA polymerases.

C-terminal domain features

Pol II NRPB1 is distinct from the largest subunits of prokaryotes, viruses, archaea, Pol I and Pol III by virtue of a long C-terminal domain (CTD) extension from the catalytic core. The Pol II CTD is composed of tandem heptad repeats bearing the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$ (Allison et al., 1985). The number of tandem repeats varies by species with 26 in yeast, 34 in Arabidopsis, 45 in Drosophila and 52 in mammals. A minimum number of heptad repeats, which varies by species, is required for *in vivo* function and viability (Allison et al., 1985; Bartolomei et al., 1988; Nonet et al., 1987). The Pol II CTD is positioned near the RNA exit pore but has not been crystallized with the complete yeast Pol II complex because of its mobility (Armache et al., 2005; Cramer et al., 2001). The heptad repeats are connected to the catalytic core by a flexible linker that forms an alpha helix binding Rpb7, which is part of a subcomplex with Rpb4 (Armache et al., 2005). The Pol II CTD is a target for post-translational

modifications and protein-protein interactions that help regulate enzyme activity and play a role in mRNA capping, splicing, cleavage and polyadenylation processing events (discussed in the next section).

Despite their Pol II evolutionary origins, Pol IV and Pol V largest subunits lack tandem heptad repeats but do have unique CTD extensions. The Pol IV CTD is well conserved among diverse plant species, whereas the Pol V CTD is still evolving between species but conserves major elements (Chapter 6). For the purposes of this introduction, the *Arabidopsis thaliana* Pol IV and Pol V CTDs will be discussed. NRPD1 and NRPE1 largest subunits share a plant-specific domain of unknown function, the Defective Chloroplast and Leaves-like (DeCL) domain. This domain is also present in three smaller Arabidopsis genome-encoded proteins that are hypothesized to play functionally similar but compartmentalized roles in ribosomal RNA (rRNA) processing and/or ribosome biogenesis events. AtDCL is chloroplast localized and required for rRNA processing and chloroplast and leaf development (Bellaoui and Gruissem, 2004; Bellaoui et al., 2003; Keddie et al., 1996); DOMINO1 is nuclear and nucleolus localized with an embryo defective mutant phenotype (Lahmy et al., 2004), and an uncharacterized DeCL-containing gene product, At3g46630, is predicted to localize to mitochondria (Lahmy et al., 2004). The presence of the plant-specific DeCL domain in the NRPD1 and NRPE1 CTDs suggests a possible RNA-associated role consistent with Pol IV and Pol V being plant-specific nuclear RNA polymerases but this has yet to be formally tested.

NRPE1 also contains two additional C-terminal domains. N-terminal of the DeCL domain are ten imperfect 16 amino acid (aa) repeats with tryptophan-glycine (WG) sequence motifs embedded within the repeats and flanking (Pontier et al., 2005). WG

sequence motifs have been demonstrated to act as protein-protein interaction domains with the Argonaute PIWI domain. Examples include *S. pombe* Ago1 interaction with Tas3 (Verdel et al., 2004), human Ago1 and Ago2 interaction with GW182 (Liu et al., 2005; Takimoto et al., 2009) and the reported Arabidopsis AGO4 interaction with NRPE1 (El-Shami et al., 2007). While the NRPE1-AGO4 interaction has been replicated *in vitro* (He et al., 2009b) (Chapter 6), *in vivo* results have not been replicated (Li et al., 2006) and thus the prevalence and significance of this interaction is still to be determined (Chapter 6). The NRPE1 WG motifs have been reported to be required for *in vivo* complementation of the *nrpe1* mutant (El-Shami et al., 2007), but these results have been found to be inaccurate under our growth and test conditions (Chapter 6).

C-terminal to the DeCL domain at the NRPE1 C-terminus is a glutamine-serine rich (QS-rich) domain unique to Arabidopsis. Spinach NRPE1 contains a proline-serine rich (PS-rich) domain in its place (Pontier et al., 2005), but a comparable domain is not detected in any other NRPE1 protein sequences (Chapter 6). The serines in the QS-rich domain are predicted to be targets of post-translational phosphorylation and glycosylation events, but this has not been experimentally determined and no functional significance has yet been assigned to this domain.

Regulation via the Pol II largest subunit C-terminal domain

As mentioned above, the Pol II CTD is composed of an array of tandem heptad repeats bearing the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$ with important regulatory roles. The Pol II CTD is a target for post-translational modifications. There are five potential phosphorylation sites in each consensus heptad repeat (Y_1 , S_2 , T_4 , S_5 , and S_7) with S_2 and

S₅ being the predominant targets (Corden et al., 1985; Zhang and Corden, 1991). This is mediated by site-specific CTD kinases and phosphatases that dynamically change the Pol II CTD phosphorylation pattern during the course of the transcription cycle. Given the number of heptad repeats in each Pol II CTD, there are many potential combinations of phosphorylation states that could be present at any one time leading to the hypothesis that there is a “CTD code” to be cracked (Egloff and Murphy, 2008).

Early studies found that purified Pol II was predominantly present in two forms that differed by the extent of phosphorylation in the Pol II CTD: a high mobility, unphosphorylated form (IIA; RNAPIIA) and a low-mobility, phosphorylated form (IIO; RNAPIIO). The more abundant, IIA form corresponds to the Pol II initiation state, whereas the IIO form corresponds to the Pol II elongation state (Payne et al., 1989). Transcription initiation begins with recognition of the unphosphorylated IIO form by the general transcription factor TATA-binding protein (TBP) and the multisubunit Mediator complex, which recruit Pol II to promoters (Myers et al., 1998; Usheva et al., 1992). The Mediator complex makes multiple contacts with Pol II subunits but requires the CTD to stimulate Pol II transcription *in vitro* (Davis et al., 2002; Myers et al., 1998; Usheva et al., 1992). Phosphorylation of S₅ by TFIIH promotes the release of Mediator (Max et al., 2007) and the binding of guanylyltransferase (Cho et al., 1997; McCracken et al., 1997), which adds a 7-methylguanosine cap to Pol II transcripts shortly after they emerge from the RNA exit pore.

The elongating form of Pol II is characterized by a hyperphosphorylated CTD with phosphorylation of both S₂ and S₅ facilitated by a host of CTD kinases (Prelich, 2002). Splicing factors such as the mammalian CA150 and yeast Prp40, Ess1 and Pin1

all preferentially bind the hyperphosphorylated CTD (Phatnani and Greenleaf, 2006). All of the heptad repeats may not be identically modified, though, as Spt6 prefers phosphorylation only at S₂ to direct splicing (Yoh et al., 2007). Towards the 3' end of the gene, phosphatases target S₂ so that the CTD is predominantly phosphorylated only at S₅. This recruits 3' polyadenylation machinery and may also signal a transcript termination signal (Licatalosi et al., 2002; Meinhart and Cramer, 2004). In the case of Pol II-mediated U2 snRNA transcription, phosphorylation of S₇ is required for *in vitro* CTD interaction with Integrator, a large complex with roles in snRNA transcription and 3' processing (Egloff et al., 2007; Jacobs et al., 2004). Finally, there is evidence that the Pol II CTD is glycosylated when the heptad serine and threonine residues lack phosphorylation in a mutually exclusive manner (Comer and Hart, 2001; Kelly et al., 1993), though the significance of this has not been determined. Thus, the Pol II CTD plays an active *in vivo* role with the regulation of Pol II transcription and the recruitment of RNA processing factors at specific stages of the transcription cycle.

As mentioned in the previous section, the Pol IV and Pol V largest subunits also have CTD extensions with a common DeCL domain in both NRPD1 and NRPE1 and NRPE1-specific 16 aa repeat elements with WG motifs and a QS-rich domain. The role of the Pol IV and Pol V CTDs is at its infancy but experiments suggest that these domains also play a vital role for full polymerase function. *nRPD1* and *nRPE1* mutants are unable to be complemented *in vivo* with *NRPD1* and *NRPE1* transgenes lacking the DeCL domain (Chapter 6). *NRPE1* transgenes bearing an internal deletion of the majority of WG motifs are partially able to complement *nRPE1* mutants suggesting the WG motifs are important but not required for full Pol V function (Chapter 6). *In vitro*

protein-protein interaction studies have also implicated interaction of ARGONAUTE4 (AGO4) with the NRPE1 CTD via the WG motifs (El-Shami et al., 2007), though it is not clear how prevalent this interaction is *in vivo* or whether it is predominantly due to AGO4 interaction with Pol V transcripts (Wierzbicki et al., 2009) (Chapter 6).

The existence of Pol IV and Pol V post-translational modifications has not yet been reported in the literature, but the Pol V largest subunit is typically detected on protein blots as two migrating bands (Pontes et al., 2006; Pontier et al., 2005) reminiscent of the IIO and IIA forms of Pol II. Deletion of the full NRPE1 CTD leads to detection of only a single band (Chapter 6). This may be suggestive of Pol V CTD post-translational modification, but does not rule out alternative splicing or proteolysis.

Pol IV and Pol V use of general transcription machinery is also a largely unexplored area, but given their Pol II evolution would not be surprising. Three labs using forward genetics (He et al., 2009b), reverse genetics (Bies-Etheve et al., 2009) and proteomics (Huang et al., 2009) approaches identified a Spt5-like transcription elongation factor named KTF1 that functions with Pol V. In the context of yeast Pol II transcription, Spt5 interacts with Pol II and RNA processing factors (Lindstrom et al., 2003) suggesting the plant-specific Pol IV and Pol V may either share Pol II transcription machinery or may have evolved functionally distinct versions of Pol II transcription machinery. We are just beginning to understand the full scope of Pol IV and Pol V regulation and activity, but, based on what is already known, Pol II will undoubtedly provide a very useful roadmap for the journey that lies ahead.

DNA-dependent RNA polymerase activity

With the lack of published Pol IV and Pol V *in vitro* activity, there has been wide speculation about whether or not they are functional polymerases and if they transcribe dsDNA, methylated dsDNA, RNA-DNA hybrids or dsRNA templates. Strong evidence exists for a conserved mechanism of nucleotide addition that applies not only to multisubunit RNA polymerases, but also single subunit DNA and RNA polymerases (Iyer et al., 2003; Joyce and Steitz, 1995; Sosunov et al., 2005; Steitz, 1998). All known DNA and RNA polymerases contain magnesium ions at their active sites bound by highly conserved chelating motifs (Dieci et al., 1995; Zaychikov et al., 1996), referred to as Metal A and Metal B in the context of Pol II (Cramer et al., 2001). Using yeast Pol II as an example (Cramer et al., 2001; Gnatt et al., 2001), the downstream DNA contacts the N-terminal “jaw domain” of Rpb5 passing between Rpb1 and Rpb2 to enter the polymerase. A transcription bubble is formed whereby the template and non-template DNA strands separate with the template strand continuing along the bottom of the “clamp” and over the “bridge helix”. Template nucleotide +1 is oriented toward the active site for recognition. Free NTPs enter the active center through a pore in the backside of the enzyme. Metal A binds the phosphate group between the nucleotide at the RNA product 3' end (position +1) and the adjacent previously incorporated nucleotide (position -1), while metal B binds the incoming NTP substrate. Both metal A and metal B act to stabilize the transition state during phosphodiester bond formation. Metal A is persistently bound, whereas metal B is transient, perhaps entering with the NTP substrate. The nucleotide at position +1 is the first of nine base pairs of DNA-RNA hybrid that travels between the Rpb1 “bridge helix” and the Rpb2 “wall”, which induces

a nearly 90-degree bend in the DNA-RNA hybrid. Once the transcript reaches 10 nucleotides in length, the RNA and DNA strands separate with the aid of the “rudder”, “lid” and “zipper” loops of Rpb1. The RNA product exits through “groove 1”, or the “RNA exit pore”, adjacent to the CTD linker and Rpb4/7 subcomplex. The template DNA strand exits through another pore re-hybridizing with the nontemplate DNA strand.

Demonstration of Pol IV and Pol V *in vivo* requirements for the Metal A and Metal B sites (Haag et al., 2009) (Chapter 4) alone does not verify that Pol IV and V are functional DdRPs. Eukaryotic single subunit RNA-dependent RNA polymerases (RdRP) including Arabidopsis RDR2 and Neurospora QDE-1 have a Metal A site with consensus sequence DxDGD (Iyer et al., 2003). Pol II has also been reported to act as an RdRP *in vitro* using a RNA template-product duplex (Lehmann et al., 2007). The reaction uses the Metal A site but is slower and less processive than Pol II DdRP activity. This specialized Pol II function may be relevant for replication of the hepatitis delta virus RNA genome (Lai, 2005; Taylor, 2003) and plant viroids (Rackwitz et al., 1981).

The conserved asparagine amino acid immediately preceding the Metal A aspartate triad, NADFDGD, has been proposed to play a role in discriminating between ribonucleotide and deoxyribonucleotide substrates in yeast Pol II transcription (Gnatt et al., 2001). Mutation of the corresponding asparagine in bacteria leads to a loss in discrimination between these two substrates (Svetlov et al., 2004). This asparagine is not conserved in any of the NRPD1 and NRPE1 proteins (Chapter 6) calling into question the specificity of Pol IV and Pol V transcription.

To date, demonstrated Pol IV or Pol V *in vitro* transcriptional activity has not been published. Arabidopsis NRPD2/NRPE2 DEAE-Sepharose column-enriched

fractions presumably containing both Pol IV and Pol V complexes failed to transcribe sheared salmon sperm DNA (Onodera et al., 2005). Cauliflower immunopurified Pol V has also failed to transcribe cauliflower total DNA and Turnip Crinkle Virus ssRNA templates (Huang et al., 2009). Run-on transcription assays in maize have also failed to identify Pol IV transcripts (Erhard et al., 2009).

Chapter 5 of this thesis demonstrates *in vitro* DNA-dependent RNA polymerase activity for Arabidopsis immunopurified Pol IV. Using a tripartite oligo scaffold that mimics a dsDNA template with an elongating RNA product, Pol IV-derived full-length RNA transcripts were obtained. *In vitro* full-length transcription was dependent on the Pol IV Metal A site. Reactions supplemented with alpha amanitin, a potent Pol II inhibitor, did not inhibit Pol IV *in vitro* activity consistent with NRPD1 lacking conserved trigger loop sequences targeted by alpha amanitin.

Pol V-dependent transcripts have been detected *in vivo* corresponding to intergenic and noncoding loci (Wierzbicki et al., 2008). Transcripts are dependent on the Pol V Metal A site and are characterized by having 5' triphosphates or 7meG caps, a lack of poly A tails, short in length (~200 nt) and can initiate from multiple sites. Pol V can be crosslinked to chromatin *in vivo*, as well as to Pol V-dependent RNA transcripts supporting the hypothesis that Pol V is a DNA-dependent RNA polymerase (Wierzbicki et al., 2008).

ROLES OF RNA POLYMERASES IV AND V IN GENE SILENCING

RNA Polymerases IV and V have roles in many plant small RNA pathways that ultimately lead to gene silencing. RNA silencing pathways can be diverse but at their core have three things in common: a double-stranded RNA trigger, Dicer-mediated cleavage of the dsRNA producing small RNAs, and incorporation of small RNA into an Argonaute-RISC (AGO-RISC) complex to bind/cleave complementary transcripts and/or direct DNA methylation and gene silencing. Double-stranded RNA (dsRNA) is produced via overlapping bi-directional transcripts, self-complementary RNA hairpin transcripts, or with the aid of a RNA-dependent RNA polymerase that transcribes single-stranded RNA (ssRNA) to produce dsRNA. The action of a Dicer protein, an endoribonuclease III-like enzyme, cleaves the dsRNA substrate into small RNA duplexes typically 21-24 nucleotides (nt) in length in Arabidopsis, with 2 nt 3'OH overhangs on both ends. An Argonaute protein binds one strand of the small RNA duplex to make a RNA-induced silencing complex (RISC) that uses the small RNA to conduct a homology search for complementary RNA transcripts. AGO-RISC can bind the target RNA preventing translation, cleave the target RNA leading to target degradation, or direct DNA methylation at target loci for gene silencing.

S. pombe, *Tetrahymena*, *Drosophila*, mammals and Arabidopsis all have RNA silencing mechanisms and are among the most studied systems. The Arabidopsis genome has greatly expanded the number of proteins involved in RNA silencing encoding four DICER-LIKE proteins (DCL1-4), six RNA-DEPENDENT RNA POLYMERASE

proteins (RDR1-6) and ten ARGONAUTE proteins (AGO1-10). In so doing, the Arabidopsis RNA silencing machinery components have become functionally specialized with varying degrees of redundancy. While evidence suggests non-plant eukaryotic systems may use RNA Polymerase II (Pol II) to transcribe the RNA silencing trigger, evidence suggests that plants have evolved functionally specialized forms of Pol II, named Pol IV and Pol V, for this role. Pol IV and Pol V have been found to be involved in gene silencing phenomena that include RNA-directed DNA methylation, paramutation, flowering and development, abiotic and biotic stress-inducible responses, and short- and long-distance silencing.

RNA-directed DNA methylation

RNA-directed DNA methylation (RdDM) is a mechanism whereby small-interfering RNA (siRNA) directs DNA methylation to homologous target chromosomal loci either in *cis* or *trans* that induces heterochromatin formation and gene silencing, primarily at highly repetitive sequences. siRNAs are able to direct cleavage of homologous mRNA transcripts when integrated into an AGO-RISC complex and are also hypothesized to recruit the factors for heterochromatin formation in a sequence-specific manner by either binding RNA transcripts still present at the originating DNA locus or by directly binding the DNA locus. Reverse genetic candidate approaches and the results of a few very successful genetic screens have identified the core players of this pathway with additional components still being discovered. Pol IV acts at the beginning of the pathway and is required for producing RNA precursors for siRNA biogenesis, whereas

Pol V acts at the downstream end of the pathway with Pol V-generated transcripts believed to act as a scaffold for the chromatin modification machinery.

siRNA biogenesis

Pol IV was originally implicated in RdDM as the result of a genetic screen to identify *silencing defective* (*sde*) mutants. Arabidopsis plants will silence expression of a green fluorescent protein (GFP) transgene when crossed with plants containing a second transgene encoding the silenced potato virus X (PVX)-GFP transgene (Dalmay et al., 2000). Plants with this GFP-silenced genetic background (GxA) were mutagenized to identify individuals that expressed GFP. The *sde4* mutant not only reactivated GFP expression, but also caused a loss of *AtSN1* retrotransposon siRNA production and DNA methylation (Hamilton et al., 2002). The *sde4* mutant was later identified as NRPD1, the Pol IV largest subunit (Herr et al., 2005). A reverse genetics approach using the GxA reporter line and RNA interference to knock down *NRPD2* expression showed that NRPD2, the Pol IV second-largest subunit, was also required (Herr et al., 2005). At the same time, the Pikaard lab was studying NRPD2 by a reverse genetics approach as it had been identified as an atypical second-largest RNA polymerase subunit in the Arabidopsis genome (2000). NRPD2 was found to be nuclear localized but functionally distinct from Pol I, II and III second-largest subunits (Onodera et al., 2005) (Chapter 3). In addition to the siRNA and DNA methylation defects described above, *nprpd2* mutant nuclei displayed dispersed H3K9 methylation, 5S rDNA and chromocenters, suggesting large-scale impacts at the heterochromatin level (Onodera et al., 2005). Pol IV was later demonstrated to localize to regions with endogenous repeat loci that are targets for

RdDM (Pontes et al., 2006). Together with the findings that Pol IV requires the catalytic Metal A site (Haag et al., 2009) (Chapter 4), has *in vitro* DNA-dependent RNA polymerase activity (Chapter 5), and is required for siRNA biogenesis and the proper localization of all known proteins in the pathway (Pontes et al., 2006), the evidence supports Pol IV acting first to generate the trigger RNA from transcribed target loci.

RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) is the only Arabidopsis RNA-dependent RNA polymerase (RdRP) demonstrated to act in the RdDM pathway. Like Pol IV, RDR2 is required for siRNA production (Kasschau et al., 2007; Lu et al., 2006; Xie et al., 2004). *In vivo* co-immunoprecipitation experiments demonstrate that Pol IV and RDR2 are physically coupled (Chapter 5). This suggests that Pol IV may immediately transfer its transcripts to RDR2 for dsRNA production and help explain the specificity of RDR2 for the RdDM pathway (Kasschau et al., 2007). *In vitro* activity for RDR2 has not yet been published, but Pol IV-RDR2 affinity purified complexes have RNA- and DNA-dependent RNA polymerase activities with single-stranded templates dependent on RDR2 (Chapter 5), corresponding with the observed *in vitro* activities of RDR6 (Curaba and Chen, 2008).

RDR2-generated dsRNA is a substrate for DICER-LIKE3 (DCL3)-mediated cleavage. DCL3 is the endoribonuclease III-like enzyme predominantly responsible for generating the 24 nt siRNA size class in Arabidopsis (Qi et al., 2005; Xie et al., 2004). DCL2 and DCL4 are able to partially compensate in *dcl3* mutants by producing 21 and 22 nt siRNAs (Henderson et al., 2006; Kasschau et al., 2007), but the siRNAs are not fully functional as *dcl3* mutants have DNA methylation and transcript suppression defects, though not as severe as *nprdl* or *rdr2* mutants (Xie et al., 2004). The moss

Physcomitrella patens DCL3 homolog appears to have a conserved role producing siRNAs predominantly corresponding to transposable elements that are targets of DNA methylation (Cho et al., 2008). While DCL1 and DCL4 require the double stranded RNA-binding proteins DRB1 and DRB4, respectively, for full function, DCL3 does not appear to require any of the DRBs for siRNA production (Curtin et al., 2008).

HUA-ENHANCER 1 (HEN1) was identified as being essential for micro RNA (miRNA) stability (Park et al., 2002) and later found to bind 21-24 nt small RNA duplexes and add a methyl group on to the 2' OH of the 3' terminal nucleotide of each strand (Yang et al., 2006; Yu et al., 2005). The 3' methylation of siRNAs and miRNAs protects them from an Arabidopsis *in vivo* 3' end uridylation activity that is biased towards the sense strand (Li et al., 2005). *Hen1* mutants display decreased accumulation of siRNAs and decreased DNA methylation at *AtSN1*, fitting with its role in the RdDM pathway (Onodera et al., 2005; Xie et al., 2004). It has been proposed that 3' uridylation of small RNAs may act as a degradation signal but this hypothesis has not been formally tested.

AGO-RISC assembly

Once DCL3 produces the siRNA duplex and HEN1 methylates the 3' ends, a single strand of the siRNA duplex is bound by ARGONAUTE 4 (AGO4). The strand bound by AGO4, the sense strand, is determined by the asymmetric thermodynamic properties of the siRNA duplex itself. The siRNA strand whose 5' end is more weakly bound to the complementary strand is unwound and bound by Argonaute to form a RISC complex (Schwarz et al., 2003). The specific Argonaute that a siRNA strand associates

with determines how silencing will be mediated as some Argonautes bind target mRNAs to block translation, others cleave the target mRNA and still others help direct DNA methylation to homologous loci. In Arabidopsis, one of the key factors determining which Argonaute the siRNA associates with is the identity of the siRNA 5' nucleotide (Mi et al., 2008; Montgomery et al., 2008). In the case of AGO4, siRNAs with a 5' adenosine are favored 79% of the time (Mi et al., 2008). Other factors likely include siRNA length and channeling of substrates through individual pathways.

Argonaute proteins are characterized by the presence of the PAZ, MID and PIWI domains (Vaucheret, 2008). The MID domain binds the 5' phosphate of small RNAs, whereas the PAZ domain binds the 3' end. The PIWI domain adopts an RNaseH-like fold (Song et al., 2004) that also acts as an interaction domain for WG motif-containing proteins (El-Shami et al., 2007; Liu et al., 2005; Verdel et al., 2004) as discussed in a previous section. Argonaute proteins come in catalytic and non-catalytic forms depending on the presence of the catalytic Asp-Asp-His (DDH) triad in the PIWI domain (Baumberger and Baulcombe, 2005; Qi et al., 2005; Qi et al., 2006; Rivas et al., 2005). AGO4 binds siRNAs *in vivo* (AGO4-RISC) and cleaves target mRNAs *in vitro* (Qi et al., 2006).

AGO4 also controls locus-specific siRNA accumulation and DNA methylation (Zilberman et al., 2003). Interestingly, mutagenesis of the AGO4 catalytic triad not only abolishes cleavage activity but also variably affects siRNA accumulation and DNA methylation (Qi et al., 2006). These results reflect what is observed in *ago4* mutants with DNA methylation reduced but not gone (Xie et al., 2004) and siRNA accumulation decreased for some, but not all, loci (Qi et al., 2006). This is likely be due to redundancy

between the ten Argonaute proteins in Arabidopsis as AGO6 has also been implicated as having a role in the RdDM pathway. *Ago4,6* double mutants show a more dramatic loss of siRNA accumulation and DNA methylation than either single mutant (Zheng et al., 2007). Still to be resolved is the observation that *nRPD1*, *rdr2* and *dcl3* mutants reduce AGO4 stability (Li et al., 2006). This is hypothesized to be due to the loss of 24 nt siRNAs, as the *dcl2,3,4* triple mutant shows decreased AGO4 stability compared to a *dcl3* single mutant (Wierzbicki et al., 2009). Also in support of this, AGO4 stability is unaffected in *nrpe1* mutants where siRNA accumulation is unaffected at most loci (Wierzbicki et al., 2009).

Pol V transcription, AGO4-RISC and DNA methylation

Corresponding with its role in helping direct DNA methylation, AGO4-RISC is hypothesized to target loci complementary to its bound sense siRNA strand for RNA-directed DNA methylation. To dissect this downstream end of the RdDM pathway, a genetic screen was employed using an inverted repeat trigger that is homologous to the seed-specific promoter that drives expression of a GFP transgene. Arabidopsis seeds in this background display silenced GFP. These plants were mutagenized and screened for mutants *defective in RNA-directed DNA methylation* (*drd* mutants) that displayed GFP activation. Two of the mutants, *drd3* and *drd2*, corresponded to the Pol V largest (NRPE1) and second-largest (NRPE2) subunits, respectively (Kanno et al., 2005a). The *nrpe1* and *nrpe2* mutants are characterized by a loss of DNA methylation and reactivation of silenced loci also affected by *nRPD1*, *rdr2*, *dcl3* and *ago4*. A reverse genetics approach and subsequent large-scale sequencing showed that NRPE1 is required

for the accumulation of only some siRNAs with many only mildly affected, if at all (Mosher et al., 2009; Pontier et al., 2005). These two studies established that Pol V was functionally distinct from Pol IV, with Pol V being more involved with the downstream end of the RdDM pathway for establishment of gene silencing.

Subsequent studies established that Pol V likely functions as a DNA-dependent RNA polymerase *in vivo* due to the requirement of the NRPE1 Metal A site for function (Haag et al., 2009) (Chapter 4) and the detection of Pol V-dependent transcripts that are made independent of siRNA production (Wierzbicki et al., 2008). Evidence suggests that Pol V transcripts may act as scaffolds for AGO4-RISC binding since AGO4 can be crosslinked to Pol V transcripts (Wierzbicki et al., 2009), but does not necessarily rule out the possibility that AGO4-RISC binds target DNA, as Pol V transcription is also required for AGO4 binding to Pol V-dependent loci (Wierzbicki et al., 2009). The act of Pol V transcription in intergenic regions is thus hypothesized to act as a roadblock preventing other RNA polymerases from initiating transcription, either directly or indirectly (Wierzbicki et al., 2008), helping address the paradox of why you need transcription to silence transcription.

Two SNF2 chromatin-remodeling proteins, CLSY1 (Smith et al., 2007) and DRD1 (Kanno et al., 2005b); a SMC hinge domain protein, DMS3 (Kanno et al., 2008); and a Spt5-like transcription elongation factor, KTF1 (Bies-Etheve et al., 2009; He et al., 2009b; Huang et al., 2009), have also been identified to act in RdDM. Based on genetic evidence and its localization pattern, CLSY1 is hypothesized to act between Pol IV and RDR2, potentially on Pol IV transcripts (Pikaard et al., 2008; Smith et al., 2007), whereas DRD1 and DMS3 are required for Pol V interaction with chromosomal loci and

transcription (Wierzbicki et al., 2008; Wierzbicki et al., 2009). Mutants in KTF1 have reduced DNA methylation and release silencing of RdDM target loci but do not affect siRNA accumulation (Bies-Etheve et al., 2009; He et al., 2009b; Huang et al., 2009). KTF1 has been found associated with an immunopurified partial Pol V complex from cauliflower (Huang et al., 2009). KTF1 interacts with Pol V transcripts and contains WG motifs that mediate interaction with AGO4 (He et al., 2009b). It is hypothesized that KTF1 may bind Pol V and/or Pol V transcripts to help recruit AGO4 via its WG motifs.

The AGO4-RISC interaction with Pol V transcripts and/or Pol V-dependent loci is hypothesized to recruit DNA methylation and chromatin modification machinery to targeted loci. Evidence suggests the siRNA sequence directs DNA methylation to complementary chromosomal sequences accounting for about 30% of the cytosine DNA methylation in Arabidopsis (Cokus et al., 2008; Lister et al., 2008). The putative *de novo* cytosine DNA methyltransferases associated with RdDM are DOMAINS REARRANGED METHYLTRANSFERASE 1 and 2 (DRM1/DRM2) that are required for the establishment of *de novo* cytosine methylation in all contexts (CG, CNG and CNN) but not the maintenance of DNA methylation (Cao and Jacobsen, 2002). DNA methylation is believed to feedback on siRNA production as *drm1,2* mutants lack siRNA production at some loci (Onodera et al., 2005; Zilberman et al., 2004).

This order and progression of the RdDM pathway has largely been deduced genetically and by predictions and/or confirmation of protein enzymatic activities. Localization of the proteins in the pathway supports this and suggests that substrates are trafficked. Pol IV localizes to regions of dense DAPI staining associated with the chromocenters and heterochromatin and co-localizes to some extent with RDR2 (Pontes

et al., 2006) (Chapter 5). RDR2 is also present around the inner periphery of the nucleolus but is disrupted in *clsyl* mutants, whereas Pol IV localization is only partially affected (Smith et al., 2007) possibly capturing moments when Pol IV and RDR2 are physically coupled (Chapter 5). RDR2, DCL3, AGO4 and NRPE1 all co-localize with one another and siRNA in a distinct compartment of the nucleolus where siRNA processing is hypothesized to occur, with partial co-localization of NRPE1, KTF1 and AGO4 at target loci (He et al., 2009b; Li et al., 2006; Pontes et al., 2006).

Paramutation

Paramutation is a heritable chromatin change induced by allele-specific interactions that affects gene expression. Reports of paramutation have been made in mice and other eukaryotes with maize the best studied. Multiple paramutable maize loci have been reported including *rl*, *bl*, and *pl1*. Each encodes a transcription factor that activates the anthocyanin pigment biosynthetic pathway. The pathways produce red/purple pigments in maize tissue-specific patterns that are easily observed with corresponding changes in RNA transcript levels. The expression of one allele sensitive to altered expression in a heterozygote (called the paramutable allele) is altered by the presence of the other allele that induces the change (called the paramutagenic allele). Thus in the case of the *bl* locus, *B-I* is the paramutable allele and has extreme purple pigmentation, whereas *B'* is the paramutagenic allele and is weakly expressed with light pigmentation. When crossed, the *B'/B-I* heterozygote has light pigmentation, effectively becoming *B'/B-I**. If the heterozygote is outcrossed to a naïve *B-I* plant, the progeny only inherit the *B'* allele and all have light pigmentation. The two alleles have the exact same

DNA sequence with the same DNA methylation patterns but have transcription rates that differ by 10- to 20-fold (Chandler et al., 2000).

Forward genetic screens have identified three proteins required for maize paramutation that are all components of the Arabidopsis RdDM pathway. REQUIRED TO MAINTAIN REPRESSION 6 (RMR6) is the maize ortholog of Arabidopsis NRPD1, the Pol IV largest subunit (Erhard et al., 2009). MEDIATOR OF PARAMUTATION 1 (MOP1) is the maize ortholog of Arabidopsis RDR2 (Alleman et al., 2006). RMR1 is a maize SNF2-like protein related to the same family as Arabidopsis CLSY1 and DRD1 proteins (Hale et al., 2009). RMR6 and MOP1 are required for the establishment and maintenance of maize paramutation, whereas RMR1 is only required for maintenance of paramutation. In addition, *rmr6* and *mop1* mutants lose production of 24 nt siRNAs (Erhard et al., 2009; Nobuta et al., 2008), with corresponding hypomethylation and a release of silencing at transposable elements. Interestingly, maize has a heterochromatic, repeat-associated class of 22 nt siRNAs that are unaffected in *mop1* mutants suggesting there may be further specialization of the maize RNA silencing pathways (Nobuta et al., 2008). While the specific mechanism for paramutation has not yet been determined, screens are ongoing to identify and map additional components required. Genes mapped thus far suggest the involvement of a heritable siRNA silencing system working *in trans* with additional RNA silencing machinery expected to be involved.

Flowering and development

The transition from vegetative to reproductive growth and flowering is controlled by endogenous and environmental signals. Longer day length and colder temperatures

are two major environmental stimuli that induce flowering in Arabidopsis. Plants defective in the RdDM pathway are viable and show no obvious morphological phenotypes except for a delay in flowering that is exacerbated when the plants are grown under short day conditions compared to long day conditions (Chan et al., 2004; Liu et al., 2007; Liu et al., 2004; Pontier et al., 2005; Ream et al., 2009). This phenotype can be measured by the number of days till flowering or by the number of rosette leaves at flowering. While this does not have significant consequences for Arabidopsis grown in normal lab conditions, it could have negative consequences for plants growing in the wild or for crops grown for agricultural production, as proper environmental conditions and timing are crucial for successful reproduction.

Networks of genes controlled by epigenetic mechanisms determine flowering time in Arabidopsis. FLOWERING LOCUS C (*FLC*) is a repressor of flowering in Arabidopsis. Cold temperature treatment (vernalization) represses the *FLC* gene by chromatin modifications dependent on Pol IV, RDR2, DCL3 and AGO4 (Liu et al., 2004; Swiezewski et al., 2007). The *FCA* and *FPA* flowering time regulators repress *FLC* expression, thus *fca* and *fpa* mutants are late flowering. *FCA* and *FPA* were also identified in a suppressor screen for transgene silencing and displayed transposon reactivation at some loci. They are hypothesized to be RNA binding proteins that may bind aberrant RNAs and recruit Pol IV (Baurle et al., 2007). The *FLOWERING WAGENINGEN* (*FWA*) locus is another repressor of flowering in Arabidopsis that contains tandem repeats in its promoter with corresponding siRNAs that direct DNA methylation and silencing. The *FWA* locus is controlled by the RdDM pathway as *nrrpd1*,

rdr2, *dcl3*, *ago4*, *nrpe1* and *drm2* mutants release *FWA* silencing and lead to late flowering (Chan et al., 2004; Pontier et al., 2005; Soppe et al., 2000).

Significantly, Pol IV and RDR2 have also been demonstrated to play developmental roles in maize. Both *rmr6* and *mop1* mutants display leaf development defects and problems with sex determination, while *mop1* is also reported to be late flowering (Alleman et al., 2006; Erhard et al., 2009). The possibility exists that with the larger genome size of maize and other crops, Pol IV has adopted additional roles beyond those present in plants with smaller genomes such as Arabidopsis.

There is also a recent link between genomic imprinting and RNA silencing in Arabidopsis as the expression of more than 100,000 Pol IV-derived siRNAs in the developing endosperm are transcribed specifically from thousands of loci on the maternal chromosomes (Mosher et al., 2009). It is proposed that a burst of Pol IV-derived siRNA expression is activated in the female gametophyte and persists in the endosperm with any epigenetic marks responsible for uniparental expression of Pol IV-derived siRNAs in developing seeds lost as the embryo develops into a mature plant. This does not have a negative impact on selfed *npr1* mutants, as they are viable, but is thought to be a possible mechanism of distinguishing self from non-self. A distant hybrid may have essential genes contributed by the pollen that are silenced by the maternally derived Pol IV siRNAs halting embryo development. It may also play a role in hybrid vigor giving rise to phenotypes not observed in either parent.

Abiotic and biotic stress-inducible responses

Plants are limited in the ways they can respond to abiotic and biotic stresses since they are not mobile. In response they have developed complex coping mechanisms that are induced upon stress stimuli. Two of these stress-inducible responses involve Pol IV in the natural antisense RNA (nat-siRNA) pathway (Borsani et al., 2005; Katiyar-Agarwal et al., 2006) and both Pol IV and Pol V in the related long siRNA (l-siRNA) pathway (Katiyar-Agarwal et al., 2007). Each pathway employs a common mechanism that activates the expression of a stress-inducible gene, producing a transcript that overlaps with a constitutively active gene transcribed in the opposite direction. The primary siRNA that results from this bidirectional transcription generates secondary siRNAs that spread into the body of the constitutively expressed transcript. Silencing of the constitutively active gene transcript releases suppression of another gene that in turn activates a stress response within the plant. The potential scope of these pathways is great as there are at least 646 potential Arabidopsis nat-siRNA loci (Jin et al., 2008).

Abiotic stresses encountered by plants include temperature, salt, flood, drought, nutrients and other environmental factors. The nat-siRNA pathway has been characterized by the Arabidopsis salt-stress response (Borsani et al., 2005). *P5CDH* and *SRO5* are convergently transcribed gene pairs with overlapping 3' ends. *P5CDH* is constitutively expressed and upon salt-stress *SRO5* gene expression is induced. 24 nt nat-siRNAs are produced that correspond to the overlapping dsRNA region and are dependent on Pol IV, RDR6, SGS3 and DCL2. Cleavage of the *P5CDH* transcript sets the phase for the production of further 21 nt *P5CDH* nat-siRNAs by DCL1. The down regulation of *P5CDH* leads to decreased proline degradation, which in turn leads to salt tolerance (Borsani et al., 2005). The dependence of Pol IV for the production of 24 nt

nat-siRNAs but not *P5CDH* or *SRO5* transcripts suggests that, in this case, Pol IV may have a DNA-independent role or, alternatively, Pol IV may be recruited to transcribe the DNA by virtue of the overlapping transcripts with the Pol IV transcript being specifically channeled into siRNA production.

Biotic stresses include bacterial and viral pathogenesis and herbivory. The two examples published thus far both involve infection of *Arabidopsis* by *Pseudomonas syringae* (Katiyar-Agarwal et al., 2007; Katiyar-Agarwal et al., 2006). In one case, infection activates *ATGB2* expression causing the production of a 22 nt nat-siRNA that targets the constitutively expressed *PPRL*, a negative regulator of pathogen resistance. The pathway requires Pol IV, RDR6 and SGS3, but unlike the salt-stress response, involves components of the micro RNA pathway (DCL1, HYL1 and HEN1) and leads to the down regulation of *PPRL* transcript (Katiyar-Agarwal et al., 2006). The second example also involves infection of *Arabidopsis* by *Pseudomonas syringae* and the detection of endogenously expressed 39-41 nt l-siRNA that match the overlapping region of the *SRRLK* and *AtRAP* gene pair with eventual *AtRAP* down regulation (Katiyar-Agarwal et al., 2007). The production of l-siRNA requires Pol IV, Pol V, and components of the trans-acting siRNA (ta-siRNA) pathway (DCL1, HYL1, HEN1, HST1, RDR6, DCL4 and AGO7). The ta-siRNA pathway initiates with production of a 21 nt miRNA called a pri-tasiRNA that targets a complementary transcript *in trans* triggering dsRNA production by RDR6 and phased 24 nt ta-siRNAs by DCL4 (Brodersen and Voinnet, 2006). It is not yet known where Pol IV and Pol V act in the pathway or if there is a self-reinforcement loop in place. The nat-siRNA and l-siRNA pathways each use a collection of RNA silencing proteins that do not always act in coordination, raising

many questions with regard to the channeling of substrates in Arabidopsis silencing pathways.

Short- and Long-distance spread of silencing

Pol IV is a required component for both short- and long-range spread of RNA silencing in plants. The two silencing systems differ in that short-range spread occurs in a non-cell-autonomous manner through plasmodesmata in the range of 10-15 cells (Himber et al., 2003), while long-range spread occurs through the phloem between tissues (Voinnet et al., 1998). The genetic requirements of these two systems are not identical but do have some overlap, suggesting modularity of the silencing pathways and their components. Two independent genetic screens have been performed to identify short-range signaling mutants using a phloem-specific promoter that expresses a silencing reporter (Dunoyer et al., 2007; Dunoyer et al., 2005; Smith et al., 2007). Both screens have shown a requirement for NRPD1 and RDR2. In addition, DCL4, DCL1, HEN1, AGO1 and CLSY1 are involved in this process, whereas HYL1, DRB4, DCL3, AGO4, NRPE1 and DRD1 are dispensable. It is unclear whether Pol IV is an upstream and/or downstream component of this pathway. While both 24 nt and 21 nt transgene-specific siRNAs are produced, the DCL4-dependent 21 nt siRNAs are believed to be the short-range RNA mobile signal (Dunoyer et al., 2007; Dunoyer et al., 2005; Smith et al., 2007).

Neither of these two screens are able address whether the identified proteins are required for the production and/or perception of the short-range RNA signal. Other work focusing on the long-distance spread of RNA silencing between tissues has made use of a GFP reporter system and grafting techniques to address this very question (Brosnan et al.,

2007). NRPD1, RDR2, DCL3, AGO4 and RDR6 are each required for the scion (shoots) to respond to an RNA silencing signal originating from the grafted rootstock, but not for signal production. Like short-range silencing, NRPE1 is dispensable for both signal production and perception. This grafting screen has not only provided insight into the proteins required for perception of the mobile signal, but also the amplification of that signal. Pol IV, RDR2, DCL3 and AGO4 produce 24 nt siRNAs corresponding to the 3' end of the silencer present in the rootstock leading to RDR6 and DCL4 production of the predominant 21 nt siRNA class corresponding to sequence 3' of the 24 nt siRNAs. While the proteins required for production of the long-distance RNA signal and the identity of the mobile RNA signal itself are not yet known, the phenomenon does not appear dependent on DCL-cleavage products since *dcl1* and *dcl2,3,4* mutant rootstocks are still silencing-competent (Brosnan et al., 2007). Given the presence of decapped RNA in the scions and the dependence on RDR6, there may be a requirement for intermediate amplification of the signal involving longer RNA species. Alternatively, a siRNA present below detection limits may be responsible for acting as a silencing trigger and setting the subsequent phase. It is hypothesized that Pol IV may be acting in an analogous manner to its role in the nat-siRNA pathway given the overlapping, complementary transcripts that lead to dsRNA production.

SCOPE OF THIS THESIS

The primary focus of my thesis has been the elucidation of Pol IV and Pol V requirements for the RNA-directed DNA methylation pathway and determination of their biochemical activities. Sequences for the Pol IV and Pol V largest and second-largest subunits were discovered with the sequencing of the *Arabidopsis thaliana* genome in the year 2000. I joined the Pikaard lab two years later and by that time Yasuyuki Onodera had generated the basic tools to study the common Pol IV and Pol V second-largest subunit, NRPD2. He had demonstrated that the *nripd2* mutant plants were late flowering with an increased frequency of abnormal floral phenotypes, but NRPD2 could not functionally substitute for the second-largest subunits of Pol I, II or III and NRPD2 column-enriched fractions failed to demonstrate *in vitro* DNA-dependent RNA polymerase activity. This called into question whether NRPD2 was part of a functional RNA polymerase complex and what role it played in the plant.

By the time that I began working with Pol IV and Pol V, the diverse network of proteins involved in plant RNA silencing pathways was just beginning to be discovered. The Pikaard lab engaged in a race with the Baulcombe lab to demonstrate NRPD2 involvement in siRNA production, DNA methylation and heterochromatin formation. It was an exciting time as I was performing phylogenetic analyses of the DNA-dependent RNA polymerases in diverse organisms, characterizing the domain structures of the Pol IV and Pol V largest and second-largest sequences, and genotyping T-DNA accession lines for *nripd1* and *nripe1* mutants. Meanwhile, Tom Ream and Pedro Costa-Nunes were

analyzing the mutants for DNA methylation and siRNA phenotypes relative to previously known players, demonstrating a role for Pol IV in the RNA-directed DNA methylation pathway, as described in Chapter 3.

These findings set the stage for the rest of my thesis work. I focused on deciphering Pol IV and Pol V biochemical activities and determining the requirements of the NRPD1 and NRPE1 C-terminal domains. This required the generation of many Arabidopsis transgenic lines for genetic analyses that could also express epitope-tagged proteins for purification and activity assays. Chapter 2 describes the GATEWAY-compatible plant transformation vectors developed by the Pikaard lab and the contribution I made in determining which epitope tags worked best in Arabidopsis. This knowledge was used by Keith Earley to engineer a collection of vectors that were instrumental in producing a large number of transgenic lines essential to my thesis.

In order to demonstrate Pol IV and Pol V activity, I needed a way to inhibit the activities of the two polymerases for control reactions. It was predicted that α -amanitin would not inhibit Pol IV and Pol V, if indeed they were functional polymerases, so I decided to mutate the invariant Metal A and Metal B sites of the Pol IV and Pol V largest and second-largest subunits, as described in Chapter 4. A failure of the mutated genes to complement *in vivo* would be a good initial indication that the Pol IV and Pol V active sites are functional and these affinity purified proteins could in turn be used as controls alongside affinity purified versions of their wild type counterparts. This work found that the invariant Metal A and Metal B sites were required for Pol IV and Pol V *in vivo* function but not subunit assembly. Performing multiple protein sequence analysis and modeling using the Pol II crystal structure, I was also able to illustrate that the majority of

divergence among Pol IV and Pol V largest and second-largest subunits is concentrated around the active center raising the question of whether they have conserved Pol II mechanistic properties.

Almost three years were spent attempting to obtain Pol IV and Pol V *in vitro* transcription activity using multiple types of DNA and RNA templates of different lengths and combinations. Chapter 5 describes the successful demonstration of Pol IV *in vitro* DNA-dependent RNA polymerase activity using a tripartite oligo template that mimics a stalled open transcription bubble. The Metal A site is required for this activity, as predicted. Using antibodies raised in the lab by Tom Ream and myself, it was also demonstrated that Pol IV physically interacts with RDR2 and that an RNA-dependent RNA polymerase activity observed in Pol IV affinity purified samples is RDR2-dependent. This interaction provides an explanation for how Pol IV transcripts are channeled specifically to RDR2 for dsRNA production. Pol V *in vitro* activity was never obtained but the NRPE1 Metal A site mutant was instrumental in Andrzej Wierzbicki's work identifying Pol V-dependent transcripts *in vivo* and also supports Pol V having DNA-dependent RNA polymerase activity.

I also worked to determine the C-terminal domain (CTD) requirements of the NRPD1 and NRPE1 largest subunits, as described in Chapter 6. Having evolved from Pol II, I hypothesized that the Pol IV and Pol V CTDs may have regulatory roles analogous to the Pol II CTD. I generated a series of twelve Arabidopsis transgenic lines to assess the *in vivo* complementation of various NRPD1 and NRPE1 genomic constructs harboring different CTD deletions. I was able to demonstrate that the Defective Chloroplast and Leaves-like (DeCL) domain at the C-terminus of both NRPD1 and

NRPE1 is required for full complementation, whereas other domains are largely dispensable. The over-expression of individual CTD domains was notably found to dominantly suppress the RNA-directed DNA methylation pathway supporting the hypothesis that the Pol IV and Pol V CTDs have regulatory roles.

References

- (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), 796-815.
- Alleman, M., Sidorenko, L., McGinnis, K., Seshadri, V., Dorweiler, J.E., White, J., Sikkink, K., and Chandler, V.L. (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature*, 442(7100), 295-298.
- Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell*, 42(2), 599-610.
- Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem*, 280(8), 7131-7134.
- Awrey, D.E., Weilbaecher, R.G., Hemming, S.A., Orlicky, S.M., Kane, C.M., and Edwards, A.M. (1997) Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem*, 272(23), 14747-14754.
- Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005) A ratchet mechanism of transcription elongation and its control. *Cell*, 120(2), 183-193.
- Bartolomei, M.S., Halden, N.F., Cullen, C.R., and Corden, J.L. (1988) Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Mol Cell Biol*, 8(1), 330-339.
- Baumberger, N., and Baulcombe, D.C. (2005) *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A*, 102(33), 11928-11933.
- Baurle, I., Smith, L., Baulcombe, D.C., and Dean, C. (2007) Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. *Science*, 318(5847), 109-112.
- Bellaoui, M., and Gruissem, W. (2004) Altered expression of the *Arabidopsis* ortholog of DCL affects normal plant development. *Planta*, 219(5), 819-826.
- Bellaoui, M., Keddie, J.S., and Gruissem, W. (2003) DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. *Plant Mol Biol*, 53(4), 531-543.
- Bergsland, K.J., and Haselkorn, R. (1991) Evolutionary relationships among eubacteria, cyanobacteria, and chloroplasts: evidence from the *rpoC1* gene of *Anabaena* sp. strain PCC 7120. *J Bacteriol*, 173(11), 3446-3455.
- Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R., and Lagrange, T. (2009) RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep*, 10(6), 649-654.
- Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell*, 123(7), 1279-1291.
- Brodersen, P., and Voinnet, O. (2006) The diversity of RNA silencing pathways in plants. *Trends Genet*, 22(5), 268-280.

- Brosnan, C.A., Mitter, N., Christie, M., Smith, N.A., Waterhouse, P.M., and Carroll, B.J. (2007) Nuclear gene silencing directs reception of long-distance mRNA silencing in Arabidopsis. *Proc Natl Acad Sci U S A*, 104(37), 14741-14746.
- Brueckner, F., and Cramer, P. (2008) Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat Struct Mol Biol*, 15(8), 811-818.
- Bushnell, D.A., Cramer, P., and Kornberg, R.D. (2002) Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 Å resolution. *Proc Natl Acad Sci U S A*, 99(3), 1218-1222.
- Cao, X., and Jacobsen, S.E. (2002) Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol*, 12(13), 1138-1144.
- Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004) RNA silencing genes control de novo DNA methylation. *Science*, 303(5662), 1336.
- Chandler, V.L., Eggleston, W.B., and Dorweiler, J.E. (2000) Paramutation in maize. *Plant Mol Biol*, 43(2-3), 121-145.
- Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev*, 11(24), 3319-3326.
- Cho, S.H., Addo-Quaye, C., Coruh, C., Arif, M.A., Ma, Z., Frank, W., and Axtell, M.J. (2008) *Physcomitrella patens* DCL3 is required for 22-24 nt siRNA accumulation, suppression of retrotransposon-derived transcripts, and normal development. *PLoS Genet*, 4(12), e1000314.
- Choder, M., and Young, R.A. (1993) A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. *Mol Cell Biol*, 13(11), 6984-6991.
- Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, 452(7184), 215-219.
- Comer, F.I., and Hart, G.W. (2001) Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry*, 40(26), 7845-7852.
- Corden, J.L., Cadena, D.L., Ahearn, J.M., Jr., and Dahmus, M.E. (1985) A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A*, 82(23), 7934-7938.
- Cramer, P. (2006) Recent structural studies of RNA polymerases II and III. *Biochem Soc Trans*, 34(Pt 6), 1058-1061.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science*, 292(5523), 1863-1876.
- Curaba, J., and Chen, X. (2008) Biochemical activities of Arabidopsis RNA-dependent RNA polymerase 6. *J Biol Chem*, 283(6), 3059-3066.
- Curtin, S.J., Watson, J.M., Smith, N.A., Eamens, A.L., Blanchard, C.L., and Waterhouse, P.M. (2008) The roles of plant dsRNA-binding proteins in RNAi-like pathways. *FEBS Lett*, 582(18), 2753-2760.

- Dalmay, T., Hamilton, A., Mueller, E., and Baulcombe, D.C. (2000) Potato virus X amplicons in Arabidopsis mediate genetic and epigenetic gene silencing. *Plant Cell*, 12(3), 369-379.
- Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.A. (2002) Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol Cell*, 10(2), 409-415.
- Dieci, G., Hermann-Le Denmat, S., Lukhtanov, E., Thuriaux, P., Werner, M., and Sentenac, A. (1995) A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *EMBO J*, 14(15), 3766-3776.
- Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005) RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev*, 19(19), 2301-2306.
- Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A., and Voinnet, O. (2007) Intra- and intercellular RNA interference in *Arabidopsis thaliana* requires components of the microRNA and heterochromatic silencing pathways. *Nat Genet*, 39(7), 848-856.
- Dunoyer, P., Himber, C., and Voinnet, O. (2005) DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nat Genet*, 37(12), 1356-1360.
- Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem*, 266(1), 71-75.
- Egloff, S., and Murphy, S. (2008) Cracking the RNA polymerase II CTD code. *Trends Genet*, 24(6), 280-288.
- Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, 318(5857), 1777-1779.
- El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev*, 21(20), 2539-2544.
- Erhard, K.F., Jr., Stonaker, J.L., Parkinson, S.E., Lim, J.P., Hale, C.J., and Hollick, J.B. (2009) RNA polymerase IV functions in paramutation in *Zea mays*. *Science*, 323(5918), 1201-1205.
- Farago, M., Nahari, T., Hammel, C., Cole, C.N., and Choder, M. (2003) Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. *Mol Biol Cell*, 14(7), 2744-2755.
- Furter-Graves, E.M., Hall, B.D., and Furter, R. (1994) Role of a small RNA pol II subunit in TATA to transcription start site spacing. *Nucleic Acids Res*, 22(23), 4932-4936.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*, 292(5523), 1876-1882.
- Goler-Baron, V., Selitrennik, M., Barkai, O., Haimovich, G., Lotan, R., and Choder, M. (2008) Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes Dev*, 22(15), 2022-2027.

- Haag, J.R., Pontes, O., and Pikaard, C.S. (2009) Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS One*, 4(1), e4110.
- Hale, C.J., Erhard, K.F., Jr., Lisch, D., and Hollick, J.B. (2009) Production and processing of siRNA precursor transcripts from the highly repetitive maize genome. *PLoS Genet*, 5(8), e1000598.
- Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002) Two classes of short interfering RNA in RNA silencing. *EMBO J*, 21(17), 4671-4679.
- He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S., and Zhu, J.K. (2009a) NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes Dev*, 23(3), 318-330.
- He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H., and Zhu, J.K. (2009b) An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell*, 137(3), 498-508.
- Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet*, 38(6), 721-725.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science*, 308(5718), 118-120.
- Himber, C., Dunoyer, P., Moissiard, G., Ritzenthaler, C., and Voinnet, O. (2003) Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *EMBO J*, 22(17), 4523-4533.
- Huang, L., Jones, A.M., Searle, I., Patel, K., Vogler, H., Hubner, N.C., and Baulcombe, D.C. (2009) An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol*, 16(1), 91-93.
- Hull, M.W., McKune, K., and Woychik, N.A. (1995) RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes Dev*, 9(4), 481-490.
- Ishihama, A. (1981) Subunit of assembly of Escherichia coli RNA polymerase. *Adv Biophys*, 14, 1-35.
- Iyer, L.M., Koonin, E.V., and Aravind, L. (2003) Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct Biol*, 3, 1.
- Jacob, S.T., Sajdel, E.M., and Munro, H.N. (1970) Specific action of alpha-amanitin on mammalian RNA polymerase protein. *Nature*, 225(5227), 60-62.
- Jacobs, E.Y., Ogiwara, I., and Weiner, A.M. (2004) Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing. *Mol Cell Biol*, 24(2), 846-855.
- Jin, H., Vacic, V., Girke, T., Lonardi, S., and Zhu, J.K. (2008) Small RNAs and the regulation of cis-natural antisense transcripts in Arabidopsis. *BMC Mol Biol*, 9, 6.
- Jokerst, R.S., Weeks, J.R., Zehring, W.A., and Greenleaf, A.L. (1989) Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosophila. *Mol Genet*, 215(2), 266-275.

- Joyce, C.M., and Steitz, T.A. (1995) Polymerase structures and function: variations on a theme? *J Bacteriol*, 177(22), 6321-6329.
- Kanno, T., Aufsatz, W., Jaligot, E., Mette, M.F., Matzke, M., and Matzke, A.J. (2005a) A SNF2-like protein facilitates dynamic control of DNA methylation. *EMBO Rep*, 6(7), 649-655.
- Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Bohmdorfer, G., Gregor, W., Kreil, D.P., Matzke, M., and Matzke, A.J. (2008) A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat Genet*, 40(5), 670-675.
- Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005b) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet*, 37(7), 761-765.
- Kaplan, C.D., Larsson, K.M., and Kornberg, R.D. (2008) The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol Cell*, 30(5), 547-556.
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol*, 5(3), e57.
- Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., and Jin, H. (2007) A novel class of bacteria-induced small RNAs in Arabidopsis. *Genes Dev*, 21(23), 3123-3134.
- Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Jr., Zhu, J.K., Staskawicz, B.J., and Jin, H. (2006) A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci U S A*, 103(47), 18002-18007.
- Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005) RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science*, 309(5733), 467-469.
- Keddie, J.S., Carroll, B., Jones, J.D., and Gruissem, W. (1996) The DCL gene of tomato is required for chloroplast development and palisade cell morphogenesis in leaves. *EMBO J*, 15(16), 4208-4217.
- Kedinger, C., Gniazdowski, M., Mandel, J.L., Jr., Gissinger, F., and Chambon, P. (1970) Alpha-amanitin: a specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus. *Biochem Biophys Res Commun*, 38(1), 165-171.
- Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993) RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem*, 268(14), 10416-10424.
- Kimura, M., Ishiguro, A., and Ishihama, A. (1997) RNA polymerase II subunits 2, 3, and 11 form a core subassembly with DNA binding activity. *J Biol Chem*, 272(41), 25851-25855.
- Kimura, M., Sakurai, H., and Ishihama, A. (2001) Intracellular contents and assembly states of all 12 subunits of the RNA polymerase II in the fission yeast *Schizosaccharomyces pombe*. *Eur J Biochem*, 268(3), 612-619.
- Kireeva, M.L., Nedialkov, Y.A., Cremona, G.H., Purtov, Y.A., Lubkowska, L., Malagon, F., Burton, Z.F., Strathern, J.N., and Kashlev, M. (2008) Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol Cell*, 30(5), 557-566.

- Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007) Functional architecture of RNA polymerase I. *Cell*, 131(7), 1260-1272.
- Lahmy, S., Guilleminot, J., Cheng, C.M., Bechtold, N., Albert, S., Pelletier, G., Delseny, M., and Devic, M. (2004) DOMINO1, a member of a small plant-specific gene family, encodes a protein essential for nuclear and nucleolar functions. *Plant J*, 39(6), 809-820.
- Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T., and Lagrange, T. (2009) PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. *Proc Natl Acad Sci U S A*, 106(3), 941-946.
- Lai, M.M. (2005) RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis delta virus. *J Virol*, 79(13), 7951-7958.
- Landick, R. (2009) Functional divergence in the growing family of RNA polymerases. *Structure*, 17(3), 323-325.
- Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006) A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J*, 25(1), 118-128.
- Larkin, R.M., and Guilfoyle, T.J. (1998) Two small subunits in Arabidopsis RNA polymerase II are related to yeast RPB4 and RPB7 and interact with one another. *J Biol Chem*, 273(10), 5631-5637.
- Larkin, R.M., Hagen, G., and Guilfoyle, T.J. (1999) Arabidopsis thaliana RNA polymerase II subunits related to yeast and human RPB5. *Gene*, 231(1-2), 41-47.
- Lehmann, E., Brueckner, F., and Cramer, P. (2007) Molecular basis of RNA-dependent RNA polymerase II activity. *Nature*, 450(7168), 445-449.
- Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. *Cell*, 126(1), 93-106.
- Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr Biol*, 15(16), 1501-1507.
- Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell*, 9(5), 1101-1111.
- Lindell, T.J., Weinberg, F., Morris, P.W., Roeder, R.G., and Rutter, W.J. (1970) Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science*, 170(956), 447-449.
- Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., 3rd, and Hartzog, G.A. (2003) Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol Cell Biol*, 23(4), 1368-1378.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, 133(3), 523-536.

- Liu, F., Quesada, V., Crevillen, P., Baurle, I., Swiezewski, S., and Dean, C. (2007) The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. *Mol Cell*, 28(3), 398-407.
- Liu, J., He, Y., Amasino, R., and Chen, X. (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. *Genes Dev*, 18(23), 2873-2878.
- Liu, J., Rivas, F.V., Wohlschlegel, J., Yates, J.R., 3rd, Parker, R., and Hannon, G.J. (2005) A role for the P-body component GW182 in microRNA function. *Nat Cell Biol*, 7(12), 1261-1266.
- Lotan, R., Bar-On, V.G., Harel-Sharvit, L., Duek, L., Melamed, D., and Choder, M. (2005) The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes Dev*, 19(24), 3004-3016.
- Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G., and Choder, M. (2007) The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *J Cell Biol*, 178(7), 1133-1143.
- Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., and Meyers, B.C. (2006) MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res*, 16(10), 1276-1288.
- Luo, J., and Hall, B.D. (2007) A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol*, 64(1), 101-112.
- Maillet, I., Buhler, J.M., Sentenac, A., and Labarre, J. (1999) Rpb4p is necessary for RNA polymerase II activity at high temperature. *J Biol Chem*, 274(32), 22586-22590.
- Martin, C., Okamura, S., and Young, R. (1990) Genetic exploration of interactive domains in RNA polymerase II subunits. *Mol Cell Biol*, 10(5), 1908-1914.
- Max, T., Sogaard, M., and Svejstrup, J.Q. (2007) Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J Biol Chem*, 282(19), 14113-14120.
- McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev*, 11(24), 3306-3318.
- Meinhart, A., and Cramer, P. (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, 430(6996), 223-226.
- Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G.J., and Qi, Y. (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, 133(1), 116-127.
- Mitsuzawa, H., Kanda, E., and Ishihama, A. (2003) Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res*, 31(16), 4696-4701.
- Miyao, T., Barnett, J.D., and Woychik, N.A. (2001) Deletion of the RNA polymerase subunit RPB4 acts as a global, not stress-specific, shut-off switch for RNA polymerase II transcription at high temperatures. *J Biol Chem*, 276(49), 46408-46413.

- Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell*, 133(1), 128-141.
- Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., and Baulcombe, D.C. (2009) Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. *Nature*, 460(7252), 283-286.
- Myers, L.C., Gustafsson, C.M., Bushnell, D.A., Lui, M., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (1998) The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev*, 12(1), 45-54.
- Nobuta, K., Lu, C., Shrivastava, R., Pillay, M., De Paoli, E., Accerbi, M., Arteaga-Vazquez, M., Sidorenko, L., Jeong, D.H., Yen, Y., Green, P.J., Chandler, V.L., and Meyers, B.C. (2008) Distinct size distribution of endogenous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. *Proc Natl Acad Sci U S A*, 105(39), 14958-14963.
- Nonet, M., Sweetser, D., and Young, R.A. (1987) Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell*, 50(6), 909-915.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, 120(5), 613-622.
- Orlicky, S.M., Tran, P.T., Sayre, M.H., and Edwards, A.M. (2001) Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J Biol Chem*, 276(13), 10097-10102.
- Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol*, 12(17), 1484-1495.
- Patel, D.D., and Pickup, D.J. (1989) The second-largest subunit of the poxvirus RNA polymerase is similar to the corresponding subunits of procaryotic and eucaryotic RNA polymerases. *J Virol*, 63(3), 1076-1086.
- Payne, J.M., Laybourn, P.J., and Dahmus, M.E. (1989) The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit IIa. *J Biol Chem*, 264(33), 19621-19629.
- Phatnani, H.P., and Greenleaf, A.L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev*, 20(21), 2922-2936.
- Pikaard, C.S., Haag, J.R., Ream, T., and Wierzbicki, A.T. (2008) Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci*, 13(7), 390-397.
- Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell*, 126(1), 79-92.
- Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev*, 19(17), 2030-2040.

- Prelich, G. (2002) RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. *Eukaryot Cell*, 1(2), 153-162.
- Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.P., Lottspeich, F., Garrett, R.A., and Zillig, W. (1989) Archaeobacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc Natl Acad Sci U S A*, 86(12), 4569-4573.
- Qi, Y., Denli, A.M., and Hannon, G.J. (2005) Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol Cell*, 19(3), 421-428.
- Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*, 443(7114), 1008-1012.
- Rackwitz, H.R., Rohde, W., and Sanger, H.L. (1981) DNA-dependent RNA polymerase II of plant origin transcribes viroid RNA into full-length copies. *Nature*, 291(5813), 297-301.
- Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. (2009) Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell*, 33(2), 192-203.
- Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol*, 12(4), 340-349.
- Saez-Vasquez, J., and Pikaard, C.S. (1997) Extensive purification of a putative RNA polymerase I holoenzyme from plants that accurately initiates rRNA gene transcription in vitro. *Proc Natl Acad Sci U S A*, 94(22), 11869-11874.
- Scafe, C., Martin, C., Nonet, M., Podos, S., Okamura, S., and Young, R.A. (1990) Conditional mutations occur predominantly in highly conserved residues of RNA polymerase II subunits. *Mol Cell Biol*, 10(3), 1270-1275.
- Schneider, G.J., Tumer, N.E., Richaud, C., Borbely, G., and Haselkorn, R. (1987) Purification and characterization of RNA polymerase from the cyanobacterium *Anabaena* 7120. *J Biol Chem*, 262(30), 14633-14639.
- Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 115(2), 199-208.
- Selitrennik, M., Duek, L., Lotan, R., and Choder, M. (2006) Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of *Saccharomyces cerevisiae* RNA polymerase II by two pathways. *Eukaryot Cell*, 5(12), 2092-2103.
- Sheffer, A., Varon, M., and Choder, M. (1999) Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4. *Mol Cell Biol*, 19(4), 2672-2680.
- Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C. (2007) An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell*, 19(5), 1507-1521.
- Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, 305(5689), 1434-1437.

- Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000) The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell*, 6(4), 791-802.
- Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. (2003) Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J*, 22(9), 2234-2244.
- Sosunov, V., Zorov, S., Sosunova, E., Nikolaev, A., Zakeyeva, I., Bass, I., Goldfarb, A., Nikiforov, V., Severinov, K., and Mustaev, A. (2005) The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. *Nucleic Acids Res*, 33(13), 4202-4211.
- Steitz, T.A. (1998) A mechanism for all polymerases. *Nature*, 391(6664), 231-232.
- Svetlov, V., Vassilyev, D.G., and Artsimovitch, I. (2004) Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. *J Biol Chem*, 279(37), 38087-38090.
- Sweetser, D., Nonet, M., and Young, R.A. (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci U S A*, 84(5), 1192-1196.
- Swiezewski, S., Crevillen, P., Liu, F., Ecker, J.R., Jerzmanowski, A., and Dean, C. (2007) Small RNA-mediated chromatin silencing directed to the 3' region of the Arabidopsis gene encoding the developmental regulator, FLC. *Proc Natl Acad Sci U S A*, 104(9), 3633-3638.
- Takimoto, K., Wakiyama, M., and Yokoyama, S. (2009) Mammalian GW182 contains multiple Argonaute-binding sites and functions in microRNA-mediated translational repression. *RNA*, 15(6), 1078-1089.
- Tan, Q., Li, X., Sadhale, P.P., Miyao, T., and Woychik, N.A. (2000) Multiple mechanisms of suppression circumvent transcription defects in an RNA polymerase mutant. *Mol Cell Biol*, 20(21), 8124-8133.
- Taylor, J.M. (2003) Replication of human hepatitis delta virus: recent developments. *Trends Microbiol*, 11(4), 185-190.
- Touloukhonov, I., Zhang, J., Palangat, M., and Landick, R. (2007) A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol Cell*, 27(3), 406-419.
- Ujvari, A., and Luse, D.S. (2006) RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol*, 13(1), 49-54.
- Ulmasov, T., Larkin, R.M., and Guilfoyle, T.J. (1996) Association between 36- and 13.6-kDa alpha-like subunits of Arabidopsis thaliana RNA polymerase II. *J Biol Chem*, 271(9), 5085-5094.
- Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D., and Aloni, Y. (1992) Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. *Cell*, 69(5), 871-881.
- Vaucheret, H. (2008) Plant ARGONAUTES. *Trends Plant Sci*, 13(7), 350-358.
- Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science*, 303(5658), 672-676.

- Voinnet, O., Vain, P., Angell, S., and Baulcombe, D.C. (1998) Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell*, 95(2), 177-187.
- Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*, 127(5), 941-954.
- Wang, Z., and Roeder, R.G. (1997) Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. *Genes Dev*, 11(10), 1315-1326.
- Weil, P.A., and Blatti, S.P. (1975) Partial purification and properties of calf thymus deoxyribonucleic acid dependent RNA polymerase III. *Biochemistry*, 14(8), 1636-1642.
- Werner, F. (2007) Structure and function of archaeal RNA polymerases. *Mol Microbiol*, 65(6), 1395-1404.
- Werner, F., and Weinzierl, R.O. (2002) A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell*, 10(3), 635-646.
- Werner, M., Hermann-Le Denmat, S., Treich, I., Sentenac, A., and Thuriaux, P. (1992) Effect of mutations in a zinc-binding domain of yeast RNA polymerase C (III) on enzyme function and subunit association. *Mol Cell Biol*, 12(3), 1087-1095.
- Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*, 119(4), 481-489.
- Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*, 135(4), 635-648.
- Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet*, 41(5), 630-634.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol*, 2(5), E104.
- Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. (2006) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res*, 34(2), 667-675.
- Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007) The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev*, 21(2), 160-174.
- Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science*, 307(5711), 932-935.
- Zaychikov, E., Martin, E., Denissova, L., Kozlov, M., Markovtsov, V., Kashlev, M., Heumann, H., Nikiforov, V., Goldfarb, A., and Mustaev, A. (1996) Mapping of catalytic residues in the RNA polymerase active center. *Science*, 273(5271), 107-109.
- Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*, 98(6), 811-824.

- Zhang, J., and Corden, J.L. (1991) Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J Biol Chem*, 266(4), 2297-2302.
- Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. (2007) Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J*, 26(6), 1691-1701.
- Zilberman, D., Cao, X., and Jacobsen, S.E. (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, 299(5607), 716-719.
- Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004) Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr Biol*, 14(13), 1214-1220.

CHAPTER 2

GATEWAY-COMPATIBLE VECTORS FOR PLANT FUNCTIONAL GENOMICS
AND PROTEOMICS

Published in *The Plant Journal* (2006) 45 (4): 616-629.

My contributions to this work:

I performed the groundwork for this project while rotating with Sigma's Plant Biotechnology group. I tested the detection of several popular epitope tags by spiking tagged recombinant proteins into the total protein extracts of several plant species and performed Western blot analysis (Figure 3). We found that some plant species have endogenous proteins that cross-react with popular commercially available antibodies making those epitopes less than desirable for use in those plants. The results were in turn used by Keith Earley to generate GATEWAY-compatible plant transformation vectors utilizing those epitope tags demonstrated to work best in a broad panel of plant species. These plant transformation tagging vectors were invaluable for the genetic and biochemical analyses performed during the course of my thesis research. I also helped write and edit the manuscript.

TECHNIQUES FOR MOLECULAR ANALYSIS

Gateway-compatible vectors for plant functional genomics and proteomics

Keith W. Earley^{1,*}, Jeremy R. Haag¹, Olga Pontes¹, Kristen Opper², Tom Juehne², Keming Song² and Craig S. Pikaard¹

¹Biology Department, Washington University, 1 Brookings Drive, St Louis, MO 63130, USA, and

²Plant Biotechnology Group, Sigma-Aldrich Company, 2909 Laclede Avenue, St Louis, MO 63103, USA

Received 20 May 2005; revised 23 August 2005; accepted 3 October 2005.

*For correspondence (fax +1 314 935 4432; e-mail kwearley@artsci.wustl.edu).

Summary

Gateway cloning technology facilitates high-throughput cloning of target sequences by making use of the bacteriophage lambda site-specific recombination system. Target sequences are first captured in a commercially available 'entry vector' and are then recombined into various 'destination vectors' for expression in different experimental organisms. Gateway technology has been embraced by a number of plant laboratories that have engineered destination vectors for promoter specificity analyses, protein localization studies, protein/protein interaction studies, constitutive or inducible protein expression studies, gene knockdown by RNA interference, or affinity purification experiments. We review the various types of Gateway destination vectors that are currently available to the plant research community and provide links and references to enable additional information to be obtained concerning these vectors. We also describe a set of 'pEarleyGate' plasmid vectors for *Agrobacterium*-mediated plant transformation that translationally fuse FLAG, HA, cMyc, AcV5 or tandem affinity purification epitope tags onto target proteins, with or without an adjacent fluorescent protein. The oligopeptide epitope tags allow the affinity purification, immunolocalization or immunoprecipitation of recombinant proteins expressed *in vivo*. We demonstrate the utility of pEarleyGate destination vectors for the expression of epitope-tagged proteins that can be affinity captured or localized by immunofluorescence microscopy. Antibodies detecting the FLAG, HA, cMyc and AcV5 tags show relatively little cross-reaction with endogenous proteins in a variety of monocotyledonous and dicotyledonous plants, suggesting broad utility for the tags and vectors.

Keywords: affinity purification, epitope tag, fusion protein, protein localization, recombinational cloning.

Introduction

Moving beyond gene discovery to understanding gene function is facilitated by the ability to easily express proteins from cloned genes in both homologous and non-homologous biological contexts. For instance, expression in plants of a protein engineered to include an oligopeptide epitope tag can allow affinity purification or immunoprecipitation of that protein and any associated proteins (Fritze and Anderson, 2000; Jarvik and Telmer, 1998). This can be an extremely useful approach for the isolation, identification and biochemical analysis of multi-protein complexes. Similarly, fusing an open reading frame to a fluorescent protein, such as green, yellow, red or cyan fluorescent proteins (GFP, YFP, RFP or CFP, respectively), can be useful for determining the

subcellular localization of a protein and for testing for interactions with other fluorescently tagged proteins within living cells (Ehrhardt, 2003; Hanson and Kohler, 2001; Haseloff, 1999; Stewart, 2001). A researcher might also find it useful to express a target protein in *Escherichia coli* or insect cells in order to test for enzymatic activities, to produce sufficient recombinant protein for raising antibodies, or to perform protein interaction studies. Engineering multiple expression vector constructs to accomplish these goals for every target gene of interest using traditional ligase-mediated cloning is time-consuming and laborious, posing a technical barrier for high-throughput functional genomics or proteomics projects. Fortunately, such barriers have been

lowered considerably by the advent of Gateway cloning technology (Hartley *et al.*, 2000).

Gateway cloning exploits the bacteriophage lambda recombination system, thereby bypassing the need for traditional ligase-mediated cloning. Once captured in a Gateway-compatible plasmid 'entry vector', an open reading frame or gene flanked by recombination sites can be recombined into a variety of 'destination vectors' that possess compatible recombination sites. Destination vectors for protein expression in *E. coli*, yeast, mammalian, and insect cells are commercially available and are marketed by Invitrogen (Carlsbad, CA, USA). Although Gateway-compatible plant destination vectors for expression of proteins in transgenic plants are not commercially available at the present time, a number of laboratories have engineered such vectors (Table 1; Figure 1). These plant destination vectors have been designed for a variety of specific purposes including protein localization, promoter functional analysis, gene overexpression, gene knockdown by RNA interference, production of epitope-tagged proteins for affinity purification, or analysis of protein/protein interactions using

fluorescence resonance energy transfer (FRET), bioluminescence resonance energy transfer (BRET) or bimolecular fluorescence complementation (BiFC).

In addition to reviewing previously described Gateway-compatible plant destination vectors, we describe a series of pEarleyGate vectors that we designed for transient or stable expression of proteins fused to a variety of oligopeptide epitope tags and/or GFP, YFP or CFP. Representative immunoblotting, affinity purification and protein localization data are provided in order to illustrate the usefulness of pEarleyGate vectors.

Gateway cloning

The Gateway cloning system exploits the accurate, site-specific recombination system utilized by bacteriophage lambda in order to shuttle sequences between plasmids bearing compatible recombination sites (Figure 2). In the Pikaard laboratory, the preferred method for initially capturing sequences of interest is to use topoisomerase-mediated cloning (Shuman, 1994), which eliminates the

Table 1 Gateway compatible plant destination vectors

References	Uses for vectors	Reporter genes/tags	Website
Karimi <i>et al.</i> (2002)	Promoter analysis Inducible expression Protein localization RNAi	GUS, GFP, YFP, CFP, Luciferase	http://www.psb.ugent.be/gateway/
Helliwell and Waterhouse (2003)	RNAi		http://www.pi.csiro.au/rnai/hithroughput.htm
Curtis and Grossniklaus (2003)	Promoter analysis Inducible expression Protein localization RNAi	GFP, GUS, His	http://www.unizh.ch/botinst/devo_website/curtisvector/
Joubes <i>et al.</i> (2004)	Inducible expression		http://www.psb.ugent.be/gateway/
Bensmihen <i>et al.</i> (2004)	Epitope tagging Activation domain addition	HA, VP16	http://www.isv.cnrs-gif.fr/jg/alligator/vectors.html
Rohila <i>et al.</i> (2004)	TAP protein purification	Protein A IgG binding domain, calmodulin	
Walter <i>et al.</i> (2004)	BiFC	Truncated C- and N-termini of YFP for BiFC	
Lo <i>et al.</i> (2005)	Inducible RNAi		
Rubio <i>et al.</i> (2005)	TAP protein purification	Protein A IgG binding domain, cMyc-His	
Tzfira <i>et al.</i> (2005)	Protein localization	GFP	
Karimi <i>et al.</i> (2005)	Multicomponent recombination		http://www.psb.ugent.be/gateway/
Albrecht von Arnim (University of Tennessee, Knoxville, TN, USA, personal communication)	BRET	Luciferase, YFP	http://www.bio.utk.edu/vonarnim/BRET/BRET-vectors.html
This article	Protein localization Affinity purification Immunolocalization	HA, FLAG, cMyc, AcV5, TAP, His, GFP, YFP, CFP	http://www.biology.wustl.edu/pikaard/pearleygate%20plasmid%20vectors/pearleygate%20homepage.html

BiFC, bimolecular fluorescence complementation; BRET, bioluminescence resonance energy transfer; GFP, YFP and CFP, green, yellow and cyan fluorescent proteins, respectively; RNAi, RNA interference; TAP, tandem affinity purification; His, histidine; HA, cMyc, FLAG and AcV5 are epitope tags (see page 11 for sequences).

Gateway vectors available for plant expression

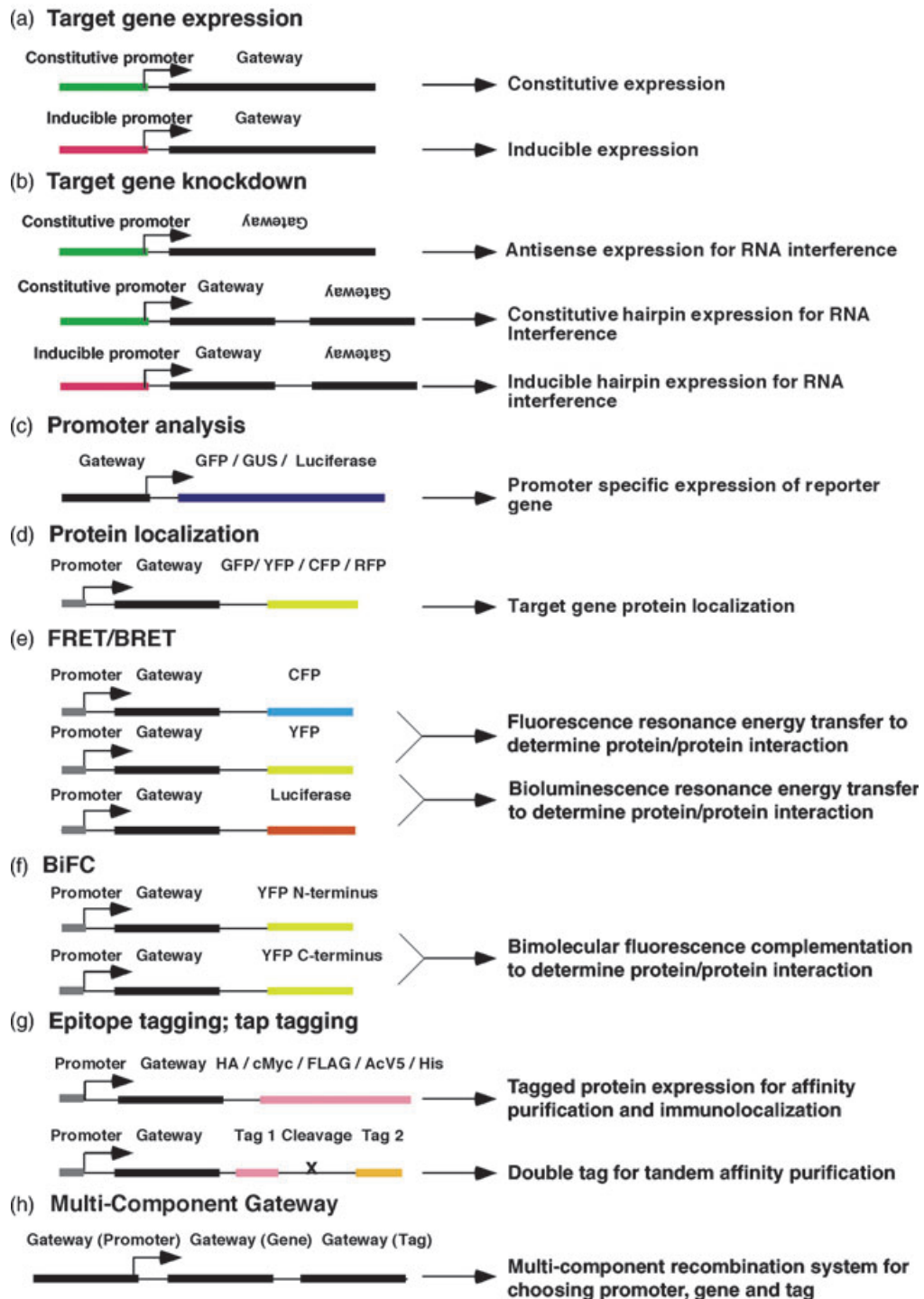
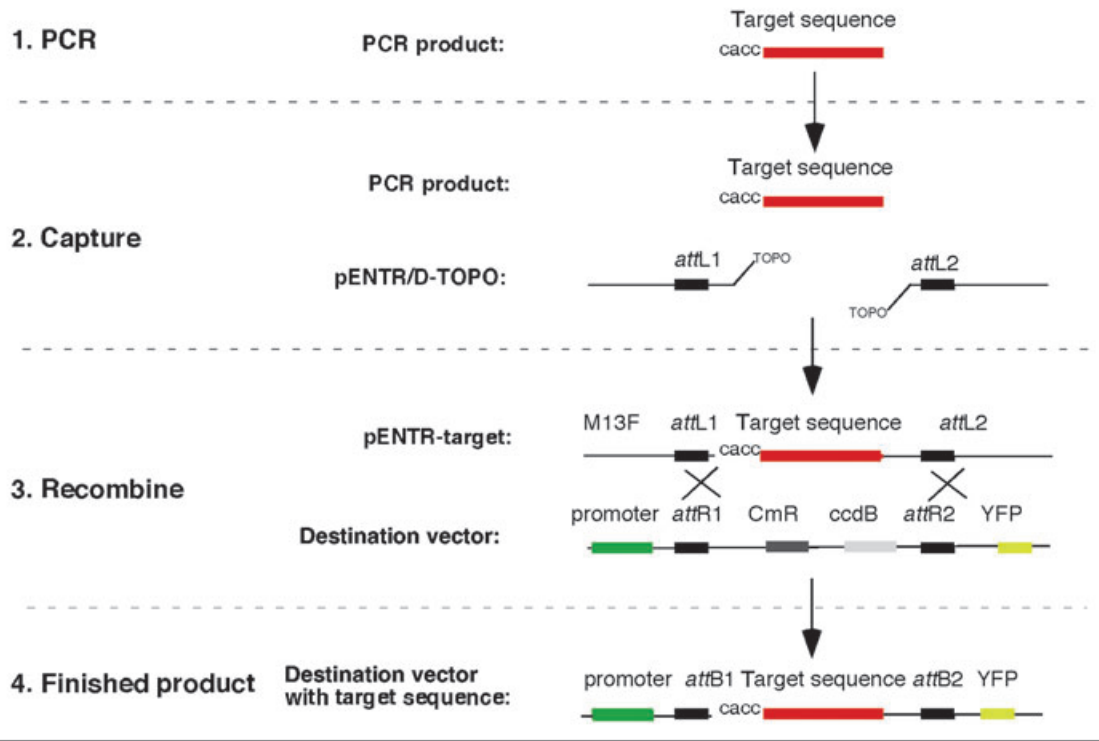


Figure 1. A summary of available Gateway-compatible vectors for use in plants. Diagrams illustrate Gateway-compatible vectors for (a) protein overexpression, (b) RNA knockdown, (c) promoter analysis, (d) protein subcellular localization, (e) fluorescence resonance energy transfer and bioluminescence resonance energy transfer, (f) bimolecular fluorescence complementation, (g) epitope tagging and tandem affinity purification, and (h) multi-component transgene assembly. All vectors contain *attR* recombination sites and a *ccdB* cassette for selection of successful recombination events. Only C-terminal fusions are illustrated in this figure but, for most constructs, N-terminal constructs are also available. Table 1 provides links by which more detailed information concerning available vectors can be obtained.

need for conventional DNA ligase-mediated molecular cloning. In this approach, one uses polymerase chain reaction (PCR) to amplify the target sequence using a forward primer that includes the sequence CACC at the 5' end.

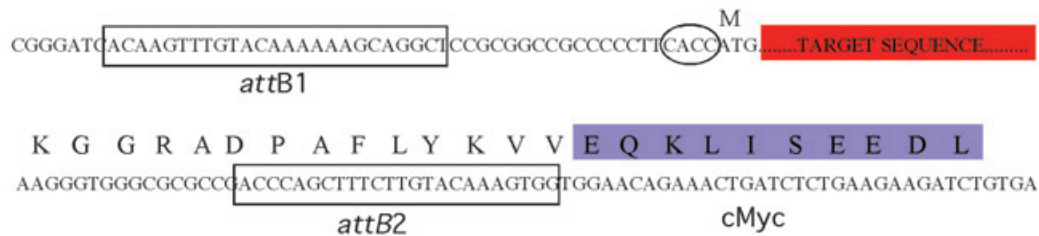
This sequence facilitates directional incorporation into Invitrogen's pENTR/D-TOPO entry vector (Figure 2a, steps 1 and 2). The resulting recombinant plasmid has the target DNA sequences flanked by *attL* recombination sequences.

(a) **Topoisomerase-mediated capture and Gateway recombination**

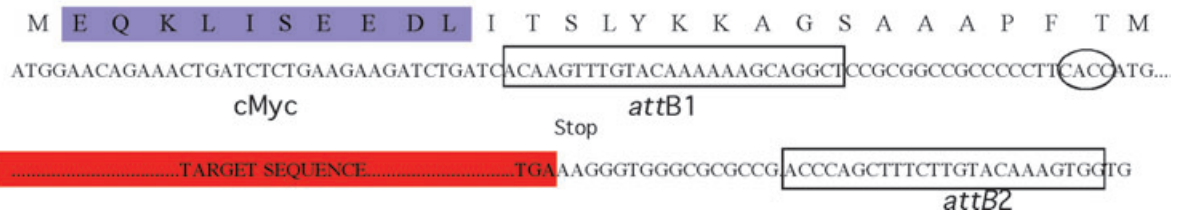


(b) **Examples of Gateway-mediated epitope tag fusions**

C-Terminal c-Myc Fusion (pEarleyGate 303):



N-Terminal c-Myc Fusion (pEarleyGate 203):



Once flanked by *attL* recombination sites, the sequence can be recombined with *attR* sites using the LR clonase reaction mix (Invitrogen). This reaction transfers the target sequence into a desired destination vector (Figure 2a, steps 3 and 4). Destination vectors contain a gene (*ccdB*) that is lethal to most strains of *E. coli*. 'Empty' destination vectors are therefore selected against upon transformation of *E. coli* cells with the recombination reaction. This negative selection, combined with positive selection for an antibiotic resistance marker, ensures that resulting colonies contain plasmids that have undergone recombination. The ease and speed with which a captured target sequence can be shuttled simultaneously into a variety of destination vectors are great advantages for high-throughput functional genomics/proteomics investigations.

Although we use topoisomerase-mediated cloning almost exclusively for capturing target sequences in entry vectors, there are other options. One option is to use traditional ligase-mediated insertion of a target sequence into an entry vector at a multiple cloning site that is flanked by *attL* sites. A second option is to use PCR primers that include *attB* sites when amplifying the target sequence. The resulting PCR products can be recombined directly into a donor vector containing *attP* recombination sites using the BP clonase reaction mix (Invitrogen). This BP recombination reaction results in the target sequence being flanked by *attL* sequences, which allows subsequent recombination with a destination vector. These options, as well as detailed protocols, are described in the Gateway cloning manual(s) available from Invitrogen's website (<http://www.invitrogen.com>).

Gateway-compatible destination vectors for use in plants

A number of laboratories have developed Gateway-compatible plant expression vectors in recent years, each designed with a specific purpose in mind (Table 1; Figure 1). Many of these plasmid vectors can replicate in both *E. coli* and *Agrobacterium tumefaciens* and possess left border and right border sequences for *Agrobacterium*-mediated T-DNA transfer. The different types of vectors, their key features and

uses, URLs for websites where more information can be obtained, and pertinent references are summarized in Table 1. In some cases, the vectors can only be obtained by interested researchers through a Materials Transfer Agreement (MTA) with the laboratory and institution that engineered the plasmids. However, some vectors, including the complete set of pEarleyGate vectors, do not require an MTA and are freely available through the Arabidopsis Biological Resource Center (Columbus, OH, USA).

Plant destination vectors for constitutive or inducible gene expression

It is often useful to express a gene or open reading frame ectopically from a constitutive promoter in order to test its function in a variety of cell types. Alternatively, one might wish to control when the gene is expressed by making use of an inducible promoter. Gateway-compatible vectors have been designed for both purposes (Figure 1a). For instance, in addition to vectors that allow the expression of cloned target sequences from the strong, constitutive 35S promoter of cauliflower mosaic virus, Curtis and Grossniklaus have engineered vectors that make use of a heat-shock gene promoter or an estrogen-responsive promoter (Curtis and Grossniklaus, 2003).

An inducible Gateway-compatible expression vector that allows tighter control of gene expression than previously designed inducible systems has recently been described. This 'double-lock' inducible system requires both heat shock induction and dexamethasone-inducible control of cellular targeting of cyclization recombination (CRE) recombinase in order to activate a promoter disrupted by a DNA fragment flanked by locus of X-over P1 sites. Specifically, heat shock is used to induce the expression of CRE recombinase fused to the hormone-binding domain of the rat glucocorticoid receptor. The resulting protein remains sequestered in the cytoplasm until dexamethasone treatment, which allows the protein to move into the nucleus, catalyze the removal of the sequence blocking transcription by the 35S promoter, and thereby allow expression of the target gene (Joubes *et al.*, 2004).

Figure 2. Overview of Gateway cloning for generation of fusion proteins

(a) Topoisomerase-mediated capture and Gateway recombinational cloning of target sequences.

(1) A sequence of interest (e.g. a cDNA open reading frame) is amplified by PCR using a forward oligonucleotide primer that has the sequence CACC preceding the sequence of interest in order to facilitate direction cloning into the pENTR/D-TOPO vector (obtained from Invitrogen). A proofreading polymerase that generates PCR products with blunt ends is required. (2) PCR products are mixed with the pENTR/D-TOPO vector, which has covalently attached topoisomerase molecules that catalyze ligation of target and vector sequences. *attL1* and *attL2* sites flanking the cloning site mediate subsequent recombination reactions. (3) Using the LR clonase reaction enzyme mix (Invitrogen), which contains the enzymes required for recombination between *attL* and *attR* sites, the target sequence is recombined into a destination vector of choice. Located between the *attR* sites of the destination vector is a chloramphenicol resistance gene (*CmR*) and a *ccdB* gene which is lethal to most strains of *Escherichia coli*. As a result, only those *E. coli* transformed with plasmids having undergone successful recombination events survive (4).

(b) Examples of Gateway-mediated addition of cMyc epitope tags to the C-terminal or N-terminal ends of a target sequence in pEarleyGate 303 or 203, respectively. The *attB* sites (boxed) result from *attL-attR* recombination. The CACC sequence added at the 5' end of the PCR-amplified target sequence is circled. Amino acids are indicated using a single-letter code. Note that additional amino acids derived from *att* sites and adjacent pENTR vector sequences are added to the translated protein.

Plant destination vectors for gene knockdown by the RNA interference (RNAi)

As first shown by Waterhouse *et al.* (Waterhouse *et al.*, 1998), expression of double-stranded RNA is sufficient to trigger the RNAi pathway in plants, leading to the degradation of homologous mRNAs (Baulcombe, 2004). Production of a double-stranded RNA trigger is relatively easy to accomplish by cloning two copies of a target gene segment, in inverted orientation relative to one another, downstream of a strong promoter. Destination vectors that make use of Gateway cloning in order to capture a given trigger RNA sequence in both the forward and reverse orientations have been designed by Helliwell and Waterhouse and are named 'pHellsgate' vectors (Helliwell and Waterhouse, 2003; Wesley *et al.*, 2001) (Figure 1b). Similar vectors have been designed by Karimi *et al.* (Karimi *et al.*, 2002). An alternative approach is to simply produce a full-length antisense transcript to a given target cDNA by cloning the gene sequence in reverse orientation relative to the promoter (Figure 1b). If the antisense transcript anneals with the endogenous mRNA, the resulting double-stranded RNA can trigger the RNAi response. Karimi *et al.* have engineered pairs of Gateway-compatible destination vectors that allow expression of either sense or antisense transcripts of a cloned target sequence (Karimi *et al.*, 2002).

Recently, an ethanol-inducible Gateway-compatible pHellsgate vector that allows reversible expression of dsRNA has been described (Lo *et al.*, 2005). Because knockdown can be induced by the addition of ethanol and reversed by removal (or evaporation) of the ethanol, transcriptional gene silencing can be controlled. This system can potentially allow the conditional knockdown of essential genes for which constitutive knockdown might be lethal. Knockdown of target genes at specific times in development is also possible using this strategy.

Plant destination vectors for promoter analysis

Expression patterns for a given gene can be investigated by fusing the promoter of that gene to a reporter coding sequence and then determining the organs, cell types and developmental stages in which the reporter protein is expressed. To simplify the making of constructs for this purpose, Gateway-compatible vectors have been designed that allow promoter sequences to be recombined into plant destination vectors upstream of B-glucuronidase (GUS) or GFP reporter genes (Curtis and Grossniklaus, 2003; Karimi *et al.*, 2002) (Figure 1c). GUS enzymatic activity converts a colorless substrate (X-Gluc) into a product that is an intense blue color and can be used in tissues cleared of chlorophyll and other natural pigments in order to achieve sensitive detection of transgene expression. A potential disadvantage, however, is that

these methods are destructive and kill the plant cells that are analyzed. By contrast, GFP or other fluorescent proteins (e.g. YFP, CFP or RFP) can be visualized in living cells and can be monitored over time. Weakly expressed fluorescent proteins may escape detection, however, as a result in part of background fluorescence from endogenous plant pigments. By fusing GUS and GFP open reading frames, some vectors allow both reporters to be simultaneously expressed, allowing one to choose which reporter assay to employ (Karimi *et al.*, 2002).

Plant destination vectors for subcellular protein localization and detection of protein/protein interactions

Unlike the vectors described above for promoter analyses, translational fusion of a protein to a fluorescent protein allows the subcellular localization of the protein to be determined. Gateway-compatible vectors that fuse GFP, YFP, CFP or RFP to either the C-terminus or the N-terminus of a target protein have been engineered by several laboratories (Curtis and Grossniklaus, 2003; Karimi *et al.*, 2002; Tzfira *et al.*, 2005) (Figure 1d–f). In some cases, the vectors have been designed such that a six-histidine tag (His tag) is added to the fluorescent protein (Curtis and Grossniklaus, 2003) to facilitate affinity purification of the protein on nickel-chelating resin. An alternative is provided by pEarleyGate vectors that have an influenza A virus haemagglutinin (HA) epitope tag fused to the fluorescent protein, allowing immunological affinity purification or immunoprecipitation (see description of pEarleyGate vectors below).

Gateway-compatible vectors that add YFP, CFP or luciferase to target proteins can also be useful for assaying protein/protein interactions *in vivo* using FRET, BRET or BiFC (Figure 1e,f). FRET makes use of photons emitted by CFP in order to excite YFP. Therefore, detection of YFP emission upon CFP excitation indicates a physical interaction between the proteins fused to CFP and YFP. BRET is a related phenomenon, which utilizes luciferase emissions to excite YFP. Gateway-compatible vectors for both of these applications are currently available (Karimi *et al.*, 2002, Albrecht von Arnim, University of Tennessee, Knoxville, TN, USA, pers. comm.). Walter *et al.* also describe Gateway-compatible vectors that facilitate BiFC assays, in which non-fluorescent N- and C-terminal fragments of YFP must dimerize to reconstitute YFP fluorescence (Walter *et al.*, 2004).

Epitope tagging vectors for protein purification

A number of groups, including ours, have created Gateway-compatible plant destination vectors that add one or more epitope tags to target proteins (Bensmihen *et al.*, 2004; Rohila *et al.*, 2004; Rubio *et al.*, 2005) (Figure 1g). Epitope tags are short, hydrophilic peptide sequences recognized by specific antibodies. Compared with larger protein fusions, the small

size of epitope tags makes them less likely to interfere with protein folding and function (Fritze and Anderson, 2000; Jarvik and Telmer, 1998). Epitope tags recognized by monoclonal or monospecific antibodies offer a means of efficient detection, affinity purification, or subcellular localization of tagged proteins. Expression of recombinant proteins bearing epitope tags can also eliminate the need to generate antibodies recognizing each new protein to be studied, which can be problematic as a result of low antigenicity or high background cross-reaction with other proteins. Single epitope or tandem affinity peptide (TAP) tags are increasingly used to facilitate large-scale, high-throughput proteomics studies (Gavin *et al.*, 2002; Ho *et al.*, 2002). Two groups have recently described Gateway-compatible TAP tagging vectors for use in plants. Rohila *et al.* described a TAP tag containing two copies of the immunoglobulin G (IgG) binding domain of *Staphylococcus aureus* protein A separated from a calmodulin-binding peptide by an intervening Tobacco Etch Virus (TEV) cleavage site (Rohila *et al.*, 2004). Rubio *et al.* described a TAP tag containing two IgG binding domains, a six-histidine metal-binding domain, a cMyc epitope tag and a protease 3C cleavage site (Rubio *et al.*, 2005). Both groups have successfully purified protein complexes from plants using these expression vectors.

Plant destination vectors for modular assembly of transgenes

Recently, Invitrogen has expanded its repertoire of recombination sites in order to allow multiple gene elements to be recombined simultaneously into a destination vector. This modular approach allows one to choose among various promoters, reporter genes or epitope tags in entry vectors and then recombine these into a destination vector that will piece the elements together in the correct order. Karimi *et al.* have embraced this new technology to generate plant destination vectors bearing multi-site Gateway cassettes (Karimi *et al.*, 2005) (Figure 1h).

pEarleyGate vectors

We have designed a large set of Gateway-compatible plant destination vectors that are useful for epitope-tagging proteins of interest. As a prelude to designing Gateway-compatible epitope-tagging vectors, we conducted an evaluation of four epitope tag/antibody combinations in a variety of commonly studied plant species. We spiked total leaf protein extracts of tobacco (*Nicotiana tabacum*), *Arabidopsis thaliana*, maize (*Zea mays*), soybean (*Glycine max*), rice (*Oryza sativa*), tomato (*Lycopersicon esculentum*), and cotton (*Gossypium hirsutum*) (Figure 3a) with proteins displaying AcV5, HA, FLAG, and cMyc epitopes. Immunoblot detection of the tagged recombinant proteins was then conducted, as shown in Figure 3b–e. We found

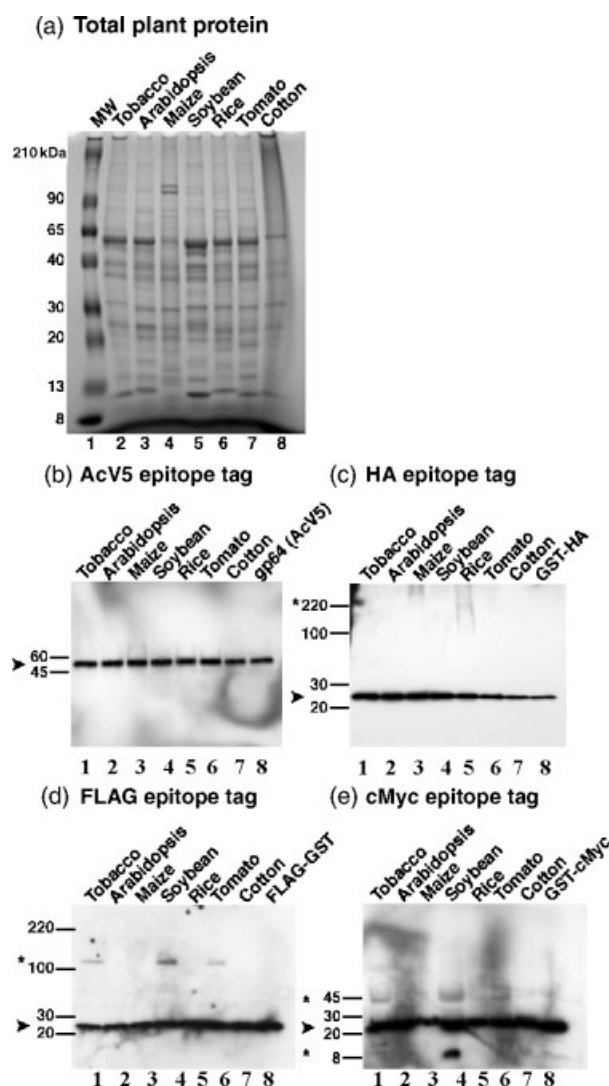


Figure 3. *In vitro* evaluation of AcV5, HA, FLAG and cMyc epitope detection in commonly studied plants.

(a) Total leaf protein (20 μ g) extracted from tobacco (*Nicotiana tabacum*), *Arabidopsis thaliana*, maize (*Zea mays*), soybean (*Glycine max*), rice (*Oryza sativa*), tomato (*Lycopersicon esculentum*) or cotton (*Gossypium hirsutum*) was loaded in adjacent lanes of a 10–20% gradient sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel (Invitrogen). Following electrophoresis, the gel was stained using EZBlue Gel Staining Reagent (Sigma-Aldrich) to demonstrate that equivalent amounts of protein were loaded in each lane.

(b–e) Immunoblot detection of epitope-bearing proteins spiked into tobacco, *A. thaliana*, maize, soybean, rice, tomato or cotton protein samples. Total leaf protein (20 μ g) was spiked with either (b) 225 ng of total viral protein from the baculovirus *Autographa californica*, which bears the AcV5 epitope on its gp64 coat protein, (c) 100 ng of glutathione S-transferase (GST) fused to an HA tag (GST-HA), (d) 100 ng of GST fused to a FLAG tag (FLAG-GST) or (e) 1 μ g of GST fused to a cMyc tag (GST-cMyc). In lane 8 of each gel, the epitope-tagged recombinant protein alone was loaded as a control. Proteins were subjected to electrophoresis, immunoblotting using commercially available antibodies recognizing the four epitopes and chemiluminescent detection. Asterisks indicate cross-reacting proteins.

that all four epitope tags were readily detected in all species tested, although in some species there was cross-reaction between the antibodies and endogenous proteins. For instance, the HA antibody (Figure 3c) interacted with some high-molecular-weight proteins in maize and rice, the FLAG M2 antibody (Figure 3d) cross-reacted with an endogenous protein of approximately 125 kDa in tobacco, soybean, and tomato, and the cMyc (Clone 9E10) antibody (Figure 3e) cross-reacted with an endogenous protein of ~10 kDa in soybean and a protein of ~45 kDa in tobacco and soybean.

Based on the results of Figure 3, we designed Gateway-compatible vectors that would add AcV5, HA, FLAG, or cMyc epitope tags to either the N- or C-termini of target proteins (see Figure 4). We also engineered a vector containing a TAP tag consisting of a calmodulin-binding peptide separated from two copies of a Protein A peptide (which will bind to IgG resin) by a TEV protease cleavage site (Rigaut *et al.*, 1999). pEarleyGate vectors 201–205 allow the addition of HA, FLAG, cMyc, AcV5 or TAP epitope tags to target proteins encoded by cloned cDNA sequences. These vectors make use of the enhanced cauliflower mosaic virus 35S promoter for strong constitutive expression of tagged proteins. A second set of pEarleyGate vectors, 301–304, allows the addition of HA, FLAG, cMyc or AcV5 sequences to the C-terminus of recombinant transgenes. Because these vectors contain no promoter, they are useful for cloning genomic fragments that include promoter sequences, introns and exons, with the tag being added to the last exon in lieu of the natural stop codon. A third set of pEarleyGate vectors were engineered to add both a fluorescent protein and an epitope or His tag to a target protein: pEarleyGate 101 will add YFP with an HA tag, pEarleyGate 102 adds CFP with an HA tag, and pEarleyGate 103 will add GFP with a His tag. The pEarleyGate 101–103 vectors generate C-terminal fusions to the fluorescent protein/epitope tag. pEarleyGate 104 adds an N-terminal YFP to targeted proteins but contains no epitope tag sequence.

All 14 pEarleyGate vectors are derived from pFGC5941 (<http://www.chromDB.org>), which was built using a pCAMBIA (<http://www.cambia.org>) binary vector backbone. pEarleyGate vectors support *Arabidopsis tumefaciens*-mediated stable transformation, and can be obtained from the Arabidopsis Biological Resource Center (<http://www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/abrchome.htm>). Detailed information for pEarleyGate vectors, including maps and sequence information, is available at the Pikaard laboratory website (<http://biology4.wustl.edu/pikaard/>).

In vivo evaluation of pEarleyGate vectors

Detection of different epitope-tagged versions of the same target protein, expressed from pEarleyGate derived T-DNAs in transgenic *A. thaliana*, is shown in Figure 5. For this comparison, the open reading frame for HDA6, an *A. thali-*

ana histone deacetylase, was recombined into pEarleyGate 200-series vectors. Resulting N-terminal HA, FLAG, cMyc, or AcV5-tagged recombinant proteins or C-terminal TAP-tagged proteins were expressed from mRNAs driven by the cauliflower mosaic virus 35S promoter. Multiple transgenic *A. thaliana* lines were generated for each pEarleyGate construct. Leaf tissue from individual primary transformants was then homogenized in sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) sample buffer and boiled, and an aliquot of the resulting lysate was loaded in a single lane of an SDS–PAGE gel. Following electrophoresis and immunoblotting, the recombinant proteins were detected using commercially available antibodies recognizing the different epitope tags. As shown in Figure 5, HA, FLAG, cMyc, AcV5 and TAP tagged HDA6 proteins were detected in multiple independent lines, with expression levels varying from line to line. Relatively low background cross-reaction with endogenous proteins was observed for all antibodies tested, consistent with the prior spiking experiments. Smaller products detected in protein extracts of plants expressing full-length tagged proteins but not detected in non-transgenic controls are presumably cleavage products or incomplete translation products derived from the transgenes.

Use of epitope tags for affinity purification

To evaluate the usefulness of pEarleyGate vectors for production of recombinant proteins that can be affinity-purified by virtue of their epitope tags, we extracted total soluble protein from *A. thaliana* lines overexpressing HDA6 tagged with FLAG, HA, or cMyc epitopes. Anti-HA, FLAG, or cMyc antibodies conjugated to agarose beads were then used to capture the tagged proteins. For each epitope tag tested, HDA6 protein was effectively affinity-captured and greatly enriched in bead-associated fractions as compared with input extracts (Figure 6a).

Interestingly, elution of the protein from the matrix using excess epitope peptides appears to be more difficult for some antibody–epitope combinations than for others. For instance, FLAG-tagged HDA6 could be eluted using a high concentration of competing peptide, but cMyc and HA (data not shown) tagged proteins were not eluted using similar conditions. The latter tagged proteins were only eluted under denaturing conditions in SDS–PAGE sample buffer (Figure 6b).

We were also interested in determining if pEarleyGate epitope-tagging vectors are useful for immunolocalization experiments. For this set of experiments we recombined the cDNA sequence for *HDT1*, a histone deacetylase known to localize to the nucleolus when fused to GFP or YFP (Lawrence *et al.*, 2004; Zhou *et al.*, 2004), into pEarleyGate 200-series vectors. As shown in Figure 7, immunolocalization of the cMyc epitope reveals that the

Organization of pEarleyGate T-DNA regions

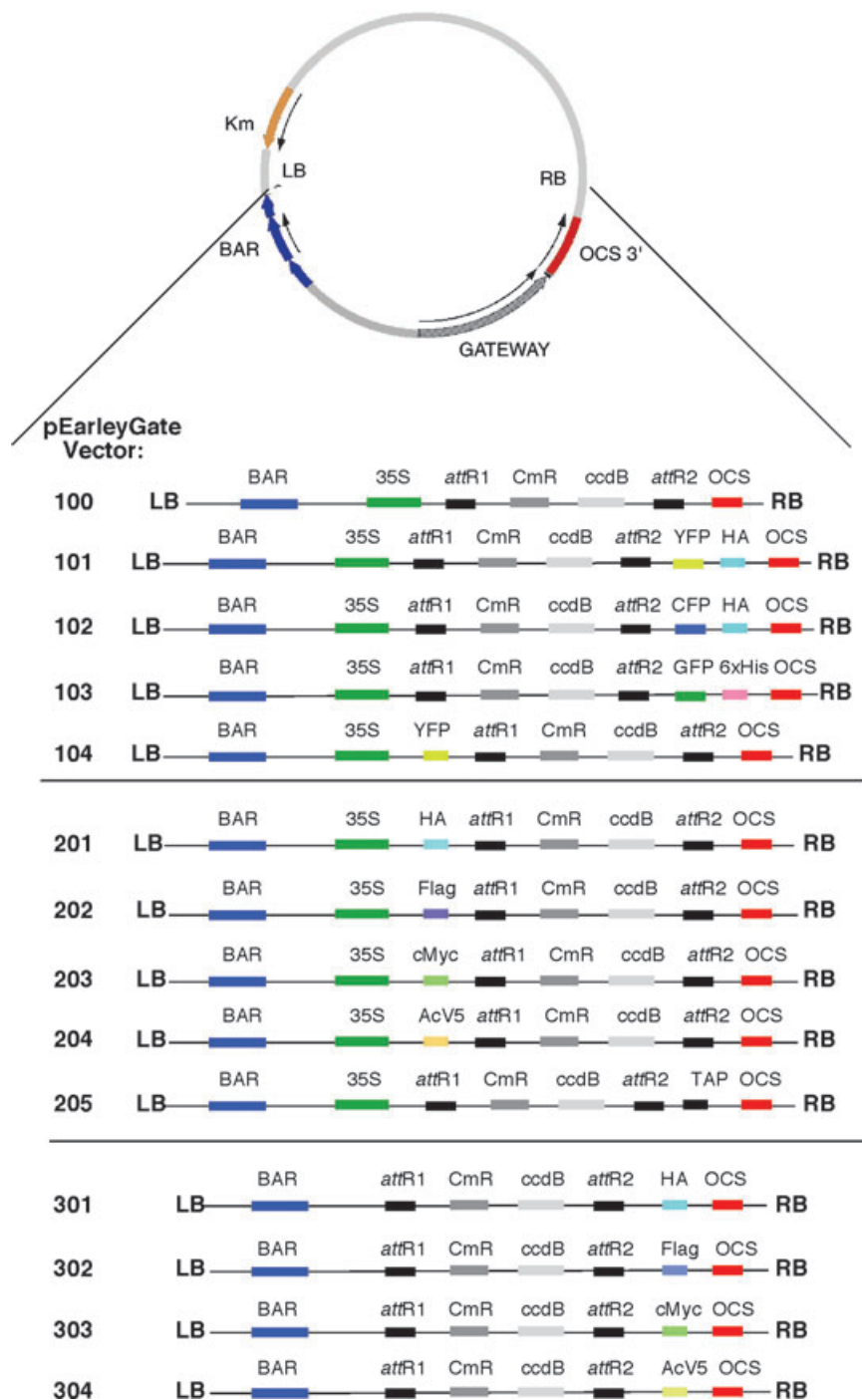


Figure 4. pEarleyGate plant transformation vectors. The pEarleyGate vectors are derived from pFGC5941 (<http://www.chromDB.org>), which was built using a pCambia (<http://www.cambia.org/>) plasmid backbone. As a result, all of the pEarleyGate plasmids are binary vectors that will replicate in both *Escherichia coli* and *Agrobacterium tumefaciens* and have left border (LB) and right border (RB) sequences for *Agrobacterium*-mediated T-DNA transfer.

The organization of the T-DNAs for each of the various pEarleyGate vectors is shown. The Gateway cassettes in each vector include *attR1*, a chloramphenicol resistance gene (*CmR*), the *ccdB* killer gene and *attR2*. 35S, the cauliflower mosaic virus 35S promoter and its upstream enhancer. OCS, the 3' sequences of the octopine synthase gene, including polyadenylation and presumptive transcription termination sequences. BAR, the Basta herbicide resistance gene for selection of transgenic plants. Km, the bacterial kanamycin resistance gene within the plasmid backbone. Different pEarleyGate vectors allow engineering and expression of proteins fused in frame with HA, FLAG, cMyc, AcV5 or tandem affinity purification (TAP) tags and/or yellow, green or cyan fluorescent proteins (YFP, GFP or CFP, respectively) at either the amino-terminal or carboxy-terminal end of the target proteins.

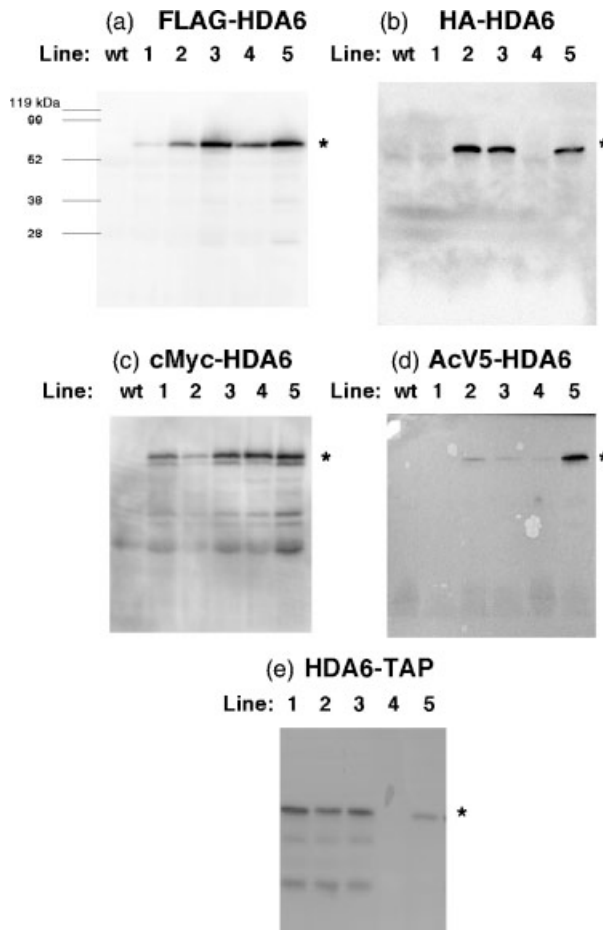


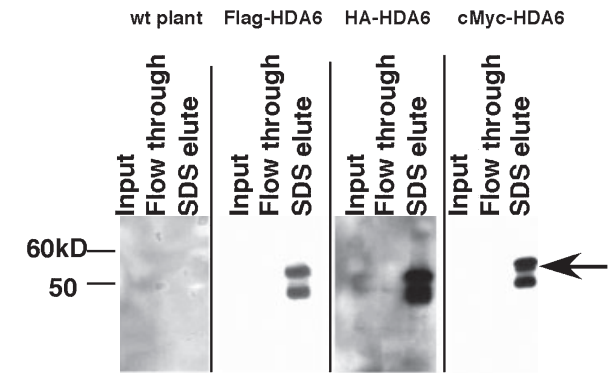
Figure 5. Immunoblot detection of epitope-tagged recombinant proteins expressed from pEarleyGate-derived T-DNAs in *Arabidopsis thaliana*. The open reading frame of *HDA6* was recombined into pEarleyGate 202, 201, 203, 204 or 205 to generate FLAG, HA, cMyc, AcV5, or tandem affinity purification (TAP)-tagged *HDA6* fusion proteins, respectively. For each construct, leaf tissue from five independent Basta-resistant T1 plants (lanes 1–5) or a non-transformed control (wt) plant was homogenized in sodium dodecyl sulfate (SDS) sample buffer and equal aliquots were subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) on a 12.5% Tris-glycine gel. Proteins were transferred to nitrocellulose or PVDF membrane and epitope-tagged proteins were detected using: (a) anti-AcV5 monoclonal antibody (diluted 1:2000) followed by anti-mouse immunoglobulin G (IgG)–horseradish peroxidase (HRP) secondary antibody (diluted 1:2000), or (b) anti-HA–HRP monoclonal antibody (diluted 1:3000), or (c) anti-FLAG–AP M2 monoclonal antibody (diluted 1:1000), or (d) anti-cMyc–alkaline phosphatase (AP) monoclonal antibody (diluted 1:1000), or (e) peroxidase-conjugated anti-IgG (diluted 1:2000). Protein–antibody complexes were visualized by chemiluminescent detection of AP or HRP activity. Asterisks indicate full-length epitope-tagged *HDA6*.

tagged HDT1 protein is detected in the nucleolus of transgenic plants, as expected.

In vivo evaluation of pEarleyGate fluorescent protein fusion vectors

pEarleyGate vectors designed for fusing target proteins to GFP, YFP or CFP include an epitope tag fused in frame with

(a) Affinity capture of epitope-tagged proteins



(b) Peptide elution of epitope-tagged proteins

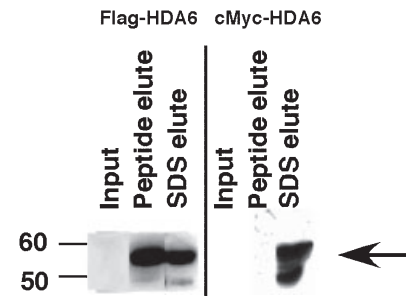


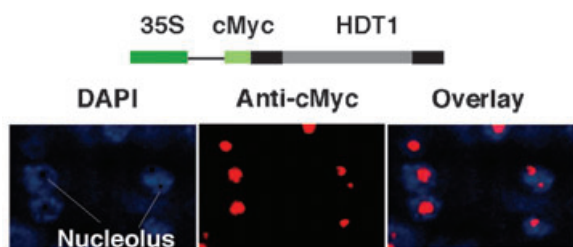
Figure 6. Affinity purification of FLAG, HA, or cMyc-tagged *HDA6* expressed in *Arabidopsis thaliana* transgenic plants.

(a) *A. thaliana* plants expressing FLAG, HA, or cMyc-tagged *HDA6* were homogenized in extraction buffer and incubated with anti-FLAG, anti-HA or anti-cMyc antibodies conjugated to agarose beads. Beads and bound proteins were then washed extensively with extraction buffer and bound proteins were eluted by boiling in sodium dodecyl sulfate (SDS) sample buffer. Equal aliquots of the input homogenate, wash (flow-through) and eluted proteins were subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and recombinant proteins were detected by immunoblotting using anti-FLAG, anti-HA or anti-cMyc antibodies. Arrows indicate full-length epitope-tagged *HDA6*.

(b) Peptide elution of affinity-captured proteins works better for some epitope tags than for others. FLAG- or cMyc-tagged *HDA6* affinity captured on agarose beads was first incubated with FLAG or cMyc peptide under non-denaturing conditions and beads were subsequently boiled in SDS sample buffer. Aliquots of the input, peptide-eluted or SDS-eluted fractions were subjected to SDS–PAGE and recombinant proteins were detected by immunoblotting using anti-FLAG or anti-cMyc antibodies. Arrows indicate full-length epitope-tagged *HDA6*. Note that FLAG-tagged *HDA6* could be peptide-eluted but cMyc-tagged protein was not eluted from beads using cMyc peptide.

the fluorescent protein. Their design allows the vectors to be used for *in vivo* localization of resulting fluorescent fusion proteins, for immunolocalization of the protein in fixed cells by virtue of the epitope tag or for affinity purification or detection of the protein on immunoblots. As a test of the pEarleyGate fluorescent protein fusion vectors, we recombined the *HDT1* cDNA into pEarleyGate 101. As expected, the HDT1-YFP-HA fusion protein localizes to the nucleolus, as can be deduced by comparing the fluorescence signal with the differential interference contrast (DIC)

(a) Immunolocalization using pEarleyGate vectors



(b) Fluorescence microscopy using pEarley Gate vectors

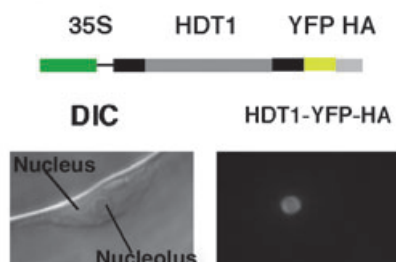


Figure 7. Use of pEarleyGate vectors for protein localization experiments. (a) Immunolocalization of cMyc-tagged HDT1 expressed using pEarleyGate 203. HDT1 localizes to the nucleolus (n), which corresponds to the 4',6-diamidino-2-phenylindole-negative region(s) of the nuclei. (b) Localization of HDT1-YFP-HA fusion protein expressed using pEarleyGate 101. The protein was localized by virtue of yellow fluorescent protein (YFP) fluorescence. The nucleus and nucleolus are clearly visible in the image obtained by differential interference contrast (DIC) microscopy.

image (Figure 7b). Upon boiling leaf tissue in SDS-PAGE sample buffer, and subjecting extracted proteins to SDS-PAGE and immunoblotting using anti-HA antibody, the HDT1-YFP-HA fusion protein is also readily detected by virtue of its epitope tag (data not shown). Collectively, these data demonstrate that pEarleyGate 101–103 can be useful for detecting proteins both *in situ* and following fractionation and immunoblotting.

Concluding remarks

Gateway technology is increasingly used to facilitate proteomic analyses (Gong *et al.*, 2004; Koroleva *et al.*, 2005; Pendle *et al.*, 2005; Reboul *et al.*, 2003; Tian *et al.*, 2004) and efforts are ongoing to clone the *A. thaliana* ORFeome (the comprehensive collection of full-length cDNAs) into Gateway pENTR vectors (Gong *et al.*, 2004; <http://www.evry.inra.fr/public/projects/orfeome/orfeome.html>). One can shuttle these ORFs into the various destination vectors now available. We anticipate that the pEarleyGate vectors will be a useful addition to the sets of Gateway-compatible vectors already available to the plant community for protein over-expression, gene silencing, protein localization and promoter analysis.

Experimental procedures

Notes on the use of pEarleyGate destination vectors

- (i) The pENTR/D-TOPO vector that we use in most of our recombination reactions contains the same bacterial selection marker as the pEarleyGate vectors (kanamycin resistance). To prevent transformation of bacteria with the pENTR plasmid following the recombination reaction, we cut the pENTR vector bearing the target sequence of interest with a restriction endonuclease that cleaves within the pENTR backbone but does not cut within the target sequence. We often use *Mlu*I, which cuts twice within the pENTR backbone. Most other Gateway-compatible destination vectors have different selectable markers, in which case the pENTR plasmid does not need to be cut before the recombination reaction. Alternatively, one could make use of a pDONR vector that has an antibiotic resistance marker other than kanamycin.
- (ii) Before recombining the sequence of interest into the pEarleyGate vectors, we typically gel-purify the digested fragment that contains the sequence of interest flanked by the *att*L sites. However, the recombination reaction also works with cleaved DNA that is purified using a commercial DNA clean-up kit.
- (iii) We recombine ~100 ng of pEarleyGate plasmid DNA with ~100 ng of pENTR fragment using the LR clonase reaction mix (Invitrogen). We find that the concentration of the two fragments can vary without disrupting the success rate of the recombination. We have also found that clonase reactions can be scaled down to half-reactions without jeopardizing successful recombination events, which reduces the cost per reaction.
- (iv) After the recombination reaction, we treat the reaction with proteinase K to digest the clonase enzymes, and transform the resulting reaction into a *ccdB*-sensitive strain of *E. coli* (we typically use DH5- α). We select for positive clones by plating transformation reactions on LB medium that contains 50 μ g ml⁻¹ kanamycin.

Detailed protocols for capturing target sequences in entry vectors and transferring them to destination vectors are available at Invitrogen's website (<http://www.invitrogen.com>).

Plant Material

Arabidopsis thaliana ecotype Columbia, *Z. mays*, *O. sativa*, *G. max* and *L. esculentum* were grown for 4 to 6 weeks under long-day conditions (16 h light/8 h dark) at room temperature using fluorescent light illumination. *N. tabacum* and *G. hirsutum* were grown for 4 weeks at ~25°C on a 14 h light/10 h dark cycle. For immunoblot analysis of epitope-tagged constructs and immunoprecipitation experiments, *A. thaliana* plants were grown for 2 to 3 weeks under long-day conditions. For fluorescent protein analyses, transgenic *A. thaliana* seeds were germinated on sterile semi-solid Murashige-Skoog medium (Sigma-Aldrich, Saint Louis, MO, USA) supplemented with 1% sucrose (pH 5.8), and plants were examined after 2 weeks of growth.

Epitope tag sequences

The FLAG epitope sequence used in this study is DYKDDDDK; the HA epitope is YPYDVPDYA; the cMyc epitope is EQKLISEEDL; the AcV5 epitope is SWKDASGWS, and the TAP tag sequence is

EKRWWKKNFIVSAANRFKISSSGALDYDIPTTASENLYFQGLKTA-ALAQHDEAVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPS-QSANLLAEAKLNDAQAPKVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKLNGAQAPKVDANSAGKST (Rigaut *et al.*, 1999).

Epitope-tagged protein spiking experiments

Recombinant proteins used in the protein spiking study were cloned and expressed in bacterial expression vectors based on the MAC vector backbone (Sigma-Aldrich). Inserts were generated by PCR and directionally cloned using the Director Universal PCR kit (Sigma-Aldrich). Recombinant epitope-tagged proteins FLAG-GST, GST-cMyc, and GST-HA were expressed in *E. coli* strain BL21-DE3 and affinity-purified using glutathione affinity resin (Sigma-Aldrich). Proteins were quantified by the method of Bradford (Bradford, 1976) using commercially available Bradford Reagent (Sigma-Aldrich).

Total leaf protein was extracted from 100 mg of fresh leaf tissue using the Plant Total Protein Extraction Kit (Sigma-Aldrich) supplemented with 1:100 [volume/volume (v/v)] diluted plant protease inhibitor cocktail (Sigma-Aldrich). The protein concentration was determined by the method of Bradford (Bradford, 1976). Total protein (20 µg) was then spiked with 100 ng of FLAG-GST, 100 ng of GST-HA, 1 µg of GST-cMyc, or 225 ng of *Autographa californica* total protein and subjected to SDS-PAGE, electroblotting to Hybond-ECL nitrocellulose (Amersham Biosciences, Piscataway, NJ, USA) or PVDF (Millipore, Billerica, MA, USA) membrane, and probing with appropriate antibodies using standard methods (Fritze and Anderson, 2000). Anti-FLAG M2[®] monoclonal antibody-alkaline phosphatase conjugate, anti-HA monoclonal antibody-peroxidase conjugate (Clone HA-7), anti-cMyc monoclonal antibody-alkaline phosphatase conjugate (clone 9E10), anti-mouse IgG (whole molecule)-alkaline phosphatase conjugate, and peroxidase-conjugated anti-peroxidase were all from Sigma-Aldrich; anti-*Autographa californica* gp64 protein monoclonal antibody (clone AcV5) was from eBioscience (San Diego, CA, USA).

For Western blot analysis of protein spiking experiments, the following dilutions of antibodies were used. Anti-AcV5 monoclonal antibody was diluted 1:2000 prior to incubation with the blot and was detected, after washing, using 1:30 000-diluted anti-mouse IgG (whole molecule)-alkaline phosphatase (AP) conjugate as the secondary antibody. Other epitopes were detected following a single incubation with AP- or horseradish peroxidase (HRP)-conjugated primary antibodies. Final dilutions for the antibodies were: anti-HA-HRP, 1:10 000; anti-FLAG M2-AP, 1:10 000; and anti-cMyc-AP, 1:50 000. Chemiluminescent detection of alkaline phosphatase (AP) or peroxidase (HRP) activity was performed using CDP-Star Chemiluminescent substrate and Chemiluminescent Peroxidase substrate, respectively (Sigma-Aldrich).

Construction of pEarleyGate plasmid vectors

pEarleyGate 100–105. To create pEarleyGate 100, the Gateway cassette was amplified by PCR from the Reading Frame B DNA fragment (purchased from Invitrogen) using the following primers: forward 5'-cgcgctcgagatcacaagttgtacaaaaagc-3' and reverse 5'-gccctaggcaccactttgtacaagaagc-3'. The resulting PCR product was digested with *XhoI* and *AvrII* and ligated (Rapid DNA Ligation Kit; Roche, Mannheim, Germany) into pFGC5941 (<http://www.ChromDB.org>), replacing its *XhoI* to *AvrII* fragment. To create pEarleyGate101 and 102, YFP and CFP were amplified by PCR using primers forward 5'-tgctagggtgagcaagggcgaggagc-3' and

reverse 5'-tcttaattaagcgaatctggaacatcgtagtggtatctagatccggtggatcc-3'. Resulting PCR products were digested with *AvrII* and *PacI* and inserted into the adjacent *AvrII* and *PacI* sites of pEarleyGate 100. To create pEarleyGate 104, YFP was excised from pCAM-35S-EYFP-C1 (Fritze and Anderson, 2000) using *BamHI* and *NcoI* and ligated into the *BamHI* and *NcoI* sites of pFGC5941, replacing its *BamHI*-*NcoI* fragment. The Gateway cassette was then added by PCR amplifying the Reading Frame B cassette using primers forward 5'-cgcgctcgagatcacaagttgtacaaaaagc-3' and reverse 5'-cgcgctcaccactttgtacaagaagc-3' and ligating the resulting PCR product into the *NcoI* and *AvrII* sites of the plasmid that had been converted to blunt ends by treatment with T4 DNA polymerase (NEB) and 10 mM dNTPs. To create pEarleyGate 103, the GFP-6 × His fragment of pCAMBIA 1302 was amplified by PCR, cut with *XhoI* and *AvrII*, and ligated into pFGC5941, replacing its *XhoI* to *AvrII* fragment. The Gateway cassette was then added by amplifying the Reading Frame B DNA fragment by PCR using the primers forward 5'-cgcgctcgagatcacaagttgtacaaaaagc-3' and reverse 5'-cgcgctcgagcaccactttgtacaagaagc-3', cutting with *XhoI* and ligating the resulting PCR fragment into the *XhoI* site of the plasmid.

pEarleyGate 201–205. Gateway cassettes with adjacent epitope tag sequences were amplified by PCR using the Invitrogen Reading Frame B sequence. Forward primers adding HA, FLAG, cMyc, or AcV5 epitope tags to Gateway cassette sequences were: HA, 5'-accatacagatgttccagattacgctatcacaagttgtacaaaaagc-3'; FLAG, 5'-gactacaagacgatgacgacacaaatcacaagttgtacaaaaagc-3'; cMyc, 5'-gaacagaagtgatctctgaagaagatctgatcacaagttgtacaaaaagc-3'; AcV5, 5'-tcttgaaagatgagcggctgtctatcacaagttgtacaaaaagc-3'. An identical reverse primer, 5'-aattaactctctgactcactaggc-3', was used for all PCR reactions. Resulting PCR products were cloned into pFGC5941 that had been digested with *NcoI* and *AvrII* and treated with T4 DNA polymerase and 10 mM dNTPs to generate blunt ends. To create pEarleyGate 205, the TAP fragment of pBM3947 was amplified by PCR using primers forward 5'-cctaggagatggaaaagagaagatg-3' and reverse 5'-gccttaattaatcaggttgactcccc-3', cut with *AvrII* and *PacI* and ligated into pEarleyGate100.

pEarleyGate 301–304. Gateway cassettes with adjacent epitope tag sequences were amplified by PCR using the Invitrogen Reading Frame B sequence. Reverse primers adding HA, FLAG, cMyc, or AcV5 epitope tags to Gateway cassette sequences were: HA, 5'-tcaagcgaatctggaacatcgtagtggtacaccactttgtacaagaagc-3'; FLAG, 5'-tcattgtcgtcatgcttctgtagtcaccactttgtacaagaagc-3'; cMyc, 5'-tcacagatctctcagagatcagttctgtccaccactttgtacaagaagc-3'; AcV5, 5'-tcaagaccagcgcctcgtcatctttccaagacaccactttgtacaagaagc-3'. An identical forward primer, 5'-gaattctgcagtcgagcgg-3', was used for all PCR reactions. Resulting PCR products were ligated into pFGC5941 which had been digested with *EcoRI* and *AvrII* and treated with T4 DNA polymerase and 10 mM dNTPs to generate blunt ends.

All ligation reactions including the Gateway cassette were transformed into *E. coli* DB3.1 cells (Invitrogen), which are resistant to the *ccdB* gene. Positive clones were selected on LB plates containing 34 µg ml⁻¹ chloramphenicol.

Recombination of target sequences into pEarleyGate plant expression vectors

HDA6 and *HDT1* coding sequences, either with or without their natural stop codon, were amplified from cloned cDNAs by PCR

using Platinum Pfx polymerase (Invitrogen) and the following primers: *HDA6* forward 5'-caccatggaggcagcagaaagc-3' and reverse 5'-ctagagagctgggacactgagc-3'; *HDT1* (no stop) forward 5'-caccatggagttctgggaattg-3' and reverse 5'-ctggcagcagcgtgctgg-3'; *HDT1* (stop) forward 5'-caccatggagttctgggaattg-3' and reverse 5'-tcactggcagcagcgtgc-3'. The resulting PCR products were captured by topoisomerase-mediated cloning into the pENTR/D-TOPO vector (Invitrogen). Entry clones containing *HDT1* and *HDA6* sequences, pENTR-*HDA6* and pENTR-*HDT1*, were cut with *MluI* to linearize the pENTR plasmid in order to prevent subsequent transformation of *E. coli* by the entry vector rather than (or in addition to) the pEarleyGate destination vector (see notes on the use of pEarleyGate vectors, above). The DNA fragment containing the *HDA6* sequence flanked by *attL* recombination sites was recombined into the pEarleyGate 201, 202, 203, 204, and 205 plasmids using LR clonase (Invitrogen). The DNA fragment containing *HDT1* without a stop codon was recombined into pEarleyGate 101 to form a C-terminal YFP-HA fusion and the DNA fragment containing pENTR-*HDT1* with a stop codon was recombined into pEarleyGate 203 to form a N-terminal cMyc fusion. Recombined plasmids were transformed into *E. coli* DH5- α cells. Positive clones were selected on kanamycin LB plates. Recombinant plasmids were then transformed into *A. tumefaciens* strain LBA 4404 for subsequent plant transformation.

Plant transformation and detection of epitope-tagged recombinant proteins

A. tumefaciens-mediated transformation of *A. thaliana* ecotype Columbia was accomplished by using the floral dip technique (Bechtold and Pelletier, 1998) as modified by Clough and Bent (Clough and Bent, 1998).

A single leaf from plants transformed with pEarleyGate vectors was homogenized in 400 μ l of SDS-PAGE sample buffer [50 mM Tris (pH 6.8), 6% glycerol, 2% SDS, 100 mM dithiothreitol (DTT), and 0.01% bromophenol blue] and boiled for 5 min. Samples were centrifuged at 16 000 *g* for 10 min. A volume of 20 μ l of supernatant was loaded onto SDS-PAGE gel and epitope-tagged proteins were detected by immunoblotting. Antibody dilutions used for detection of *in planta* expressed epitope-tagged proteins by Western blot analysis are included in the legend of Figure 5.

Affinity purification experiments

Above-ground tissues of 3-week-old *A. thaliana* plants expressing HA, FLAG, cMyc, or AcV5 tagged *HDA6* transgenes were harvested and ground to a fine powder in liquid nitrogen. Two volumes [weight/volume (w/v)] of Cell Lytic P (Sigma) solution, amended to include 1:100 (v/v) diluted plant-specific protease inhibitor cocktail (Sigma-Aldrich) and 1 mM phenylmethylsulfonyl fluoride (PMSF), was then mixed with the powder. Homogenates were filtered through four layers of miracloth (Calbiochem, San Diego, CA, USA) and subjected to centrifugation at 6000 *g* for 15 min. The supernatant containing epitope-tagged *HDA6* was incubated with anti-HA, anti-cMyc or anti-FLAG-conjugated agarose (all from Sigma-Aldrich) for 1 h at 4°C. The conjugated agarose resins were washed twice with Cell Lytic P extraction buffer and proteins were eluted with SDS-PAGE sample buffer (50 mM Tris-HCl, pH 6.8, 6% glycerol, 2% SDS, 100 mM DTT and 0.01% bromophenol blue) or Cell Lytic P buffer containing 3 \times FLAG peptide (200 μ g ml⁻¹). Samples were subjected to electrophoresis on an SDS-PAGE gel, transferred to PVDF membrane and analyzed by immunoblotting with the appropriate antibody.

Analysis of fluorescent tags and immunolocalization experiments

Root tissue expressing HDT1-YFP-HA was imaged using a Zeiss M2Bio microscope equipped with a Zeiss Axiocam digital camera and a Nikon Eclipse E600 fluorescence microscope with a Q Imaging Retiga EX digital camera. Fluorescence microscopy and immunolocalization experiments were performed as previously described (Lawrence et al., 2004; Onodera et al., 2005).

Acknowledgements

We apologize to any researchers whose Gateway-compatible plant expression vectors were overlooked for this review. We thank Mike Dyer (Washington University, St Louis, MO, USA) for expert care of the tobacco, cotton, and *A. thaliana* plants; Monsanto (St Louis, MO, USA) for providing inbred maize, rice and cotton seeds; Perry Kim (Queen's University, Ontario, Canada) for supplying protein samples of the baculovirus *Autographa californica* (AcMNV); Mark Johnston (Washington University) for supplying the pBM3947 vector, and Erik Nielsen (Donald Danforth Plant Science Center, St Louis, MO, USA) for supplying the pCAM-35S-EYFP-C1 vector. Sigma-Aldrich Company generously provided many of the reagents needed for this study. Construction and testing of pEarleyGate vectors were supported by the United States National Science Foundation (NSF; grants DBI-9975930 and DBI-0421619) and by the United States National Institutes of Health (NIH; grant R01-GM60380). All opinions expressed are those of the authors and do not necessarily reflect the views of the NSF or NIH.

References

- Baulcombe, D. (2004) RNA silencing in plants. *Nature*, **431**, 356–363.
- Bechtold, N. and Pelletier, G. (1998) In planta *Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Methods Mol. Biol.* **82**, 259–266.
- Bensmihen, S., To, A., Lambert, G., Kroj, T., Giraudat, J. and Parcy, F. (2004) Analysis of an activated ABI5 allele using a new selection method for transgenic *Arabidopsis* seeds. *FEBS Lett.* **561**, 127–131.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.
- Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.
- Curtis, M.D. and Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469.
- Ehrhardt, D. (2003) GFP technology for live cell imaging. *Curr. Opin. Plant Biol.* **6**, 622–628.
- Fritze, C.E. and Anderson, T.R. (2000) Epitope tagging: general method for tracking recombinant proteins. *Methods Enzymol.* **327**, 3–16.
- Gavin, A.C., Bosche, M. and Krause, R. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141–147.
- Gong, W., Shen, Y.P. and Ma, L.G. (2004) Genome-wide ORFeome cloning and analysis of *Arabidopsis* transcription factor genes. *Plant Physiol.* **135**, 773–782.
- Hanson, M.R. and Kohler, R.H. (2001) GFP imaging: methodology and application to investigate cellular compartmentation in plants. *J. Exp. Bot.* **52**, 529–539.

- Hartley, J.L., Temple, G.F. and Brasch, M.A.** (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**, 1788–1795.
- Haseloff, J.** (1999) GFP variants for multispectral imaging of living cells. *Methods Cell Biol.* **58**, 139–151.
- Helliwell, C. and Waterhouse, P.** (2003) Constructs and methods for high-throughput gene silencing in plants. *Methods*, **30**, 289–295.
- Ho, Y. and Gruhler, A. and Heilbut, A.** (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, **415**, 180–183.
- Jarvik, J.W. and Telmer, C.A.** (1998) Epitope tagging. *Annu. Rev. Genet.* **32**, 601–618.
- Joubes, J., De Schutter, K., Verkest, A., Inze, D. and De Veylder, L.** (2004) Conditional, recombinase-mediated expression of genes in plant cell cultures. *Plant J.* **37**, 889–896.
- Karimi, M., Inze, D. and Depicker, A.** (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**, 193–195.
- Karimi, M., De Meyer, B. and Hilson, P.** (2005) Modular cloning in plant cells. *Trends Plant Sci.* **10**, 103–105.
- Koroleva, O.A., Tomlinson, M.L., Leader, D., Shaw, P. and Doonan, J.H.** (2005) High-throughput protein localization in Arabidopsis using *Agrobacterium*-mediated transient expression of GFP-ORF fusions. *Plant J.* **41**, 162–174.
- Lawrence, R.J., Earley, K., Pontes, O., Silva, M., Chen, Z.J., Neves, N., Viegas, W. and Pikaard, C.S.** (2004) A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Mol. Cell*, **13**, 599–609.
- Lo, C., Wang, N. and Lam, E.** (2005) Inducible double-stranded RNA expression activates reversible transcript turnover and stable translational suppression of a target gene in transgenic tobacco. *FEBS Lett.* **579**, 1498–1502.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O. and Pikaard, C.S.** (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, **120**, 613–622.
- Pendle, A.F., Clark, G.P. and Boon, R.** (2005) Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Mol. Biol. Cell*, **16**, 260–269.
- Reboul, J., Vaglio, P. and Rual, J.F.** (2003) *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* **34**, 35–41.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B.** (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030–1032.
- Rohila, J.S., Chen, M., Cerny, R. and Fromm, M.E.** (2004) Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. *Plant J.* **38**, 172–181.
- Rubio, V., Shen, Y. and Saijo, Y., Liu, Y., Gusmaroli, G., Dinesh-Kumar, S.P. and Deng, X.W.** (2005) An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. *Plant J.* **41**, 767–778.
- Shuman, S.** (1994) Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. *J. Biol. Chem.* **269**, 32678–32684.
- Stewart, C.N. Jr.** (2001) The utility of green fluorescent protein in transgenic plants. *Plant Cell Rep.* **20**, 376–382.
- Tian, G.W., Mohanty, A. and Chary, S.N.** (2004) High-throughput fluorescent tagging of full-length Arabidopsis gene products in planta. *Plant Physiol.* **135**, 25–38.
- Tzfira, T., Tian, G.W. and Lacroix, B.** (2005) pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. *Plant Mol. Biol.* **57**, 503–516.
- Walter, M., Chaban, C. and Schutze, K.** (2004) Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* **40**, 428–438.
- Waterhouse, P.M., Graham, M.W. and Wang, M.B.** (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl. Acad. Sci. USA*, **95**, 13959–13964.
- Wesley, S.V., Helliwell, C.A. and Smith, N.A.** (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J.* **27**, 581–590.
- Zhou, C., Labbe, H. and Sridha, S.** (2004) Expression and function of HD2-type histone deacetylases in Arabidopsis development. *Plant J.* **38**, 715–724.

CHAPTER 3

PLANT NUCLEAR RNA POLYMERASE IV MEDIATES siRNA AND DNA-METHYLATION DEPENDENT HETEROCHROMATIN FORMATION

Published in *Cell* (2005) 120 (5): 613-622.

My contributions to this work:

I performed phylogenetic analysis on the NRPD1a/NRPD1, NRPD1b/NRPE1 and NRPD2/NRPE2 subunits establishing that they are distinct from eukaryotic DNA-dependent RNA Polymerases I, II and III largest and second-largest subunits as well as those found in eubacteria, archaea, cyanobacteria, chloroplasts and viruses (Figures 1A, 1B and 2A). I genotyped and identified homozygous *nrd1a-3/nrd1-3* and *nrd1b-11/nrpe1-11* T-DNA insertion lines ordered from the ABRC stock center that were used in genetic analyses (Figure 4). I also generated multiple protein sequence alignments of the proteins and identified an annotation error for NRPD1b/NRPE1 in the Arabidopsis genome (Supplemental Figures 1, 2, 3 and 4). This prediction was demonstrated to be accurate by the later publications of Kanno et al., 2005 and Pontier et al., 2005. Lastly, I assisted in the editing of the paper and responding to reviewer comments.

Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation

Yasuyuki Onodera,^{1,2,4} Jeremy R. Haag,^{1,4}
Thomas Ream,^{1,4} Pedro Costa Nunes,³
Olga Pontes,¹ and Craig S. Pikaard^{1,*}

¹Biology Department
Washington University
1 Brookings Drive
St. Louis, Missouri 63130

²Graduate School of Agriculture
Faculty of Agriculture
Hokkaido University
Kita 9, Nishi 9, Kita-ku
Sapporo 060-8589
Japan

³Secção de Genética
Centro de Botanica e Engenharia Biologica
Instituto Superior de Agronomia
Tapada da Ajuda
1349-017 Lisboa
Portugal

Summary

All eukaryotes have three nuclear DNA-dependent RNA polymerases, namely, Pol I, II, and III. Interestingly, plants have catalytic subunits for a fourth nuclear polymerase, Pol IV. Genetic and biochemical evidence indicates that Pol IV does not functionally overlap with Pol I, II, or III and is nonessential for viability. However, disruption of the Pol IV catalytic subunit genes *NRPD1* or *NRPD2* inhibits heterochromatin association into chromocenters, coincident with losses in cytosine methylation at pericentromeric 5S gene clusters and *AtSN1* retroelements. Loss of CG, CNG, and CNN methylation in Pol IV mutants implicates a partnership between Pol IV and the methyltransferase responsible for RNA-directed de novo methylation. Consistent with this hypothesis, 5S gene and *AtSN1* siRNAs are essentially eliminated in Pol IV mutants. The data suggest that Pol IV helps produce siRNAs that target de novo cytosine methylation events required for facultative heterochromatin formation and higher-order heterochromatin associations.

Introduction

In eukaryotes, three nuclear DNA-dependent RNA polymerases (RNAPs) transcribe genomic DNA into RNA. RNA polymerase I (Pol I) transcribes the ribosomal RNA (rRNA) genes clustered at nucleolus organizer regions (Grummt, 2003); RNA polymerase II (Pol II) transcribes the vast majority of genes, including protein-coding genes (Woychik and Hampsey, 2002), and RNA polymerase III (Pol III) transcribes genes encoding short

(<400 nt) structural RNAs that include tRNAs and 5S rRNA (Schramm and Hernandez, 2002).

RNA polymerases I, II, and III are composed of 12–17 proteins, including subunits sharing sequence and structural homology with the eubacterial RNA polymerase subunits β' , β , α' , α'' , and ω (Archambault and Friesen, 1993; Cramer et al., 2001; Zhang et al., 1999). RNA Pol I, II, and III (designated RPA, RPB, and RPC in yeast and N [nuclear] RPA, NRPB, and NRPC in *Arabidopsis*) largest subunits are homologous to eubacterial β' and are encoded by different genes, (*N*)RPA1, (*N*)RPB1, and (*N*)RPC1. Likewise, the second-largest subunits of Pol I, II, and III are β homologs encoded by (*N*)RPA2, (*N*)RPB2, and (*N*)RPC2. Together, the largest and second-largest subunits form the catalytic center in which RNA synthesis occurs (Cramer et al., 2000; Zhang et al., 1999), with α' , α'' , and ω serving regulatory or assembly functions.

Surprisingly, analysis of the *Arabidopsis thaliana* genome sequence revealed evidence for a fourth class of RNA polymerase in addition to Pol I, II, and III (CSP and Jonathan Eisen, discussed in *Arabidopsis Genome Initiative* [2000]). Specifically, two class IV largest and second-largest subunit genes were predicted, implying the existence of a nuclear RNA polymerase IV (Pol IV) distinct from eubacterial-type RNAPs of chloroplasts, from mitochondrial polymerase, or from RNA-dependent RNA polymerases (RdRP).

Here, we present evidence that RNA Pol IV is located within the nucleus and plays a role in heterochromatin formation. Dispersal of chromocenters in Pol IV mutants is correlated with the loss of cytosine methylation from pericentromeric 5S gene clusters and *AtSN1* retroelements. By contrast, methylation of constitutively heterochromatic 180 bp centromere core repeats is not appreciably affected in Pol IV mutants. We propose that Pol IV is required for the production of siRNAs that direct de novo methylation of repetitive elements that are subject to facultative heterochromatin formation, thereby facilitating higher-order heterochromatin associations.

Results

Genes for RNA Pol IV

An unrooted phylogenetic tree of DNA-dependent RNA polymerase (RNAP) largest subunits (Figure 1A) reveals distinct clades for eubacteria, cyanobacteria and chloroplasts, archaea, DNA viruses, and eukaryotic RNA polymerases I (RPA1), II (RPB1), and III (RPC1). *Arabidopsis thaliana* (At) Pol I, II, and III largest subunits group with their orthologs from rice (Os), yeast (Sp and Sc), *C. elegans* (Ce), *Drosophila* (Dm), and human (Hs). Unlike other eukaryotes, *Arabidopsis* and rice have additional genes (*NRPD1a* and *b*) that form a clade for a putative Pol IV.

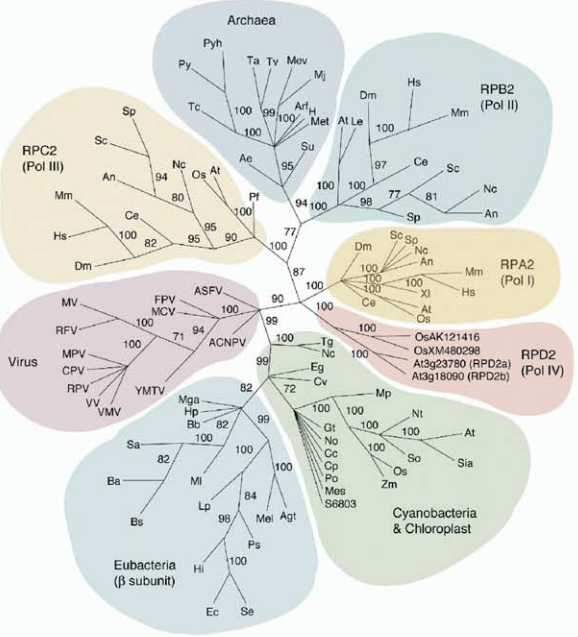
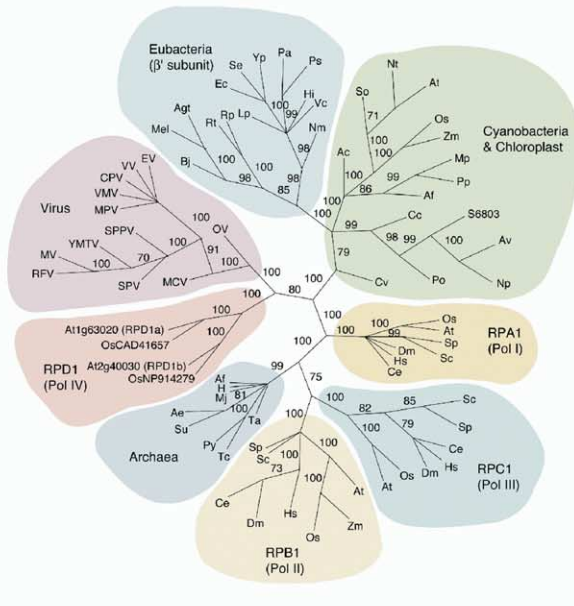
An unrooted tree of RNAP second-largest subunits resembles the tree for the largest subunits (Figure 1B). Again, in addition to clades for RPA2 (Pol I), RPB2 (Pol II), and RPC2 (Pol III), a plant-specific NRPD2 (Pol IV)

*Correspondence: pikaard@biology.wustl.edu

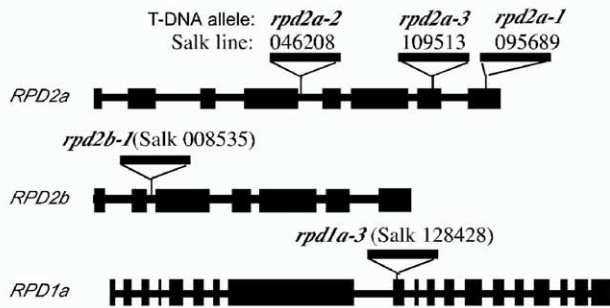
⁴These authors contributed equally to this work.

A RNAP largest subunits

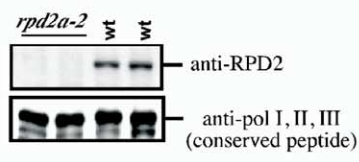
B RNAP second-largest subunits



C T-DNA disrupted alleles



D Immunoblot



E RPD2 immunolocalization

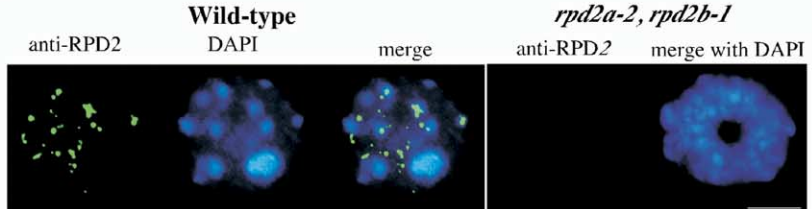


Figure 1. Evidence for RNA Pol IV in Plants

(A and B) Unrooted neighbor-joining phylogenies based on conserved domains A, C, D, and F of DNA-dependent RNA polymerase largest subunits and conserved domains A, C, D, F, G, H, and I of DNA-dependent RNA polymerase second-largest subunits. Bootstrap values are given for branch nodes. Species designations and GenBank accession numbers for the sequences analyzed are provided in [Tables S1](#) and [S2](#). (C) Diagrams of T-DNA-disrupted *nrdp2* and *nrdp1* alleles. Exons are denoted by black rectangles. (D) Immunoblot showing no detectable NRPD2 protein in two *nrdp2a-2* mutant individuals, unlike their wild-type siblings. A control immunoblot utilized an antibody raised against a peptide conserved in Pol I, II, and III second-largest subunits. (E) NRPD2 localizes to the nucleus. On the left is a wild-type interphase nucleus showing immunolocalization of NRPD2 relative to ten DAPI-positive chromocenters. On the right is a homozygous *nrdp2a-1 nrdp2b-1* nucleus. The dark, DAPI-negative region is the nucleolus. The wild-type and mutant plants were progeny of homozygous siblings. The size bar corresponds to 5 μ m. Arabidopsis pol IV subunit names are abbreviated from NRPD to RPD in this and all subsequent figures.

clade exists. In both *Arabidopsis* and rice, there are two *NRPD2* genes (*NRPD2a* and *NRPD2b*) that were apparently duplicated after monocots and dicots diverged.

Multiple alignments revealed that *NRPD2* proteins closely resemble their Pol I–III homologs, whereas *NRPD1* sequences frequently lack amino acids that are invariant in Pol I–III largest subunits, including amino acids near the active site (see [Figures S1–S4](#) in the [Supplemental Data](#) available with this article online). Therefore, we focused our studies on *NRPD2* but also subjected *nRPD1a* mutants to a subset of the same assays. *NRPD1b* was ignored because existing annotation suggested that this gene lacks essential C-terminal domains.

Only *NRPD2a* appears to be expressed in *Arabidopsis*, based on existing EST (cDNA) sequences and by our inability to amplify *NRPD2b* RNA using RT-PCR or 5' RACE. By contrast, *NRPD2a* sequences were readily amplified by PCR and by primer extension ([Figure S5](#)) to yield a full-length mRNA sequence (GenBank accession number AY862891).

Salk lines 046208, 109513, and 095689 contain the T-DNA-disrupted mutant alleles *nRPD2a-2*, *nRPD2a-3*, and *nRPD2a-1*, respectively. Salk lines 008535 and 128428 contain the *nRPD2b-1* and *nRPD1a-3* alleles ([Figure 1C](#)). Plants homozygous for these alleles were identified by PCR or Southern blot analysis of segregating families. The *nRPD2a* and *nRPD1a* alleles are all recessive and cause equivalent molecular phenotypes (data below and data not shown).

NRPD2 Expression and Nuclear Localization

RNA and protein blot analyses showed that *NRPD2a* is expressed throughout the plant but is most highly expressed in flowers and roots (data not shown). In homozygous *nRPD2a-2* mutants, no *NRPD2* protein is detectable ([Figure 1D](#)), indicating that *nRPD2a-2* is a null allele. Immunolocalization of *NRPD2* showed it to be a nuclear protein that is concentrated in numerous distinct foci ([Figure 1E](#)). Examination of 56 interphase nuclei revealed 10–15 *NRPD2* signals in 71% of the nuclei and fewer than ten signals in 29% of the nuclei. In the nucleus shown, there are ten prominent DAPI-positive heterochromatic chromocenters, which are made up of centromeric repeats for the ten chromosomes, dispersed pericentromeric repeats, and four NORs (nucleolus organizer regions) ([Fransz et al., 2002](#)). Approximately 15 *NRPD2* signals of varying size are apparent in [Figure 1E](#), five of which are located at chromocenters and five of which are at the edges of chromocenters. Similar association of *NRPD2* with chromocenters was observed in all nuclei.

Genetic Analysis of *NRPD* Mutants

To rule out any possible functional redundancy of *NRPD2a* and *NRPD2b*, we generated lines homozygous for both the *nRPD2a-2* and *nRPD2b-1* alleles, which was laborious, because the genes are linked (~10 cM genetic distance). We first crossed *nRPD2a-2* and *nRPD2b-1* homozygotes to generate F1 individuals that were hemizygous for each allele. The F1 was then outcrossed with a wild-type plant such that all resulting progeny had a wild-type chromosome 3 and either an *nRPD2a-2*

or an *nRPD2b-1* allele but not both, unless a meiotic recombination event occurred between the two genes. We then identified the latter rare recombinants that had one wild-type chromosome 3 and one chromosome 3 bearing both the *nRPD2a-2* and *nRPD2b-1* alleles, allowed these to self-fertilize, and genotyped their progeny. Plants homozygous for both *nRPD2a-2* and *nRPD2b-1* (referred to as *nRPD2* double mutants or simply *nRPD2* in the remainder of the paper) were recovered, demonstrating that *NRPD2* is nonessential for viability. Siblings that were homozygous for the wild-type *NRPD2* gene were also identified and used as controls in subsequent assays. This genetic strategy is likely to have segregated away any potential T-DNAs unlinked to *NRPD2*, but, if such T-DNAs persist, they are as likely in the wild-type control plants as in their double mutant siblings.

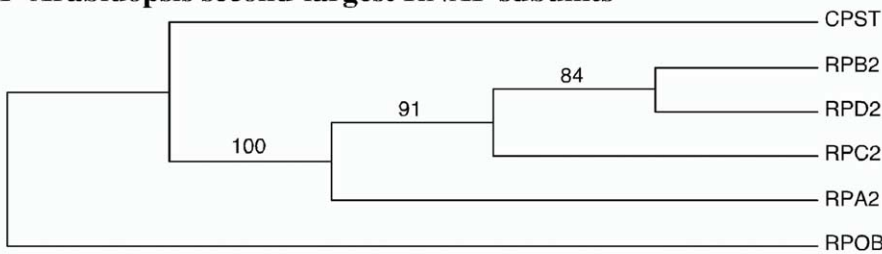
We tested whether *NRPD2* might be functionally redundant with the *NRPA2*, *NRPB2*, or *NRPC2* subunits of Pol I–III by asking if any of these subunits were nonessential. We identified hemizygous individuals bearing T-DNA insertions in *NRPA2*, *NRPB2*, or *NRPC2* and genotyped 60–80 of their progeny. Only homozygous wild-type and hemizygous progeny were obtained; no homozygous mutants were recovered (data not shown). These results indicate that *NRPA2*, *NRPB2*, and *NRPC2* are essential genes, unlike *NRPD2a* and *NRPD2b*, and that *NRPD2* genes do not complement *nRPA2*, *nRPB2*, or *nRPC2* mutations. The *nRPD2* double mutation also failed to induce haploinsufficiency in plants hemizygous for *nRPA2*, *nRPB2*, or *nRPC2* mutations, consistent with the interpretation that *NRPD2* does not overlap functionally with Pol I, II, or III.

***NRPD2* Does Not Copurify with DNA-Dependent RNA Polymerases I–III**

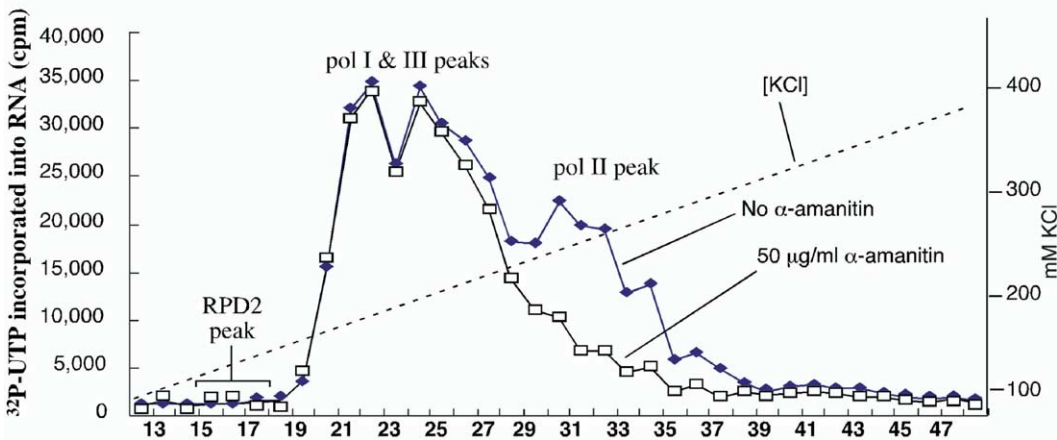
Among *Arabidopsis* RNAP second-largest subunits, *NRPD2* is most similar to *NRPB2* ([Figure 2A](#)). Therefore, we asked if *NRPD2* copurified with RNA Pol II activity, as might be expected if *NRPD2* is an alternative Pol II subunit. Nuclear extract was fractionated by anion exchange chromatography, and fractions were tested for DNA-dependent RNA polymerase activity ([Figure 2B](#)) and for the presence of *NRPD2*, *NRPB2*, or a 24 kDa polymerase subunit (RPB5) that is shared by Pol I, II, and III ([Larkin et al., 1999](#); [Saez-Vasquez and Piikaard, 2000](#)).

The DNA-dependent RNA polymerase assay measures the incorporation of radioactive nucleotide triphosphates into RNA using sheared template DNA, which allows polymerase initiation from broken DNA ends in a promoter-independent fashion ([Schwartz and Roeder, 1974](#)). Duplicate reactions were performed with and without α -amanitin, a potent inhibitor of RNA Pol II, and mean values were plotted ([Figure 2B](#)). Comparison of the RNA polymerase activity profiles reveals a peak of activity that is inhibited by α -amanitin (fractions 29–37), indicative of Pol II ([Figure 2B](#)). As expected, *NRPB2* eluted in these fractions ([Figure 2C](#)). By contrast, *NRPD2* eluted in fractions 15–18, suggesting that *NRPD2* is not an alternative Pol II subunit. Immunoblotting of column fractions using an antibody against the 24 kDa subunit that is shared by Pol I, II, and III revealed a good correspondence between the presence of the

A Arabidopsis second-largest RNAP subunits



B DEAE chromatography of DNA-dependent RNAP activity



C Immunoblotting of column fractions

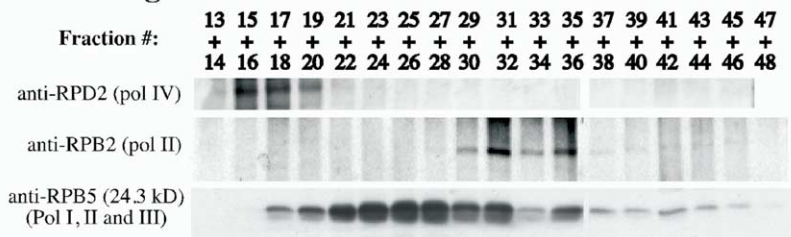


Figure 2. NRPD2 Does Not Cofractionate with Pol II or with DNA-Dependent RNA Polymerase Activity

(A) Neighbor-joining tree (with bootstrap values based on 1000 replications) for second-largest subunits of *Arabidopsis* chloroplast RNAP and RNA polymerases I, II, and III. The *E. coli* RpoB subunit serves as the outgroup.

(B) Fractionation of DNA-dependent RNA polymerase activity by DEAE-Sepharose chromatography. Fractions eluted with a linear KCl gradient were tested for RNA polymerase activity both with and without α -amanitin.

(C) Immunoblot detection of NRPD2, NRPB2, and NRPB5 in fractions eluted from the DEAE column.

Arabidopsis pol IV subunit names are abbreviated from NRPD to RPD in this and all subsequent figures.

24 kDa subunit and RNAP activity. Surprisingly, the peak fractions for NRPD2a displayed no detectable RNAP activity. We conclude that NRPD2 is not an alternative subunit of a conventional DNA-dependent RNA polymerase.

Heterochromatin Association Is Impaired in *nRPD2* Mutants

In *nRPD2* mutants, we noted an increased number and decreased size of DAPI-positive heterochromatic foci in interphase nuclei relative to wild-type siblings (Figure 1E), prompting further investigation. Histone H3 dimethylated on lysine 9 (H3^{dimethyl}K9) is a marker of heterochromatin (Richards and Elgin, 2002) that colocal-

izes with chromocenters in wild-type nuclei (Figure 3A). However, in *nRPD2* mutant siblings, the H3^{dimethyl}K9 signals are dispersed and colocalize with the numerous, small DAPI-positive foci (Figure 3A; Table S3).

Chromocenters involving NORs are relatively resistant to dispersal (Figure 3B). It is noteworthy that there are four NORs in a diploid nucleus, located at the tips of chromosomes 2 and 4. However, 36% of wild-type and 19% of *nRPD2* interphase nuclei show only two NOR fluorescence in situ hybridization (FISH) signals (as in Figure 3B) due to association of pairs of NORs and their linked centromeres. Nuclei with either three or four NOR FISH signals are also observed in wild-type and *nRPD2* mutants, but only *nRPD2* mutants frequently

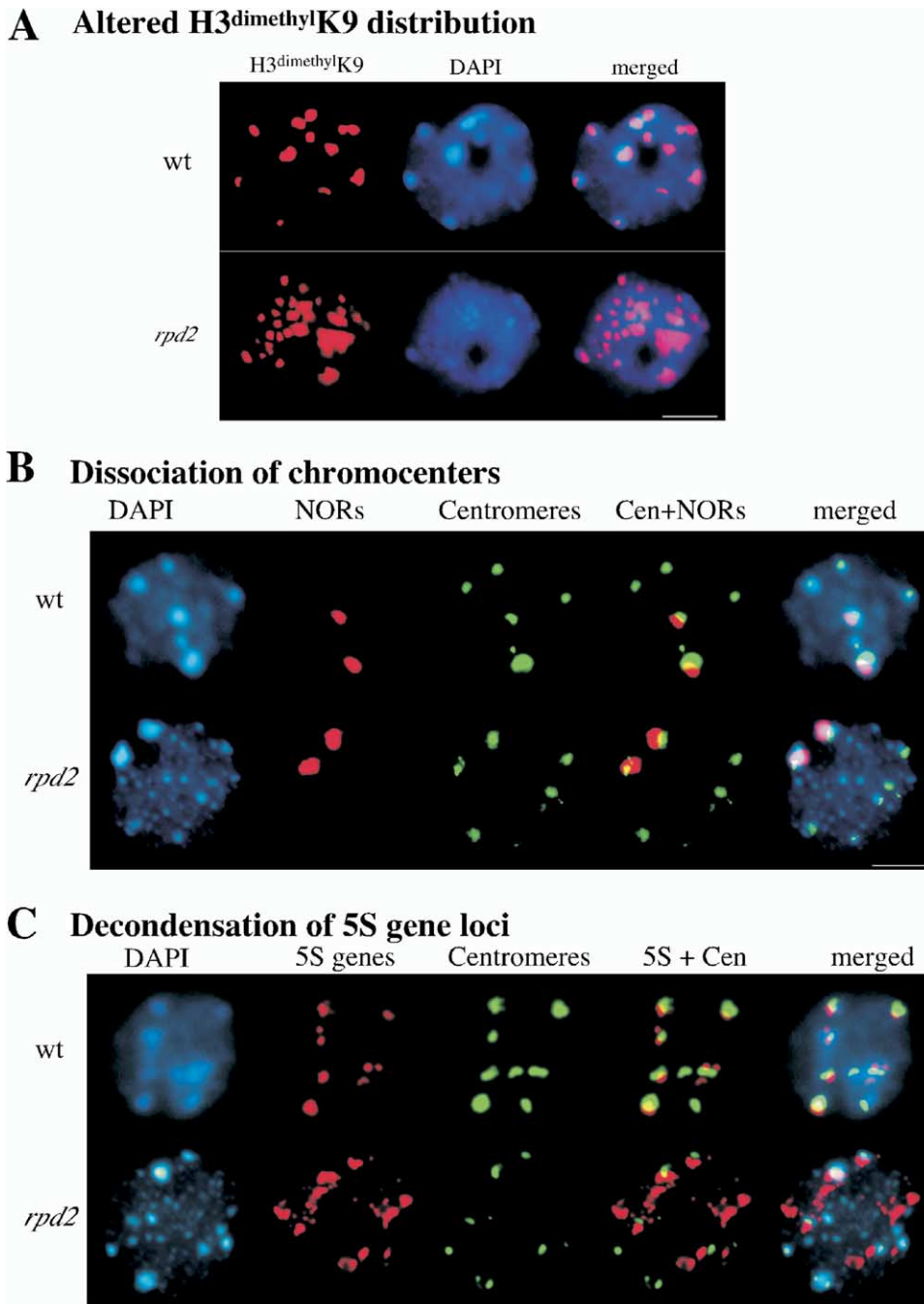


Figure 3. Heterochromatin Is Disrupted in *nprpd2* Mutants

(A) Immunolocalization of histone H3 dimethylated on lysine 9 in interphase cells of wild-type and the *nprpd2a-2 nprpd2b-1* mutant. Chromatin was counterstained with DAPI.

(B) Chromocenters containing NORs are relatively resistant to dispersal in *nprpd2a-2 nprpd2b-1* mutants. Centromeres and NORs (45S rRNA gene loci) were detected by FISH. Chromatin was counterstained with DAPI.

(C) 5S gene loci become decondensed and dissociated from centromeres in *nprpd2a-2 nprpd2b-1* double mutants. 5S genes and centromeres were detected by FISH. Wild-type and mutant plants were progeny of homozygous siblings. Size bars in all panels correspond to 5 μ m.

Arabidopsis pol IV subunit names are abbreviated from NRPD to RPD in this and all subsequent figures.

(23%) show >4 NOR signals (Table S3), presumably due to dissociation of facultative heterochromatin subdomains of the ~4 Mbp NORs.

5S rRNA gene repeats are tandemly arranged in peri-

centromeric regions of chromosomes 3, 4, and 5 in *Arabidopsis* ecotype Col-0 such that dual FISH typically reveals substantial overlap of 5S and 180 bp centromere repeat signals in wild-type cells (Figure 3C).

However, in *nRPD2* double mutant siblings, the 5S genes are typically decondensed and show significantly less ($p = 0.0012$) colocalization with centromeres, consistent with the interpretation that pericentromeric facultative heterochromatin is dispersed away from the constitutively heterochromatic centromeres (see Table S3 for quantitation).

Pol IV Participates in the siRNA-Chromatin Modification Pathway

Heterochromatin disruption and 5S gene dispersal in Pol IV mutants suggested a possible loss of cytosine methylation (Soppe et al., 2002). To determine if *nRPD2* or *nRPD1a* mutants affect 5S gene cytosine methylation, we performed Southern blotting using methylation-sensitive restriction endonucleases. HpaII and MspI cut CCGG motifs, but HpaII will not cut if the inner C is methylated, and MspI will not cut if the outer C is methylated (McClelland et al., 1994). HaeIII recognizes GGCC but won't cut if the inner C is methylated. Digestion of 5S genes with these three enzymes reports on methylation at CG (HpaII), CNG (MspI), and CNN (in the ecotype Col-0, the 5S HaeIII site is a CNN site). The Southern blots reveal ladders of bands at ~500 bp intervals (Figure 4A), the size of a 5S gene repeat (Campbell et al., 1992). High levels of methylation cause most of the hybridization signal to be near the top of the ladder, whereas loss of methylation results in more signal near the bottom.

5S gene methylation at HpaII, MspI, and HaeIII sites is decreased in *nRPD1a-3* and *nRPD2* mutants (Figure 4A, lanes 3, 5, 18, 20, 22, and 24) relative to their wild-type siblings (lanes 2, 4, 19, 21, 23, and 25), with HaeIII digestion showing the largest effect. Comparison of *nRPD1* and *nRPD2* to the DNA methylation mutants *ddm1*, *met1*, *cmt3*, and *drm1drm2* showed that HpaII digestion of 5S genes in *nRPD1* and *nRPD2* mutants occurred to the same extent as in a *drm1drm2* double mutant (compare lanes 3, 5, and 6) but to a lesser extent than in a *ddm1* (lane 10) or *met1* (lane 11) mutant. DRM2 is responsible for de novo methylation in all sequence contexts (CG, CNG, and CNN); DDM1 is involved in maintenance of methylation in all sequence contexts, and MET1 is primarily responsible for maintenance of CG methylation (reviewed in Bender [2004]). DRM1 has no known function. CMT3 is primarily responsible for maintenance of CNG methylation, so a *CMT3* mutant has little effect on HpaII digestion (lane 7) but has a profound effect on MspI digestion (lane 16). Collectively, the results indicate that Pol IV affects 5S gene methylation in all sequence contexts (CG, CNG, and CNN). Interestingly, the highly methylated 180 bp centromere repeats are unaffected by *nRPD1* and *nRPD2* mutations (Figure 4B), suggesting that Pol IV does not affect global cytosine methylation levels but acts on only a subset of methylated genomic sequences.

Methylation of *AtSN1*, a well-characterized retroelement family (Hamilton et al., 2002; Xie et al., 2004), was assayed using HaeIII digestion followed by PCR (Figure 4C) (Hamilton et al., 2002). If HaeIII sites are methylated, the DNA is not cut and can be amplified. However, if CNN methylation is lost at any of three HaeIII sites (see

diagram), HaeIII digestion precludes PCR amplification. In wild-type Col-0, Ler, or Ws (the genetic backgrounds for the mutants tested), *AtSN1* elements are heavily methylated and resistant to HaeIII cleavage. Methylation is unaffected by *met1* or *cmt3* mutants but is substantially reduced in a *drm1 drm2* double mutant, as expected for CNN methylation. HaeIII methylation is also disrupted in mutants of the heterochromatic siRNA pathway, including *rdr2* (RNA-dependent RNA polymerase 2), *hen1* (Hua enhancer 1), or *dcl3* (Dicer-like 3), consistent with published results (Xie et al., 2004). By contrast, *AtSN1* methylation is not diminished in a mutant of *DCL1*, the dicer responsible for miRNA production. Importantly, *AtSN1* methylation is also reduced in both *nRPD1* and *nRPD2* mutants. The loss of *AtSN1* methylation in both siRNA pathway mutants and *nRPD* mutants suggests that Pol IV might also affect siRNAs. Consistent with this hypothesis, 5S gene and *AtSN1* siRNAs are significantly reduced or eliminated in *nRPD2* and *nRPD1* mutants (Figures 4D and 4E) as in *hen1*, *rdr2*, *drm*, or *ago4* mutants, confirming prior studies (Herr et al., 2005; Xie et al., 2004; Zilberman et al., 2004). By contrast, mutations of the RNA-dependent RNA polymerases *rdr1* or *rdr6* (*sgs2*, also known as *sde1*) had no effect, though *rdr6* is known to function in RNA silencing of transgenes (Baulcombe, 2004). Interestingly, 5S siRNA levels were actually increased in *ddm1* and *met1* mutants (Figure 4D), indicating that disrupted maintenance of cytosine methylation is not the explanation for loss of 5S siRNAs in *nRPD1* and *nRPD2* mutants.

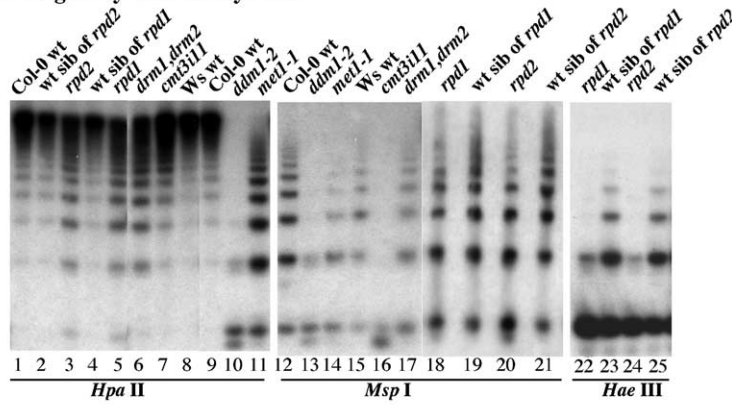
Importantly, miRNA levels are unaffected in *nRPD* mutants, as shown by comparison of miR163, 159, 164, 171, and 172 levels in mutant and wild-type siblings (Figure 4F), indicating that Pol IV acts only in the siRNA pathway and not in the miRNA pathway.

Discussion

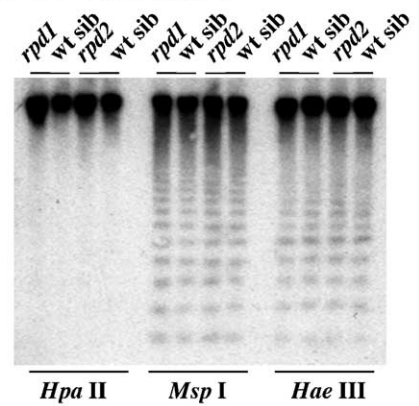
Loss of NRPD1 or NRPD2 function causes the loss of cytosine methylation at pericentromeric 5S genes and *AtSN1* retroelements yet has no discernible effect on centromere repeat methylation. These observations suggest that Pol IV primarily affects facultative heterochromatin rather than constitutive heterochromatin, consistent with the localization of NRPD2 at foci that overlap or are adjacent to chromocenters but are not fully coincident with chromocenters. We propose that Pol IV acts on genes that cycle between decondensed, euchromatic states and condensed, chromocenter-associated heterochromatic states, playing a key role in the amplification of siRNAs that direct cytosine methylation to these genes when they become activated (Aufsatz et al., 2002; Wassenegger, 2000).

Interestingly, the total amount of H3^{dimethyl}K9, a reliable marker of heterochromatin, does not appear to be reduced in Pol IV mutant nuclei. Instead, the H3^{dimethyl}K9 is simply dispersed into a larger number of heterochromatic foci. Collectively, these data, combined with data showing disruption of chromocenters in *ddm1* and *met1* mutants (Soppe et al., 2002), suggest that loss of cytosine methylation from either pericentromeric repeats or centromeric repeats is sufficient to disrupt

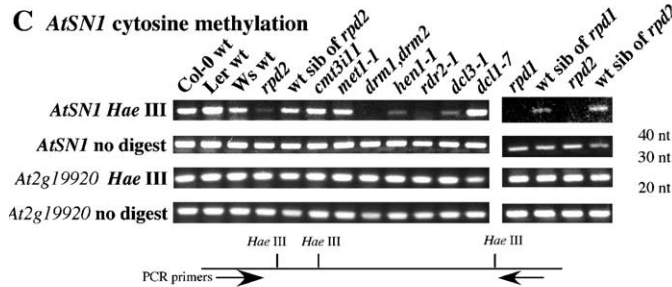
A 5S gene cytosine methylation



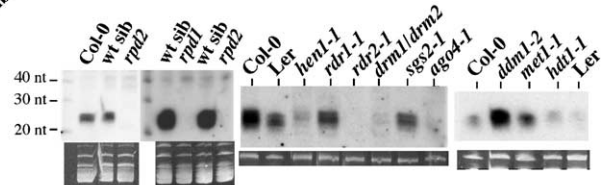
B Centromere repeats



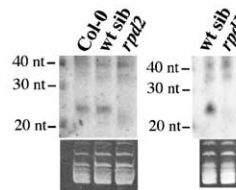
C *AtSN1* cytosine methylation



D 5S siRNAs



E *AtSN1* siRNAs



F miRNAs

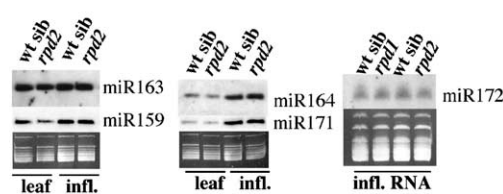


Figure 4. NRPD1 and NRPD2 Are Required for 5S Gene and *AtSN1* Cytosine Methylation and siRNA Accumulation

(A) Analysis of 5S gene repeats in *nrdp1a-3* and *nrdp2a-2 nrdp2b-1* double mutants relative to wild-type siblings and methylation mutants. Genomic DNA digested with HpaII, MspI, or HaeIII was hybridized to a 5S gene probe. *nrdp1*, *nrdp2*, *drm1*, and *met1* mutants are in the Col-0 genetic background; *drm1drm2* and *cmt3* are in the WS background.

(B) Methylation of 180 bp centromere repeats is apparently unaffected in *nrdp1* and *nrdp2* mutants relative to wild-type siblings.

(C) *nrdp1* and *nrdp2* mutations cause decreased *AtSN1* cytosine methylation. PCR was used to amplify a portion of an *AtSN1* retroelement that includes three HaeIII sites. Undigested DNA and a gene lacking HaeIII sites served as PCR controls.

(D) 5S siRNAs in *nrdp1*, *nrdp2*, and mutants affecting siRNA production. Small RNA blots were probed for 5S siRNA sequences. Ethidium-stained gel bands serve as loading controls. The *hdt1* mutant is an ecotype Col-0 line with a T-DNA insertion in a nucleolar histone deacetylase; it serves as a T-DNA control in the blot at far right.

(E) *AtSN1* siRNAs are reduced or eliminated in *nrdp1* and *nrdp2* mutants.

(F) miRNAs 159, 163, 164, and 171 are unaffected in *nrdp1* and *nrdp2* mutants.

Arabidopsis pol IV subunit names are abbreviated from NRPD to RPD in this and all subsequent figures.

higher-order heterochromatin association into chromocenters. One possibility is that methylcytosine binding domain proteins and/or their associated proteins might act as linkers or bridges that help bring together dispersed heterochromatin domains.

At 5S genes, Pol IV affects cytosine methylation in all sequence contexts (CG, CNG, and CNN). Importantly, CG, CNG, and CNN de novo methylation is accomplished by DRM methyltransferase activity (Cao et al., 2003; Cao and Jacobsen, 2002). DRM is also responsible for siRNA-directed DNA methylation (in all sequence contexts) in *Arabidopsis* (Cao et al., 2003). We

have shown that Pol IV and DRM activities are both needed for CNN methylation at *AtSN1* retroelements, as are genes of the siRNA pathway. These facts, combined with our demonstration that 5S and *AtSN1* siRNAs are essentially eliminated in Pol IV mutants, are most parsimonious with the hypothesis that Pol IV is involved in production of siRNAs that guide DRM-mediated cytosine methylation to repeated sequences complementary to the siRNAs (Chan et al., 2004). This would explain why loss of cytosine methylation in Pol IV mutants is most apparent at CNN (HaeIII in our experiments) sites, which would be dependent on continuous de novo

methylation due to the lack of a dedicated CNN maintenance methyltransferase (reviewed in Bender [2004]). By contrast, preexisting methylation at CG and CNG sites would be perpetuated by the MET1 and CMT3 maintenance methyltransferases, explaining the lesser effect of Pol IV or *drm* mutations on HpaII and MspI-sensitive 5S gene methylation (Figure 4A).

One could argue that DNA methylation is upstream of siRNA production, as suggested by the decrease in *AtSN1* siRNAs in *ddm1* and *met1* mutants (Lippman et al., 2003). However, this hypothesis does not fit with the fact that *ddm1* and *met1* cause dramatic decreases in 5S gene methylation yet actually increase 5S siRNA levels, possibly due to derepression of silenced 5S genes, thereby increasing the number of transcripts from which to generate dsRNAs and siRNAs. By contrast, Pol IV and *drm* mutations cause only modest decreases in total methylation yet essentially eliminate 5S siRNAs.

So how can loss of de novo methylation in a *drm* mutant eliminate siRNAs (Figure 4D) if siRNAs are upstream of de novo methylation? This apparent paradox might be explained if initial, primary siRNAs direct de novo methylation events that then trigger a massive amplification of siRNAs, and more extensive methylation, by a mechanism requiring Pol IV. Presumably, it is this second wave that yields the high levels of siRNAs and methylation that we detect. One possibility is that methylated DNA serves as the template for Pol IV-mediated transcription of aberrant RNAs. Another possibility is that methylation stalls elongating polymerases, as suggested by studies in *Neurospora* (Rountree and Selker, 1997), providing RDR2 with an opportunity to make dsRNAs from incomplete transcripts and leading to local production of aberrant RNAs or siRNAs that prime Pol IV transcription. Testing such hypotheses will be priorities for future studies.

Experimental Procedures

Plant Strains

Arabidopsis mutants *hen1-1*, *rdm2-1*, *dcl3-1*, and *dcl1-7* were provided by Jim Carrington. *met1-1* was provided by Eric Richards. *cmt3i11* was provided by Judith Bender. *sgs2-1* (alias *sde1*; *rdm6*) was provided by Herve Vaucheret. Salk T-DNA insertion lines and other mutants were obtained from the *Arabidopsis* Biological Resource Center (ABRC).

RNA and Immunoblot Analysis of NRPD2

RNA was isolated as described previously (Chen et al., 1998). RNA blots were hybridized to a probe generated by random priming of the *NRPD2a* 5' RACE cDNA product using standard methods (Sambrook and Russell, 2001). For immunoblotting, plant tissue was homogenized in SDS sample buffer (125 mM Tris-HCl [pH 6.8], 2% SDS, 10% glycerol, and 0.7 M β -mercaptoethanol) and 40 μ g of protein, determined using a BCA (bicinchoninic acid) protein assay kit (PIERCE), subjected to SDS-PAGE on a 7.5% gel, and electroblotted to a PVDF membrane. Anti-NRPD2 and anti-NRPB2 antisera were raised in rabbits against peptides DMDIDVKDLEEFEA and MEYNEYEPEEPQYVE of NRPD2a (At3g23780) and *A. thaliana* NRPB2 (At4g21710), respectively. Anti-Pol I+II+III rabbit antiserum was raised against peptide GDKFSSRHGQKG, which is conserved in Pol I, II, and III second-largest subunits. Sera were affinity purified using peptides covalently linked to NHS-activated Sepharose resin (Pharmacia Biotech). Columns were washed with 3–5 column volumes of PBS (pH 7.0), 0.05% Tween-20; antibodies were eluted using 0.1 M glycine-HCl (pH 3.0) neutralized by addition of Tris-HCl

(pH 8.0) and stored at -80°C . Antisera were diluted 1:250 for probing immunoblots. The secondary antibody, diluted 1:5000, was peroxidase-linked donkey anti-rabbit IgG (Amersham). Immunoblots were visualized by chemiluminescence (ECL Western Blotting Detection kit; Amersham).

Screening of T-DNA Knockout Lines

T-DNA insertions in *NRPD2a*, *NRPD2b*, and *NRPD1a* were verified by PCR and sequencing using a T-DNA left border primer (5'-CGTCCGCAATGTGTTAAG-3') and primers specific for *NRPD2a*, *NRPD2b*, or *NRPD1a* as suggested by the suppliers of the Salk lines. Screening by Southern blot analysis was according to standard methods (Sambrook and Russell, 2001).

Anion Chromatography and DNA-Dependent RNA

Polymerase Assay

Arabidopsis plants were grown for 10 days at 25°C in 3 liter flasks containing 1 liter of liquid 1 \times Gamborg B5 medium, 1 \times Gamborg vitamins (Sigma), and 2% sucrose shaken at moderate speed. Tissue (200 g) was homogenized, and crude nuclear proteins were fractionated by DEAE-Sepharose chromatography and tested for RNA polymerase activity as described previously (Saez-Vasquez and Pikaard, 1997).

Phylogenetic Analyses

RNAP subunits were identified by blastp searches using *E. coli* RPOC and RPOB, *S. cerevisiae* RPB1 and RPB2, and *A. thaliana* NRPD1a and NRPD2a protein sequences. Sequences were aligned, using Clustal X (version 1.81). Conserved sequences were highlighted using BOXSHADE. (<http://bioweb.pasteur.fr/seqanal/interfaces/boxshade.html>). Phylogenetic analysis was by the neighbor-joining method, with 1000 bootstrap replications, using PAUP (version 4.0b10).

Cytosine Methylation Assays

Genomic DNA (100 ng) was digested with HpaII, MspI, or HaeIII. Following agarose gel electrophoresis, DNA was blotted to uncharged nylon membranes. Probes were generated by random priming, and blots were hybridized using standard methods (Sambrook and Russell, 2001).

AtSN1 methylation assays used ~ 100 ng of DNA digested with HaeIII (or undigested for controls). Approximately 5% of digestion reaction DNA was then used for each PCR reaction. PCR conditions were 2 min at 94°C , followed by 35 cycles of 94°C for 30 s, 53°C for 30 s, and 72°C for 30 s. Primer sequences for *AtSN1* were the following: 5'-ACTTAATTAGCACTCAAATTAACAAAATAAGT-3' and 5'-TTTAAACATAAGAAGAGTTCCTTTTCATCTAC-3'. The *At2g19920* control was amplified using 5'-TCACCCGAACAGTTGGAAGAA GAG-3' and 5'-GTGAGGAACCGTCCATTATTGCT-3'. PCR products were subjected to agarose gel electrophoresis.

In Situ Hybridization and Immunolocalization

Emerging leaves of 21-day-old plants were fixed in ethanol:acetic acid (3:1, v/v). Nuclei were prepared as described (Schwarzacher and Mosgoeller, 2000). FISH using biotin-dUTP or digoxigenin-dUTP labeled 180 bp *A. thaliana* pericentromeric repeat, 5S gene or 45S rRNA gene intergenic spacer sequence probes was as described previously (Pontes et al., 2004).

For immunolocalization experiments, nuclei were fixed in 4% paraformaldehyde. H3^{dimethyl}K9 was localized using published methods (Houben et al., 1996) with antibody purchased from Upstate Biotechnology. For NRPD2, slides were permeabilized with 10% DMSO, 3% NP-40 in PBS, before blocking with 1% BSA in PBS. Primary antibodies were diluted 1:100 in PBS, 1% BSA, and slides were incubated overnight at 4°C . Secondary antibodies were conjugated to rhodamine or fluorescein (Sigma). Chromatin was counterstained with DAPI in antifade buffer (Vector Laboratories). Nuclei were examined using a Nikon Eclipse E600 epifluorescence microscope and images collected using a Q-Imaging Retiga EX digital camera.

siRNA and miRNA Detection

RNA was isolated using the mirVana miRNA isolation kit (Ambion). RNA (2–6 μ g) was resolved by denaturing polyacrylamide gel electrophoresis on a 20% (w/v) gel. Gels were electroblotted (20 mA/cm² for 2 hr) to Magnacharge nylon membranes (0.22 μ m; Osmonics) using a semidry transfer apparatus. An end-labeled RNA ladder was used as a molecular weight marker (Decade Marker System, Ambion). The ATSN1 riboprobe was synthesized from a NdeI-linearized plasmid DNA template (Zilberman et al., 2003). All other riboprobes were generated according to the mirVana probe construction kit (Ambion) using oligonucleotides specific for a given small RNA and labeling by T7 polymerase transcription in the presence of α -³²P CTP. DNA oligonucleotides for 5S and miRNA probes were the following: siR1003T7 (5S) (5'-AGACCGTGAGGCCAACTTGG CATctgtctc-3'); small letters are complementary to the T7 promoter oligonucleotide), miR159T7 (5'-TTTGGATTGAAGGGAGCTC TAcctgtctc-3'), miR163T7 (5'-TTGAAGAGGACTTGAACCTTGGAT cctgtctc-3'), and miR164T7 (5'-TGGAGAAGCAGGGCACGTGCA cctgtctc-3'). Unincorporated nucleotides were removed using Perfora DTR Gel Filtration Cartridges (Edgebiosystems). Blot hybridization was in 50% formamide, 0.25 M Na₂HPO₄ (pH 7.2), 0.25 M NaCl, 7% SDS at 42°C (14–16 hr) followed by two 15 min washes at 37°C in 2 \times SSC, two 15 min washes at 37°C in 2 \times SSC, 0.1% SDS, and a 10 min wash in 0.5 \times SSC, 1% SDS.

Supplemental Data

Supplemental Data include five figures, three tables, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at <http://www.cell.com/cgi/content/full/120/5/613/DC1/>.

Acknowledgments

We are indebted to David Baulcombe for sharing his lab's unpublished evidence that *sde4* is *nrpd1a*. We thank the Carrington, Richards, Vaucheret, and Bender labs for mutant seed stocks; the Carrington lab for cloned *AtSN1*; Thomas Guilfoyle for the anti-RPB5 antibody; M.E. Gifford for phylogenetic analysis advice; Douglas Chalker for use of his fluorescence microscope; and our labmate Keith Earley for RNAP assay advice. We thank Dr. Wanda Viegas for comentoring P.C.N. This research was supported by NIH grant R01-GM60380 and USDA grant 99-35301-7865 (to C.S.P.). P.C.N. was supported by a predoctoral fellowship from the Portuguese Fundação para a Ciência e Tecnologia (SFRH/BD/6520/2001) and the Luso-American Foundation (grant 237/2004).

Received: December 23, 2004

Revised: January 24, 2005

Accepted: February 4, 2005

Published online: February 10, 2005

References

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

Archambault, J., and Friesen, J.D. (1993). Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol. Rev.* 57, 703–724.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J., and Matzke, M. (2002). RNA-directed DNA methylation in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA Suppl.* 99, 16499–16506.

Baulcombe, D. (2004). RNA silencing in plants. *Nature* 431, 356–363.

Bender, J. (2004). Chromatin-based silencing mechanisms. *Curr. Opin. Plant Biol.* 7, 521–526.

Campbell, B.R., Song, Y., Posch, T.E., Cullis, C.A., and Town, C.D. (1992). Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. *Gene* 112, 225–228.

Cao, X., and Jacobsen, S.E. (2002). Role of the *Arabidopsis* DRM

methyltransferases in de novo DNA methylation and gene silencing. *Curr. Biol.* 12, 1138–1144.

Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* 13, 2212–2217.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. *Science* 303, 1336.

Chen, Z.J., Comai, L., and Pikaard, C.S. (1998). Gene dosage and stochastic effects determine the severity and direction of uniparental rRNA gene silencing (nucleolar dominance) in *Arabidopsis* allopolyploids. *Proc. Natl. Acad. Sci. USA* 95, 14891–14896.

Cramer, P., Bushnell, D.A., Fu, J., Gnat, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288, 640–649.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* 292, 1863–1876.

Fransz, P., De Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. (2002). Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. USA* 99, 14584–14589.

Grumt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev.* 17, 1691–1702.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. *EMBO J.* 21, 4671–4679.

Herr, A., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science*, in press. Published online February 3, 2005. 10.1126/science.1106910.

Houben, A., Belyaev, N.D., Turner, B.M., and Schubert, I. (1996). Differential immunostaining of plant chromosomes by antibodies recognizing acetylated histone H4 variants. *Chromosome Res.* 4, 191–194.

Larkin, R.M., Hagen, G., and Guilfoyle, T.J. (1999). *Arabidopsis thaliana* RNA polymerase II subunits related to yeast and human RPB5. *Gene* 231, 41–47.

Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* 1(3) e67 DOI: 10.1371/journal.pbio.0000067.

McClelland, M., Nelson, M., and Raschke, E. (1994). Effect of site-specific methylation on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res.* 22, 3640–3659.

Pontes, O., Neves, N., Silva, M., Lewis, M.S., Madlung, A., Comai, L., Viegas, W., and Pikaard, C.S. (2004). Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. *Proc. Natl. Acad. Sci. USA* 101, 18240–18245.

Richards, E.J., and Elgin, S.C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* 108, 489–500.

Rountree, M.R., and Selker, E.U. (1997). DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*. *Genes Dev.* 11, 2383–2395.

Saez-Vasquez, J., and Pikaard, C.S. (1997). Extensive purification of a putative RNA polymerase I holoenzyme from plants that accurately initiates rRNA gene transcription in vitro. *Proc. Natl. Acad. Sci. USA* 94, 11869–11874.

Saez-Vasquez, J., and Pikaard, C.S. (2000). RNA polymerase I holoenzyme-promoter interactions. *J. Biol. Chem.* 275, 37173–37180.

Sambrook, J., and Russell, D.R. (2001). *Molecular Cloning: A Laboratory Manual*, Third Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* 16, 2593–2620.

Schwartz, L.B., and Roeder, R.G. (1974). Purification and subunit structure of deoxyribonucleic acid-dependent ribonucleic acid polymerase I from mouse myeloma, MOPC 315. *J. Biol. Chem.* 249, 5898–5906.

Schwarzacher, H.G., and Mosgoeller, W. (2000). Ribosome biogenesis in man: current views on nucleolar structures and function. *Cytogenet. Cell Genet.* 97, 243–252.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J.* 21, 6549–6559.

Wassenegger, M. (2000). RNA-directed DNA methylation. *Plant Mol. Biol.* 43, 203–220.

Woychik, N.A., and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. *Cell* 108, 453–463.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kassachau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* 2(5), e104 DOI: 10.1371/journal.pbio.0020104.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98, 811–824.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299, 716–719.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* 14, 1214–1220.

Accession Numbers

The GenBank accession number for the NRPD2a mRNA sequence determined for this paper is AY862891.

Note Added in Proof

In the early online version of the article, the genes NRPD1a, NRPD1b, NRPD2a, and NRPD2b were named RPD1a, RPD1b, RPD1a, and RPD2b, respectively. We have changed the names due to a nomenclature conflict.

Supplemental Data

Plant Nuclear RNA Polymerase IV Mediates

siRNA and DNA Methylation-Dependent

Heterochromatin Formation

Yasuyuki Onodera, Jeremy R. Haag,
Thomas Ream, Pedro Costa Nunes,
Olga Pontes, and Craig S. Pikaard

I. Phylogenetic Analyses

Species whose subunit sequences are included in the unrooted trees of Figure 1 are the following: Ac, *Adiantum capillus-veneris*; ACNPV, *Autographa californica nucleopolyhedrovirus*; Af, *Anthoceros formosae*; Agt, *Agrobacterium tumefaciens*; An, *Aspergillus nidulans*; Ap, *Aquifex pyrophilus*; Arf, *Archaeoglobus fulgidus*; ASFV, *African swine fever virus*; At, *Arabidopsis thaliana*; Av, *Anabaena variabilis*; Ba, *Bacillus anthracis*; Bb, *Borrelia burgdorferi*; Bj, *Bradyrhizobium japonicum*; Bs, *Bacillus subtilis*; Cc, *Cyanidium caldarium*; Ce, *Caenorhabditis elegans*; Cp, *Cyanophora paradoxa*; CPV, *Cowpox virus*; Cv, *Chlorella vulgaris*; Dm, *Drosophila melanogaster*; Ec, *Escherichia coli*; Eg, *Euglena gracilis*; EV, *Ectromelia virus*; FPV, *Fowlpox virus*; Gt, *Guillardia theta*; H, *Halobacterium salinarum*; Hi, *Haemophilus influenzae*; Hp, *Helicobacter pylori*; Hs, *Homo sapiens*; Le, *Lycopersicon esculentum*; Lp, *Legionella pneumophila*; MCV, *Molluscum contagiosum virus*; Mel, *Mesorhizobium loti*; Mes, *Mesostigma viride*; Met, *Methanothermobacter thermautotrophicus*; Mev, *Methanococcus vannielii*; Mg, *Mycoplasma genitalium*; Mga, *Mycoplasma gallisepticum*; Mj, *Methanocaldococcus jannaschii*; Ml, *Mycobacterium leprae*; Mm, *Mus musculus*; Mp, *Marchantia polymorpha*; MPV, *Monkeypox virus*; Mt, *Mycobacterium tuberculosis*; MV, *Myxoma virus*; Nc, *Neurospora crassa*; Nca, *Neospora caninum*; Nm, *Neisseria meningitidis*; No, *Nephroselmis olivacea*; Np, *Nostoc punctiforme*; Nt, *Nicotiana tabacum*; Os, *Oryza sativa*; OV, *Orf virus*; Pa, *Pseudomonas aeruginosa*; Pf, *Plasmodium falciparum*; Po, *Porphyra purpurea*; Pp, *Physcomitrella patens*; Ps, *Pseudomonas syringae*; Py, *Pyrococcus abyssi*; Pyh, *Pyrococcus horikoshii*; RFV, *Rabbit fibroma virus*; Rp, *Rickettsia prowazekii*; RPV, *Rabbitpox virus*; Rt, *Rickettsia typhi*; S6803, *Synechocystis sp. PCC 6803*; Sa, *Staphylococcus aureus*; Sc, *Saccharomyces cerevisiae*; Se, *Salmonella enterica*; Sia, *Sinapis alba*; So, *Spinacia oleracea*; Sp, *Schizosaccharomyces pombe*; SPPV, *Sheeppox virus*; SPV, *Swinepox virus*; Su, *Sulfolobus acidocaldarius*; Ta, *Thermoplasma acidophilum*; Tc, *Thermococcus celer*; Tg, *Toxoplasma gondii*; Tv, *Thermoplasma volcanium*; Vc, *Vibrio cholerae*; VMV, *Variola major virus*; VV, *Vaccinia virus*; Xl, *Xenopus laevis*; YMTV, *Yaba monkey tumor virus*; Yp, *Yersinia pestis*; Zm, *Zea mays*.

Additional Methods for Phylogenetic Analyses

Second-largest subunits in some of the archaea and largest subunits in archaea and chloroplasts display a split domain architecture (Bergslund and Haselkorn, 1991; Puhler et al., 1989; Schneider and Hasekorn, 1988). In these cases, sequences were joined and aligned in Clustal X (version 1.81) to fit the domain architecture of *E. coli* and *S. cerevisiae* protein sequences in order to facilitate phylogenetic comparisons. The annotated sequence for At2g40030 (*RPD1b*) present in Genbank lacks conserved C-terminal domains G and H, and was not studied functionally due to the presumption that it would be non-functional. However, our own analysis of the genomic sequence using TWINSCAN (<http://www.genes.cs.wustl.edu>) revealed part of domain G in what is currently annotated as an intergenic region and the remainder of the predicted protein can be found in a predicted neighboring gene, *At2g40040*, suggesting that the existing annotation is incorrect. We used our own annotation for *A. thaliana* RPD1b in the phylogenetic analysis shown in Figure 1. The annotated sequence for *O. sativa* RPD1a (CAD41657) also appeared to be

inaccurate in parts after alignment, so the genomic sequence was analyzed using FGENESH+ (www.softberry.com) with *O. sativa* RPD1b as a reference sequence in order to perform gene finding with similarity. The sequences were aligned and a final prediction for *O. sativa* RPD1a was used in the phylogenetic analyses.

Arabidopsis RPD1a is 30% identical (42% similar) to rice OsCAD41657, but only 14% identical (23% similar) to *Arabidopsis RPD1b*. The higher similarity among orthologs between species than among paralogs within a species indicates that two *RPD1* genes existed prior to the divergence of monocots and dicots ~200 million years ago (Wolfe et al., 1989).

The *Arabidopsis* RPD2a protein is 84% identical to the predicted *Arabidopsis* RPD2b open reading frame and 55% identical to rice OsAK121416.

Tables S1 and S2. GenBank Accessions for the DNA-Dependent RNA Polymerase Largest Subunits Analyzed in Figure 1**Supplemental Table 1 - RNAP Largest Subunit Sequences**

Category	Genbank Accession	Abbreviation	Organism	Gene/Locus	Protein
Pol IV	NM_104980	At1g63020	Arabidopsis thaliana	At1g63020	RPD1a
	NM_129561	At2g40030	Arabidopsis thaliana	At2g40030	RPD1b
	XP_473570	OsXP473570	Oryza sativa	CAD41657	RPD1a
	NP_914279	OsNP914279	Oryza sativa	AP004365	RPD1b
Pol I	NM_115626	AtRpal	Arabidopsis thaliana	At3g57660	RPA1
	J03530	ScRpal	Saccharomyces cerevisiae	YSCPOLA1	
	NM_079019	DmRpal	Drosophila melanogaster		
	AAC99959	HsRpal	Homo sapiens		
	NP_496872	CeRpal	Caenorhabditis elegans		
	NP_496872	OsRpal	Oryza sativa		
J50080	SpRpal	Schizosaccharomyces pombe			
Pol II	NM_119746	AtRpb1	Arabidopsis thaliana	At4g35800	RPB1
	NM_078569	DmRpb1	Drosophila melanogaster		
	X03128	ScRpb1	Saccharomyces cerevisiae	SCRPO21	
	CAA45125	HsRpb1	Homo sapiens		
	NP_500523	CeRpb1	Caenorhabditis elegans		
	AAQ08515	ZmRpb1	Zea mays		
	XP_493925	OsRpb1	Oryza sativa		
NP_595673	SpRpb1	Schizosaccharomyces pombe			
Pol III	NP_595673	AtRpc1	Arabidopsis thaliana	At5g60040	RPC1
	X03129	ScRpc1	Saccharomyces cerevisiae	SCRPO31	
	AF021351	HsRpc1	Homo sapiens		
	NM_132843	DmRpc1	Drosophila melanogaster		
	NP_501127	CeRpc1	Caenorhabditis elegans		
	NP_501127	OsRpc1	Oryza sativa		
	O94666	SpRpc1	Schizosaccharomyces pombe		
Eubacteria	AAC43086	EcRpoC	Escherichia coli K12	rpoC	RPOC
	NP_457916	SeRpoC	Salmonella enterica		
	NP_252959	PaRpoC	Pseudomonas aeruginosa		
	YP_026389	BaRpoC	Bacillus anthracis		
	NP_215182	MtRpoC	Mycobacterium tuberculosis		
	NP_073010	MgRpoC	Mycoplasma genitalium		
	NP_438672	HiRpoC	Haemophilus influenzae		
	NP_220532	RpRpoC	Rickettsia prowazekii		
	CAA61517	SaRpoC	Staphylococcus aureus		
	CAA52958	ApRpoC	Aquifex pyrophilus		
	NP_994402	YpRpoC	Yersinia pestis biovar Medievalis str. 91001		
	NP_229983	VcRpoC	Vibrio cholerae O1 biovar eltor str. N16961		
	ZP_00123798	PsRpoC	Pseudomonas syringae pv. syringae B728a		
	YP_094367	LpRpoC	Legionella pneumophila subsp. pneumophila str. Philadelphia 1		
	NP_282991	NmRpoC	Neisseria meningitidis Z24		
	NP_102111	MelRpoC	Mesorhizobium loti MAFF303099		
	NP_772049	BjRpoC	Bradyrhizobium japonicum USDA 110		
	NP_354930	AgtRpoC	Agrobacterium tumefaciens str. C58		
	YP_067097	RtRpoC	Rickettsia typhi str. Wilmington		
	Archaea	CAA47723	Tc	Thermococcus celer	rpoA1
CAA47724		Tc	Thermococcus celer	rpoA2	
NP_126306		Py	Pyrococcus abyssi	rpoA1	
NP_126307		Py	Pyrococcus abyssi	rpoA2	
NP_248036		Mj	Methanocaldococcus jannaschii DSM 2661	rpoA1	
NP_248037		Mj	Methanocaldococcus jannaschii DSM 2661	rpoA2	
NP_444249		H	Halobacterium	rpoA1	
P15354		H	Halobacterium	rpoA2	
NP_148215		Ae	Aeropyrum pernix	rpoA1	
NP_148214		Ae	Aeropyrum pernix	rpoA2	
NP_070713		Af	Archaeoglobus fulgidus	rpoA1	
NP_070714		Af	Archaeoglobus fulgidus	rpoA2	
CAA48281		Ta	Thermoplasma acidophilum	rpoA1	
CAA48282		Ta	Thermoplasma acidophilum	rpoA2	
P11512		Su	Sulfolobus acidocaldarius	rpoA1	
P11514		Su	Sulfolobus acidocaldarius	rpoA2	
Virus		NP_044030	MCV	Molluscum contagiosum virus	
	O57204	VV	Vaccinia virus		
	AAF14956	MV	Myxoma virus		
	AAF17950	RFV	Rabbit fibroma virus		
	AAR07427	YMTV	Yaba monkey tumor virus		
	T28521	VMV	Variola major virus		
	CAD90647	CPV	Cowpox		
	AAL69807	SPV	Swinepox virus		
	NP_659643	SPPV	Sheeppox virus		
	AAL40548	MPV	Monkeypox virus		
	AAM92386	EV	Ectromelia virus		
	NP_957833	OV	Orf virus		

Archaea	CAA32924	Su	Sulfolobus acidocaldarius	rpoB		
	CAA47722	Tc	Thermococcus celer	rpoB		
	NP_248034	Mj	Methanocaldococcus jannaschii DSM 2661	rpoB2		
	NP_248035	Mj	Methanocaldococcus jannaschii DSM 2661	rpoB1		
	NP_281214	H	Halobacterium	rpoB2		
	NP_281213	H	Halobacterium	rpoB1		
	NP_148216	Ae	Aeropyrum pernix K1	rpoB		
	NP_126305	Py	Pyrococcus abyssi GE5	rpoB		
	NP_143407	Pyh	Pyrococcus horikoshii OT3	rpoB		
	CAA51726	Mev	Methanococcus vannielii	rpoB2		
	CAA51727	Mev	Methanococcus vannielii	rpoB1		
	NP_070711	Arf	Archaeoglobus fulgidus DSM 4304	rpoB2		
	NP_070712	Arf	Archaeoglobus fulgidus DSM 4304	rpoB1		
	NP_276179	Met	Methanothermobacter thermautotrophicus str. Delta H ⁺	rpoB2		
	NP_276180	Met	Methanothermobacter thermautotrophicus str. Delta H ⁺	rpoB1		
	NP_111701	Tv	Thermoplasma volcanium GSS1	rpoB		
	NP_393870	Ta	Thermoplasma acidophilum DSM 1728	rpoB		
	Viruses	AAC55257	MCV	Molluscum contagiosum virus		
		AAO89423	VV	Vaccinia virus		
		AAF15002	MV	Myxoma virus		
AAF17997		RFV	Rabbit fibroma virus			
AAR07472		YMTV	Yaba monkey tumor virus			
T28566		VMV	Variola major virus			
AAM13599		CPV	Cowpox virus			
S78061		ASFV	African swine fever virus			
AAL40593		MPV	Monkeypox virus			
YP_006777		RPV	Rabbitpox virus			
CAE52727		FPV	Fowlpox virus (isolate HP-438[Munich])			
AAA66680		ACNPV	Autographa californica nucleopolyhedrovirus			
Cyanobacteria & Chloroplast		BAA84377	AtCPST	Arabidopsis thaliana (CPST)	rpoB	RPOB
		Q9TL06	NoCPST	Nephroselmis olivacea (CPST)	rpoB	
	P11703	SoCPST	Spinacia oleracea (CPST)	rpoB		
	P06271	NiCPST	Nicotiana tabacum (CPST)	rpoB		
	P46818	SiaCPST	Sinapis alba (CPST)	rpoB		
	CAA60276	ZmCPST	Zea mays (CPST)	rpoB		
	NP_039373	OsCPST	Oryza sativa (CPST)	rpoB		
	RNLVB	MpCPST	Marchantia polymorpha (liverwort) CPST	rpoB		
	Q9MUS5	MesCPST	Mesostigma viride (CPST)	rpoB		
	BAA57969	CvCPST	Chlorella vulgaris (green algae) CPST	rpoB		
	CAA50138	EgCPST	Euglena gracilis (CPST)	rpoB		
	AAC35676	GiCPST	Guillardia theta (CPST)	rpoB		
	AAC08138	PoCPST	Porphyra purpurea (CPST)	rpoB		
	NP_045031	CcCPST	Cyanidium caldarium (CPST)	rpoB		
	NP_043230	CpPST	Cyanophora paradoxa (PST)	rpoB		
	AAD17842	TgPST	Toxoplasma gondii (PST)	rpoB		
	AAF14261	NcaPST	Neospora caninum (PST)	rpoB		
	NP_440685	S6803	Synechocystis sp. PCC 6803	rpoB		
	Other	NP_701431	Pf	Plasmodium falciparum 3D7		

RNA polymerase subunits are categorized according to clade designations.

II. Protein Alignments

Supplemental Figure 1. Multiple Alignment of RPD1 with DNA-Dependent RNA Polymerase Largest Subunits of *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec)

Supplemental Figure 1. Alignment of RNAP Largest Subunits

```

At_RPD1  -----M--EDDCEELQVPVGTLTLSIGFSISNNNDRDKMSVLEV-----
At_RPB1  -----MDTRFPFSPAEVSKVRVVFQFGLSPDEIRQMSVIHVEHSETTEK--GK
At_RPC1  ---METKMEIEFTKKPYIEDVGPLKIKSINFSVLSDLVEMKAAEVQWNIGLYDHS-FK
At_RPA1  MAHAQTTEVCLSFHRSLLPMPGASQVVESVRFMSFTEQDVRKHSFLKVTSEFILHDNV-GN
Sc_RPB1  -----MVGQQYSSAPLRTVKEVQFGLFSPEEVRAISVAKIRFPETMDETQTR
Ec_RPOC  -----MKDLLKFLKAQTKTEEFDAIKIALASPD MIRSWSFGEVKKPETINRYTFK
consensus      -   K   F   A   V   VKSIQFSILSPDEVKMSVL V   PET D   K

Conserved domain A
At_RPD1  EAPNQVTLDSRLGLENPDSVCRITCGSKDRKVC EGHFGVINFAYSI INPYFLKEVAALLNKI
At_RPB1  PKVCGLSLTRLGTIDRKVKCEETCMAN-MAECPGHFGYLELAKPMYHVGFMKTVLSIMRCV
At_RPC1  PYENGLLDPRMCPENKKSICTTCEGN-FQNCPGHYGYLKLDPVYNVGYFNFILDLKCI
At_RPA1  EFPGLYDLKLGPKDDKQACNSCGQL-KLACPGHCGHIELVFPYIYHPLLFNLLFNFLQRA
Sc_RPB1  AKIGGLNDPRLGSIDRNLCQTCQEG-MNECPGHFGHIDLAKPVFHVGFIAKIKKVCQV
Ec_RPOC  EERDGLFCARIFGEVVKDYELCGKYK-RLK---HRGVICEKCGEVTQTKVRRERMGHIE
consensus P   GGL D RLG PDKK   C TC   R   CPGHFG IELA PVYHVGFI   I   IL CI

At_RPD1  CPGCKYIRKKQFQITEDQPERCRYCT-----LNTGYPLMKFRVTTKEVF
At_RPB1  CFNCSKIIADEVCRSLFRQAMKIK-----NPKNRLKKILDACKNKTKCDGGD
At_RPC1  CKRCSNMILLDEKLYEDHLRKMNRNPRM-----EPLKKTTELAKAVVKCSTMASQRRI
At_RPA1  CFFCHHFMAKPEDVERAVSQLKLIKGDIVSAKQLESNTPTKSKSDESCEVVTIDSSE
Sc_RPB1  CMHCGKLLLDLDEHN-ELMRQALAIKDS-----KKRFAAIWTLCKTKMVCET--
Ec_RPOC  LASPTAHTWFLKS-LPSRIGLLLDMP-----LRDIERVLYFESYVVIEGGMTNL
consensus C   CS IL DE   E R ALKI                               K RL   LE CKSKM TDE

At_RPD1  RRSGLVVEVNEESLMKLLKRGVLTLP-----
At_RPB1  DIDLVQSHSTDEPVKKSRRGCCGAQQPKLTIEG-----
At_RPC1  TCKKCGYLNGMVKKIAAQFGIGISHDRSKIHG-----
At_RPA1  ECELSDVEDQRWTSLQFAEVTAVLKNFMRLSSKSCSRCKGINPKLEKPMFGWVRMRAMKD
Sc_RPB1  ---DVPSE-DDPTQLVSRGCCNTQPTIRKDC-----
Ec_RPOC  ERQQLLTHEQYLDALIEFGDEFD-----
consensus E   DI SE QD T L   RGG GIT P IKI G

At_RPD1  -----PDYWSFLPQDSNIDESCCLKPTRRII
At_RPB1  -----MKMIAEYKIQRKKNDEPDQLPAPAER
At_RPC1  -----GEIDECKSAISHTKQST---AAINPL
At_RPA1  SDVGANVIRGLKLLKSTSSVENPDGFDDSGIDALSEVEDGDKETREKSTEVAAEFFEHNS
Sc_RPB1  -----LKLVGSWKKDRATGDAD----EPELR
Ec_RPOC  -----AKMGAEAIQALLKSMDLQCEQL
consensus                               I   FK R   DE   E   L

```

Conserved domain B

```

At_RPD1  THAQVYALLGIDQRLIKKDIP-----MFNSLGLTSPVTFNGYRVTEI
At_RPB1  KQTLGADRVLSVLKRISDADCOLLGFNPKFA-----RPDWMILEVLPIDPPFVRPSVM
At_RPC1  TYVLDPNLVLGLFKRMSDKDCELL---YIAY-----RPENLIIITCMLVPPLSIRPSVM
At_RPA1  KRDLLPSEVRNLIKHLWQNEHEFCSFIGDLWQSGSEKIDYSMFFI  ESLVLPPTKFRPPTT
Sc_RPB1  --VLSTEEILNIFKHISVKDFTSLGFNEVFS-----RPEWMILTCLPVPPPVRPSIS
Ec_RPOC  REELNETNSETKRKIKTKRIKLEAFVQSGN-----KPEWMILTVLPVLPDLRLVLP
consensus  K L VL I KRLS KD LLGF RPEWMILT LPVPPP VRPSVM
    
```

```

At_RPD1  VHQFN GARLIFDERTR IYKKLVGFEGNTLELSSRVM ECMQYSRIFSETVSSSKDS-----
At_RPB1  MDATSRSEDDLT HQ LAMIIRHNENLKRQEKNGAPAHISEFTQLLQFHIAIYFDNELPGQ
At_RPC1  IGGIQSNENDLTARLKQIILGNASLHKILSQPTSSPKNMQVMDTVQIEVARYINSEVRG-
At_RPA1  GGD-SVM EHPQIVGLNKVIESNNILGNACTNKL DQSKVIFWRN LQESVNVLFDSKTAT-
Sc_RPB1  FNESQRGE DDLTFK LADILKANISLETLEHNGAPHHAIEEAESLLQFHVATYMDNDIAGQ
Ec_RPOC  LDGGRFATS D LNDLYRRVINRNRLKRLDLAAPDIIVRNEKRM LQEA VDALLDNGRRGR
consensus  I G Q AE DLT RLR IIK N L RIL NGAP IMQ RLLQE VATYFDSEI G
    
```

Conserved domain C

```

At_RPD1  -----ANPYQKKS DTPKLCGLR-FMKDVL L GKRS DHTFRTV VVGDP SLKINEIGIPESIA
At_RPB1  PRATQKSGRPIKSI CSRLKAKEGRIRGNLMGKRVDFSARTVITPDPTINIDELGVFWSIA
At_RPC1  -CQNQPEEHLPSGILQRLKKGGRFRANLSGKRVEFTGRTVISPDNLKITEVGIPI LMA
At_RPA1  -----VQSQRDSSGICQLLEKKEGLFRQKMMGKRVNHACRSVISPDPIAVNDIGIPCF A
Sc_RPB1  PQALQKSGRPVKSIRARLKGKEGRIRGNLMGKRVDFSARTVISGDPNLELDQVGVKFSIA
Ec_RPOC  -AITGSNKRP LKSLADMIKKGQCRFRQNL L GKRV DYSGRSVITVGPYLRHQCGLEKKMA
consensus  Q S RPLKSI RLKGKEGRFRGNLMGKRVDFSARTVISPD LKL EIGIP SIA
    
```

```

At_RPD1  KRLQVSEHLNQCNKERLVTSFVPTLLDNKE-----MHVRRGDRIVAI
At_RPB1  LNLTYPEVTVPYNIERLRELVDYGP HPPPGK-----TGAKYIIRDGQRLDLRYLKK S
At_RPC1  QILTFPECVSRHNI EKLRQCVRNGPNKYPG-----ARNVRYPDGSSRTL VG DYRKR
At_RPA1  LKLTYP ERVTPWNVEKLR EAIINGPDIHPGATHYSKDSSTMKLPST EKARATAFKLLSS
Sc_RPB1  KTLTYPEV VTPYNI DRLTQLVRNGPNEHPG-----AKYVIRDSGDRIDLRYSKRA
Ec_RPOC  LELFKFFIYGKLELRGLATTIKA AKK MVER-----
consensus  L LTYPE VTPYNIERL R VRNGP PG K DGR LR LKK
    
```

Conserved domain D

```

At_RPD1  QVNDLQ TG-----DKIFRSLMDGDTVLMNRPPSIHQHSLIAMTVRILPTTSVVS LN
At_RPB1  SDQHLELG-----YKVERHLQDGFVLFNRQPSLHKMSIMGHRIRIMP-YSTFR LN
At_RPC1  IADELAIG-----CIVDRHLQEGDVVLFNRQPSLHRMSIMCHRARIMP-WRTLRFN
At_RPA1  RGATTELGKTCDINFE GKTVHRHMRDGDIVLVNRQPTLHKPSLMAHKVRVLKGEKTLRLH
Sc_RPB1  GDIQLQYG-----WKVERHIMDNDPVLFN RQPSLHKMSIMMAHRVKVIE-YSTFR LN
Ec_RPOC  -EEAVVWD-----ILDEVIREHPVLLNRAPT LHLRLGIQAFEPVLI E-GKAIQLH
consensus  D L LG KVERHLM DGD VLFNRQPSLHKMSIMAHVRRIIP YSTLRLN
    
```

*** * * (active site)**

```

At_RPD1  PICCLPFRGDFDGDCLHG YVPSIQAKVELDELVALDKQLINRQNGRNLLSLGQDSITAA
At_RPB1  LSVTSPYNADFDGD EMMNHVPQSFETRAEVLELMMV PKCIVSPQANRPVMGIVQDTLLGC
At_RPC1  ESVCNPNADFDGD EMMNHVPQTEEARTEAITLMG-----
At_RPA1  YANCTYNADFDGD EMMNVHFPQDEISRAEAYNIVNANNQYARPSNGEPLRALIQDHIVSS
Sc_RPB1  LSVTSPYNADFDGD EMMNLHVPQSEETRAELSQLCAVPLQIVSPQSNKPCMGLVQDTLLCGI
Ec_RPOC  PLVCAAYNADFDGDQMAVHVPLTLEAQLEARALMMSTNNILSPANGEP IIVPSQVVLG-
consensus  SVCSPYNADFDGD EMMNHVPQSEEARAEA LMAV QIVSPQNGRPLMGIVQDTLLG
    
```

Conserved

At_RPD1 YLVNVEKNCYLNR AQMQQLQM-----YCPFQLP PPAITKA
 At_RPB1 RKI-**IKRDTFIEKDVFMNTLM**-----WWE**DFD**GKVPAPAILKP
 At_RPC1 -----DTFYDRAAFSLICS-----YMGDGMDSID**LPTPTILKP**
 At_RPA1 VLL-**IKRDTFLDKDHFNQLLFSSGVTDMVLSTFSGRSGKVMVSAS**AE**LLTVTPAILKP**
 Sc_RPB1 RKL-**ILRDTFIELDQVLNMLY**-----WVP**WD**GV**IP**TPAILKP
 Ec_RPOC -----LY**Y**MTRDCVN-----AKGEGMVL**TG**PK**EAERLYR**
 consensus L T RDTFIDRD FNNLL D D LPTPAILKP

domain E

At_RPD1 SPSSTEFQ**WTGMQLFGMLFPPGFD**-YTYPLNNVV-----
 At_RPB1 -----RPLWTGKQVFNLI**IPKQINLLRYS**SAWHADTETG-----
 At_RPC1 -----IELWTGKQIFSVLLRPNASIRV**YVTLNVKEKNFKKG**-----
 At_RPA1 -----VPLWTGKQVITAVLNQITKGHPPTVEKATKLPVDFFKCRSREVKPNSGDLTKKK
 Sc_RPB1 -----KPLWSGKQILSVAIPNGIHLQRF-----DEGTT-----
 Ec_RPOC -----SGLASLHARVKVRI**TEYEK**DANG-----ELV-----
 consensus PLWTGKQIFGVLP L Y D

At_RPD1 SNGELLSF**SEGS**AWLRDGE**CNFIERLLKHKD**GKVLD---II**YSAQEMLSQWLLMR**GLSVS
 At_RPB1 -----FITPGDTQ**VRIER**GELLAGT**LCKKTLGT**-----SNGSL**VHVIWEEV**GPDA
 At_RPC1 EHGFD**ETMCINDGWVYFRNSELISGQLGKATLALDIFPLGNGNKDGLYSILLRDYN**SHAA
 At_RPA1 EIDESWKQNL**NEDKIHIRKNEFVCGVIDKAQFAD**-----YGLV**HVTV**HELYCS**NA**
 Sc_RPB1 -----LLSPK**NGMLIIDGQIIFGVVEKKT**VGS-----SNGGL**IHVVTREK**GPQVC
 Ec_RPOC AKTSLKDT**TVGRAILWMIVPKGLPYSIVNQALGK**K-----AIS**KMLNTCYRILGLK**P
 consensus ISIGDA L I GELI GVL K TLG S GLLHVV RD G AA

At_RPD1 LAD**LYLSSDLQSRKNL**TEEIS**YGLREAEQVCNKQQLMVESWRD**FLAVNGED**KEE**DSVSD**L**
 At_RPB1 RKF**L**GH**TQWL**VNY**WLLQNGFTIGIGDTIADSSTMEKINETISNAKTA**V**KDLIRQ**FQ**GKEL**
 At_RPC1 **AVCMNRLAKLSARWIGIHGFSIGIDDVQ**PEELS**KERKDSIQFGYDQCHRKI**EEFN**RGNL**
 At_RPA1 **G**N**LLSVFSRLFTVFIQTHGFTCGVDDLIILKDMDEERTKQLQ**ECEN**VGERVLRKTF**GID**V**
 Sc_RPB1 AKL**FGNIQKV**VNF**WLLHNGFSTGIGDTIADGPTMREITETIAEAKKKVLDVTK**IAQAN**LL**
 Ec_RPOC VIFADQIM**YTGFA**YAARS**GASVGIDDMVIP**-----E**KKHEIIS**EAEAE**VAIQ**Q**FQS**---
 consensus A L I KL WLL GFSIGIDDLI EEI ESI EA V DVIEEFQ G DL

At_RPD1 ARFCYE-----R**QKSATLSE**LAVSA**FKDAYR**-----DV**QALAYRYG**DQSN
 At_RPB1 DPEP-----G**RTMRDTFENRVNQVLN**KAR-----DD**AGSSAQKSL**AE**TN**
 At_RPC1 QLKA-----G**LDGAKSLEAEITGILNTIR**-----E**ATGKACMSGL**HWRN
 At_RPA1 DVQID**PQDMRSR**IERIL**YEDGESALASLDRSIVNYLNQCSSKGMNDLLSDGLL**KTPGRN
 Sc_RPB1 TAKH-----G**MTLRESFEDNVVRF**LNEAR-----D**KAGRLAEVNL**KDLN
 Ec_RPOC -----G**LVTAGERYNKVIDIWAAN**-----D**RVSKAMM**---D**SFN**
 consensus GLT A S E VV FLN AR DDVGK AL L N
 18 aa deleted

Conserved domain F

At_RPD1 **SFLIMS**KA**GSKGNIGKLVQHS**MC**IGLONS**AVSLS**FGFPRE**L**CAAWNDPNSPLR**GAK**GKD**
 At_RPB1 **NLKAMV**TAG**SKGSFINISQMTACV**Q**QNV**EG**KRIPFGFDGRTLP**PH**TKDDYGF**ESR----
 At_RPC1 **SPLIMS**Q**CGSKGSPINISQMVACV**Q**QTVN**GH**RAPDGFIDRSLPHF**PR**MSKSPA**AK----
 At_RPA1 **CISLMT**IS**GAKGSKVNFQCISSHL**Q**QDLE**GR**VPRM**VSG**KTLP**CF**HPWDS**FRAG----
 Sc_RPB1 **NVKQVM**AG**SKGSFINIAQMSACV**Q**QSV**EG**KRIAFGFVDRTLP**PH**SKDDYS**ESK----
 Ec_RPOC **SIYMA**D**SARGSAAQIRQLAGMRGL**MA**KPDGS**-----I**IE**-----
 consensus SI IMS AGSKGS INI QMSACVQQ VEGKRIP GF DRTLPHF K DYSP AK

Bridge helix

At_RPD1 STTTESYVPPYGV IENSFLTGLNPLESFVHVSVTSRDSSFSGNADLP--GTL SRRLMFFMRD
 At_RPB1 -----GFVENSYLRLGLTPQEFFFHAMGGREGLIDTAVKTS ETGYIQRRIVKAMED
 At_RPC1 -----GFVANSFYSGLTATEFFFHTMGGREGLVDTAVKTA STGYMSRRIMKALED
 At_RPA1 -----GFISDRFLSGLRPQEEYFHC MAGREGLVDTAVKTSRSGYLQRCIMKNLES
 Sc_RPB1 -----GFVENSYLRLGLTPQEFFFHAMGGREGLIDTAVKTAETGYIQRRIVKAMED
 Ec_RPOC -----TPI TANFREGLNVLQYFISTHGARKGLADTALKTANSGYLTRRIVDVAQD
 consensus GFIENSFLSGLTPQEFFFHAMGGREGLIDTAVKTA TGYLQRRIMKALED

At_RPD1 IYAAVDGTVRNSFGNQLVQFTYETDGPVEDITG-----
 At_RPB1 IMVKYDGTVRNSLG-DVIQFLYGEDGMDAVWIE SQKLDLTKMKKSEFDRTFKYEIDDENW
 At_RPC1 LLVHYDNTVRNLSG-CILQFTYGDGMDPALME-----
 At_RPA1 LKVNVDCTVRDADG-SIIQFQYGEDGVDVHRSS-----
 Sc_RPB1 IMVHYDNTVRNSLG-NVIQFIYGEDGMDAAHIEKQSLDTIGGSDAAFEKRYRVDLLNTDH
 Ec_RPOC LVVTEDDCGTHEGI-MMTPVIEGGDVKEPLRDR-----
 consensus IMV YD TVRNS G IIQFIYGEDGMD IE

At_RPD1 -----EALGSL SACALSE
 At_RPB1 NPTYLSDEHLEDLKGIRELRDVFDAEYSKLETDRFQLGTEIATNGDSTWPLPVMIKRHIW
 At_RPC1 -----GKDGAPLNFNRLFLKV
 At_RPA1 -----FIEKFKELTINQDMVLQ
 Sc_RPB1 TLDPSLLESSEILGDLKLQVLLDEEYKQLVKDRKFLR-EVFVDGEANWPLPVMIRRIIQ
 Ec_RPOC -----VLGRVTAEDVLK
 consensus GE PL VN LI

At_RPD1 AAYSALDQPIS-----LLETSP LLNLKNVLECGSKKGQREOTMSLYLSEYLSK--
 At_RPB1 NAQKTFKIDLRKISDMHPVEIVDAVDKLRLLVVPGDALSVEACKNATLFFNILLRST
 At_RPC1 QATCPPRSHHTYLS-----SEELSQKFEELVRHDKSRVCTDAFVKSLREFFVSLLG--
 At_RPA1 KCSEDMLSG-----ASSYISDLPI SLKKGAEKFEAMPMNERIASKFVR--
 Sc_RPB1 NAQQTFFHIDHTKPSDLTIKDIVLGVKDLQENLLVLRGKNEIIQNAQRDAVTLFCCLLRSR
 Ec_RPOC PGTADILVPRN-----TLLHEQWCDLLEENSVDAVKVRSVSCDTEFGVCAH--
 consensus NAQ I I T S V ALS L E LLVL V VEAQ L LF LLR

Conserved domain G

At_RPD1 -----KKHGF EYGSLEIKNHLEKLSFSEIVSTSMIFSPSSNTKVP LSPWVCH
 At_RPB1 LASKRVLEEYKLSREAF EFWVIGELESRFLOSLVAPGEMIGCVAAQSIGEPATQMT--LNT
 At_RPC1 -----VKSASPPQVLYKASGVTDKQLEAGTAIGTIGAQSIGEPGTQMT--LKT
 At_RPA1 -----QEELLKLVKSKFFASLAQPGEPVGVLAQAQSVGEPSTQMT--LNT
 Sc_RPB1 LATRRVLQ EYRLTKQAFDWVLSNIEAQFLRSVVHPGEMVGVLAQAQSIGEPATQMT--LNT
 Ec_RPOC -----CYGRDLARGHIIINKGEAIGVIAAQSIGEPGTQMT--MRI
 consensus K AFEWVL IKS F SLV PGE IGVIAAQSIGEPATQMT LNT

At_RPD1 FHI SEKVLKRKQLSAESV VSSLN-EQYKSRNRELKLDIVDLDIQNTNHCSSDDQAMKDDN
 At_RPB1 FHYAGVSAKNVT LGVPRLREIIN-VAKRIKTPSLSVYLTPEASKSKEGAKTVQCALEYTT
 At_RPC1 FHFAGVASMNIITQGVPRINEIIN-ASKNISTPVISAELENPLELTS--ARWVKGRIEKT
 At_RPA1 FHLAGRGEMNVT LGIPRIQEILMTAAANIKTPIMTCPLLKG--KTKEDANDITDRIRKII
 Sc_RPB1 FHFAGVASKKVTISGVPR LKEILN-VAKNMKTPSLTVYLEPGHAADQEQAKLIRSAIEHTT
 Ec_RPOC FHI GGA--DIITGGLPRVADLFE--ARRPKEPAILAEISGI-----
 consensus FHFAGVA KNVT LGVPRL EILN AKNIKTP LSVEL T E AK I AIE TT

192 aa deleted

At_RPD1 VCITVTVVEASKHSVLELDAIRLVLIPFL-----LDSPVKG---

At_RPB1 LRSVTQATEVWYDEDPMSTIIIEEDFEFVVR-----SYEMPDEDV

At_RPC1 LGQVAESIIEVLMTSTSASVRIILDN-----KIIEEACLSI

At_RPA1 VADIIKSMELSVVPEYTVYENEVCSIHKLKINLYKPEHYPKHTDITEEDWEETMRAVFLRK

Sc_RPB1 LKSVTIASEIIYYDEDPRESTVIPLEDEEIIQLHFS-----LLDEEAEQ

Ec_RPOC -----VSFGKETK GK-----

consensus L V IEVSY PDP S I D I I

At_RPD1 -----

At_RPB1 SPDKISPWLLIRIELNREMMVDDKKL SMADIAEKINLEFDDDLTCIFNDNAQKLIIRIRIM

At_RPC1 TPWSVKNSILKTPRIKLNNDIRVLDTG-----

At_RPA1 LEDAIEETHMKMLHRIRGIHNDVTGPIAGNETDNDDSVSGKQNEDDGDDGEGTEVDDLGS

Sc_RPB1 SFDQQSPWLLRLELDRAAMNDKDLTMGQVGERIKQTFKNDLFVIWSEDNDEKLIIRCRVV

Ec_RPOC -----

consensus S D I LLRL R ND L MA DD I

At_RPD1 -----DQGIK

At_RPB1 NDE-----GPKGELQDESAEDDVFLKRIE SNML

At_RPC1 -----LDITPVVDRSRAHFN

At_RPA1 DAQKQKQETDEMDYEENSEDETNEPSSISGVEDPEMDSENEDETEVSKEDTPEPQIESME

Sc_RPB1 R-----PKSLDAEAEAEEDHMLKRIENTML

Ec_RPOC -----

consensus D E D L K E M

At_RPD1 KVNILWTD RPKAPKRNGNHLAGELYLKVTM-----

At_RPB1 TEMALRGIPDINKVFIKQVRKSRFDEEGGF-----

At_RPC1 LHNLNKGIKTVERVVVAEDMDKSKQIDG-----

At_RPA1 PQKEVKGVKNVKEQSKKKRRKRVRAKSDRHIFVKGEKGEKFEVHFKFATDDPHILLAQIAQ

Sc_RPB1 ENITLRGVENIERVVMKYDRKVPSPTEYVK-----

Ec_RPOC -----RRLVITPV DGS DP-----

consensus LRGIK I RVVI K G

At_RPD1 YGDRGKRNCWTA-----LLET

At_RPB1 KTSEEWMLDTEG-----VNLL

At_RPC1 --KTKWKL FVEG-----TNLL

At_RPA1 QTAQKVYIQNSGKIERCTVANCGDPQVIYHGDNPKERREISNDEKKASPALHASGVDFPA

Sc_RPB1 --EPFWVLET DG-----VNLS

Ec_RPOC --YEE MIPKWRQ-----LNVF

consensus EWML EG LNL

At_RPD1 CLPIMDMIDWGRSHPDNIRQCCSVYGIDAGRSIFVANLESASDTCKEILREHILLVADS

At_RPB1 AVMCHEDVDPKRTTNSNHLIEIIEVLGIEAVRRALLDELRVVISFDGSYVNYRHIAILCDT

At_RPC1 AVMGTPGINGRTTTSNNVVEVSKTLGIEAARTTIIDEIGTVMGNHGMSIDIRHMLLADV

At_RPA1 LWFEQDKLDVRYLYSNSIHDMLNIFGVEAARETIIREINHVFKSYGISVSIRHLNLIADY

Sc_RPB1 EVMIVPGIDPTRIYTN SFIDIMEVLGIEAGRAALYKEVYNVIASDCSYVNYRHMALIVDV

Ec_RPOC EGERVERGLVISDGPEAPHDILRLRGVHAVTRYIVNEVQDVYRLQGVKINDKHIEVIVRQ

consensus VM HD ID RRT SN IIDIL VLGIEAAR II EI VI GI IN RHL LLAD

Conserved domain H

At_RPD1 LSVTGEFVALNAKGWSKQRQVESTPAPFTQACFSSPSQCFLKAAKEGVRDDIQCSIDALA
 At_RPB1 MTYRCHLMATRHCIN-----RNDTGPLMRCSFEETVDILLDAAAYAETDCLRGVTENIM
 At_RPC1 MTYRGEVLGIQRTGIQ-----KMDKSVLMQASFERTGDHLFSAAAAGKVDNIEGVTECVI
 At_RPA1 MTFSGYRPMRMMGGI-----AESTSPFCRMTFETATKFIVQAATYGEKDTLETPSARIC
 Sc_RPB1 MTIQGGLTSVTRHGFN-----RSNTGALMRCSFEETVEILFEAGASAELEDCRGVSENV
 Ec_RPOC ML---SRDLLGITKAS-----LATESFISAASFQETTRVLEAAVAKRDELRLKENV
 consensus MTY G LLAITR G N R TTSPLMRASFEETTDILLDAAA GERDDL RGVSENV
 41 aa deleted

At_RPD1 WQKVPGFGTGDQFEI I I SPKVHGF TTPVDVYDLLSSTKTMRR TNSAPKSDKATVQPFGLL
 At_RPB1 LGQLAFIGTGDCELYLNDE-MLKNAIELQLPSYMDGLEFGMT PARSVSGTPYHEGMMS
 At_RPC1 MGIPMKLGTGILKVLQRTDDLPK-----LKYGDP IIS-----
 At_RPA1 LGLPALSGTGCDFLMQRVEL-----
 Sc_RPB1 LQMAFIGTGAFDVMIDEESLVKYMPEQKI TEIEDGQDGGVT EYSN-----ESGLVNA
 Ec_RPOC VGRLEIAGTGYAYHQDMRRAAG-----
 consensus LG LAPIGTG DLMIR E L K I G P

At_RPD1 HSAFLKDIKVL DKGIPMSLLRTIFTWKNIELLSQSLKRILHSYEINELLNERDEGLVKM
 At_RPB1 NYLLSPNMR LSPMSDAQFSPYVGGMAFSPSSSPGYSPTSPGYSPTSPGYSPT
 At_RPC1 -----
 At_RPA1 -----
 Sc_RPB1 DLLVKDELIMFSPLVDSGSNDAMAG-GFTAYGGVDYG-----EATSP---FAAYGEAETS
 Ec_RPOC -----EAPAAEQV
 consensus I F P

At_RPD1 VLQLHPNSVEKIGPGVKGIRVAKSKHGDSCCFEVVRIDGTFEDFSYHKCVLGATKIIAPK
 At_RPB1 PGYSPTSP TYSPSSPGYSPTSPAYSPTSPSYSP TSPSYSP TSPSYSP TSPSYSP TSPSYSP
 At_RPC1 -----
 At_RPA1 -----
 Sc_RPB1 PGFGVSSPGFSPTSP TYSP TSPAYSPTSPSYSP TSPSYSP TSPSYSP TSPSYSP TSPSYSP
 Ec_RPOC TAEDASASLAELLNAGLGSDNE-----
 consensus G P T

At_RPD1 KMNFKYSKYLKNGTLES GGFSEN P-----
 At_RPB1 PTSPSYSP TSPAYSPTSPAYSPTSPAYSPTSPSYSP TSPSYSP TSPSYSP TSPSYSP TSP
 At_RPC1 -----
 At_RPA1 -----
 Sc_RPB1 PTSPSYSP MSPSYSP TSPSYSP TSPSYSP TSPSYSP TSPSYSP TSPSYSP TSPSYSP TSP
 Ec_RPOC -----
 consensus S S S YS

At_RPD1 -----
 At_RPB1 SYSPTSPAYSPTSPGYSPTSPSYSP TSPSYGPTSPSYNPQSAKYSPSIAYSNARLSPA
 At_RPC1 -----
 At_RPA1 -----
 Sc_RPB1 AYSPTSPSYSP TSPSYSP TSPSYSP TSPSYSP TSPNYSP TSPSYSP TSPGYS PGSPAYSP
 Ec_RPOC -----
 consensus


```

At_RPD1 -----
At_RPB1 SPYSP TSPNYSPTSPSYSPTSPSYSPSSPTYSPPSSPYSSGASPDYSPSAGYSPTLPGYSP
At_RPC1 -----
At_RPA1 -----
Sc_RPB1 -----
Ec_RPOC -----
consensus -----

```

```

At_RPD1 -----
At_RPB1 SSTGQYTPHEGDKKDKTGKKDASKDDKGNP
At_RPC1 -----
At_RPA1 -----
Sc_RPB1 -----KQDEQKHNEENENSR-----
Ec_RPOC -----
consensus -----

```

The alignment was performed using ClustalX and then edited by hand using MacClade 4.03 prior to being exported to BOXSHADE for shading. Positions with identical amino acids are indicated by green shading, whereas similar amino acids are indicated by yellow shading. Previously published (Cramer et al., 2001) alignments and structural features were considered during the editing process. Regions of the *E. coli* β' subunit that do not align with the eukaryotic RNAPs were deleted, as indicated below the alignments. Conserved domains (Jokerst et al., 1989) are indicated with letters and bold lines above the alignments. The active site (metal A site; Cramer et al., 2001), is indicated by asterisks. Also noted is the bridge domain, which traverses the cleft in the polymerase near the active site. Domain assignments are according to Cramer et al. (2001). Protein sequences compared are: At_RPD1 (Pol IV), At_RPB1 (Pol II), At_RPC1 (Pol III), At_RPA1 (Pol I), Sc_RPB1 (Pol II), and Ec_RPOC (β' subunit).

Figure S2. Multiple Alignment of RPD2 with DNA-Dependent RNA Polymerase Second-Largest Subunits of *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec)

Supplemental Figure 2. Alignment of RNAP 2nd Largest subunits

```

At_RPD2    MPDMDIDVKDLEEFEEATTGEINLSELGEGFLQSFCKKAATSFFDKYGLISHQINNSYNYFI
At_RPB2    -----MEYNEYEPEP-QYVEDDDDEEITQEDAWAVISAYFEKGLVRRQQLDSFDEFI
At_RPC2    ---MGLDQEDLDLTNDHDFIDKEKLSAPIKSTADKFQLVPEFLKVRGLVKQHLDNFNYFI
At_RPA2    -----MNVNAKDSTVPTMEDFKELHNLVTHHLESFDYMT
Sc_RPB2    ----MSDLANSEKYYDED-PYGFEDESAPITAEDSWAVISAFFREKGLVSRQQLDSFNQFV
Ec_RpoB    -----MVYSYTEKKRIRKDFGKRPQVLDVPY-LLSIQLDSFQKFI
consensus      D D E Y E           D       T D W V I S F F E K G L V S Q Q L D S F N Y F I
    
```

```

At_RPD2    EHGLQNVFQSFGEMLVEPSFDVVK--KKDNDWRYATVKFGEVTVKPTFFSD--KELEFL
At_RPB2    QNTMQEIVDEHSADIEIRPESQHNPGHQSDFAETIYKISFGQIYLSKPMTESDGETATLF
At_RPC2    NVGIHKIVKANSRITS-----TVDPISIYLRFKKVRVGEPSIINVN-TVENIN
At_RPA2    LKGLDVMFNRIKPVSVYDPN-----TENELSIWLENPLVFAPQKESFKSSTRKEPLL
Sc_RPB2    DYTLDLICHEDSTLILEQLAQHTT--ESDNISRKYEISFGKIYVTKPMVNESDGVTHALY
Ec_RpoB    EQDPEGQYGLEAAFRSVFPIQSYS-----GNSELQYVSYRLGEPV-----FD
consensus      GLQ I                               D E       L F G V Y V K P       S D       L
    
```

Conserved domain A

```

At_RPD2    FWHARLQNMITYSARIKVNVOVEVFKNTVVKSDKFKTGQDNVVEKKILDVVKQDILIGSIF
At_RPB2    FKAARLRNLTYSAPLYVDVTKRVIK-----KGHDG--EVTETQDFTKVFVIGKVP
At_RPC2    PHMCRLADMTYAAPIFVNIIEYVHGS-----HGKNAKSAKDNVILGRMP
At_RPA2    FFCRCQAKISYTGTFMADVCFKYND-----GVVVRDKVDFDQGF
Sc_RPB2    FQEARLRNLTYS SGLFVDVKKRTYE AIDVPGRELKYELIA--EESDDSESCKVFIGRLP
Ec_RpoB    VQECQIRGVTYSAPLRVKLRLVYEREAPEGT-----VKDIKEQEVVMGEIP
consensus      P EARLRNLTYSAPLFVDV RVFD                               E       DV K KVFVIGRIP
    
```

Conserved domain B

```

At_RPD2    VMVKSILCKTSEKG-KENCKKGDCAFDDGGYFVIKGAEKVFIAQEQMCTKRLWISNSP--
At_RPB2    IMLRS SYCTLFQNSEKDLTELGECPYDQGGYFIINGSEKVLIAQEKMSNHHVYVFKRQP
At_RPC2    IMLRSCRCVLHGKDEEELARLGECPDLPGGYFIIKGTAKVLLIQEQLSKNRIIIDSCK--
At_RPA2    IMLMSKLCSLKGADCRKLLKCKESTSEMGGYFILNGIERVFRCVIAPKRNHPTSMIRNSF
Sc_RPB2    IMLRSKNCYLSEATE SDLYKLLKECPFDMGGYFIINGSEKVLIAQERSAGNIQVFFKKAAP
Ec_RpoB    LMTDN-----GTFVINGTERVIVS QLHRSPGVFFDSDKGKT
consensus      IMLRS C L           EKDL KLGECPFD GGYFIINGSEKVLIAQE MS N VFI K
    
```

```

At_RPD2    -----WTVSFRSENKRNR FIVRLSENEKAEDYKRREKVLIVYFLSTELPWWLLFF
At_RPB2    NKYAYVGEVRSMAENQNPSTMFVRLARASAKGSSGQYIRCTLPYIRTEIPIIIVFR
At_RPC2    -----KGNINASVTSSTEMTKSKTVIQMEKEKTYLFRHFRVKKIPIIIVLK
At_RPA2    RDRKEGYSSKAVVTRCVRDDQSSVTVKLYLNRNGSARVGFVIVGREYLLPVGLVVKALTN
Sc_RPB2    SPISHVAEIRSALEKGRFISTLQVKLYGRE----GSSARTIKATLPYIKQDIPVIVIFR
Ec_RpoB    HSSGKVLYNARIIPYRGSWLDFEFDPKDN-----LFVRIDRRR--KLEATILR
consensus      V R I V F R ST FV KL R           G G I V T L Y I E I P I I I F R
    
```

At_RPD2 ALGVSSDKRAMDLIAFDGDDASITNSLIASIHVADAVCEAFRCG---NNALTYVEQQIKS
 At_RPB2 ALGFVADKDIILEHICYDFADTQMMELLRPSLEEFVIQNQLVALDYIGKRGATVGVTKKEK
 At_RPC2 AMGMESEQEIIVQMVGRDPRFSASLLPSIEECVSEGVNTQKQALDYIEAKVVKISYGTPEE
 At_RPA2 SCDEEIIYESLNCCYSEHYGRGDGAIGTQLVREAKIILDEVVDLGLFTREQCRKHLG-QH
 Sc_RPB2 ALGIIPDGEILEHICYDVNDWQMLEMLKPCVEDGFIQDRETAIDFIGRRGTALGIKKEK
 Ec_RpoB ALNYTTEQIILDLFFFEKV----LFTNDLDHGPIYISETLRVDPNTNDRISALVIEYRM-MRPG
 consensus ALGI SD EILE I YD D ML L IE A VI D L L AK V I K
 107 aa deleted

Conserved domain C

At_RPD2 TKFPPAESVDECLHLYLFPGLQSLKKKARFLGYMVKCLNSYAGKRRKCNRDSFRNKRIE
 At_RPB2 RIKYARDIILQKEMIPHVGI GEHCETKKAYYFGYIIHRLILCALGRRPEDDRHDYGNKRLD
 At_RPC2 KDGRALSILRDLFLAHVVPVDDNFRQKCFYVGVMLRRMI EAMLNKDAMDDKYVGNKRLE
 At_RPA2 FQPVLGDVVAEAVLRDYLFLVHLDNDHDKFNLLIFIIQKLYSLVDQTSLPDNPISLQIQEIL
 Sc_RPB2 RIQYAKDIIQKEFLPHITQLEGFESRKAFFLGYMINRLLCALDRKDQDDRDFGKRRLD
 Ec_RpoB EPTTREAASLFFENLFFSEDYDL--KDDIIDVMKKLIDIRNGKGEVDDIDHLGNRRIR
 consensus R A DIL LL HL V E E KKAFFLGYMIKRL L GKR DDRDFGNKRID
 24 aa deleted

At_RPD2 LAGELLEREIRVHLAHARRKMTRAMQKHLSGDG-----DLKPIEHYLDASVITNGL
 At_RPB2 LAGPLLGGFLFRMLFRKLRDVRSYVQKCVDNK-----EVN-LQFAIKAKTITISGL
 At_RPC2 LSCQLISLLFEDLFKTMLSEAIKNVDHILNKPIRAS----RFDFSQCLNKDSRYSISLGL
 At_RPA2 VPGHVITTYLKEKLEEWLRKCKSLKDELNTNSKFSFESLADVKKL INKNPPRSITGSI
 Sc_RPB2 LAGPLLAQLFKTIFKKLTKDIFRYMQRTVEEAH-----DFN-MKLANAKTITISGL
 Ec_RpoB SVGEMAENQFRVGLVRVERAVKERLSLGD-----DTLMPQDMNAKPIISAAY
 consensus LAG LL LFRVLFKKL RDVKR LQK LD DV L I AKSITISGL

At_RPD2 SRAFSTGAWSH-PFRKMERSVGVVANLGRANPLQTLIDLRRTROQ----VLYTGKVGDAE
 At_RPB2 KYSLATGNWG--QANAAGTRAGVSQVLNRLTYASTLSHLRRLNSP----IGREGKLAKPR
 At_RPC2 ERTLSTGNFDI-KRFRMHRKG-MTQVLTRLSFIGSMGFITKISPQ----FEKSRKVSQPR
 At_RPA2 ETLKLTGALKTQSGLDLQQRAGYTVQAERLNFLRFLSFFRAVHRGA---SFAGLRTTTVR
 Sc_RPB2 KYALATGNWGE-QKKAMSSRAGVSQVLNRYTYSTLSHLRRTNTP----IGRDGKLAKPR
 Ec_RpoB KEFFGSSQ-----LSQFMDQNNPLSEITHKRRISALGPGGLTRERAGFEVR
 consensus K LATGNW M RAGVSQVL RLNFLTLSHLRRI I RDGKLA PR

Conserved domain D

At_RPD2 YPHPSHWGRVCFLSTPDGENCGLVKNMSLLGLVSTQSLES--VVEKLFACGMEEELMDDTC
 At_RPB2 QLHNSQNGMMCPAETPEGQACGLVKNLALMVYITVGSAAYPILFLEEWGTENFHEISPS
 At_RPC2 SLQPSQWMLCPDTPEGESCGLVKNLALMTHVTTDEEGPLVAMCYKLVTDLEVL SAE
 At_RPA2 KILLESWGFVHTPDGTPCGLLNHMTRTSRITSQFDSKGNIRDFLKRKSVVDVLTGA
 Sc_RPB2 QLHNTHWGLVCPAETPEGQACGLVKNLSLMSCISVGTDPMPPIITFLSEWMEPLEDYVPH
 Ec_RpoB DVHPTHYGRVCFPIETPEGPNIGLINSLSVYAQTNEY-----
 consensus LHPSHWGMVCPAETPEG CGLVKNLSLMG ITT SD PII G EEVLS

At_RPD2 TPL--FGKHVLLNGDHWGLCADSESFVAELKSRRRQSELPREMEIKRDKDDNIVRIFTD
 At_RPB2 VI---EQATKIFVNGMVGVRHDPDMLVKTLLRRLRRRVDVNTVEVGVVRIIRLKELRITYD
 At_RPC2 ELHTPDSFLVILNGLILGKHSRPQYFANSLRRLRRAGKIGEFVSVFTNEKQHCVYVASDV
 At_RPA2 GMV--ESLPKLVRAQPPKVIHVLLDQVVGTLSSNLVTKVVSYIRRLKVEAPSVIPEDLE
 Sc_RPB2 QS---EDATRVFVNGVWHGVHRNPARMETLRTLRRKGDINPEVSMIRDIREKELKIFTD
 Ec_RpoB -----
 consensus I P KILVNGIW GVHR D V LRS RR DV EV IIRD ELRIFTD

At_RPD2 AGRLLRPLLVVEN-----LQKLKQEKPSQYP-----FDHLLDHGI
 At_RPB2 YGRCSRPLFIVDN-----QKLLIKKRDYALQQRSAEEDG-----WHHLVAKGF
 At_RPC2 GRVCRPLVIADKG-----ISRVKQHMKELQDGVR-----TFDDFIRDGL
 At_RPA2 VGVVPTSMGGSYPG-----LYLASC PARFIRPVKN-----ISIPSDN
 Sc_RPB2 AGRVYRPLFIVEDDDESLGHKELKVRKGHIAKLMATEYQDIEGGFEDVEEYTWSSLLNEGL
 Ec_RpoB -GFLETFYRKV-----TDGV
 consensus AGRL RPL IVE I RE D F LI DGL

At_RPD2 LELIGIEEEEDCNTAWGIKQLLKEPK-----IYTHCELDLS
 At_RPB2 IEYIDTEEEETTMTISMTISDLVQARLRPEE-----AYTENYTHCEIHPS
 At_RPC2 IEYLDVNEENNALVCLRAEAAK-----ADTTHIEIEFF
 At_RPA2 IELIGPFFQVANPINIIFISTFP-----ATHEEIHPT
 Sc_RPB2 VEYIDAEESLSILIAMQPEDLEPAEANEENDLDVDPK---RIRVSHHATTFTHCEIHPS
 Ec_RpoB VTDEIHYSALIEEGNYVLAQANSNLDEEGHFVEDLVTCRSKGESSLFRDQVDYMDVSTQ
 consensus IEYID EEEE LI MI L YTHCEIHPS

Conserved domain E

At_RPD2 FLLGVSCAVVPPANHDHGRVLYQSQKHCQQAIGFSSINPNIRCDTISQQLFYQKPLFK
 At_RPB2 LILGVCASIIIPFPDHNQSPRNTYQS-AMGKQAMGIYVINYQFRMDTLAVLYYPQKPLVT
 At_RPC2 TILGVVAGLIPYPHHNQSPRNTYQC-AMGKQAMGNIAYNQLNRMDTLLYLLVYPQRPILT
 At_RPA2 GMSIVVANITPWSHDHNQSPRNTYQC-QMAKQTMAYSTQALQFRADQKIYHLQTPQSPVVR
 Sc_RPB2 MLLGVAASIIIPFPDHNQSPRNTYQS-AMGKQAMGVFLINYNVRMDTMANILYYPQKPLGT
 Ec_RpoB QVVSIGASLIPFLEHDDANRALMGA-NMQRQAVPT-----LRAD-----KPLVVG
 consensus ILGV ASLIPFPDHNQSPRNTYQS AMGKQAMG TN N RMDTL YLLYYPQKPLVT

Conserved domain F ** (active site)****

At_RPD2 TLASECLKKEVLFNGQNAIVAVNVHLGYNQEDSIVMNKASLERGMFRSEQIRSYKAEVDA
 At_RPB2 TRAMEHLHFRQLPAGINAIVAISCYSGYNQEDSVIMNQSSIDRGFFRSYRDEEKK
 At_RPC2 TRTIELVGYDKLGAGQNAIVAVMSFSGYDIEDAIVMNKSSLDRGFGRCIVMKKIVAMSQK
 At_RPA2 TKTYTYSIDENPTGTNAIVAVLAHTGDFMEDAMILNKSSVERGMCHGQIYQ TENIDLSD
 Sc_RPB2 TRAMEYLFKRELPAGQNAIVAIAICYSGYNQEDSMIMNQSSIDRGLFRSIFFRSYMDQEKK
 Ec_RpoB TGMERAVAV-ELALGQNMVAFMPWNGYNFEDSILVSEVVQEDRFTTTHI--QELACVS
 consensus TRAME L FDELPAQNAIVAVL YSGYNQEDSIIMNKSSIDRGMFRSI FRSY E K
 82 aa deleted

Conserved domain G

At_RPD2 KDSEKRKKMDELVQFGKTHSKI GKVDSLEDDGFPFIGANMSTGDIVIGRCTESG-----
 At_RPB2 MGTLVKEDFGRFDRGSTMGRHGSYDKLDDDG LAPPGTRVSGEDVIIGKTPISQDEAQQ
 At_RPC2 YDNCTADRILIPQR---TGPDAEKMQILDDDG L ATPGEIIRPNDIYINKQVVDVTVKFT
 At_RPA2 QNS----RFDSGSKSFRRSTNKAHFRIADGLPSVGGKLYPDEPYCSIYDEVTN-----
 Sc_RPB2 YGMSITETFEKQRTNTLRMKHGT YDKLDDDG L IAPGVRVSGEDVIIGKTPISPDEEEL
 Ec_RpoB RDTKLGPEEITADIPNVG---EAALS KLDES GIVYIGAEVTEGGDILVGVLPKGETQL
 consensus DS I ERFD P R K G LDKLDDDG L PG R VSGEDIIIGK TPIS
 9 aa deleted

Conserved domain H

At_RPD2 -----ADHSIKLKHTERGIVQKVVLSS-NDEGKNFAAVSLRQVRSFCLGDKFSSM
 At_RPB2 --QS-SRYTRRDHSISLRHSETGMVDQVLLTT-NADGLRFVKVVRVRSVRIPOIGDKFSSR
 At_RPC2 SALSDSQYRPAREYFKGPEGETQVVDVALCS-DKKQQLCIKYIIRHTRRPELGDKFSSR
 At_RPA2 -----KTRHMKRKGIDPVIIVDFVSVDMKSKKHPQRANIRFRHARNPIIGDKFSSR
 Sc_RPB2 GQRT-AYHSKRDASTPLRSTENGIVDQVLVTT-NQDGLKFKVVRVRTTKIPQIGDKFASR
 Ec_RpoB -IFGEKASDVKDSLLRVPNGVSGTVIDVQV-----LKIIVKYLAVKRIQPGDKMAGR
 consensus S KD SIKLK TETGIVD VLLTS N DGLKFKVVR LR R PQIGDKFSSR
 113 aa deleted

```

At_RPD2  HGQKGVLGYLEEQNFPPFT-IQGIVPDIVINPHAFPSRQTPGQLLEAALS*KGIACP---I
At_RPB2  HGQKGTVGMTYTQEDMPWT-IEGVTPDIIVNPHAI*PSRMTIGQLIECIMGK-----
At_RPC2  HGQKGVCGIIIQQEDFPFS-ELGICPDLIMNPHGFPSRMTVGKMI*ELLGSKAG-----
At_RPA2  HGQKGVCSQLWPDIDMPFNGVTGMRPDLIINPHAFPSRMTIAMLESIAAKGGS*LHGK*FV
Sc_RPB2  HGQKGTIGITYRREDMPFT-AEGIVPDLIINPHAI*PSRMTVAHLIECLLSK-----
Ec_RpoB  HGNKGVISKINPIEDMPYD-ENGTVPDIVLNPLGVPSRMNIGQILETHLGM*AAKGIGDKI
consensus HGQKGVIGIIY QEDMPFT I GI PDIIINPHAFPSRMTIGQLIE ILSKAG I

```

Conserved domain I

```

At_RPD2  QKEGSSAA*YTKLTRHATPFSTPGVTEITEQLHRAGFSRWGNERNVYNGRS*GEMMRS*MI*FMG
At_RPB2  -----VAAHMKEGDATPFTDVTVDNISKALHKCGYQMRGFERYNNGHTGRPLTAMIFLG
At_RPC2  --VSCGRFHYGSAFGERSGHADKVETISATLVEKGFSSYSCKDLLYSGISGEPVEAYIFMG
At_RPA2  DATPFRDAVKKTN*GEESKSSLLVDDLGSMLEKGFNHYGTETLYSGYL*VELKCEIFMG
Sc_RPB2  -----VAALSGNEGDA*SPFTDITVEGISKLLREHGYSRGGFVYNGHTGK*LM*QIFFG
Ec_RpoB  NAMLKQQQEVAKLREFIQ-----LLKLGDLPTS*QIRLYDGR*TGE*QFER*VTVG
consensus AA G GDATPFS I VD IS LLHE GFQ G ERLYNG TGE L A IFMG
                    51 aa deleted

```

```

At_RPD2  PTFYQRLVHMS*EDKVKFRNTG*PVHPLTRQPVADRKRFGG*IKFGEMERDC*LIAHGASANLH
At_RPB2  PTFYQRLKHMVDDKIH*SRGRGPVQILTRQPAEGRSRD*GGLRFGEMERDC*MLAHGAAHFLK
At_RPC2  PTFYQRLKHMVLDKMHARGSGPRVMMTRQPT*EGKSKNGGLRVGEMERDC*LIAYGASMLIY
At_RPA2  PTFYQRLRHMVSDKFQVRSTGQVDQLTHQPIKGRKRG*GGIRFGEMERDS*LLAHGAS*YLLH
Sc_RPB2  PTFYQRLRHMVDDKIHARARGPMQVLT*RQPV*EGRSRD*GGLRFGEMERDC*MLAHGAA*SFLK
Ec_RpoB  YMLKLNHLVDDKMHARSTGSYS*LV*TQOPLG*GKAQFG*QRFGEMEV*WAL*EAYGAAYTLQ
consensus PTFYQRLKHMVDDKIHARGTGPV ILTRQPV*EGRSR GGLRFGEMERDC*LIAHGAS L

```

```

At_RPD2  ERLFTLSDSSQM*HICRCK*TYANVIERTPSSG-----RKIRGPYCRV*CVSSDH
At_RPB2  ERLFDQSDAYRVHVC*EVCG-LIAIANLKKNS-----FE*CRGCKN*KTD
At_RPC2  ERLMISSDPFEVQVCRACGLLGYNYK*LKKA-----VCTCKN*GDN
At_RPA2  DRLHTSSDHHIADVCSL*CGSLTSSVNVN*QKKLIQEIGKLPGRTPK*VTQ*YSCK*TSKG
Sc_RPB2  ERLMEASDAFRVHICGICGLMTVI*AKLNHNQ-----FE*CKG*DNKID
Ec_RpoB  EMLTVKSDDVNGR*TKMYKNIVDGNHQMEP-----
consensus ERL SD F VHVC ICGLL I L N CR CKN

```

```

At_RPD2  VVRVYVPYGA*KLLCQELF*SMGITLNFD*TKLC-----
At_RPB2  IVQYIPYA*CKLLFQELMSMAIAPRMLTKHLKSAKGRQ
At_RPC2  IATMKLPYA*CKLLFQELQSMNVVPR*LKLTEA-----
At_RPA2  METVAMPYVFRYLAAELASMNIKMTLQLSDREGVTD--
Sc_RPB2  IYQIHIPYA*AKLLFQELMAMNITPRLYTDRSRDF----
Ec_RpoB  ---GMPESFNVL*LKEIRSLGINIELEDE-----
consensus I V IPYA KLLFQEL SMNI PRL T

```

The alignment was performed as described previously for the largest subunits. Positions with identical amino acids are indicated by green shading, while similar amino acids are indicated by yellow shading. The last line in the alignment indicates the consensus sequence. Conserved domains (Sweetser et al., 1987) are indicated with letters and bold lines above the alignments. The active site (metal B site; Cramer et al., 2001) is indicated with asterisks. Protein sequences examined are: At_RPD2 (Pol IV), At_RPB2 (Pol II), At_RPC2 (Pol III), At_RPA2 (Pol I), Sc_RPB2 (Pol II) and Ec_RpoB. Regions of the *E. coli* β subunit that do not align with the eukaryotic RNAP proteins were deleted, as indicated below the alignments.

Figure S3. Comparison of Conserved Domains A–H in RPD1a and DNA-Dependent RNA Polymerase Largest Subunits in *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec)

Supplemental Figure 3. Domain Alignments for DNA-dependent RNA Polymerase Largest Subunits

DOMAIN	GENE	AMINO ACIDS	SEQUENCE	
A	At RPD1	37–96	EAPNOVTSRLGLNPDSVCRFCGSKDRKVCCEGHFVINFYSLINPYLKEVAALINKI	
	At RPB1	47–105	PKVGLSSTRLETIIRVVKETCMAN-MAECPGHFVYLELAKMVMHVGEMKTVLSIMRIV	
	At RPC1	56–114	RYENGLLDPMPNPKRSIITTCBGN-FQNCPEHYCYLKLDELVYVNGVYFNFLDILKCI	
	At RPA1	60–118	PFPGGLYDLKLPKIDKQAINSCGQL-KLACCPGHFCHIELVFPIYPLLNLLNFNFORA	
	Sc RPB1	48–106	AKICGLNPRLLGSIIRNLKQCEQEG-MNECPGHFCHIDLKVFVHGHIAKIKKVCCEIV	
	Ec RPOC	51–106	LERDGLFCALIFGVKDYELCGKYK-RLK---LRVYCEKCGEVTQTKVRRERMGHIE	
	consensus		P GGL D RLG PDKK C TC R CPGHFG IELA PVYHVGF I I L I C I	
	B	At RPD1	214–260	MFNSLGLTSPFVITNGYVTEIVHGFNGALIFPDERTRLYKKLVGF
		At RPB1	242–288	RPDWMLLEVLTPPEPVRPSVMDATSRSEDDLTQCLAMTRHNE
		At RPC1	249–295	RPEMLITICMLVPLSRPSVMIGGIQSNNDLAKARQKILGNAS
At RPA1		337–382	DYSMFFIESLVPPTKFRPPTTGGD-SVMHPCVGNKLVESNII	
Sc RPB1		230–276	RPEWMLITCLPVPPEPVRPSISFNESQRGDDLTFLKADLTKANISL	
Ec RPOC		233–279	KPEWMILTLPVLPEDLRRLPLDGGRFATSLNDLYRRVNRNR	
consensus			RPEWMILT LPVPPP VRPSVMI G Q AE DLT RLR I I K N L	
C		At RPD1	301–356	PKICLR-FMKDVLKRSHTFRVTVVGSKINIGIFESIAKRQVSHLNO
		At RPB1	339–395	SRKAREGIFGNLMKRVDFSAITVITDPTINIDLGVFSLANLTYBETVITY
		At RPC1	344–400	QRLKGGGRFRAMSGKRVEITGRVISPDPAKITEVGIILMAQILTFPECVSRH
	At RPA1	428–484	QLHEKKEGLRQKMGKRVNHACRSVISDPYIAVNDIGIPCFAKLITYPERVTFW	
	Sc RPB1	327–383	ARLRKKEGIRGNLMKRVDFSAITVISGPNTEIDQVGVKSTAKTLTYPEVITY	
	Ec RPOC	329–385	DMIKGQGRFRQNLGKRVVYSGHVSITVGVYLRHQCLLKKMAELFKFIYKGL	
	consensus		RLKKEGGRFRGNLMKRVDFSAITVISDPP LKL EIGIP SIAL LTYPE VTFY	
	D	At RPD1	407–468	VLMNPFSSIQHSLIAMTVRILFTTVVSLNPICLLFRGDFDGCCLHGYVQSIQAKVLD
		At RPB1	451–511	VLENRQPSLHKMSIMGHRIIMF-YSTRNLNLSVTSYPNADFGEEMNHVQSFTRAEVL
		At RPC1	460–520	VLENRQPSLHKMSIMCHRAIMF-WRLLFRFSVGNYPNADFGEEMNHVQSFTEARTEFI
At RPA1		562–623	VLVNRQQLTHHEFLMVKRVILKGEKILRHLYANFTYNADFGEEMNHVQSFQDISRAEAY	
Sc RPB1		442–502	VLENRQPSLHKMSIMAHRRKVIY-YSTRNLNLSVTSYPNADFGEEMNLHVQSEFTRAEVL	
Ec RPOC		421–481	VLLNAAFLHRLGQAFEPVLEE-GKALQHLPLCAAYNADFGEQMAVHVPLTLEQLAR	
consensus			VLFNQRQPSLHKMSIMAHRRVRIIP YSTRNLN SVCSYPNADFGEEMNHVQSEARAEA	
E		At RPD1	524–560	LPTPAILKASPSSTEEQNTGMQLGMLFPGFD-YTYPLNNVV
		At RPB1	569–607	VPAPAILK-----RPLWTGKQVNLIIKQINLLRYSAWHADT
		At RPC1	549–587	LPTPAILK-----IELWTGKQIESVLLRPNASIRVYVTLNVE
	At RPA1	703–741	TVTPAILK-----VELWTGKQIVAVLNQITKGHPPTVEKAT	
	Sc RPB1	560–593	LPTPAILK-----KELWSKQILVALNGLHQR-----DE	
	Ec RPOC	530–563	PKEAERLYR-----SGLASLHARVKRITTEYKEDANG-----EL	
	consensus		LPTPAILKP PLWTGKQIFGVLIIP L Y D	
	F	At RPD1	728–817	MSKAGSKGNIGRLVTHMCTGLONSAVLSFGFPRLLCAAWNDPNSPLRGAAGKGDSTTTESYVPGVGIENSFLTGLNLESLVSVTSR
		At RPB1	760–835	MVTAGSKGSEFINISDMTACVCGQONVEGKRIFPGDGRITLPHPTKDYCESR-----GFVENSYRLTLPQEPFFHAMGGR
		At RPC1	757–832	MSQCSKSGSEFINISDMVACVCGQNTNHHAPDGHIDSLPHFPRMSKSEAAK-----GFVANSSYSLTATPFFPFFHAMGGR
At RPA1		941–1016	MTLSAKGSKVINFOQISHLGQODLEGRVFRMVSCKTLPCHFWPWSERAG-----GFISDRFLSGLRPEYFFPFFHAMGGR	
Sc RPB1		746–821	MVMAGSKGSEFINIADMSACVCGQSVGKRIAFSGVDTLPHEKDDYSESK-----GFVENSYRLTLPQEPFFHAMGGR	
Ec RPOC		725–780	MADSARSSAACRDLAGMRGLMAKPDGS-----IIE-----TPIITANRELNVLQYISTHGA	
consensus			MS AGSKGS INI QMSACVGGQ VEGKRIP GF DRTLPHF K DYSP AK GFVENSFLSGLTLPQEPFFHAMGGR	
At RPD1		818–843	DSSFSGNADLP--TISRLMPPFRITY	
At RPB1		836–863	EGLIDTAVKTSRSGYIQRRLVKALEDIM	
At RPC1		833–860	EGLVDTAVKTSRSGVMSRRLMKALEDLL	
At RPA1	1017–1044	EGLVDTAVKTSRSGYLQRCIMNLESLK		
Sc RPB1	822–849	EGLIDTAVKTAETGYIQRRLVKALEDIM		
Ec RPOC	781–788	KGLADTALKTANSGYLFRRLVDVAQLV		
consensus		EGLIDTAVKTA TGYLQRRLMKALEDIM		
G	At RPD1	945–1006	LEKNHLEKLSFSEIVSTSMIFSPSSNTKVPVLSPWVCHITISEKVLKQKSAESVSSLN	
	At RPB1	1062–1121	GELEFRLLQLVADGEMICVAAQSIGEPATQMT--LNTFHYAGVSAKNTLVGVRRLRILIN	
	At RPC1	967–1026	LYKAGVTDKQLEATAICTGAQSIGEPATQMT--LKTFHFAGVASMILIQGVPHINRILIN	
	At RPA1	1041–1100	KLIVKRFARLAAQGEVGVVLAQSIGEPATQMT--LNTFHLAAGEMVILGIDRLOEILM	
	Sc RPB1	1047–1106	SNIEACQLRNVVHGEVGVVLAQSIGEPATQMT--LNTFHFAGVASKRVISGVPRILKILIN	
	Ec RPOC	898–1146	CYGRDLARGHILNKREAGVIAAQSIGEPATQMT--MRTFHIGA---DITGLLEFVADLFE	
	consensus		IKS F SLV PGE IGVIAAQSIGEPATQMT LNTFHFAGVA KNVTLVGVRRL EILN 192 aa deleted	
	At RPD1	1214–1268	PAFTQCSSPSCFLKAAKEVREDDQSSIDALAWKVPVGFSTGQDFEIIISP	
	At RPB1	1410–1464	LGELMRCSEFEVVDILLDANAYATDCLRGVFNEMIGOLABITGDCLEVLNDE	
	At RPC1	1261–1315	KVLMCASFERGDFHFSAAASFKVONIEVTECHMGPMPKLGTRILKVLQRTD	
At RPA1	1515–1569	TSFCHMTFFATKFIYQAATYGEKTIETPSARICGLPLSITGCFDLMQVFE		
Sc RPB1	1394–1448	LGALMRCSEFEVVEIFEAASALLDRCGVSENVILGOMABITGAFDVMIDE		
Ec RPOC	1317–1371	EFISASRQETRVITPAVAVKRELRGLKENVIVRLIAGTYAYHODMR		
consensus		TSPLMRASFETDILLDAAA GERDRLRGVSENVILG LAPITG DLMIR E		

The alignment for each conserved domain, determined using ClustalX, was exported to BOXSHADE. Positions with identical amino acids are indicated by green shading; similar amino acids are indicated by yellow shading. The last line in the alignment indicates the consensus for all sequences. Proteins whose domains are aligned are: At_RPD1 (Pol IV), At_RPB1 (Pol II), At_RPC1 (Pol III), At_RPA1 (Pol I), Sc_RPB1 (Pol II), and Ec_RPOC (β' subunit)

Figure S4. Comparison of Conserved Domains A–I in RPD2 and DNA-Dependent RNA Polymerase Second-Largest Subunits in *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec)

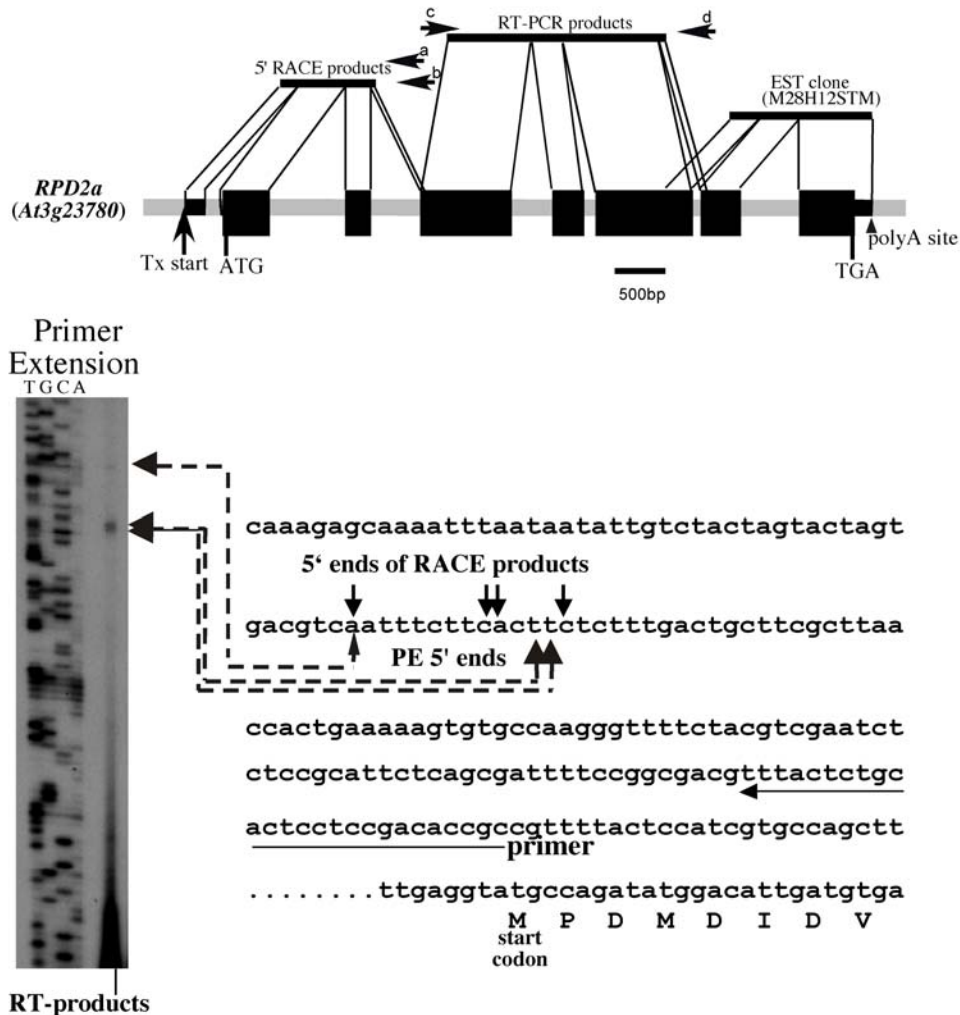
Supplemental Figure 4. Domain Alignments for RNAP Second-Largest Subunits

DOMAIN	GENE	AMINO ACIDS	SEQUENCE
A	At RPD2	119-142	WHARLQNMITYSARIKUNVQVEVFK
	At RPB2	113-136	KAARLENLTYSAFLYVDVTKRVIK
	At RPC2	105-128	HMCRLADMITYAIFVNIIEYVHGS
	At RPA2	38-61	FECQAKISYTGTFMAIVCFKYND
	Sc RPB2	115-138	QEARLRNLTYSSGLFVDVKKRITYE
	Ec RPOB	83-106	QECQIRGVITYSAPLRKLRVVIYE
	consensus		e rlrnvtYsaplyvdv riye
B	At RPD2	206-220	GVFVIKGAEKVFLIAQ
	At RPB2	189-203	GVFIINGSEKVLIAQ
	At RPC2	176-190	GVFIIKSTKVLIIQ
	At RPA2	105-119	GVFIIINGIERVFERCV
	Sc RPB2	201-215	GVFIINGSEKVLIAQ
	Ec RPOB	134-148	GVFVIINGTHERVIVSQ
	consensus		GyFiingtEkvliaq
C	At RPD2	385-400	GRKCENTRISFNFKRI
	At RPB2	383-398	GRRPEDDRDHYGNKRL
	At RPC2	354-369	NNDAMDDKIVYGNKRL
	At RPA2	298-313	QTSLFPNPFSLQNGEI
	Sc RPB2	391-406	DRKQDDDRDHFQKRL
	Ec RPOB	438-453	GRGEVDDIDHLENRRI
	consensus		gkr ddrDh gnkri
D	At RPD2	507-536	FYPHPSHWRVCFLESTPDCENCGLVKNMSL
	At RPB2	503-532	RQLHNSQWGMCFAPSTPEQAGGLVKNLAL
	At RPC2	480-509	RSIQPESQWMLCPDTPPEGESCGLVKNLAL
	At RPA2	431-460	RKLLHESWGFPCVHTPDCTPCGLLNHMTR
	Sc RPB2	512-541	RQLHNTHWGLVCPAETPEQAGGLVKNLAL
	Ec RPOB	548-577	RDVHPTHYGRVCPETPEQPNIIGLINSLSV
	consensus		R lhpshwGmvCpieTPeG cGLvknls1
E	At RPD2	696-714	LLGVSCAVVFFANHDGFR
	At RPB2	711-729	ILGNVCSILIPFDHNQSPR
	At RPC2	675-693	ILGNVAGLIPYEHHNQSPR
	At RPA2	620-638	MISVAVNTPWSDHNQSPR
	Sc RPB2	748-766	ILGVAASITIPFDHNQSPR
	Ec RPOB	660-678	VVSWGASLIPPLEHDDANR
	consensus		ilgV asliPfpdHnqspR
F	At RPD2	765-800	VIFNGQNAIVAIVNVHLSYNOEDSIVMKNASLERGM
	At RPB2	779-814	OLEPAGINAIVAISCYSGYNOEDSIVMKNQSSIDRGGF
	At RPC2	743-778	KLAGQNAIVAIVMSFSGYDIEDALVMKNSSLDKRGFG
	At RPA2	688-723	ENFTTINAIVLAHTGFDMEDAMLNKSSVERGMC
	Sc RPB2	816-851	ELPAGQNAIVAIAICYSGYNOEDSMHNQSSIDRGL
	Ec RPOB	793-828	ELALGQNMRAVAFMPWNSYNFEDSLVSEKRVQEDRR
	consensus		e1paGqNaIvAvm wsGynqEdsIimnkssvdrgmf
G	At RPD2	836-866	IGKVDSEDDGFFPFIKANMSTGDIVIGRTE
	At RPB2	850-880	HGSYDKLDDGGLAPPCTRVSGEDVILGKTFE
	At RPC2	811-841	AEKMQILDDGGLATPGEIIRPNDIYNKQVE
	At RPA2	755-785	KAHFRIADGLPSVQKLYPDEPYCSIYDE
	Sc RPB2	887-917	HGTYDKLDDGGLIAGVVRVSGEDVILGKTFE
	Ec RPOB	859-889	EAALSGLDESGLIVYIAEATHGDLVGVVTE
	consensus		hg ldkldddgl pG rvsge diligk tp
H	At RPD2	895-966	KNFAASLQVRSCLGDKFSSMHGQKGVLYLEEQNFPPTIQGVFDIVINPHAPPSRQTPQLLEAALS
	At RPB2	922-993	LRFVKVRSVRIPIQIGDKFSSRHGQKGTVMGTYTQEDMPTLIEGVTPDITVNPHATPSRMTIGLIECTMG
	At RPC2	886-957	QLCKIYIIRHTRFELGDKFSSRHGQKGVCHIIIQDEFPFS-ELGICPDLIMNPHGPPSRMTVKKMIELLGS
	At RPA2	816-888	PQRANIRFHARNEIVGDKFSSRHGQKGVCSQLWPDIDMPEVNGTMRPDLIINPHAPPSRMTIAMLLESIAA
	Sc RPB2	961-1032	LRFVKVRSVRIPIQIGDKFASRHGQKGTIIGITRRREDMPT-AGGVFDLIIINPHATPSRMTVAHLIECLLS
	Ec RPOB	1047-1118	LKIVKIVLAVKRIQPPDEKMAGRHGKGVISKINPIEDMPEVD-ENSTPVDIVLNLGVESRNIIGLIECTMG
	consensus		lkfvkvrlr r pqlGDKfssrHGqKGVigmiy qedmPft i Gi pdiInPhafPSRmtigqllE ils
I	At RPD2	1003-1101	QLHRAGSRWGNRERVINGRSGEMMRSMIFMGPTFYQRIVHSEDKVKFRNTGFPVHPLTRQFVADKRFGGIKFGEMERDCLIAGASANLHERLFTLSD
	At RPB2	1019-1117	ALHKCYQMRGFPERMYNGHTGRPTAMIFLGPITYQRIKHMVDDKIHSGRSPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGA AHFKERIFDQSD
	At RPC2	988-1086	TIVKGEFSYSKDLNSGISGEPVEYIFMGPIYQRIKHMVLDKMHARGSGFRVMMTROPTGKSKNGGLRVGEMERDCLIAYGASMLIYERIMISSD
	At RPA2	928-1026	MLKKEGNSHYCTETLNSGYLVEIKCEIFMGPIYQRIKHMVDDKIQFVRSVTCQDQLTHQIKGKKGGLRFGEMERDCMIAHGAASFLKRIEASD
	Sc RPB2	1058-1156	LURHEGYSRGEFVMYNGHTGKLMQIIFGPTFYQRIKHMVDDKIHARSGEMQVLTROPVGRSRDGGLRFGEMERDCMIAHGAASFLKRIEASD
	Ec RPOB	1198-1296	LKLGDLPTSGQIRLYDGRTEGQFERPVTVGYMMLKLNHLVDDKMHARSGYSYSLVTCQPLGSKAQFGGRFGEMERDCMIAHGAASFLKRIEASD
	consensus		l1kekqfG e rlyNgrtGe l a ifmgptyyqrLkHmvdDKmhrRgtGpv llTrQPlgrsr GGLrFGEMERdccliAhGA lherL SD

The alignments were conducted and displayed as described for Supplemental Figure 3. Proteins whose domains are aligned are: At_RPD2 (Pol IV), At_RPB2 (Pol II), At_RPC2 (Pol III), At_RPA2 (Pol I), Sc_RPB2 (Pol II), and Ec_RPOB (β' subunit).

III. Determination of RPD2a Full-Length mRNA Sequence

Figure S5. Determination of the Full-Length mRNA Sequence for RPD2a by RT-PCR, 5' RACE, and Primer Extension



The diagram shows the relative positions of the eight exons, depicted as black rectangles with coding regions expanded in size relative to the 5' and 3' untranslated regions. The transcription (Tx) start site, initiation codon (ATG), stop codon (TGA) and poly A addition sites are indicated. Also shown are the relative positions of a pre-existing partial cDNA (EST M28H12STM) and the clones obtained by RT-PCR and 5' RACE that were sequenced as part of this study. Shown at the lower left is an autoradiogram displaying primer extension products run adjacent to a sequencing ladder generated using the same primer. Minor and major start sites were detected by primer extension, corresponding closely to the 5' ends of sequenced 5' RACE products.

Supplemental Experimental Procedures

The 5' portion of the RPD2a mRNA sequence was amplified by 5' RACE (rapid amplification of cDNA ends) using Invitrogen's GeneRacer kit with nested-PCR primers 5'-CGGACCTGAAGGAGACTGTCCATG-3' and 5'-TCCGAGAGGCGCACAATGAA-3' (primers a and b, respectively in the diagram). The central region of the RPD2a mRNA sequence was amplified by reverse transcription followed by PCR (RT-PCR) using primers 5'-ATGCCAGATATGGACATTGATGTGAAGGAT-3' and 5'-ATCAGCATAGCTTGGTGTCTGAAGTTGAG-3' (primers c and d, respectively in the figure). The resulting cDNA fragments were cloned using the TOPO TA Cloning Kit (Invitrogen) and sequenced using an ABI automated sequencer and big dye terminator technology. To verify the 5' ends determined by 5' RACE, primer extension was performed according to standard methods (Sambrook and Russell, 2001). A 30 nt antisense oligonucleotide (5'-AACGGCGGTGTCGGAGGAGTGCAGAGTAAA-3') that was 5' end-labeled using T4 polynucleotide kinase and [γ - 32 P] ATP was used as the primer. The reverse transcription reaction was performed using ~1.0 ug Poly(A)⁺ RNA and SuperScript RNase H⁻ reverse transcriptase (GIBCO BRL). Primer extension products were subjected to electrophoresis on a denaturing polyacrylamide sequencing gel alongside sequencing reactions generated using the same end-labeled primer. The resulting gel was vacuum dried onto filter paper and exposed to X-ray film.

IV. Supporting Data for Cytological Observations

Table S3. Cytological Changes in *rpd2* Mutants

Chromocenters (CCs)

Genotype	Number of cells analyzed	Patterns observed	
		6-10 large, diffuse CC's	≤4 CCs
Wild-type	80	93%	7%
<i>rpd2</i> double mutant	120	34%	66% (>20 small DAPI foci)

$\chi=68.56, p<0.001$

NORs

Genotype	Number of cells analyzed	Number of FISH signals per nucleus				
		1	2	3	4	>4
Wild-type	60	0%	36%	25%	39%	0%
<i>rpd2</i> double mutant	46	0%	19%	30%	28%	23%

$\chi=17.95, p<0.001$

5S rRNA genes

Genotype	Number of cells analyzed	Patterns observed	
		Substantial colocalization with centromeres	Substantial dispersal away from centromeres
Wild-type	65	69%	31%
<i>rpd2</i> double mutant	72	42%	58%

$\chi=10.5, p=0.0012$

Supplemental References

Bergsland, K.J., and Haselkorn, R. (1991). Evolutionary relationships among eubacteria, cyanobacteria, and chloroplasts: evidence from the *rpoC1* gene of *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* *173*, 3446–3455.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* *292*, 1863–1876.

Jokerst, R.S., Weeks, J.R., Zehring, W.A., and Greenleaf, A.L. (1989). Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*. *Mol. Gen. Genet.* *215*, 266–275.

Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.P., Lottspeich, F., Garrett, R.A., and Zillig, W. (1989). Archaeobacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc. Nat. Acad. Sci. USA* *86*, 4569–4573.

Sambrook, J., and Russell, D.R. (2001). *Molecular Cloning: A Laboratory Manual, Third Edition* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schneider, G.J., and Hasekorn, R. (1988). RNA polymerase subunit homology among cyanobacteria, other eubacteria and archaeobacteria. *J. Bacteriol.* *170*, 4136–4140.

Sweetser, D., Nonet, M., and Young, R.A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc. Nat. Acad. Sci. USA* *84*, 1192–1196.

Wolfe, K.H., Gouy, M., Yang, Y.W., Sharp, P.M., and Li, W.H. (1989). Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Nat. Acad. Sci. USA* *86*, 6201–6205.

CHAPTER 4

METAL A AND METAL B SITES OF NUCLEAR RNA POLYMERASES POL IV AND POL V ARE REQUIRED FOR siRNA-DEPENDENT DNA METHYLATION AND GENE SILENCING

Published in PLoS ONE (2009) 4 (1): e4110.

My contributions to this work:

I designed and performed all experiments in the main text and supplemental materials including identification of substituted “invariant” amino acids in NRPD1, NRPE1 and NRPD2/NRPE2, the generation of both wild type and active site mutant transgenic lines, co-immunoprecipitation experiments, flowering time analysis, and DNA methylation, siRNA production and transcription analysis of selected loci. These experiments demonstrated that the Metal A and Metal B sites of Pol IV and Pol V are required for *in vivo* activity. Olga Pontes performed immunolocalization analysis of the NRPD1 and NRPE1 Metal A mutants in Figure 5 demonstrating a disruption of wild type nuclear localization patterns in the majority of analyzed nuclei. I wrote and edited the paper, with the assistance of Craig Pikaard. I contributed significantly to the intellectual value of the paper, with the assistance of Craig Pikaard.

Metal A and Metal B Sites of Nuclear RNA Polymerases Pol IV and Pol V Are Required for siRNA-Dependent DNA Methylation and Gene Silencing

Jeremy R. Haag, Olga Pontes, Craig S. Pikaard*

Department of Biology, Washington University, St. Louis, Missouri, United States of America

Abstract

Plants are unique among eukaryotes in having five multi-subunit nuclear RNA polymerases: the ubiquitous RNA polymerases I, II and III plus two plant-specific activities, nuclear RNA polymerases IV and V (previously known as Polymerases IVa and IVb). Pol IV and Pol V are not required for viability but play non-redundant roles in small interfering RNA (siRNA)-mediated pathways, including a pathway that silences retrotransposons and endogenous repeats via siRNA-directed DNA methylation. RNA polymerase activity has not been demonstrated for Polymerases IV or V *in vitro*, making it unclear whether they are catalytically active enzymes. Their largest and second-largest subunit sequences have diverged considerably from Pol I, II and III in the vicinity of the catalytic center, yet retain the invariant Metal A and Metal B amino acid motifs that bind magnesium ions essential for RNA polymerization. By using site-directed mutagenesis in conjunction with *in vivo* functional assays, we show that the Metal A and Metal B motifs of Polymerases IV and V are essential for siRNA production, siRNA-directed DNA methylation, retrotransposon silencing, and the punctate nuclear localization patterns typical of both polymerases. Collectively, these data show that the minimal core sequences of polymerase active sites, the Metal A and B sites, are essential for Pol IV and Pol V biological functions, implying that both are catalytically active.

Citation: Haag JR, Pontes O, Pikaard CS (2009) Metal A and Metal B Sites of Nuclear RNA Polymerases Pol IV and Pol V Are Required for siRNA-Dependent DNA Methylation and Gene Silencing. PLoS ONE 4(1): e4110. doi:10.1371/journal.pone.0004110

Editor: Frederic Berger, Temasek Life Sciences Laboratory, Singapore

Received: October 8, 2008; **Accepted:** November 30, 2008; **Published:** January 1, 2009

Copyright: © 2009 Haag et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Work in the Pikaard lab is supported by National Institutes of Health grant GM077590. The Pontes laboratory is supported by a grant from the Edward Mallinckrodt Foundation. The content of this paper is solely the responsibility of the authors and does not necessarily reflect the views of the NIH or the Edward Mallinckrodt Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pikaard@biology.wustl.edu

Introduction

The largest and second-largest subunits of eukaryotic multi-subunit nuclear RNA polymerases are homologs of the β' and β subunits of *E. coli* RNA polymerase, respectively, and of the equivalent largest subunits of eukaryotic RNA polymerases I, II and III. These subunits interact to form the entry and exit channels for the DNA template, the catalytic center for RNA polymerization and the exit channel for the RNA transcript [1]. The largest and second-largest subunits of RNA polymerases IV and V (abbreviated Pol IV and Pol V) were initially identified upon analysis of the *A. thaliana* genome sequence, which led to the identification of two genes for an atypical fourth class of largest subunit and two genes for an atypical fourth class of second-largest subunit in addition to the canonical Pol I, II and III subunits [2,3]. Phylogenetic analyses suggest that the atypical subunits arose from duplicated Pol II subunit genes in a multi-step process that began in green algae prior to the evolution of land plants [4] more than 500 million years ago.

For purposes of subunit nomenclature, nuclear RNA polymerases I, II and III in Arabidopsis are designated NRPA, NRPB and NRPC and their largest subunits are NRPA1, NRPB1 and NRPC1. Extending this convention to the atypical polymerases, their largest subunits have been designated either NRPD1a and NRPD1b [5,6] or RPD1 and RPE1 [4]. The latter nomenclature has been adopted, in modified form [7], to allow the naming of Pol

IV-specific subunits using the NRPD prefix and the naming of Pol V-specific subunits (formerly Pol IVb) using the NRPE prefix. There are two atypical second-largest polymerase subunit genes, but only one is functional in Arabidopsis and is used by both Pol IV and Pol V, as shown by co-immunoprecipitation, colocalization [8] and genetic evidence [9,10]. This second-largest subunit gene has the synonymous names NRPD2a (NRPD2 for simplicity) and NRPE2.

The *NRPD1* (*NRPD1a*), *NRPE1* (*NRPD1b*) and *NRPD2/NRPE2* genes are not essential for viability [5,6,9,10], unlike the genes encoding the equivalent subunits of Pol I, II and III [5,11]. However, Pol IV and Pol V subunits localize within the nucleus [5,8,12] and are required for the silencing of transgenes, retrotransposons and other endogenous repeats via a 24 nt siRNA-dependent DNA methylation pathway [13]. Pol IV appears to act at the beginning of the RNA-directed DNA methylation pathway because Pol IV colocalizes with endogenous repeat loci that give rise to abundant 24 nt siRNAs and because mutation of Pol IV catalytic subunits causes the loss of 24 nt siRNAs and the mislocalization of other proteins in the pathway [8]. RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) acts downstream of Pol IV, presumably using single-stranded Pol IV transcripts as templates for the production of complementary RNAs. Resulting double-stranded RNAs (dsRNA) are then thought to serve as substrates for DICER-LIKE 3 (DCL3), an RNase III-like endonuclease that cleaves the dsRNAs into 24 nt

siRNA duplexes, one strand of which associates with ARGONAUTE 4 (AGO4) to form an RNA-induced silencing complex (RISC). AGO4-RISC presumably uses each siRNA as a guide, targeting cytosine methylation to DNA sequences complementary to the siRNA in a process catalyzed by the *de novo* DNA methyltransferase, DRM2 [14]. Pol V is required for the methylation of target sequences, generating RNA transcripts at target loci that are hypothesized to basepair with AGO4-RISC siRNAs and facilitate the recruitment of DRM2 to the adjacent chromatin [7].

In a previous report, we showed that column fractions enriched for Arabidopsis NRPD2/NRPE2, and therefore presumably containing Pol IV and Pol V complexes, lack detectable promoter-independent RNA polymerase activity using sheared template DNA whereas activity was readily detected in fractions enriched for Pol I, II and III [5]. To explain this negative result, it has been proposed that Pol IV and Pol V may require specialized templates, such as methylated DNA or dsRNA, or may even lack transcriptional activity altogether [5,6,8,10,15,16]. However, the NRPD1, NRPE1 and NRPD2/NRPE2 subunits possess minimal Metal A and Metal B motifs typical of RNA polymerase active sites. The Metal A and Metal B sites bind magnesium ions that guide free nucleoside triphosphates into the active site for RNA synthesis, stabilize the transition state of the growing RNA chain and participate in transcript cleavage events during polymerase backtracking, a process which helps prevent polymerase arrest at pause sites [17,18]. The Metal A site within the largest subunit of multi-subunit RNA polymerases permanently binds a magnesium ion and is formed by three invariant aspartate residues within a nearly invariant NADFDGD motif [19]. The Metal B site is formed by an invariant glutamate and aspartate pair in the second-largest subunit that, in cooperation with one of the aspartates of the Metal A site, transiently binds a second magnesium ion [19]. Mutation of the amino acids that comprise the Metal A or Metal B sites is sufficient to abrogate transcriptional activity in bacteria [20], archaea [21] and eukaryotes [22].

We hypothesized that if RNA Polymerases IV and V function as RNA polymerases, their Metal A and Metal B consensus sequences should be essential for their known biological activities. To test this hypothesis, we conducted site-directed mutagenesis of the Metal A and Metal B motifs within the NRPD1, NRPE1 and NRPD2/NRPE2 subunits, stably incorporated the engineered genes into transgenic plants that were defective for the corresponding endogenous genes and tested for the restoration of Pol IV and Pol V functions *in vivo*. We show that the Metal A and Metal B sites are required for the biological functions of Pol IV and Pol V including siRNA production, RNA-directed DNA methylation and transposon silencing. Additionally, the active sites are required for the distinctive punctate nuclear localization patterns observed for Pol IV and Pol V [5,8], suggesting that these foci represent Pol IV and Pol V transcription factories [23].

Results

Pol IV catalytic subunits retain core sequences of polymerase active sites

Pol IV and Pol V are rapidly-evolving enzymes, with Arabidopsis NRPD1 (formerly NRPD1a) and NRPE1 (formerly NRPD1b) having amino acid substitution rates 20 times greater than the NRPE1 subunit of Pol II, and NRPD2/NRPE2 having a substitution rate 10 times greater than the Pol II NRPE2 subunit [4]. Based on multiple sequence alignments, we identified the amino acid positions that are invariant among Arabidopsis Pol I, II and III and *S. cerevisiae* Pol II, implying that these amino acids are

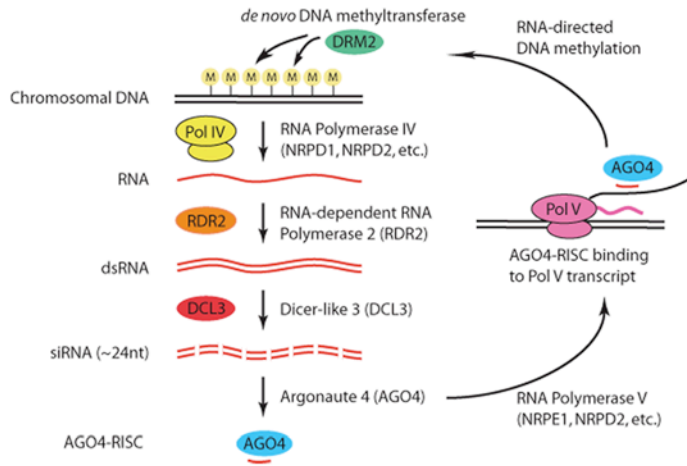
critically important for polymerase structure and function. Interestingly, numerous amino acids that are invariant among the canonical polymerases (i.e. Pol I, II and III) are substituted by other amino acids in Pol IV and Pol V {Herr, 2005 #2694}{Onodera, 2005 #2695}. In Figures 1B and C we mapped the positions of amino acids that are invariant among the canonical polymerases but different in NRPD1, NRPE1 or NRPD2/NRPE2 onto the *S. cerevisiae* Rpb1 and Rpb2 subunit structures in the context of a yeast Pol II elongation complex crystal structure. Interestingly, a large proportion of the “invariant” amino acids that have been substituted in NRPD1, NRPE1 (NRPD1b) and NRPD2/NRPE2 cluster in the vicinity of the catalytic center. In particular, sequences surrounding the Metal A binding site, bridge helix, cleft and funnel domains of NRPD1 and NRPE1 and the hybrid binding region of NRPD2 [1,19,24] are hotspots of Pol IV divergence relative to the invariant amino acids of the canonical polymerases (see also Figure S1 and Table S2). These regions govern interactions with the DNA template and the RNA/DNA hybrid that forms between the template and nascent transcript [25].

Multiple sequence alignment in the vicinity of the Metal A and Metal B sites of RNA polymerase largest and second-largest subunits illustrates the sequence divergence that has occurred in Pol IV and Pol V subunits relative to other RNA polymerases (Figures 1D and E). Immediately surrounding the Metal A site in the largest subunit, the sequence NADFDGD is invariant among *E. coli*, chloroplast, archaeal (*Pyrococcus*), viral, and eukaryotic Pol I, II and III polymerases. This sequence motif is part of an extended sequence, YNADFDGDEM that is conserved in eukaryotic Pol I, II and III and archaeal polymerases. However, despite having apparently evolved from a duplicated Pol II largest subunit, the NRPD1 subunit of Pol IV has only the core DFDGD sequence that includes the three magnesium-coordinating aspartates. In the NRPE1 (NRPD1b) subunit of Pol V, this core sequence consensus is extended by only one amino acid: the alanine preceding the first aspartate (ADFDGD). Importantly, the consensus sequence DxDGD occurs at the active sites of single-subunit RNA-dependent RNA polymerases, such as Arabidopsis RDR2 and RDR6 or *Neurospora* QDE-1. Therefore, the conservation of the minimal DFDGD sequence in NRPD1 and NRPE1 is consistent with the hypothesis that these subunits have minimal Metal A sites. Likewise, the NRPD2 subunit utilized by both Pol IV and Pol V contains the core ED motif of the Metal B site as part of an extended G(Y/F)NQEDS motif also present in the second-largest subunit of Pol II. Collectively, these observations suggest that Pol IV and Pol V have Metal A and Metal B sites at their presumptive active sites.

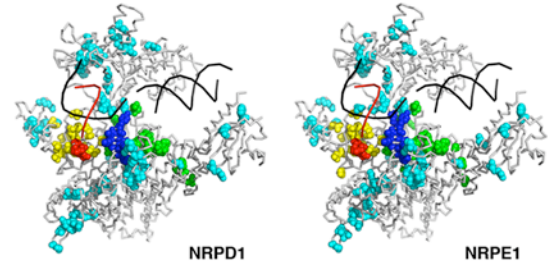
Pol IV and Pol V Metal A and Metal B motifs are required for siRNA accumulation

To address whether the presumptive active sites of Pol IV and Pol V are required for their functions, we performed site-directed mutagenesis to change the acidic residues of the Metal A and Metal B sites to alanines. Three amino acid substitutions were performed in the largest subunits of Pol IV and Pol V: for NRPD1 these were D447A, D449A and D451A and for NRPE1 (NRPD1b) they were D449A, D451A and D453A. For NRPD2/NRPE2, E785A and D786A mutations were introduced (Figure 2A). Full-length genomic clones bearing these mutations, expressed using the endogenous promoters and containing their complete intron-exon structures, were fused at the C-terminus to a FLAG peptide epitope tag, as were equivalent wild-type (non-mutant) constructs. Resulting *NRPD1* transgenes were introduced into the *nrdp1a-3* null mutant, *NRPE1* (*NRPD1b*) transgenes were

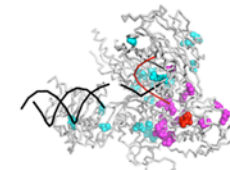
A. Model of RNA-directed DNA Methylation Pathway in Arabidopsis



B. Substituted "Invariant" Amino Acids in NRPD1 and NRPE1



C. Substituted "Invariant" Amino Acids in NRPD2



D. Conservation of the Metal A Site among RNA Polymerase Largest Subunits

At_NRPD1	T	S	V	S	L	N	I	C	L	L	F	R	G	S	S	C	L	H	G	V	F	P	O	S	R	V	L		
Rice_NRPD1	Q	S	A	V	S	I	N	P	C	H	D	R	F	K	G	F	D	D	C	L	H	G	Y	I	C	L	S	I	L
Poplar_NRPD1	P	S	V	L	A	N	P	C	H	D	R	F	K	G	F	D	D	C	L	H	G	Y	F	P	O	S	R	V	L
At_NRPE1	D	N	T	V	K	N	P	M	S	L	S	L	S	D	F	P	C	C	V	H	L	F	P	O	S	L	R	A	V
Rice_NRPE1	D	H	T	K	N	P	M	S	L	S	L	S	D	F	P	C	C	V	H	L	F	P	O	S	L	R	A	V	
Poplar_NRPE1	D	H	A	V	K	N	P	M	S	L	S	L	S	D	F	P	C	C	V	H	L	F	P	O	S	L	R	A	V
At_NRP1	E	K	L	R	H	Y	A	N	S	T	Y	N	D	F	P	S	S	E	N	N	V	F	P	O	S	E	S	A	A
At_NRP1	Y	S	F	R	L	S	V	T	S	Y	N	D	F	P	S	S	E	N	N	V	F	P	O	S	E	S	A	A	
At_NRP1	W	R	L	R	F	R	E	S	V	N	Y	N	D	F	P	S	S	E	N	N	V	F	P	O	S	E	S	A	A
At_NRP1	Y	S	F	R	L	S	V	T	S	Y	N	D	F	P	S	S	E	N	N	V	F	P	O	S	E	S	A	A	
Yeast_Rpb1	G	K	A	Y	C	H	P	V	A	A	Y	N	D	F	P	S	S	E	N	N	V	F	P	O	S	E	S	A	A
Ecoli_RpoC	G	R	T	C	H	P	V	K	G	E	N	D	F	P	S	S	E	N	N	V	F	P	O	S	E	S	A	A	
At_CPST	G	D	T	K	S	I	G	V	N	S	N	D	F	P	S	S	E	N	N	V	F	P	O	S	E	S	A	A	
Vaccinia_Virus	G	D	T	K	S	I	G	V	N	S	N	D	F	P	S	S	E	N	N	V	F	P	O	S	E	S	A	A	
Pyrococcus	Y	R	F	R	L	A	V	T	S	Y	N	D	F	P	S	S	E	N	N	V	F	P	O	S	E	S	A	A	

↑↑↑

At_RDR2	F	P	O	N	G	R	P	R	P	R	C	S	E	L	L	P	F	S	D	E	L	S	S	E	M	D	
At_RDR6	F	P	O	N	G	R	P	R	P	R	A	S	S	E	L	L	P	F	S	D	E	L	S	S	E	M	D
QDE-1	S	T	K	G	V	L	A	K	L	K	L	S	S	E	L	L	P	F	S	D	E	L	S	S	E	M	D

E. Conservation of the Metal B Site among RNA Polymerase 2nd Largest Subunits

At_NRPD2	F	N	G	N	A	I	V	N	V	H	L	G	N	C	E	I	V	N	K	A	L	E	S	R	F	S	E	O					
Rice_NRPD2	F	N	G	N	A	I	V	N	V	H	L	G	N	C	E	I	V	N	K	A	L	E	S	R	F	S	E	O					
Zm_NRPD2	F	N	G	N	A	I	V	N	V	H	L	G	N	C	E	I	V	N	K	A	L	E	S	R	F	S	E	O					
Poplar_NRPD2	F	N	G	N	A	I	V	N	V	H	L	G	N	C	E	I	V	N	K	A	L	E	S	R	F	S	E	O					
At_NRP2	P	T	T	N	A	I	V	N	V	L	A	R	T	F	D	M	E	A	M	I	L	A	S	V	E	R	M	C	H	G	Q	I	
At_NRP2	P	A	T	T	N	A	I	V	N	V	L	A	R	T	F	D	M	E	A	M	I	L	A	S	V	E	R	M	C	H	G	Q	I
At_NRP2	R	H	K	G	V	C	G	L	L	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	
Yeast_Rpb2	F	A	G	N	A	I	V	N	V	H	L	G	N	C	E	I	V	N	K	A	L	E	S	R	F	S	E	O					
Ecoli_RpoB	A	L	G	N	A	I	V	N	V	H	L	G	N	C	E	I	V	N	K	A	L	E	S	R	F	S	E	O					
At_CPST	A	L	G	N	A	I	V	N	V	H	L	G	N	C	E	I	V	N	K	A	L	E	S	R	F	S	E	O					
Vaccinia_Virus	C	F	S	R	V	T	I	N	S	Y	K	L	L	S	G	L	I	K	Q	T	F	C	L	G	L	D	I	V	T				
Pyrococcus	R	H	K	G	V	C	G	L	L	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q		

Figure 1. Catalytic residues that comprise the Metal A and Metal B binding sites of DNA-dependent RNA polymerases are conserved in the NRPD1, NRPE1/NRPD1b and NRPD2 subunits. A) Model for the RNA-directed DNA methylation pathway in Arabidopsis. B and C) Positions of NRPD1, NRPE1 and NRPD2 divergence at sites that are invariant in canonical RNA polymerases. The image shows the yeast Pol II Rpb1 and Rpb2 subunits (gray) in complex with the dsDNA substrate (black) and RNA product (red) within Protein Data Bank crystal structure 1R9T (Kornberg laboratory). Amino acids that are invariant among the Arabidopsis Pol I, II and III subunits and yeast Rpb1 or Rpb2, but that are different in NRPD1, NRPE1 or NRPD2, are displayed as spheres. Red spheres highlight the positions of the invariant Metal A and Metal B sites in the largest and second-largest subunits, respectively. Substituted amino acids in the left, bridge helix, and active site domains of the largest subunit are colored green, blue and yellow, respectively. Substituted amino acids in the hybrid binding domain of the second-largest subunit are colored magenta. Substituted amino acids in the largest and second-largest subunits that are located outside of these domains are colored cyan. For a complete listing of the highlighted amino acids refer to Table S2. D and E) Multiple protein sequence alignments of RNA polymerase largest and second-largest subunit active site regions. Amino acids highlighted in red and designated by arrows represent the invariant Metal A and Metal B sites. Identical amino acids are highlighted in green and similar amino acids are highlighted in yellow. doi:10.1371/journal.pone.0004110.g001

introduced into the *nrdp1b-11* mutant and *NRPD2* transgenes were introduced into the *nrdp2a-2 nrdp2b-1* double mutant. Note that the *NRPD2b* gene is a pseudogene due to a frameshift mutation, such that the double mutant is used only as a precaution. The double mutant is hereafter referred to simply as *nrdp2*. Six or more independent transformants for each transgene construct were analyzed to determine the ability of the transgenes to genetically rescue their respective null mutants and all lines for a given construct were found to display the same phenotypes. The active site mutant transgenic lines are abbreviated as *NRPD1^{DDD-AAA}-FLAG*, *NRPE1^{DDD-AAA}-FLAG* or *NRPD2^{ED-AA}-FLAG* in Figures 2, 3, 4, 5.

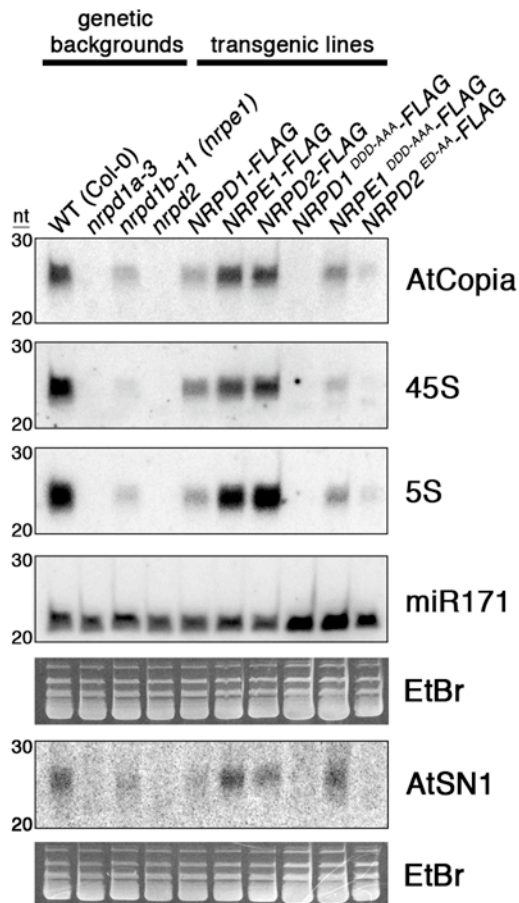
The requirement for the presumptive Pol IV and Pol V active sites was first tested by comparing the abilities of wild-type or mutant transgenes to rescue the accumulation of siRNAs corresponding to 45S or 5S rRNA gene repeats or *AtCopia* or *AtSNI* retrotransposons (Figure 2B). siRNAs corresponding to these repetitive sequences are predominantly 24 nt in size and are readily

detectable in wild-type (WT; ecotype Col-0) plants. However, the siRNAs are eliminated in *nrdp1* or *nrdp2* mutants and are substantially reduced in *nrdp1* (*nrdp1b*) mutants, in agreement with prior studies [6,8,9,10]. In transgenic lines expressing wild-type *NRPD1-FLAG*, *NRPE1-FLAG* or *NRPD2-FLAG* transgenes in their respective mutant backgrounds, siRNA production is restored, albeit to lower than wild-type levels in the case of the *NRPD1* transgene. A delay in flowering time observed in the *nrdp1* mutant, and other mutants affecting the siRNA-directed DNA methylation pathway, is also not fully restored by the *NRPD1* transgene (Figure S2), suggesting a correlation between siRNA levels and more rapid flowering. Importantly, no rescue of siRNA levels is observed in transgenic lines expressing the *NRPD1^{DDD-AAA}-FLAG* or *NRPE1^{DDD-AAA}-FLAG* transgenes; in these lines, siRNA levels are the same as in the *nrdp1a-3* or *nrdp1b-11* mutant parental lines. These results indicate that the Metal A sites of Pol IV (*NRPD1*) and Pol V (*NRPE1/NRPD1b*) largest subunits are required for small RNA biogenesis or accumulation. Trace siRNA signals were detected in

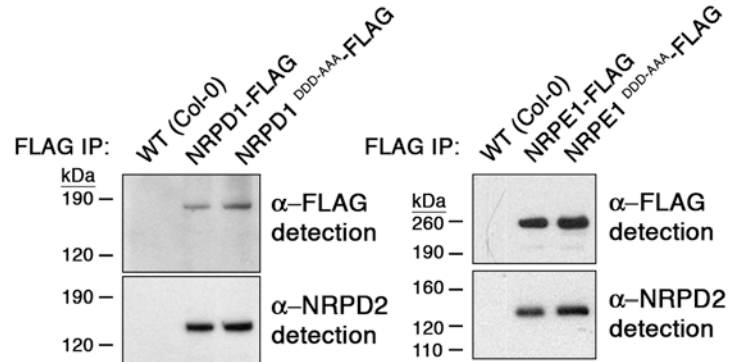
A. Site-directed Mutagenesis of Pol IV and Pol V Active Sites

NRPD1 WT	443	PFRGDFDGDCLHG	455
NRPD1 ^{DDD-AAA}	443	PFRGAFAGACLHG	455
NRPE1 WT	445	PLSADFDCVHVL	457
NRPE1 ^{DDD-AAA}	445	PLSAAFAGACVHVL	457
NRPD2 WT	781	GYNQEDSIVM	790
NRPD2 ^{ED-AA}	781	GYNQAASIVM	790

B. Small RNA Accumulation in Active Site Mutants



C. NRPD2 co-IP with NRPD1 and NRPE1



D. NRPD2 Expression

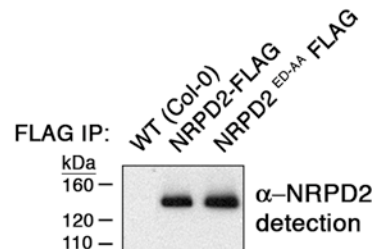


Figure 2. Pol IV and Pol V active site amino acids are required for rescue of small RNA production but not Pol IV or Pol V subunit assembly. A) Acidic amino acids of the Metal A and Metal B sites were mutated to alanines by site-directed mutagenesis. Resulting full-length genomic transgenes were transformed into Arabidopsis *nrpd1a-3*, *nrpd1b-11* (*nrpe1*) and *nrpd2a/2b* (*nrpd2*) homozygous mutants, respectively, as were wild-type versions of each genomic construct. B) RNA blot analysis of small RNAs purified from Arabidopsis inflorescence. Membranes were sequentially probed with body-labeled RNA probes specific for *AtCopia*, 45S rRNA gene intergenic spacer, 5S rRNA gene intergenic spacer, miR171 or *AtSN1* small RNAs. Images of ethidium-bromide stained gels are displayed below the relevant autoradiograms to show that equal amounts of RNA were loaded in each lane. Migration of the 20-nt and 30-nt RNA markers is indicated at the left of each autoradiogram. C and D) Pol IV and Pol V largest subunits bearing active site mutations are indistinguishable from wild-type versions of the proteins in terms of expression level or ability to assemble with the NRPD2 subunit. FLAG-tagged recombinant proteins immunoprecipitated from total protein extracts using anti-FLAG antibodies were detected on immunoblots using FLAG M2 antibody. Membranes were then stripped and re-probed using a polyclonal antibody specific for NRPD2.

doi:10.1371/journal.pone.0004110.g002

the *NRPD2^{ED-AA}-FLAG* transgenic plants but not in the *nrpd2* mutant parental line (Figure 2B). This suggests that the NRPD2 contribution to the Metal B site is not absolutely required for siRNA biogenesis, but is clearly important.

Two trivial explanations for the results of Figure 2B could be that the Pol IV and Pol V active site mutant proteins are not expressed at levels comparable to their wild-type counterparts or

that mutation of the active site region disrupts Pol IV or Pol V subunit assembly. To test these possibilities, anti-FLAG co-immunoprecipitation (co-IP) experiments were performed using equal amounts of total protein extracted from transgenic plants expressing either the wild-type or active site mutant versions of the *NRPD1-FLAG* and *NRPE1-FLAG* transgenes (Figure 2C). Equivalent amounts of the wild-type or mutant large subunits were

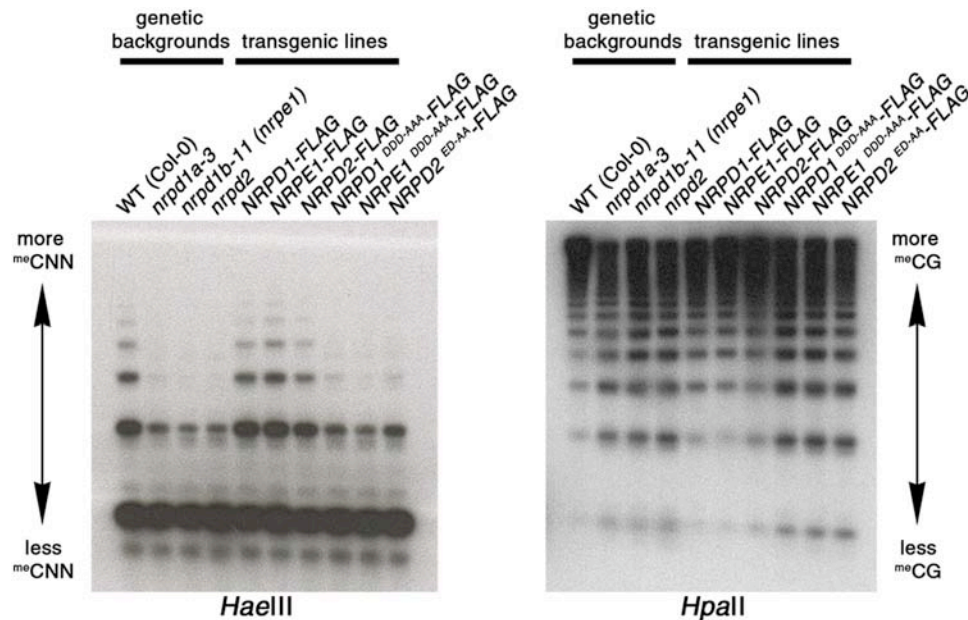


Figure 3. Pol IV and Pol V active site amino acids are required for the RNA-directed methylation of 5S rRNA gene repeats. Southern blot comparison of *HaeIII* or *HpaII*-digested genomic DNA of wild-type (WT), *nrpd1a*, *nrpe1/nrpd1b*, and *nrpd2* mutants or of transgenic lines generated by transforming these mutants with *NRPD1*, *NRPE1/NRPD1b* or *NRPD2a* full-length transgenes whose sequences are either wild-type or are mutated at the Metal A or Metal B sites. Both wild-type and mutant recombinant proteins have FLAG epitope tags at their carboxyl termini. doi:10.1371/journal.pone.0004110.g003

immunoprecipitated, indicating that they are expressed at similar levels. Moreover, equivalent amounts of the NRPD2/NRPE2 subunit were co-immunoprecipitated by the wild-type or mutated versions of the Pol IV or Pol V largest subunits, suggesting that mutation of the largest subunit active sites does not affect assembly with other subunits. Likewise, the wild-type and active site mutant versions of the *NRPD2-FLAG* transgenes were expressed at similar levels (Figure 2D).

Pol IV and Pol V active site requirements for DNA methylation

The requirement for the presumptive Pol IV and Pol V active sites in RNA-directed DNA methylation at 5S rRNA gene repeats was tested by Southern blot analysis using the methylation sensitive restriction endonucleases, *HaeIII* and *HpaII* (Figure 3). In this assay, *HaeIII* reports on cytosine methylation in CNN motifs whereas *HpaII* reports on CG methylation. The 5S genes are organized in tandem repeat such that ladders of bands are observed following digestion with methylation-sensitive restriction endonucleases and Southern blot hybridization. Larger bands reflect a relatively high degree of methylation and smaller bands reflect reduced methylation and therefore increased susceptibility to digestion by the enzymes. In *nrpd1*, *nrpe1* (*nrpd1b*) and *nrpd2/nrpe2* mutants, similar losses of CNN or CG methylation occur relative to wild-type (WT) controls. In these mutant backgrounds, methylation is restored to wild-type levels by the corresponding wild-type transgenes (*NRPD1-FLAG*, *NRPE1-FLAG* or *NRPD2-FLAG*, respectively). However, the equivalent transgenes bearing the active site mutations (*NRPD1^{DDD-AAA}-FLAG*, *NRPE1^{DDD-AAA}-FLAG* and *NRPD2^{ED-AA}-FLAG*) fail to rescue the defects in DNA methylation caused by the *nrpd1a-3*, *nrpd1b-11* (*nrpe1*) and *nrpd2/nrpe2* mutations.

Like 5S rRNA gene loci, *AtSN1* retrotransposons are subjected to siRNA-directed DNA methylation in a Pol IV and Pol V-dependent manner [6,9,10]. We tested *AtSN1* methylation using

chop-PCR (Figure 4A). In this assay, genomic DNA is digested (chopped) with *HaeIII* and PCR primers flanking the three *HaeIII* restriction enzyme sites are then used to amplify the intervening region. If any of the three sites are unmethylated, *HaeIII* cuts the template and PCR amplification fails. Only if all three *HaeIII* sites are methylated does PCR amplification occur. In wild-type (Col-0) plants, *AtSN1* elements are methylated, rendering them resistant to *HaeIII* digestion (Figure 4B). However, in the *nrpd1a-3*, *nrpd1b-11* (*nrpe1*) or *nrpd2/nrpe2* mutants, methylation is lost, resulting in *HaeIII* susceptibility and the loss of PCR product. Whereas wild-type *NRPD1-FLAG*, *NRPE1-FLAG* and *NRPD2-FLAG* transgenes rescue their respective null mutants and restore DNA methylation at the *AtSN1* loci, the corresponding active site mutants fail to do so (Figure 4B). We conclude that the active sites of NRPD1, NRPE1 and NRPD2/NRPE2 are required for RNA-directed DNA methylation.

Pol IV and Pol V active site requirements for transcriptional silencing

Consistent with the losses in *AtSN1* siRNA accumulation (Figure 2B) and DNA methylation at *AtSN1* retrotransposons (Figure 4B), silencing of *AtSN1* elements and a retrotransposon-derived solo LTR element [26] are lost in Pol IV and Pol V mutants (Figure 4C). *AtSN1* and solo LTR transcripts are not detected by RT-PCR in wild-type (WT) plants but are apparent in *nrpd1a-3*, *nrpd1b-11* (*nrpe1*) or *nrpd2* mutants. Transforming these mutants with the *NRPD1-FLAG*, *NRPE1-FLAG* or *NRPD2-FLAG* transgenes, respectively, restores *AtSN1* and solo LTR silencing. However, the active site mutant versions of the transgenes fail to restore *AtSN1* or solo LTR silencing in the mutant backgrounds.

The NRPD1, NRPE1 and NRPD2 active sites are required for the distinctive localization patterns of Pol IV and Pol V. Although NRPD1, NRPE1 and NRPD2/NRPE2 proteins mutated at their presumptive active sites lack detectable *in vivo*

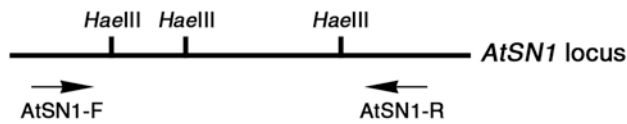
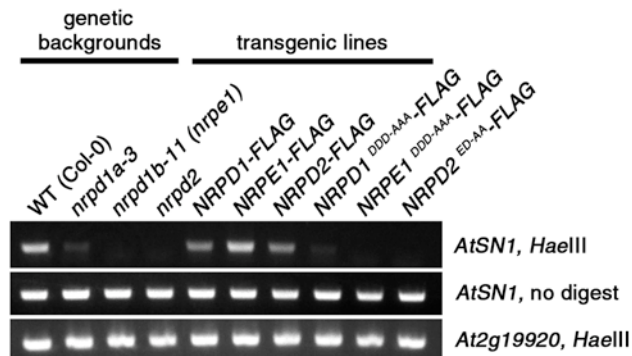
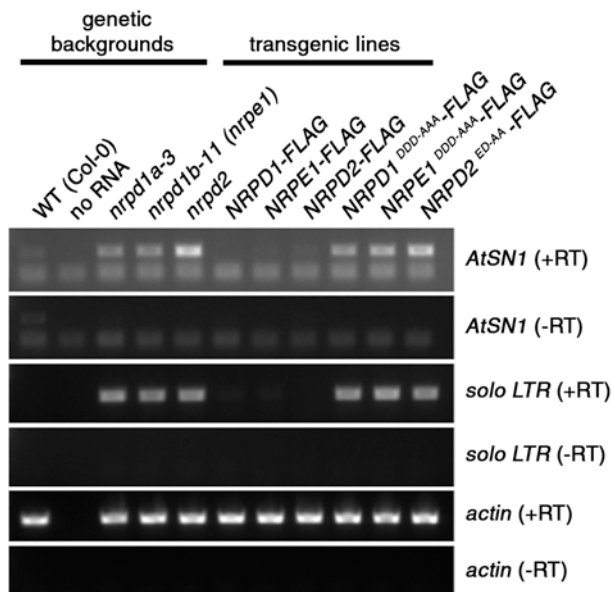
A. *AtSN1* chop-PCR assayB. *AtSN1* DNA MethylationC. *AtSN1* and *solo LTR* Transcription

Figure 4

Figure 4. DNA methylation and transcriptional silencing of *AtSN1* retrotransposons requires the Pol IV and Pol V active sites. A) Schematic of an *AtSN1* retroelement locus showing the locations of *Haelll* restriction enzyme sites and flanking PCR primers. B) *AtSN1* DNA methylation analysis using the chop-PCR assay. *AtSN1* loci were PCR amplified from *Haelll* digested or undigested genomic DNA and samples were then subjected to agarose gel electrophoresis and staining with ethidium bromide. Locus *At2g19920* lacks *Haelll* restriction sites and was used as a control. C) RT-PCR analysis of retrotransposon transcription. Random-primed cDNA was used as the template for PCR amplification of *AtSN1* and *solo-LTR* transcripts. Reactions were then subjected to agarose gel electrophoresis and staining with ethidium bromide. For each genotype, reactions from which reverse transcriptase was omitted (-RT) or for which actin RNA was PCR-amplified serve as controls.

doi:10.1371/journal.pone.0004110.g004

function, as shown by their failure to genetically rescue their corresponding null mutants, the proteins are expressed at the same levels as their wild-type counterparts and the mutated largest subunits assemble with the NRPD2/NRPE2 subunit, as shown by co-immunoprecipitation and immunoblotting (Figures 2C, D). Therefore, we investigated the nuclear localization patterns of the proteins mutated at the Metal A and Metal B sites relative to the wild-type proteins (Figure 5). As reported previously [5,8,12], immunolocalization of non-mutant NRPD1 and NRPE1 FLAG-tagged proteins reveals that the proteins are localized within punctate foci dispersed throughout the nucleoplasm, with NRPE1 also being found in a "nucleolar dot" [8] that we have interpreted to be a center for siRNA-processing and RISC assembly [8,12]. Interestingly, the NRPD1 and NRPE1 proteins mutated at their Metal A sites fail to display the distinctive nucleoplasmic puncta or foci. Instead, weak and highly dispersed signals are detected throughout the nucleoplasm. A nucleolar dot signal is observed in 14% of nuclei expressing the NRPE1^{DDD-AAA}-FLAG protein despite the lack of detectable nucleoplasmic puncta in these nuclei. Although 83% of wild-type nuclei display an NRPE1 nucleolar dot, these observations suggest that the Metal A site is not required for NRPE1 to associate with the putative siRNA processing center.

Discussion

Although RNA polymerase activity has not yet been demonstrated *in vitro* for Pol IV or Pol V, our results show that their predicted Metal A and Metal B sites, which are essential for multi-subunit RNA polymerase activity, are required for Pol IV and Pol V biological functions *in vivo*. These results suggest that both Pol IV and Pol V are catalytically active as RNA polymerases. Supporting evidence is that low-level intergenic transcripts that are dependent on Pol V can be detected *in vivo* by using RT-PCR; Pol V physically associates with these loci and production of the intergenic RNAs is abolished in the NRPE1 Metal A site mutant lines we developed in the current study [7]. Although we tested NRPD1 or NRPE1 subunits mutated at all three aspartates of their Metal A sites, genetic evidence suggests that mutation of even one of these aspartates is sufficient to disrupt Pol V function. Specifically, one of nine mutant alleles of NRPE (NRPD1b) identified by Kanno et al. in a screen for mutants disrupting silencing due to RNA-directed DNA methylation [10] results from a single amino acid substitution in the Metal A site (allele *drd3-3*: D451N).

In the vicinity of the Pol IV and Pol V active sites, numerous amino acids that are invariant in Pol I, II and III are missing or replaced by other amino acids. Many of these amino acids occur in regions that influence the predicted template channel, including the bridge helix of the largest subunit, a highly conserved structure from bacterial to eukaryotic polymerases over which the template strand passes en route to the active site [19,27,28]. The bridge helices of Arabidopsis Pol I, II and III are approximately 75% identical overall, yet more than half of their invariant amino acids are replaced in NRPD1 and NRPE1 (see Figure S1 and Table S1) [24]. Such alterations in the vicinity of the template channel and active site may facilitate the use of non-conventional templates, including the possible transcription of double-stranded RNA (dsRNA) templates rather than DNA templates. Pol IV is required in several small RNA pathways in which dsRNAs are apparently produced independent of Pol IV action, including a pathway in which siRNA production is triggered by the overlap of RNA transcripts from convergently-transcribed genes [29]. Therefore, transcription of dsRNA by Pol IV is a distinct possibility [3,24]. Moreover, there is precedent for multi-subunit DNA-dependent RNA polymerases transcribing RNA, including the replication of

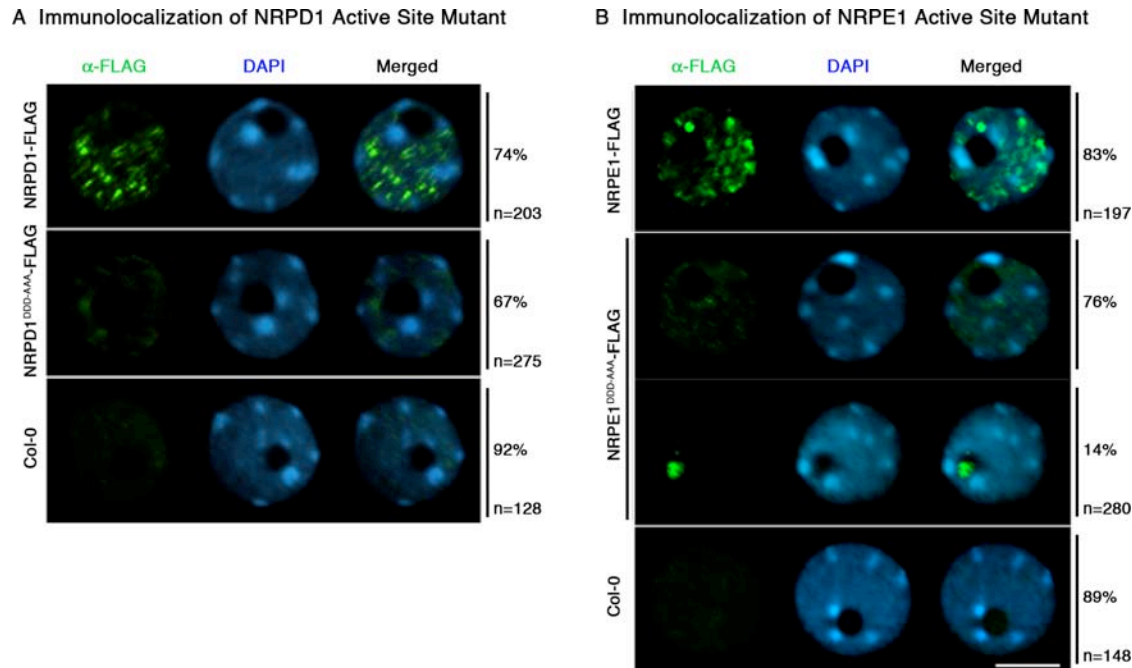


Figure 5. NRPD1 and NRPE1/NRPD1b proteins mutated at their active sites fail to display characteristic Pol IV and Pol V punctate localization patterns in Arabidopsis nuclei. FLAG epitope-tagged NRPD1 and NRPD1^{DDD-AAA} (panel A) or NRPE1 and NRPE1^{DDD-AAA} (panel B) recombinant proteins were immunolocalized (green signal) using anti-FLAG M2 antibody. Nuclei were counterstained with DAPI (blue signal). The percentage of nuclei showing a given localization pattern and the number of nuclei (n) analyzed are indicated to the right of each panel. doi:10.1371/journal.pone.0004110.g005

Hepatitis Delta Virus (HDV) or plant viroid RNAs by Pol II transcription [30,31]. Yeast Pol II has also been demonstrated to have RNA-dependent RNA polymerase (RdRP) activity although it synthesizes RNA transcripts more slowly than when transcribing DNA and is less processive [32]. It is plausible that the amino acid sequence changes in Pol IV and Pol V largest subunits at sites that are invariant in Pol I, II or III may improve catalytic activity or processivity on alternative templates, such as RNA.

Accumulation of 24 nt siRNAs requires the Metal A consensus sequences of NRPD1 and NRPE1 (Figure 2B). Interestingly, trace amounts of siRNAs are restored in *nrpd2* null mutants transformed with the NRPD2 active site mutant. One explanation for this observation may be that the second-largest subunit's contribution to magnesium ion binding at the Metal B site is slightly less critical than the magnesium binding coordinated by the largest subunit. Consistent with this interpretation, single amino acid substitutions in the Metal B site of an archaeal RNA polymerase were shown to substantially decrease, but not completely abrogate, transcriptional activity [21]. However, the trace amounts of siRNA production that are detected in *NRPD2^{ED-AA}-FLAG* lines are apparently not sufficient for rescue of RNA-directed DNA methylation at 5S rRNA genes or *AtSN1* retroelements or for restoration of *AtSN1* or solo LTR silencing.

It is noteworthy that the non-mutant *NRPD1-FLAG* transgene did not fully rescue delayed flowering time in the *nrpd1a-3* mutant background to that of wild-type plants (see Supplemental data), nor did the transgene fully rescue siRNA levels (see Figure 2B). Nonetheless, 5S rRNA gene and *AtSN1* DNA methylation levels were fully rescued by the *NRPD1-FLAG* transgene. Collectively, these observations suggest that tissue-specific differences in transgene expression, or different siRNA level thresholds, may explain the different degrees of transgene effectiveness in the various assays.

Despite evidence that NRPD1 and NRPE1/NRPD1b active site mutants are expressed at the same levels as non-mutant

recombinant proteins and are not impaired in their ability to assemble with the NRPD2 subunit, the active site mutants fail to display the characteristic punctate nucleoplasmic localization patterns typical of wild-type NRPD1 or NRPE1. One possibility could be that active site mutants are unable to bind their template(s) and thus never localized to chromatin. Although we cannot rule out this possibility, *E. coli* RNA polymerase that is mutated at the Metal A site is still able to bind DNA and form an open-promoter complex, despite being transcriptionally inactive [20]. Therefore, it is plausible that Pol IV or Pol V complexes bearing active site mutations can bind and occupy their templates. Individual loci bound by single Pol IV or Pol V molecules would likely escape detection in our immunolocalization assays. Therefore, we think it most likely that the nucleoplasmic foci at which Pol IV and Pol V are concentrated in wild-type nuclei represent transcription factories in which Pol IV or Pol V-transcribed sequences coalesce, analogous to the transcription factories observed for *E. coli* RNA polymerase or eukaryotic RNA Polymerases I, II or III [23]. If so, heterochromatic regions that are subject to Pol IV or Pol V-dependent chromatin modifications may coalesce as a result of Pol IV or Pol V transcription.

Methods

Mutant plant strains

Arabidopsis thaliana mutants *nrpd1a-3*, *nrpd2a-2 nrpd2b-1* (abbreviated as *nrpd2a/2b*) and *nrpd1b-11* were described previously [5,8]. All are apparent null mutants resulting from *Agrobacterium tumefaciens*-mediated, multi-kb insertions that disrupt the genes [33].

Multiple sequence alignment

GenBank sequences for largest and second-largest RNA polymerase subunit alignments were those described previously (see supplemental material of reference [5]), with the addition of

Zea mays NRPD2 (AAY45706), *Arabidopsis* RDR2 (NP_192851), *Arabidopsis* RDR6 (NP_190519) and *Neurospora crassa* QDE-1 (CAB42634). NRPD1 (LG_I, 8313188-8324531), NRPE1 (LG_III, 17406212-17419838) and NRPD2 (LG_XVIII, 6286719-6297405) sequences from poplar were identified using the *Populus trichocarpa* unmasked genome assembly v1.1 by JGI and the tBLASTn tool with *Arabidopsis* protein queries. Sequences were aligned using ClustalW2 and colored using BOXSHADE.

Site-directed mutagenesis

Site-directed ligase independent mutagenesis (SLIM) [34] was performed to change aspartates to alanines at the Metal A sites of *Arabidopsis* NRPD1 (NRPD1a) (D447A, D449A, D451A) and NRPE1 (NRPD1b) (D449A, D451A, D453A) and to mutate the Metal B site of NRPD2a (E785A, D786A). Nucleotides 910-2232 of the NRPD1a genomic sequence were PCR amplified from pENTR-NRPD1a with NRPD1a active site-F and NRPD1a active site-R primers (see Table S1 for primer sequences) and Pfu Ultra (Stratagene). The resulting PCR product was cloned into the pCR4-TOPO vector (Invitrogen) for subsequent mutation using primers NRPD1a DDD/AAA-F, NRPD1a mut-F, NRPD1a DDD/AAA-R and NRPD1a mut-R (see Table S1). The resulting mutated sequence within plasmid pCR4-NRPD1a^{DDD-AAA} was then subcloned back into the pENTR-NRPD1a genomic clone by digesting pENTR-NRPD1a and the pCR4-NRPD1a^{DDD-AAA} active site region PCR clone with *SacI*, gel purifying the desired fragments and performing a standard ligation reaction. The pENTR-NRPD1b (NRPE1) genomic clone was mutated with primers NRPD1b DDD/AAA-F, NRPD1b mut-F, NRPD1b DDD/AAA-R and NRPD1b mut-R (see Table S1). The pDONR-NRPD2a genomic clone was mutated with primers NRPD2a ED/AA-F, NRPD2a mut-F, NRPD2a ED/AA-R and NRPD2a mut-R (see Table S1). Proper ligation at cloning junctions and at mutated active sites was confirmed by DNA sequencing.

Generation of transgenic lines

The cloning of *NRPD1* (*NRPD1a*) and *NRPE1* (*NRPD1b*) genomic sequences and generation of *NRPD1-FLAG* and *NRPE1* (*NRPD1b*)-*FLAG* transgenic lines that rescue the *nripd1a-3* or *nripd1b-11* null mutants, respectively was described previously [8]. The full-length *NRPD2a* genomic sequence, including 1310 bp upstream of the translation start site, was amplified by PCR from *A. thaliana* (ecotype Col-0) genomic DNA using NRPD2a BP-F and NRPD2a BP-R primers (see Table S1) and Pfu Ultra (Stratagene), cloned into the pDONR221 vector using BP Clonase (Invitrogen) and confirmed by DNA sequencing. The pDONR-NRPD2a, pENTR-NRPD1^{DDD-AAA}, pENTR-NRPE1^{DDD-AAA} and pDONR-NRPD2a^{ED-AA} full-length genomic clones were recombined into pEarleyGate 302 [35] in order to add a C-terminal FLAG epitope tag in lieu of the normal stop codon; LR Clonase (Invitrogen) was used for these recombination reactions. Resulting plasmids were transformed into *Agrobacterium tumefaciens* strain GV3101 and homozygous *nripd1a*, *nrip1/nripd1b* or *nripd2* mutant plants were transformed with the corresponding transgenes using the floral dip method [36]. Seeds of dipped plants were sown and transformants were selected by spraying seedlings with BASTA herbicide. BASTA-resistant primary transformants (T1 generation plants) were then assayed by Southern blot analysis to test their 5S rRNA gene repeat methylation status. All lines displayed equivalent levels of rescue, in the case of wild-type transgenes, or lack of rescue in the case of mutant transgenes (Figure S1). T2 generation transgenic plants were used for all experiments depicted in the figures, unless indicated otherwise.

Small RNA blot hybridization

RNA was isolated from 300 mg of inflorescence tissue using the mirVana miRNA isolation kit (Ambion). RNA samples (9.5 µg each) were resolved by gel electrophoresis, transferred to nylon membrane and hybridized to radioactive probes as described previously [5]. The *AtSN1* RNA probe, body-labeled with α -³²P-CTP, was prepared according to [37]. *AtCopia*, 45S rRNA gene and 5S rRNA (siR1003) probes were prepared according to [8]. The miR171 riboprobe was generated using the mirVana probe construction kit (Ambion) in conjunction with DNA oligonucleotide miR171T7: 5'TGATTGAGCCGCGCCAATATCctgtctc3'.

DNA methylation assays

Southern blot analysis was performed using 250 ng of *HaeIII* or *HpaII*-digested genomic DNA isolated from leaves of 3 to 4-week old plants. Digested DNA was subjected to agarose gel electrophoresis and transferred to uncharged nylon membranes. The 5S rRNA gene probe, labeled with a³²P-dCTP, was generated by random priming of a full-length 5S gene repeat amplified by PCR from clone pCT4.2 [38]. Probe hybridization and autoradiography were according to standard methods [39]. The *AtSN1* DNA methylation assay involving PCR amplification of undigested or *HaeIII*-digested genomic DNA was performed as described previously [6].

RT-PCR

RNA (~1 µg) isolated from 3 to 4-week old leaf tissue was treated with RQ1 DNase (Promega) and used to generate random-primed cDNA using degenerate dN6 primers (NEB) and Superscript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. *AtSN1* RT-F and *AtSN1* RT-R primers were used to amplify *AtSN1* transcripts from the cDNA with GoTaq Green (Promega) and samples were analyzed by agarose gel electrophoresis.

Immunoprecipitation and detection of epitope-tagged proteins

Immunoprecipitation and immunoblot detection of Pol IV and Pol V proteins was performed using 4.0 g of 3-week old leaf tissue from T3 generation plants, as described previously [8]. Immunolocalization of FLAG-tagged proteins was performed using nuclei of 28-day old leaves, as previously described [8].

Supporting Information

Table S1 DNA oligonucleotides used in this study

Found at: doi:10.1371/journal.pone.0004110.s001 (0.07 MB DOC)

Table S2 Positions of amino acids that are invariant among *Arabidopsis* Pol I, II and III and yeast Pol II but have diverged in *Arabidopsis* Pol IV and Pol V largest and second-largest subunits. The table lists amino acids, numbered according to the PDB:1R9T crystal structure for yeast Pol II, and the changes at these positions in NRPD1, NRPE1 or NRPD2. These are the amino acids highlighted in Figures 1B and 1C. Amino acid substitutions are based on the multiple alignments shown in Figure S1 for the RNAP largest subunits and in the supplemental material of Onodera et al (2005) for the RNAP second-largest subunits. Major structural features, according to Cramer et al (2001), are designated to the left of the tables.

Found at: doi:10.1371/journal.pone.0004110.s002 (0.18 MB DOC)

Figure S1 Multiple alignment of *A. thaliana* RNAP Largest Subunits and the Yeast Pol II Largest Subunit. Full-length protein

sequences for *A. thaliana* NRPA1 (At3g57660), NRPB1 (At4g35800), NRPC1 (At5g60040), NRPD1 (At1g63020), NRPE1 (At2g40030) and *S. cerevisiae* Rpb1 were aligned using ClustalW2 (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>) in conjunction with final editing by hand. Alignments were colored using BOXSHADE v3.21 (http://www.ch.embnet.org/software/BOX_form.html). DNA-dependent RNA polymerase conserved domains A to H are underlined and designated to the right of the alignments. Yeast Pol II structural features, according to Cramer et al (2001), are designated below the alignments. Regions that make contact with other RNAP subunits are designated in italics above the alignments. The Metal A site is designated with asterisks above the alignment.

Found at: doi:10.1371/journal.pone.0004110.s003 (0.18 MB DOC)

Figure S2 Flowering time control is dependent upon the Pol IV and Pol V active sites. *nrpd1a*, *nrpe1/nrpd1b* and *nrpd2* mutants, or transgenic lines generated by transforming these mutants with wild-type or active site mutant versions of NRPD1, NRPE1/NRPD1b or NRPD2a full-length transgenes, were grown side-by-side under short day conditions (8 hours light/16 hours dark). The positions of pots were changed every 4–6 days according to a randomized plot design. The total number of rosette leaves for each plant was counted when the bolt (flower stalk) achieved a height of 5 cm. The histograms show the average number of leaves

at flowering+/-the standard error of the mean. Asterisks denote mean values that are significantly different ($p < 0.05$) from the wild-type (WT; ecotype Col-0) control population as determined by using the Student t-Test; a double asterisk denotes a value that is significantly different from both the WT and *nrpd1a-3* controls. The number of individual plants analyzed for each genotype is denoted by the numeric value inside each vertical bar. As expected, based on prior studies [1,2], *nrpd1a-3*, *nrpd1b-11* (*nrpe1*) and *nrpd2* mutant plants were significantly delayed in flowering relative to wild-type plants. Flowering time of the mutants was unaffected by transforming them with the NRPD1, NRPE1 or NRPD2 active site mutant transgenes. However, wild-type flowering time was restored by the non-mutant NRPE1-FLAG or NRPD2-FLAG transgenes. It is noteworthy that the non-mutant NRPD1-FLAG transgene did not fully restore flowering time in the *nrpd1a-3* mutant background to that of wild-type plants, perhaps reflecting the incomplete rescue of siRNA levels shown in Figure 2B.

Found at: doi:10.1371/journal.pone.0004110.s004 (0.45 MB TIF)

Author Contributions

Conceived and designed the experiments: JRH CSP. Performed the experiments: JRH OP. Analyzed the data: JRH OP CSP. Wrote the paper: JRH CSP.

References

- Cramer P (2002) Multisubunit RNA polymerases. *Curr Opin Struct Biol* 12: 89–97.
- The-Arabidopsis-Genome-Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Pikaard CS, Haag JR, Ream T, Wierzbicki AT (2008) Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci* 13: 390–397.
- Luo J, Hall BD (2007) A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol* 64: 101–112.
- Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, et al. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120: 613–622.
- Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* 308: 118–120.
- Wierzbicki A, Haag JR, Pikaard CS (2008) Noncoding transcription by RNA Polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135: 635–648.
- Pontes O, Li CF, Nunes PC, Haag J, Ream T, et al. (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* 126: 79–92.
- Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, et al. (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev* 19: 2030–2040.
- Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, et al. (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* 37: 761–765.
- Onodera Y, Nakagawa K, Haag JR, Pikaard D, Mikami T, et al. (2008) Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. *Genetics*; In press.
- Li CF, Pontes O, El-Shami M, Henderson IR, Bernatavichute YV, et al. (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. *Cell* 126: 93–106.
- Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJ (2006) RNA-directed DNA methylation and Pol IVb in Arabidopsis. *Cold Spring Harb Symp Quant Biol* 71: 449–459.
- Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, et al. (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr Biol* 13: 2212–2217.
- Vaughn MW, Martienssen RA (2005) Finding the right template: RNA Pol IV, a plant-specific RNA polymerase. *Mol Cell* 17: 754–756.
- Vaucheret H (2005) RNA polymerase IV and transcriptional silencing. *Nat Genet* 37: 659–660.
- Cramer P (2006) Recent structural studies of RNA polymerases II and III. *Biochem Soc Trans* 34: 1058–1061.
- Sosunov V, Sosunova E, Mustaev A, Bass I, Nikiforov V, et al. (2003) Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *Embo J* 22: 2234–2244.
- Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292: 1863–1876.
- Zaychikov E, Martin E, Denissova L, Kozlov M, Markovtsov V, et al. (1996) Mapping of catalytic residues in the RNA polymerase active center. *Science* 273: 107–109.
- Werner F, Weinzierl RO (2002) A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* 10: 635–646.
- Dieci G, Hermann-Le Denmat S, Lukhtanov E, Thuriaux P, Werner M, et al. (1995) A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *Embo J* 14: 3766–3776.
- Carter DR, Eskiw C, Cook PR (2008) Transcription factories. *Biochem Soc Trans* 36: 585–589.
- Herr AJ (2005) Pathways through the small RNA world of plants. *FEBS Lett* 579: 5879–5888.
- Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292: 1876–1882.
- Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJ, et al. (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *Embo J* 25: 2828–2836.
- Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, et al. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98: 811–824.
- Hirata A, Klein BJ, Murakami KS (2008) The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451: 851–854.
- Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* 123: 1279–1291.
- Greco-Stewart VS, Miron P, Abraham A, Pelchat M (2007) The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. *Virology* 357: 68–78.
- Ding B, Itaya A (2007) Viroid: a useful model for studying the basic principles of infection and RNA biology. *Mol Plant Microbe Interact* 20: 7–20.
- Lehmann E, Brueckner F, Cramer P (2007) Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* 445: 445–449.
- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, et al. (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* 301: 653–657.
- Chiu J, March PE, Lee R, Tillett D (2004) Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res* 32: e174.
- Earley KW, Haag JR, Pontes O, Opper K, Juehne T, et al. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* 45: 616–629.
- Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* 16: 735–743.

37. Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299: 716–719.
38. Campell BR, Song Y, Posch TE, Cullis CA, Town CD (1992) Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. *Gene* 112: 225–228.
39. Sambrook J, Russell DR (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Table S1.

Target	Primer	Sequence (5' to 3')	Application
A13g23780	NRPD2a BP-F NRPD2a BP-R	GGGGACAAAGTTTGTACAAAAAGCAGGCTAAAAGATCAGTTCCAAAGTTGGTTGGC GGGGACCACTTGTGACAAAGAAAGCTGGGTGCGCATAGCTTGGTGTCCGAAAGTTGAGAGTG	Clone NRPD2a into pDONR221
A1l63020	NRPD1a active site-F NRPD1a active site-R	CACCGGCAATAATAACGCATGCACAGG GAATAGCTGCAATCCCGTCCATTG	Amplify the NRPD1a active site region
A1l63020	NRPD1a DDD/AAA-F NRPD1a mut-F NRPD1a DDD/AAA-R NRPD1a mut-R	GGTGTCTTTGTGGAGCTTGTCTCCACGGTTACGTTCTTCAGTC CTCCACGGTTACGTTCTTCAGTC ACAAGCTCCAGCAAAAGCACCACCGAACGGCAACACAGCAGATC ACGGAAACGGCAACACAGCAGATC	Mutagenesis
A12g40030	NRPD1b DDD/AAA-F NRPD1b mut-F NRPD1b DDD/AAA-R NRPD1b mut-R	GCTGTCTTTGTGGTCTTGTGTCCATTTGTCTACCCCTCAGTCTCTTAGTG GTCCATTTGTCTACCCCTCAGTCTCTTAGTG ACAAGCACAGCAAAAGCAGCACTGAGGGGGCTACACATCAGAG ACTGAGGGGGCTACACATCAGAG	Mutagenesis
A13g23780	NRPD2a ED/AA-F NRPD2a mut -F NRPD2a ED/AA-R NRPD2a mut-R	CAACCAAGCGGCTTCCATTTGTATGATGAACAAGGCTTCAATTGGAACGTG TGATGAACAAGGCTTCAATTGGAACGTG CAATGGAAAGCCGCTTGGTTGTACCCGAGATGAACATTCACAGCAAC TACCCGAGATGAACATTCACAGCAAC	Mutagenesis
A1sNI	A1sNI-F A1sNI-R	AGGATTTATTCAAATCCACGAACTT CGACTCCATAAGTAAACGAGTTG	Chop-PCR (Herr et al., 2005)
A12g19920	A1sNI control-F A1sNI control-R	CTCTGGGTTACCTTTTCAGGAAATCAG CTAAATGAAAGACTTACCTGCTTG	Chop-PCR control (Herr et al., 2005)
A1sNI	A1sNI RT-F A1sNI RT-R	ACCAACGTGCTGTGGCCCAATGGTAAATC AAAAATAAGTGGTGTGTACAAGC	RT-PCR (Herr et al., 2005)
solo LTR	solo LTR-F solo LTR-R	ATCAATTAATATGTCATGTTAAAACCGAATTG TGTTTCGAGTTTATCTCTCTAGTCTTCATT	RT-PCR (Wierzbicki unpublished)
Actin	Actin-F Actin-R	TCATACTAGTCTCGAGAGATGACTCAGATCATGTTTGGAG TCATTTAGAGGGCGGCCACAATTTCCCGTTCTGCGGTAG	RT-PCR (Herr et al., 2005)

Clamp head	Clamp core	ScRpb1:NRPD1	ScRpb1:NRPE1	External 1	Protrusion	ScRpb2:NRPD2				
		Clamp core	Active			Dock	Active site	Pore	Hybrid binding	Fork
		G52Q	G52Q			V44I				
		L53V	-			F51Y				
		-	D55N			I172V				
		P78E	P78E			L174V				
		L86F	-			L181T				
		P89S	-			L189N				
		L202Q	-			E194D				
		P242T	-			I204V				
		-	R247S			N221K				
		P248V	P248V			I269K				
		E259R	E259S			D396E				
		T263D	T263R			L461E				
		L266T	-			L514P				
		N273V	N273V			P524F				
		L276E	L276I			I743L				
		Q297S	Q297F			P745L				
		I325S	I325D			A753C				
		L329K	L329S			N762D				
		K332G	K332W			Q763H				
		G334R	G334E			S764G				
		R337K	-			P765R				
		-	G342R			N767V				
		-	R344G			M773H				
		V345S	V345S			K775Q				
		I353V	-			M778I				
		-	P357A			S844A				
		-	L374I			D894E				
		T375Q	-			L898F				
		P377S	P377E			D951Q				
		V380L	-			R983M				
		G395-	G395-			D998N				
		P396-	P396-			I1011V				
		P400-	P400-			M1021Q				
		G401-	G401-			Y1091F				
		R412V	R412T			V1099S				
		V432I	-			G1121D				
		H435S	H435R			R1129K				
		-	L443F			G1167K				
		Q447P	Q447P							
		L450I	L450T							
		M456I	M456Q							
		H458M	H458L							
		T467V	-							
		R469S	R469K							
		Y478F	Y478L							
		N479R	N479S							
		A480G	-							
		E486C	E486C							
		M487L	M487V							
		N488H	N488H							
		H490Y	H490F							
		R498K	R498K							
		D538N	D538R							
		T539C	T539V							

	ScRpb1:NRPD1	ScRpb1:NRPE1	
Pore	F540Y	-	
	-	I565L	
	P568A	P568S	
	L571Q	L571A	
	-	G574V	
	K575M	K575F	
	G615F	G615F	
	K619S	K619V	
	L629N	L629I	
	F662L	-	
Funnel	G665S	G665S	
	G707A	G707-	
	-	L722-	
	N723A	N723-	
	-	N741Y	
	-	M746L	
	-	G750K	
	-	K752N	
	-	G753S	
	S754N	S754A	
	N575K	N757K	
	Q767L	Q767L	
	G772V	G772K	
	R774L	R774K	
	L784C	L784M	
	P785A	P785A	
	F787W	-	
	P794L	P794R	
	F799V	F799I	
	F815V	F815A	
Bridge helix	M818V	M818I	
	G820S	G820A	
	E822D	-	
	G823S	G823V	
	L824S	L824I	
	D826S	D826R	
	T827G	T827S	
	A828N	A828S	
	V829A	V829R	
	K830D	K830G	
	T831L	T831L	
	Y836T	Y836T	
	-	R839K	
	K843F	K843A	
	Cleft	E846R	E846R
		V850A	V850I
-		Y852N	
-		G861S	
G869E		-	
Foot	-	G872S	
	D874V	D874R	
	L956-	L956-	
	N959-	N959-	
	Q1070C	Q1070T	
Cleft	S1071A	S1071A	
	G1073S	G1073S	

	ScRpb1:NRPD1	ScRpb1:NRPE1
	-	E1074N
	P1075A	-
	T1077Y	T1077Y
	Q1078S	Q1078K
	M1079A	M1079A
	T1080L	T1080V
	L1081D	-
	T1083P	T1083S
	F1084I	F1084S
	H1085S	H1085P
	A1087L	A1087S
	G1088E	G1088N
	T1095L	T1095K
	G1097N	G1097V
	P1099L	P1099L
	R1100E	R1100C
	E1103S	E1103N
	I1104K	I1104F
	T1113S	T1113I
	P1114L	P1114L
	-	L1120H
	A1131S	-
	T1142S	T1142S
	E1151M	E1151L
	W1191S	W1191I
	R1199Q	R1199K
	V1282I	V1282I
	L1306V	-
	-	G1310V
	N1330D	N1330Y
	E1342D	E1342S
	-	A1343C
	-	R1345F
	E1351N	E1351R
	V1355A	V1355S
	-	G1360S
	R1366E	R1366E
	-	D1373N
	M1375L	-
	T1376S	-
	R1386A	R1386S
	-	F1402L
	E1403S	E1403I

Clamp
core



Figure S1. Multiple alignment of *A. thaliana* RNAP Largest Subunits and the Yeast Pol II Largest Subunit. Full-length protein sequences for *A. thaliana* NRPA1 (At3g57660), NRPB1 (At4g35800), NRPC1 (At5g60040), NRPD1 (At1g63020), NRPE1 (At2g40030) and *S. cerevisiae* Rpb1 were aligned using ClustalW2 (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>) in conjunction with final editing by hand. Alignments were colored using BOXSHADE v3.21 (http://www.ch.embnet.org/software/BOX_form.html). DNA-dependent RNA polymerase conserved domains A to H are underlined and designated to the right of the alignments. Yeast Pol II structural features, according to Cramer et al (2001), are designated below the alignments. Regions that make contact with other RNAP subunits are designated in italics above the alignments. The Metal A site is designated with asterisks above the alignment.



NRPD1 295 -----QKSDTPKLCGLRFMKDVLGKRSLHTFRTVVVGDPSLKLNEIGI
 NRPE1 284 KAARN--IDMRYGVSKISDSSSSKAWTEKMRTLFIRKGSGFSSRSVITGDAYRHVNEVGI
 NRPA1 417 -----VQSQRDSSGICQLIEKKEGLFRQKMGKRVNHACRSVISPDPYIAVNDIGI
 NRPB1 323 ----QPRATQKSGRPIKSIKSRLLAKEGRIRGNLMGKRVDFSARTVITPDPITINIDELGV
 NRPC1 330 -----CQNQPEEHPLSGILQRLKGGGRFRANLSGKRVEFTGRTVISPDENLKITEVGI
 ScRpb1 311 ----QPQALQKSGRPFVKSIRARLKGKEGRIRGNLMGKRVDFSARTVISGDPNLELDQVGV



C

NRPD1 340 PESIAKRLQVSEHLNQCNERLIVTSFVP-----TLLDNKEMHVRRGD
 NRPE1 342 PIEIAQRITFEERSVHNRYLQKLVDD-----KLCLSYTQGSTTYSL
 NRPA1 468 PPCFALKLTYPERVTPWNVEKLRERAIINGPDIHPGATHYSKSSMTKLPSTKARRAIAR
 NRPB1 379 PWSIALNLTYPETVTFYNIERLKLVDYGPHPKPK-----TGAKYIIRIDGQRLDLR
 NRPC1 384 PILMAQILTFPECVSRHNIEKLRQCVRNGPNKYFG-----ARNVRYPOGSSRTLVG
 ScRpb1 367 PKSIAKTLTYPEVVTFYNIDRLTQLVRNGPNEHPG-----AKYVIRVSGDRIDLRL



C

NRPD1 382 RLVAIQVNDIQTG-----DKIFRSLMDGDTVLMNRPESIHOHSLIAMTVRILFTTS
 NRPE1 385 RDGSKGHTELKPG-----QVHHRVMDGDVVFINRPPTTHKHSLQALRVVHE-DN
 NRPA1 528 KILSSRGATTELKGTCDINFEKTVHRHMRDGDIVLVNRQPTLHKPSLMAHKVRVLKGEK
 NRPB1 432 YLKKSSDQHELG-----YKVERHLQDGFVLFNRPQSLHKMSIMGHRIIRIMP-YS
 NRPC1 435 DYRKRIADELAIG-----CIVDRHLQEGDVVLFNRPQSLHRMSIMCHRARIMP-WR
 ScRpb1 417 YSKRAGDIQIQYG-----WKVERHIMDNDPVLFNRPQSLHKMSMMAHRVKVIE-YS



D

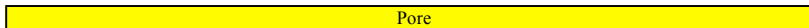
Metal A
* * *

NRPD1 433 VVSLNPICQLFFRGDFDGDCLHGYVPOSIQAKVELDELVALDKQLINRONGRNLLSLGOD
 NRPE1 435 TVKINPLMCSHLSADFDDGDCVHLIFYPOSLSAKAEVMELEFSVEKQLLSSHTQQLIQMGSD
 NRPA1 588 TLRHLHYANCSTYNADFDGDEMNVHFQDEISRABAYNIVNANNQYARPSNGEFLRALIOD
 NRPB1 482 TFRNLNSVTSPYNADFDGDEMNMHVPOSFETRAEVLELMVFKCIVSPOANRFVMGIVOD
 NRPC1 485 TLRFNESVQNPYNADFDGDEMNMHVPOTEEARTEAITLMG-----
 ScRpb1 467 TFRNLNSVTSPYNADFDGDEMNLHVPOSEETRAELSQLCAVPLQIVSPOSNKFCMGIVOD



D

NRPD1 493 SLTAAYLVNVEKNCYLNRAQMQLQV-----YCPFQLLFPF
 NRPE1 495 SLLSLRVM--LERVFLDKATAQQLAM-----YGSISLLEPP
 NRPA1 648 HIVSSVLL-TKRDTFLDKDHFNQLLFSSGVTDMVLSTFSGRSGKKVMVSASDAELLTVTP
 NRPB1 542 TLLGCRKI-TKRDTFIEKDVFMNTLM-----WWEDFDGKVEAF
 NRPC1 525 -----DTFYDRAAFLSICS-----YMGDGMDSIDLPTP
 ScRpb1 527 TLCGIRKL-TLRDTFIELDQVLNMLY-----WVPLWDGVIPTP



E

Rpb8 Interaction

NRPD1 528 AIIKASPSSTEPQWTGMOLFGLMFFPGFD-YTYPLNNVV-----
 NRPE1 528 ALRKSSES--GPAWTFQILQLAFPERLS-CKGDRFLVIG-----
 NRPA1 707 AIIKF----VPLWTGKQVITAVLNQITKGHPPTVEKATKLPVDFFKCRSREVKPNSGD
 NRPB1 579 AIIKF----RPLWTGKQVFNLIIEKQINLLRYSAWHAETETG-----
 NRPC1 553 TIIKF----IELWTGKQIFSVLLRPNASIRVYVTLNVKEKNFKKG-----
 ScRpb1 564 AIIKF----KPLWSGKQILSVAIPNGIHLQRF----DEGTT-----



E

NRPD1 567 -----SNGELLSFSEGSAAWRD-----GENFIERLLKH
 NRPE1 565 -----SDLLKFDFFVDAMGS-----IINEIVTSIFLEK
 NRPA1 762 LTKKKEIDESWKQNLNEDKLRKNEFVCGVIDKAQFAD-----YGLVHTVHELY
 NRPB1 617 -----FITPGDTQVREERGELLAGTLCRKTLLGT-----SNGSLVHVIVHEEV
 NRPC1 594 -----EHGFDETMCINDGWYFRNSELISGQLGKATLALDIFPLNGNKGDLYSILLRDY
 ScRpb1 597 -----LLSPKNGMLTIDQIIFGVVEKTVGS-----SNGGLIHVVTRERK



Rpb2 Interaction

NRPD1 596 DKGKVLDTIYSAQEMLSQWLLMRGLSVSIADLYLSSDLQSRKNLTFEISYGLREAEQVCN
 NRPE1 593 GFKETLGFDFSLQPLLMESLFAEGFSLSEDLISMSRADN--DVTHNLIIREISPMVSRLE
 NRPA1 812 GSNAAAGNLLSVFSRLEFVFLQTHGFTCGVDDLLILKDM--EERIKQLQECENVGERVLR
 NRPB1 658 GPDAAARKFLGHTQWLVNYWLLQNGFTTIGIDTADSSM--EKINFTISNAKTAVKDLIR
 NRPC1 649 NSHAAAVCMNRLAKLSARWIGIHGFSIGIDDVQGEELS--KERKDSIQFGYDQCHRKIE
 ScRpb1 638 GEQVCAKLFGNIQKVVNFWLLHNGFSTIGIDTADGPTV--REITFTTIAEAKKKVLDVTK



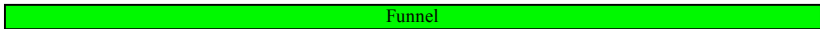
Rpb9 Interaction

NRPD1 656 KQQLMVESWRDFLAVNGEDKEEDSVSDLARFCYERQKSATLSELAVSAFKDAYRDVQALA
 NRPE1 651 LSYRDELQLEN-----SIHKVKEVA
 NRPA1 870 KTFGIDVDVQIDPQDM-RSRIERILYEDGESALASLDRSIVNYLNQCSSKGMNDLLSDG
 NRPB1 716 QFQCKELDPEP-----GRTMRDTFENRVNQVLNKAR----DDAGSSA
 NRPC1 707 EFNREGNLQKA-----GLDGAKSLEAETGILNTR----EATGKAC
 ScRpb1 696 EAQANLLTAKH-----GMTLRESFEDNVVRFLEAR----DKAGRIA



Rpb9 Interaction

NRPD1 716 YRYGDQSNFLIMSKAGSKGNIGKLVHSMCIGLONSAVLSLFGFPRELCAAWNDPNSP
 NRPE1 671 ANFMLKSYSTRNLIDIKSNSAITKLVQOTGFLLQLSLDKKIFYTKTLVEDMAIFCKRKYG
 NRPA1 929 LLKTPGRNCISLMTISGAKSKVNFQITSHLGOQDLEGRVPRMVSGKTLFCFHPWWS
 NRPB1 764 QKSLAETNNLKAMVTAGSKGSFINISOMTACVGOQNVGKRIPFGFDGRTLPHFTKDDYG
 NRPC1 745 MSGLHWRNSPLIMSQCSSKGSFINISQVACVGOQTVNGHRAFDGFDRLSPLPHFPRMSKS
 ScRpb1 734 EVNLKDLNIVKQVMVAGSKGSFINIAQMSACVGOQSVGKRIFAFGEFVDRTLPHFSKDDYS



Rpb2 Interaction *Bridge helix*

NRPD1 776 LRGAKGKDDTTESYVPGVIEENSFLTGLNPLESIVHVSVTSRDSSFSGNADLP--GTISR
 NRPE1 731 RISSSGDF-----GIVKGCFFHGLDPYEMAHSIAAREVIVRSSRGLAEPGTIFK
 NRPA1 989 PRAG-----GFI SDRFLSGLRPOEYFFHCMAGREGLVDTAVKTSRSYGLQR
 NRPB1 814 PESR-----GFVENSYLRLTLPQEFFFHAMGGREGLIDTAVKTSSETGYIQR
 NRPC1 805 PAAK-----GFVANSFYSLTATEFFFHMTGGREGLVDTAVKTAETGIMSR
 ScRpb1 794 PESK-----GFVENSYLRLTLPQEFFFHAMGGREGLIDTAVKTAETGYIQR



Rpb5 Interaction

NRPD1 834 RIMFFMRDIYAAYDGTVRNSFNQLVQFTYETDGPVEDIT-----
 NRPE1 781 NLMAVLRDVIITNDGTVRNTCSNSVIQFKYGVDSERGHQG-----
 NRPA1 1035 CLMNLLESKVNNDCTVRDADG--SIIIQFYGEDGV DVHRSS-----
 NRPB1 860 RLVKAMEDIMVKYDGTVRNSLG--DVIQFELYGEDGMDAVWIF--SQKLDLTKMKKSEFDRTFK
 NRPC1 851 RLMKALEDLIVHYDNTVRNASG--CIIQFITYGDDGMDPALME-----
 ScRpb1 840 RLVKALEDIMVHYDNTVRNSLG--NVIQFELYGEDGMDAAHIEKQSLDTIGGSDAAFEKRYR



NRPD1 874 -----
 NRPE1 821 -----
 NRPA1 1075 -----FIEKFKEIT
 NRPB1 919 YEIDDENWNPTYLSDEHLEDLKGIRELRDVFDAEYSKLETDRFQLGTEIATNGDSTWELP
 NRPC1 891 -----GKDGAPLN
 ScRpb1 899 VDLLNTDHTLDPSLLESSEIIGDLKLQVLLDDEEYKQLVKDRKFLR-EVFDGEANWELP



NRPD1 874 -----
 NRPE1 821 -----
 NRPA1 1084 INQDMVLQKCEDMLSG-----ASSYIS-----DLPIS
 NRPB1 979 VNIKRHIWNAQKTFKIDLRKISDMHPVEIVDAVDKLOERLLVVPGDALISVEAQKNATLF
 NRPC1 899 FNRFLFKVQATCPPRSHHTYLS-----SEELSQKFEEELVRHDKSRVCTDAFVKSLE
 ScRpb1 958 VNIIRRIIQNAQQTFFHIDHTKPSDLTIKDIVLGVKDLQENLLVLRGKNEIQNAQRDAVTL



F

F

F

Rpb6 Interaction

NRPD1 874 -----GALGSLSAICALSAAAY
 NRPE1 821 -----LFEAGEFVGVLAATAMSNPAY
 NRPA1 1112 LKKGAEKFVEAMPNERIAASKFVRQEEELKLVKSKFFASLAQPGFVGVLAQSVGEEST
 NRPB1 1039 FNTLLRSTLASKRVLEEYKLSREAFEWVIGETESRFLQSLVAPGEMIGCVAAQSIGEPAT
 NRPC1 952 FVSLLG-----VRSASPPQVLYKASGVTDKQLEACTAICTIGAQSIGEPCT
 ScRpb1 1018 FCCLLRSLATRRLVQFYRLTKQAFDWWLNSLTAQFLRSVVHPGEMVGVLAQVSIGEPAT



NRPD1 891 SALLQPISSLLETSPLLNLKNLECGSKKG-QREQTMSLYLSEYLSKSKKHGFEGSLETKN
 NRPE1 842 KAVLDSSPNSNSSWELMKEVLLCKVNFQNTTNDRRVILYNECHCGKRFQENAACTVRN
 NRPA1 1172 QMTLNTFHLAQRGEMNVTLGIPLRQETLMTAAANIKTPIMTCPILKG--KTKFDANDITD
 NRPB1 1099 QMTLNTFHYAGVSAKNVTLGVPRRLREIIN-VAKRIKTPSLSVYLTPEASKSKEGAKTVQC
 NRPC1 998 QMTLNTFHFAGVASMNIITQGVPRINEIIN-ASKNIISTFVISAELN--PLELTSARVWVG
 ScRpb1 1078 QMTLNTFHFAGVASKKVTSGVPRLKEIIN-VAKNMKTESLTVYLEPGAADQEQAKLIRS



Rpb9 Interaction

NRPD1 950 HLEKLSFSEIVSTSMIIFSS-NTKVPLSPWVCHFISEKVLKRRQLSAESVSSIN--
 NRPE1 902 KLNKVSCLKDTAVEFLVEYRQKQITISELFGIDSCSLHGHIHNLKTLQDWNISMQDIHQKCE
 NRPA1 1230 RLKRIITVADIIKSMELSVVYTYVYENEVCSIHKLKINLYKPEHYPKHTDITEEDWEETMR
 NRPB1 1158 ALEYTTLRSVQATEVWVDPDEMSTIIEEDFEFVRSYYEMPDEDVSP--DKISFWLLR--
 NRPC1 1055 RIEKTTLQVAESIIEVLMTSTASVRIIDNKIIEEACLS-----ITPWSVKN--
 ScRpb1 1137 AIEHTTLKSVTIASEIYYDDEIRSTVPEDEEIIQLHFSLLDEEAEQSFDQOSPWLLR--



NRPD1 1007 -----EQYKSRNRELK-----
 NRPE1 962 DVIN-----SFGQKKKAKATDFK-----
 NRPA1 1290 AVFLRKLEDAIETHMKMLHRIIRGIHNVVTGPIAGNETDNDSDSVSGQNEDDGGDDGEGTE
 NRPB1 1214 -----TELNREMMDVKKLS-----
 NRPC1 1103 -----SIKTPRIKLNNDNIR-----
 ScRpb1 1195 -----LELDRAAMNDKDI-----



NRPD1 1018 ---LDIVDLDIQTNTNHCSSDDQAMKDDNVCITVTVVEAS-----KHSVLELDAIRL
 NRPE1 981 ---RTSLSVSECCSFRDPCGSKGSMPCITFSYNATDP-----DLERTLDVLCN
 NRPA1 1350 VDLGSDAQKQKQETDEMDEENSELETNEPSSI SGVEDPEMDSENEDTEVSKEDTPEP
 NRPB1 1228 ---MADIAEKINLEFDDDLTCIFNDNAQKLIIRIRIMNDEGPKGELQDESAAEDDVFKK
 NRPC1 1119 ---VLDTGLDITPVVD-----KSRAHFNLHN
 ScRpb1 1209 ---MGQVGERIKQTFKNDLFIWSEDNDEKLIIRCRVVR---PKSLDAETEAEEDHMLKK



Rpb9/Rpb2 Interaction

NRPD1 1066 VLIPFLLDSPVKGDQCIKKVN-----
 NRPE1 1027 TVYFVLEIVIKGDSRIICSAN-----
 NRPA1 1410 QEESMEPQKEVKGKVNVEQSKKRRKFRVRAKSDRHIFVKGEKEFEVHFVFATDDPHIL
 NRPB1 1285 IESNMLTEMALRGIPDINK-----
 NRPC1 1142 LKN-----GIKTVR-----
 ScRpb1 1263 IENTMLENTLIRGVENIER-----



NRPD1 1087 -----ILWTRPKAPKRNGHLAGELYLKVIMY
 NRPE1 1048 -----IWNSSDMTWTIRNRHASRRGEVVLVDT
 NRPA1 1470 LAQIAQQTAKVYIQNSGKIERCTVANCGDPQIYHGDNPKERREISNDEKKASPAHAS
 NRPB1 1304 -----VFIKQVRKSRFDEEGFKTSEEWMLDTE
 NRPC1 1152 -----VVVAEDMKSKQIDG---KTKKLFVE
 ScRpb1 1282 -----VVMKYDRKVPSPTEYVKEFEWVLETD



Rpb5 Interaction

NRPD1 1115 GDRGKR---NCWTALLETCLEPIMDMIDWGRSHPDNIRQCCSVYGI DAGRSIFVANLESA
 NRPE1 1076 VEKSAVKQSGDAWRVIDSCLSVLHLIDTKRSIPYSVKQVQELLGLSCAFEQAVQRISAS
 NRPA1 1530 G-----VDFPALWEFQDKLDVRYLYSNSIHDMLNIFGVFAARETIIREINHV
 NRPB1 1332 G-----VNLLAVMCHED-VDFKRTTNSNHLIEIEVLGIEAVRRALLDELRVV
 NRPC1 1176 G-----TNLLAVMGTPG-TNGRITTSNNVVEVSKTLGIEAARTTIIDEIGTV
 ScRpb1 1310 G-----VNLSEVMTVPG-LIPTRIIYTNSTFIDIMEVLGIEAGRAALYKEVYV



NRPD1 1171 VSDTGKEILREHLIVADSLSVTGEFVALNAKWSKQRQVESTPAHFTQACHSSPSQCFL
 NRPE1 1136 VRMYSKGVLKEHIIILLANNMTCSGTMLGFNSGGYKALTRSLNIKAFTEATLIAPRKCFE
 NRPA1 1577 FKSYGISVSI RHINLIADYMTFSGGYRPMSPRMGGIA-----ESTSFRCRMTFFATKFTV
 NRPB1 1378 ISFDGSYVNYRHLLAICDTMTYRGHMLAITRHGINR-----NDTGFILMRCSEETVDILL
 NRPC1 1222 MGNHGMSTDIRHMMLLADVMTYRGEVLGIQRTGIQK-----MDKSVLMQASFERTGDHLF
 ScRpb1 1356 IASDGSYVNYRHMLLVDMVTQGGILTSVTRHGFNR-----SNTGALMRCSEETVEILF



H

Rpb2 Interaction Rpb6 Interaction

NRPD1 1231 KAAKEGVRDDLQGSIDALAWKVPGFGTGDQFETIISPKVHGF-----
 NRPE1 1196 KAAEKCHTDSLSTVVGSCSWGKRVDVGTGSQFELLWNQKETGL-----
 NRPA1 1632 QAATYGEKDTLETSPARICLGLPALSGTGCFDLMQORVEL-----
 NRPB1 1443 DAAAYAEETCLRGVTENIMLQQLAPIGTGDCELYLN-DEMLKNAIELQLPSYMDGLEFGM
 NRPC1 1277 SAAASGKVDNIEGVTECVIMGIPMKLGTGILKVLQRTDDLPK-----LKYGP
 ScRpb1 1411 EAGASAEFLDCRGVSENVILGQMAPIGTGAFDVMIDEEISLVKYMPEQKITEIEDGQDGGV



H

NRPD1 1274 -----
 NRPE1 1239 -----
 NRPA1 1671 -----
 NRPB1 1492 TEARSPVSGTPYHEGMMSPNYLLSPNMRLSPMSDAQFSPYVGGMAFSPSSSPGYSPPSPG
 NRPC1 1324 DEIIS-----
 ScRpb1 1471 TPEYSN-----ESGLVNADLDVKDELMTSPLVDSGSNDAMAG-GFTAYGGVDYG-----

NRPD1 1274 -----
 NRPE1 1239 -----
 NRPA1 1671 -----
 NRPB1 1552 YSPTSPGYSPTSPGYSPTSPGYSPTSPYSPSSPGYSPTSPAYSPTSPYSPTSPYSPT
 NRPC1 1329 -----
 ScRpb1 1518 -EATSP-----FAAYGEAPTSPFGVSSPGFSPTSPYSPSTSPAYSPTSPYSPT

Pol II heptad repeats

NRPD1 1274 -----
 NRPE1 1239 -----
 NRPA1 1671 -----
 NRPB1 1612 SPSYSPTSPYSPTSPYSPTSPYSPTSPAYSPTSPAYSPTSPAYSPTSPYSPTSPSY
 NRPC1 1329 -----
 ScRpb1 1567 SPSYSPTSPYSPTSPYSPTSPYSPTSPYSPTSPYSPTSPYSPTSPYSPTSPYSPTSPSY

Pol II heptad repeats

NRPD1 1274 -----
 NRPE1 1239 -----
 NRPA1 1671 -----
 NRPB1 1672 SPTSPYSPTSPYSPTSPYSPTSPAYSPTSPGYSPTSPYSPTSPSYGPTSPSYNPQS
 NRPC1 1329 -----
 ScRpb1 1627 SPTSPYSPTSPYSPTSPYSPTSPAYSPTSPYSPTSPYSPTSPYSPTSPYSPTSPSY-----

Pol II heptad repeats

NRPD1	1312	-----AFLKDIKVLDGK--	Pol V repeats
NRPE1	1616	<u>SDAAAWGSRNKKTSEIESGAGAWG</u> SWGQPSPTAEDKDTNEDDRNPWVSLKETKSREKDDK	
NRPA1	1671	-----	
NRPB1	1841	-----	
NRPC1	1329	-----	
ScRpb1	1727	-----	
NRPD1	1324	----GIPMSLLRTIFTWKN-----	
NRPE1	1676	ERSQWGNPAKKFPSSGGWSNGGGADWKGNRNHTPRPPESEDNLAPMFTATRQLDSFTSE	
NRPA1	1671	-----	
NRPB1	1841	-----	
NRPC1	1329	-----	
ScRpb1	1727	-----	
NRPD1	1339	-----IELLSQSLKRILHSYEIN---ELLNERDEGLVKMVLQLHPNSVEKIGPGVKGI	DeCL-like domain
NRPE1	1736	EQELLSDVPEVMRTLRLKIMHPSAYPDGDFISDDDKTFVLEKILNFHPQKETKLGSGVDFI	
NRPA1	1671	-----	
NRPB1	1841	-----	
NRPC1	1329	-----	
ScRpb1	1727	-----	
NRPD1	1389	RVAKS-KHGDSCCFEVVRIDGTFEDFSYHKCVLGATKIIAPKKMNFYKSKYLKN----GT	DeCL-like domain
NRPE1	1796	TVDKHTIFSDSRCFFVSTDGAKQDFSYRKSLNNYLMKKYPDRAEEFIDKYFTKPRPSGN	
NRPA1	1671	-----	
NRPB1	1841	-----	
NRPC1	1329	-----	
ScRpb1	1727	-----	
NRPD1	1444	LESGGFSENP-----	QS-rich domain
NRPE1	1856	RDRNNQDATPPGEEQSQPPNQSIGNGGDDFQTQTQSQSPSQTRAQSFSQAQAQSPSQTS	
NRPA1	1671	-----	
NRPB1	1841	-----	
NRPC1	1329	-----	
ScRpb1	1727	-----	
NRPD1	1454	-----	QS-rich domain
NRPE1	1916	QS	
NRPA1	1671	-----	
NRPB1	1841	-----	
NRPC1	1329	-----	
ScRpb1	1727	-----	
NRPD1	1454	-	
NRPE1	1976	T	
NRPA1	1671	-	
NRPB1	1841	-	
NRPC1	1329	-	
ScRpb1	1727	-	

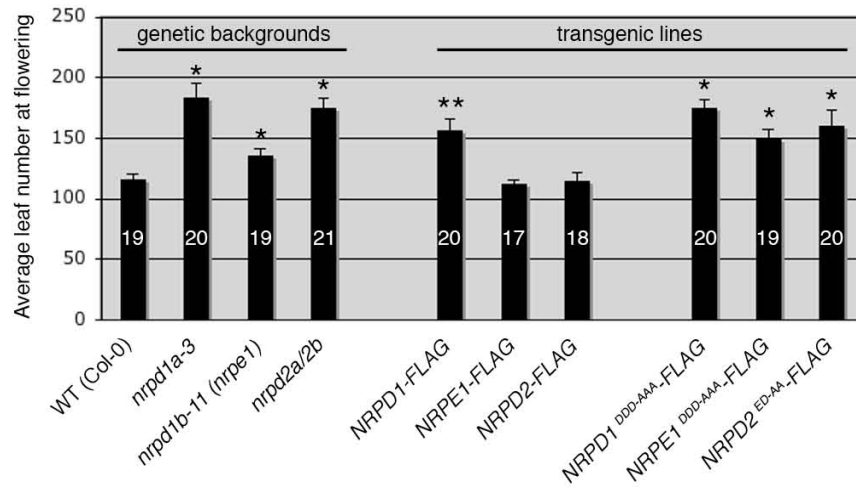


Figure S2

CHAPTER 5

DNA-DEPENDENT RNA POLYMERASE IV AND RNA-DEPENDENT RNA POLYMERASE 2 ARE PHYSICALLY COUPLED TO PRODUCE siRNA PRECURSORS

A manuscript in preparation

My contributions to this work:

In this work, I designed and performed all experiments unless otherwise noted. Tom Ream raised and affinity purified antibody against the *Arabidopsis thaliana* RDR2 protein as well as performed the Pol IV affinity purification confirming Pol IV-RDR2 interaction *in vivo* with Carrie Nicora, Angela Norbeck and Ljiljana Pasa-Tolic performing the LC-MS/MS analysis at the Pacific Northwest National Laboratory. Olga Pontes performed immunolocalizations. I raised and affinity purified antibody against the NRPD1 and NRPD2 proteins and generated all transgenic lines and crossed lines used in this study. Three key findings are demonstrated here: (1) Pol IV and RDR2 physically associate and (2) Pol IV has DNA-dependent RNA polymerase activity and (3) RDR2 is capable of transcribing both single-stranded RNA and single-stranded DNA. I wrote, edited and contributed significantly to the intellectual value of the paper, with the assistance of Craig Pikaard.

DNA-DEPENDENT RNA POLYMERASE IV and RNA-DEPENDENT RNA POLYMERASE 2 are physically coupled to produce siRNA precursors

Jeremy R. Haag¹, Thomas S. Ream¹, Olga Pontes¹, Carrie D. Nicora², Angela Norbeck², Ljiljana Pasa-Tolic² and Craig S. Pikaard^{1,*}

¹ Department of Biology, Washington University, St. Louis, MO 63130, USA

² Pacific Northwest National Laboratory, Richland, WA 99352, USA

*corresponding author: pikaard@biology.wustl.edu; phone: 314-935-7569; FAX 314-935-4432

Running title: Pol IV is a DNA-dependent RNA polymerase bound to RDR2

Abstract

In Arabidopsis, the nuclear DNA-dependent RNA polymerase, Pol IV, and the RNA-dependent RNA polymerase, RDR2, are required for the biogenesis of 24 nt small interfering RNAs (siRNAs) that direct DNA methylation and transcriptional silencing of corresponding heterochromatic loci. We show that Pol IV and RDR2 are physically associated *in vivo*. *In vitro*, Pol IV displays DNA-dependent RNA polymerase activity on templates that mimic paused transcription bubbles and RDR2 transcribes single-stranded RNA or DNA templates in a primer-independent fashion. Mechanistic coupling of Pol IV and RDR2 transcription can account for the channeling of RNA precursors in the initial steps of the 24 nt siRNA-directed DNA methylation pathway.

Pol IV and RDR2 Are Physically Coupled

Pol IV and RDR2 play key roles early in the RNA-directed DNA methylation pathway (Figure 1A). This pathway is responsible for the transcriptional silencing of repeated genomic sequences that include transposable elements, foreign transgenes and excess 5S and 45S rRNA genes (Matzke et al., 2009). Recently, we determined the subunit composition of *A. thaliana* Pol IV by LC-MS/MS (Ream et al., 2009). In addition to peptides corresponding to twelve core subunits of Pol IV, ten peptides that collectively represent 12% of the RDR2 protein sequence were identified in affinity purified Pol IV (Figure 1B). Physical association of RDR2 with Pol IV was confirmed by reciprocal co-immunoprecipitation (co-IP) experiments, exploiting transgenic lines expressing epitope-tagged Pol IV or RDR2 in conjunction with antibodies recognizing the native proteins. An RDR2-HA transgenic line was generated by rescuing the *rdr2-1* null mutation with a transgene expressing RDR2 fused to a HA epitope tag at the C-terminus (Figures 1C and Supplemental Figure 1). A tagged Pol IV line was generated by rescuing a null mutant defective for largest subunit, *nRPD1-3*, with a transgene expressing FLAG-tagged NRPD1 (Pontes et al, 2006). Following anti-HA immunoprecipitation (IP) of RDR2-HA, RDR2 is readily detected by immunoblotting and probing with an anti-RDR2 native protein antibody, as expected (Figure 1D, lane 2). This antibody also detects native RDR2 on immunoblots following anti-FLAG immunoprecipitation of NRPD1-FLAG, supporting the mass spectrometry evidence that RDR2 co-purifies with Pol IV (Figure 1D, lane3). In reciprocal experiments, anti-HA immunoprecipitation of RDR2-HA was followed by immunoblotting and probing for the Pol IV catalytic subunits, NRPD1 and NRPD2. Both catalytic subunits are detected in RDR2 IP fractions

(Figure 1E, lane 2). Collectively, the mass spectrometry and immunological results show that Pol IV and RDR2 physically associate *in vivo*.

RNA-dependent RNA polymerases interact with Dicer endonucleases in *C. elegans*, *S. pombe* and *T. thermophila* (Colmenares et al., 2007; Duchaine et al., 2006; Lee and Collins, 2007). However, co-IP analysis failed to reveal a physical association between RDR2 and DCL3 (Figure 1D, lane 5), the principle Dicer of the RNA-directed DNA methylation pathway (see Figure 1A) (Kasschau et al., 2007; Xie et al., 2004). RDR2 was also not detected in Pol V fractions obtained by immunoprecipitation of NRPE1, the Pol V largest subunit (Figure 1D, lane 4), in keeping with the absence of RDR2 peptides in Pol V fractions analyzed by LC-MS/MS. RDR2 was also absent in fractions of IPed RDR6 (Figure 1D, lane 6), an RNA-dependent RNA polymerase involved in several post-transcriptional gene silencing pathways (Borsani et al., 2005; Dalmay et al., 2000; Mourrain et al., 2000; Muangsan et al., 2004; Peragine et al., 2004; Vazquez et al., 2004). Collectively, the data of Figure 1 indicate that RDR2 specifically interacts with Pol IV.

Pol IV and RDR2 could potentially associate through protein-protein contacts or via an RNA intermediate. To test whether RDR2 might be tethered to Pol IV via a Pol IV transcript, we exploited the *NRPD1^{DDD-AAA}-FLAG* transgenic line (Haag et al., 2009). This line was generated by transforming the *nripd1-3* null mutant with a full-length *NRPD1* transgene in which the three conserved aspartates of the catalytic center's Metal A motif are mutated to alanines. This *NRPD1^{DDD-AAA}-FLAG* protein fails to complement the *nripd1-3* mutant, lacks all known Pol IV biological activity, and is expected to be transcriptionally inactive (Sosunov et al., 2005; Werner and Weinzierl, 2002; Zaychikov

et al., 1996). However, the mutated NRPD1^{DDD-AAA}-FLAG recombinant subunit appears to be unaffected in its assembly into Pol IV complexes, as indicated by its association with NRPD2, the Pol IV second-largest subunit, to the same extent as non-mutant and biologically active NRPD1-FLAG (Figure 2A). Importantly, the wild-type and active site mutant versions of Pol IV both co-immunoprecipitate RDR2 to an equivalent degree, suggesting that Pol IV does not have to be transcriptionally competent in order to interact with RDR2 (Figure 2A, lanes 4 and 5). Likewise, RDR2 is detected in immunoprecipitated NRPD1 samples treated with RNase A, which is expected to degrade any RNA molecules that might potentially tether Pol IV and RDR2 (Figure 2B). Based on these results, a physical association of Pol IV with RDR2 that is not mediated by RNA seems most likely.

Cytological studies suggest that only a fraction of the Pol IV and RDR2 present in the nucleus colocalizes and is potentially associated. Pol IV is typically detected as numerous puncta distributed throughout the nucleoplasm, but is absent from the nucleolus, which appears as a black hole in nuclei stained with the DNA-binding fluorescent dye, DAPI (Figure 1C). By contrast, RDR2 typically displays a prominent crescent, or ring, along the inner perimeter of the nucleolus in addition to being present in the nucleoplasm (Figure 1C, top row, red signals). In most nuclei (74%, n=501), there is no obvious overlap in the Pol IV and RDR2 signals. However, in a subset of the nuclei (26%, n=501), in which RDR2 tends to be more abundant in the nucleoplasm relative to the nucleolus, some overlap in the Pol IV and RDR2 signals is apparent. Taken together, the mass spec, IP and cytological results suggest that Pol IV and RDR2 can stably

associate with one another, but it is likely that only a fraction of the Pol IV and RDR2 pools participate in these interactions.

Affinity-purified Pol IV fractions display DNA and RNA-dependent RNA polymerase activities

Partially purified Pol IV and Pol V fractions have not yielded detectable DNA-dependent RNA polymerase activity *in vitro* in conventional assays that employ bulk DNA as the source of potential templates (Huang et al., 2009; Onodera et al., 2005). In initial tests using alternative templates, we used broccoli (*Brassica oleracea*) chromatin and incorporation of alpha-labeled ^{32}P -CTP as a measure of RNA synthesis. Weak RNA polymerase activity was detected from Pol IV immunoprecipitated samples compared to Pol II immunoprecipitated samples that robustly programmed the incorporation of ^{32}P -CTP into RNA polymers in an alpha-amanitin sensitive manner (data not shown). No *in vitro* activity was observed in this assay from Pol V immunoprecipitated samples (data not shown).

To explore further the weak polymerase activity detected with Pol IV samples, we turned to templates assembled by annealing defined DNA and RNA oligonucleotides (Figure 3A and B). The annealed oligos create a tripartite template, or scaffold, that mimics a transcription bubble, complete with a 8 bp RNA-DNA hybrid, single stranded DNA and RNA upstream of the hybrid and double-stranded DNA downstream of the RNA. Previous studies have shown that RNA polymerases I and II will associate with such DNA-RNA scaffolds, positioning the respective nucleic acids correctly relative to the catalytic center and the DNA and RNA exit channels such that the RNA can be

extended by addition of nucleotides templated by the downstream duplex DNA (Brueckner et al., 2007; Kuhn et al., 2007).

Using the tripartite oligonucleotide scaffold, we tested the ability of Pol II and Pol IV-RDR2 complexes to catalyze the incorporation of alpha ³²P-CTP into RNA extension products that could be resolved by denaturing polyacrylamide gel electrophoresis and visualized by autoradiography or phosphorimaging. The initial RNA strand present in the artificial transcription bubble was 16 nt; full-length extension of this RNA in a templated fashion would yield an RNA of 32 nt. In agreement with previously published studies of yeast Pol II, immunoprecipitated Arabidopsis Pol II catalyzes the synthesis of alpha ³²P-CTP- labeled RNA extension products of up to 32 nt in length using the scaffold template (Figure 3C, lane 6). As expected, this Pol II-mediated activity is inhibited by the fungal inhibitor, alpha-amanitin added at a concentration of 5 ug/ ml (Figure 3C, lane 7). Using immunoprecipitated Pol IV-RDR2, very abundant reaction products that were 12-16 nt in size and weaker, but distinctive, longer RNA extension products up to 32 nt in size were detected (lanes 3 and 4, respectively). All of these reaction products were insensitive to alpha-amanitin. Notably, multiple amino acids known to coordinate the binding of alpha amanitin to Pol II are substituted or absent in Pol IV (Supplemental Figure 2). To try to distinguish Pol IV transcripts from possible RDR2 transcripts, reactions were also conducted using immunoprecipitated Pol IV assembled using NRPD1^{DDD-AAA}, the largest subunit whose Metal A site is mutated so as to render the catalytic center inactive. Using mutant Pol IV fractions (Figure 3C, lane 5), the 12-16 nt RNA products were still abundantly produced, but the longer RNA products analogous to the Pol II extension products were absent. These results suggest that the longest RNA extension products are

Pol IV transcripts whereas the 12-16 nt products, which are absent in Pol II-containing reactions, are potential RDR2 transcripts.

To investigate the template requirements for the activities detected in Figure 3C, the tripartite scaffold template was dissected and its various components were tested in Pol IV-RDR2 or Pol II transcription reactions (Figure 3D). The full tripartite template (lanes 3 and 4) yielded both 12-16 nt and longer products, consistent with the previous results (see Figure 3C, lanes 3 and 6).

Assays performed with the annealed template and non-template DNA oligos (dsDNA) yielded long transcription products (Figure 3D, lanes 5-7), but not the highly abundant 12-16 nt RNA products (compare to lane 3), suggesting that the latter transcripts require the presence of the 16 nt RNA oligonucleotide that is used to generate the tripartite scaffold.

Assays performed with the annealed template DNA and RNA strands (RNA-DNA hybrid, lanes 8-10), but missing the non-template DNA oligo yielded products similar to those obtained with the tripartite scaffold, including the highly abundant 12-16 nt products, suggesting that the non-template DNA strand is not essential.

Transcription using the template DNA strand alone (lanes 11-13) yielded relatively long products with both Pol IV-RDR2 and Pol II, but the 12-16nt RNA products were again absent. Using the non-template DNA oligo only, extremely weak transcription products were observed (lanes 14-16), suggesting that this oligo may be too short to serve as an effective template.

Lastly, using the RNA oligo alone, abundant 12-16nt RNA products were obtained, but only for the Pol IV-RDR2 reactions (lane 18); no such products were

generated using Pol II (lane 19). We conclude that the 12-16 nt products are generated by RDR2 using the 16 nt RNA oligo as the template, consistent with these products being insensitive to mutation of the Pol IV active site (refer to Figure 3C, lane 5). As a test of this hypothesis, we crossed *NRPD1-FLAG* and *NRPD1^{DDD-AAA}-FLAG* transgenic lines with the *rdr2-1* mutant and identified *rdr2-1* homozygous mutants bearing the Pol IV transgenes by genotyping F2 families. Immunoprecipitation of NRPD1-FLAG and NRPD1^{DDD-AAA}-FLAG proteins confirms the absence of RDR2 in these genetic backgrounds (Figure 4A). Following IP of Pol IV from these plants, 12-16 nt transcription products were no longer produced *in vitro* using the full tripartite template that included the RNA oligo (Figure 4B).

Immunopurified Arabidopsis RDR6 transiently expressed in tobacco has been demonstrated to transcribe not only ssRNA but ssDNA templates *in vitro* (Curaba and Chen, 2008). To try to distinguish between potential Pol IV and RDR2 *in vitro* ssDNA transcription activities (refer to Figure 3D), reactions were conducted using immunoprecipitated Pol IV, Pol IV mutant, Pol IV (*rdr2-1*) and Pol IV mutant (*rdr2-1*) protein samples with a 76 nt ssDNA template (Figure 4C). ³²P-GTP labeled products were observed in Pol IV-RDR2 and Pol IV mutant-RDR2 IP samples (Figure 4C, lanes 2 and 3), whereas no labeled product was observed in Pol IV and Pol IV mutant IP samples in the *rdr2-1* mutant background (Figure 4C, lanes 4 and 5). This suggests that RDR2 is responsible for the observed ssDNA transcription activity.

Conclusions

Our results show that Pol IV and RDR2 associate *in vivo* and can be isolated as a complex that is transcriptionally active *in vitro*. The fact that two different RNA polymerases are present in the same reaction complicates the analyses. However, tripartite scaffolds resembling paused transcription bubbles are utilized by Pol IV-RDR2 to program the production of extension products that are insensitive to alpha amanitin but that require the Metal A motif of the polymerase active site, indicating that these are Pol IV transcripts. Compared to Pol II, this extension activity by Pol IV is very weak, perhaps helping explain the inability by several groups, including ours, to detect Pol IV transcripts in previous biochemical assays. By contrast, strong RDR2 activity on ssRNA and ssDNA templates is detected in immunoprecipitated Pol IV-RDR2 fractions. Our results suggest that low abundance Pol IV transcripts generated using DNA templates might be acted upon by RDR2, thereby generating and amplifying the RNA precursors that are subsequently diced into 24 nt siRNAs. The physical and mechanistic coupling of these activities can account for the fact that Pol IV and RDR2 are both required for the biogenesis of the vast majority of 24 nt siRNAs and for the monopoly enjoyed by RDR2 in the 24 nt siRNA pathway, to the exclusion of the other five RDRs encoded by the Arabidopsis genome.

Acknowledgements

We thank Mike Dyer and the greenhouse staff for expert plant care and members of the Pikaard lab for helpful criticism. This research was supported by National Institutes of Health grant GM077590. The Pontes laboratory is supported by a grant from

the Edward Mallinckrodt Foundation. The authors have declared that no competing interests exist.

Materials and Methods

Plant materials

Arabidopsis thaliana mutant line *nRPD1-3* has been described previously (Onodera et al., 2005). The *rdr2-1* mutant line was obtained from Jim Carrington. Transgenic lines *NRPD1-FLAG (nRPD1-3)*, *NRPE1-FLAG (nrpe1-11)*, *DCL3-FLAG (dcl3-1)*, *NRPB2-FLAG (nrpb2-1)* and *NRPD1^{DDD-AAA}-FLAG (nRPD1-3)* have been previously described (Haag et al., 2009; Onodera et al., 2008; Pontes et al., 2006).

Generation of transgenic lines and crosses

The full-length RDR2 genomic sequence, including 525 bp upstream of the translation start site, was amplified by PCR from *Arabidopsis thaliana* (ecotype Col-0) genomic DNA using gRDR2-F and gRDR2-R primers (see Supplemental Table 1) and Pfu Ultra (Stratagene). The PCR product was gel purified and cloned into the pENTR-TOPO S/D vector (Invitrogen) and confirmed by DNA sequencing. The pENTR-RDR2 full-length genomic clone was recombined into pEarleyGate 301 with LR Clonase (Invitrogen) to add a C-terminal HA epitope tag in lieu of the normal stop codon. Resulting plasmids were transformed into *Agrobacterium tumefaciens* strain GV3101 and wild-type (ecotype Col-0) plants were transformed using the floral dip method. Seeds of dipped plants were sown and transformants were selected by spraying seedlings with

BASTA herbicide. BASTA-resistant primary transformants (T1 generation plants) were crossed with *rdr2-1* homozygous mutant plants. Heterozygous *rdr2-1* mutant F1 individuals bearing the genomic *RDR2-HA* transgene were selected by PCR genotyping (LBb1 and RDR2 down-R / RDR2 up-F and RDR2 down-R / RDR2-HA-F and HA-R). Heterozygous F1 plants were selfed and PCR genotyping repeated on the resulting F2 generation to select homozygous *rdr2-1* mutant plants bearing the *RDR2-HA* transgene.

The *NRPD1-FLAG (nrpd1-3, rdr2-1)* transgenic line was the product of a *NRPD1-FLAG (nrpd1-3)* and *rdr2-1* cross. PCR genotyping was used to confirm the presence of the *NRPD1-FLAG* transgene (NRPD1 FLAG-F and FLAG-R) and the homozygous state of the *nrpd1-3* (LBa1 and NRPD1 down-F / NRPD1 up-F and NRPD1 down-R) and *rdr2-1* (LBb1 and RDR2 down-R / RDR2 up-F and RDR2 down-R) mutations. Identical methods and PCR primer sets were utilized to generate and genotype the *NRPD1^{DDD-AAA}-FLAG (nrpd1-3, rdr2-1)* transgenic line.

DNA methylation analysis

The AtSN1 DNA methylation assay involving PCR amplification of undigested or *Hae*III-digested genomic DNA was performed as previously described (Herr et al., 2005).

Antibodies

Affinity purified anti-NRPD1 and anti-NRPD2 have been described previously (Onodera et al., 2005; Ream et al., 2009). Anti-FLAG M2-HRP and anti-HA are commercially available (Sigma). Anti-RDR2 was raised against bacterially expressed

6xHis-RDR2-C (amino acids 786-1133) in rabbit (Sigma Genosys). The cloned RDR2 cDNA had a conservative V1106I substitution.

Affinity purification of RDR2 antibody

About 2 mg of 6xHis-RDR2-C protein was separated by SDS-PAGE and transferred to PVDF membrane using standard protocols. After a brief wash in TBST, the membrane was stained with Ponceau S and the region corresponding to 6xHis-RDR2-C was excised and completely destained in several exchanges of TBST over a 10 min period. The membrane was then blocked in TBST+ 5% milk for 1 hr followed by incubation with 2 mL of crude RDR2 antisera and 8 mL of TBST+ 5% milk on an orbital shaker at 4 °C overnight. Membranes were washed in TBST and cut into small strips 1 cm x 0.5 cm that were transferred to 2 mL microcentrifuge tubes. Membrane-bound antibody was eluted in 1 mL of 100 mM glycine, pH 2.5 (enough to cover the membrane strips) and the tubes were mixed thoroughly. The solution containing the eluted antibody was removed and added to a new tube containing 100 uL of 1 M Tris, pH 8.0. 1 volume of glycerol was added to a final concentration of 50%. Antibody was stored at -20 °C until needed.

Immunoprecipitation

Frozen leaf tissue (4.0g) was ground in mortar and pestle and protein extracted as in (Pontes et al., 2006). Supernatant was incubated with 35 uL anti-FLAG-M2 or anti-HA resin (Sigma) for 2 hours to overnight at 4 °C with rotation. Resin was washed two times with extraction buffer supplemented with 0.5% NP-40.

Immunoblotting

Washed immunoprecipitates were eluted from the resin with two bed volumes of 2x SDS sample buffer and boiled 5 min. Protein samples were run on 7.5% Tris-glycine gels by SDS-PAGE and transferred to nitrocellulose or PVDF membrane. Antibodies were diluted in TBST + 5% (w/v) nonfat dried milk (Schnucks) as follows: 1:250 anti-NRPD1, 1:500 NRPD2, 1:250 anti-RDR2, 1:3,000 anti-HA and 1:2,000 anti-FLAG-HRP. 1:5,000 to 1:10,000 anti-rabbit-HRP (Amersham) was used as secondary antibody. ECL Plus (GE Healthcare) was used for chemiluminescent detection of proteins. Membranes were stripped with 1% SDS, 25 mM glycine, pH 2.0 and re-equilibrated with TBST prior to subsequent blocking and immunoblotting.

In vitro transcription reactions

Each transcription reaction used the immunoprecipitate from 4.0 g leaf tissue prepared as described above. Washed immunoprecipitates were washed two additional times with CB100 buffer (100mM KCl, 25mM HEPES, pH7.9, 20% glycerol, 0.1 mM EDTA, 0.5 mM DTT and 1 mM PMSF). In vitro transcription reactions were performed essentially as (Kuhn et al., 2007). Washed immunoprecipitate still bound to the resin was resuspended to 50 uL total volume with CB100 buffer and supplemented with 50 uL 2x transcription reaction buffer (120 mM ammonium acetate, 40 mM HEPES, pH 7.6, 16 mM magnesium sulfate, 20 uM zinc sulfate, 20% glycerol, 0.16 U/uL RNaseOUT and 20 mM DTT with 2 mM ATP, 2 mM UTP, 2 mM GTP, 0.08 mM CTP, 0.2 mCi/mL alpha 32P-CTP and 4 pmol oligo template). Only one-eighth of the Pol II immunoprecipitate

was used to compensate for the lower Pol IV protein levels. In vitro transcription reactions were incubated at RT for 1.5 hrs on an orbital shaker with occasional tapping of the tubes. Reactions were stopped with the addition of 80ug RNA-Grade Proteinase K (Invitrogen) and incubated at 65 °C for 15 min followed by 3 min at 95 °C followed by phenol:chloroform extraction and precipitation with 1/10 volume 3M sodium acetate, pH 5.2, 20 ug glycogen and 2 volumes isopropanol. Precipitated RNA was then resuspended in 5 uL 1x RNA loading buffer, incubated at 80 °C for 5 min and loaded on a 15% polyacrylamide sequencing gel containing 8M urea for gel electrophoresis. Gels were transferred onto Whatman paper and dried under vacuum for 2 hrs at 80 °C prior to phosphorimager or film exposure.

The RNA extension assay utilized oligos preannealed at a final concentration of 10 uM each in 1x PNK buffer (NEB) and 50 mM NaCl for 2 min in a 95 °C water bath that was then removed from the flame and allowed to return to room temperature. Annealed oligos were stored at -20 °C. The RNA strand (5'-UGCAUAAAGACCAGGG-3'), DNA template (5'-CAGTCTGACTGTGTACGCCTGGTCCGACTCG-3') and DNA nontemplate (5'-CACACAGTCAGACTG-3') oligos were ordered from IDT.

The ssDNA transcription assay utilized a 76 nt ssDNA oligo (5'CCCTCTCCACTCCTCTCCTATTCCTATATACTCTACTCATCCCTCATAACCCACTCATCCCCACTATCCCTACTC-3') ordered from IDT. The reaction and analysis was performed as the RNA extension assay, except 0.2 miC/mL alpha 32P-GTP was used as label with 2 mM ATP, 2 mM UTP, 2 mM CTP, and 0.08 mM GTP.

Immunofluorescence

Interphase nuclei were isolated as described previously (Jasencakova et al., 2000). Upon 4% paraformaldehyde post-fixation, the nuclei were incubated overnight at 4°C with primary antibodies for RDR2 (1:100) and anti-FLAG (1:200, Sigma). Secondary antibodies anti-rabbit Alexa 488 (Invitrogen) and anti-mouse Alexa 594 were diluted at 1:500 in PBS and incubated for 3 hrs at 37 °C. DNA was counterstained with 1 µg/ml DAPI in Prolong Gold mounting medium (Invitrogen).

Microscopy and Imaging

The preparations were inspected with a Nikon Eclipse E800i epifluorescence microscope equipped with a Photometrics Coolsnap ES Mono digital camera. Images were acquired by the Phylum software and pseudocolored and merged in Adobe Photoshop.

References

- Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell*, 123(7), 1279-1291.
- Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007) CPD damage recognition by transcribing RNA polymerase II. *Science*, 315(5813), 859-862.
- Colmenares, S.U., Buker, S.M., Buhler, M., Dlakic, M., and Moazed, D. (2007) Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Mol Cell*, 27(3), 449-461.
- Curaba, J., and Chen, X. (2008) Biochemical activities of Arabidopsis RNA-dependent RNA polymerase 6. *J Biol Chem*, 283(6), 3059-3066.
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*, 101(5), 543-553.
- Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte, D., Jr., Pang, K., Brownell, D.R., Harding, S., Mitani, S., Ruvkun, G., Yates, J.R., 3rd, and Mello, C.C. (2006) Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell*, 124(2), 343-354.
- Haag, J.R., Pontes, O., and Pikaard, C.S. (2009) Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS One*, 4(1), e4110.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science*, 308(5718), 118-120.
- Huang, L., Jones, A.M., Searle, I., Patel, K., Vogler, H., Hubner, N.C., and Baulcombe, D.C. (2009) An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol*, 16(1), 91-93.
- Jasencakova, Z., Meister, A., Walter, J., Turner, B.M., and Schubert, I. (2000) Histone H4 acetylation of euchromatin and heterochromatin is cell cycle dependent and correlated with replication rather than with transcription. *Plant Cell*, 12(11), 2087-2100.
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol*, 5(3), e57.
- Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007) Functional architecture of RNA polymerase I. *Cell*, 131(7), 1260-1272.
- Lee, S.R., and Collins, K. (2007) Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat Struct Mol Biol*, 14(7), 604-610.
- Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009) RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol*, 21(3), 367-376.
- Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jouette, D., Lacombe, A.M., Nikic, S., Picault, N., Remoue, K., Sanial, M., Vo, T.A., and

- Vaucheret, H. (2000) Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*, 101(5), 533-542.
- Muangsan, N., Beclin, C., Vaucheret, H., and Robertson, D. (2004) Geminivirus VIGS of endogenous genes requires SGS2/SDE1 and SGS3 and defines a new branch in the genetic pathway for silencing in plants. *Plant J*, 38(6), 1004-1014.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, 120(5), 613-622.
- Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., Ream, T., Ito, Y., and Pikaard, C.S. (2008) Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. *Genetics*, 180(1), 207-218.
- Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev*, 18(19), 2368-2379.
- Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell*, 126(1), 79-92.
- Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. (2009) Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell*, 33(2), 192-203.
- Sosunov, V., Zorov, S., Sosunova, E., Nikolaev, A., Zakeyeva, I., Bass, I., Goldfarb, A., Nikiforov, V., Severinov, K., and Mustaev, A. (2005) The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. *Nucleic Acids Res*, 33(13), 4202-4211.
- Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gascioli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell*, 16(1), 69-79.
- Werner, F., and Weinzierl, R.O. (2002) A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell*, 10(3), 635-646.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol*, 2(5), E104.
- Zaychikov, E., Martin, E., Denissova, L., Kozlov, M., Markovtsov, V., Kashlev, M., Heumann, H., Nikiforov, V., Goldfarb, A., and Mustaev, A. (1996) Mapping of catalytic residues in the RNA polymerase active center. *Science*, 273(5271), 107-109.

Figure Legends

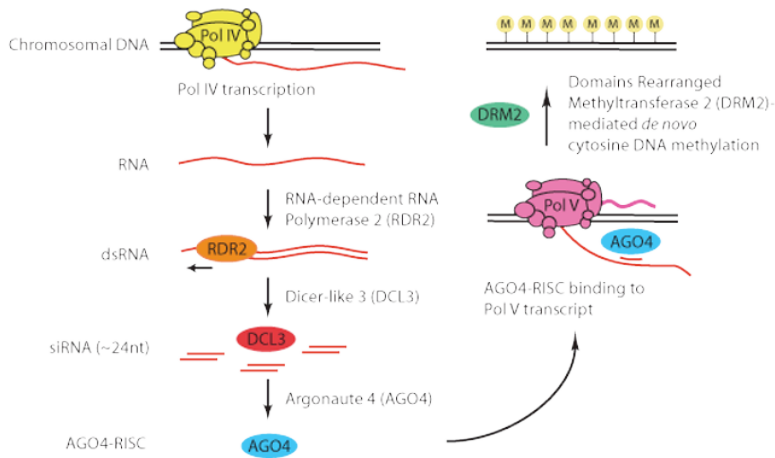
Figure 1. Pol IV and RDR2 interact *in vivo*. (A) Model of RNA-directed DNA methylation pathway in *Arabidopsis thaliana*. (B) LC-MS/MS RDR2 peptide coverage in affinity purified Pol IV. Identified peptides are highlighted in yellow; overlapping peptides are highlighted in green. (C) RDR2-HA transgene rescues 5S siRNA (siR1003) in *rdr2-1* mutant as analyzed by small RNA Northern blot. miR173 and ethidium bromide-stained rRNA are shown as loading controls. (D) RDR2 co-immunoprecipitates with NRPD1-FLAG (lane 3) but not NRPE1-FLAG, DCL3-FLAG or RDR6-FLAG (lanes 4-6) affinity purified proteins demonstrated by Western blot using a native RDR2 antibody. FLAG-tagged proteins were confirmed to be affinity purified by anti-FLAG Western detection. (E) NRPD1 and NRPD2 co-immunoprecipitate with RDR2-HA in a reciprocal IP using native antibodies for Western detection.

Figure 2. Pol IV and RDR2 interaction is independent of Pol IV transcripts. (A) RDR2 co-immunoprecipitates with NRPD1-FLAG and NRPD1^{DDD-AAA}-FLAG by Western blot detection. (B) RDR2 co-immunoprecipitates with RNaseA treated NRPD1-FLAG by Western blot detection. (C) Pol IV and RDR2 co-localize in the nucleoplasm of a subset of *Arabidopsis* interphase nuclei.

Figure 3. Pol IV displays DNA-dependent RNA polymerase activity. (A) Model of a Pol II open transcription bubble modeled after Gnatt et al, 2001. (B) Oligo RNA extension template that mimics a Pol II open transcription bubble used for *in vitro* activity assays modeled after Kuhn et al, 2007. (C and D) Phosphorimages of dried denaturing polyacrylamide gels containing *in vitro* activity assays programmed by affinity purified Pol II, Pol IV and Pol IV ASM complexes supplemented with transcription buffer, template, $\alpha^{32}\text{P}$ -CTP label and a full complement of unlabeled NTPs. (C) *In vitro* reactions used the full tripartite RNA extension template illustrated in (B). (D) *In vitro* reactions using dissected components of the tripartite RNA extension template. Affinity purified Pol II and Pol IV complexes incubated with the full tripartite template (lanes 3 and 4), dsDNA (lanes 5-7), RNA-DNA hybrid (lanes 8-10), DNA template strand (lanes 11-13), DNA nontemplate strand (lanes 14-16) and RNA strand (lanes 17-20).

Figure 4. RDR2 transcribes single-stranded RNA and DNA. (A) Co-IP and Western blot analysis of NRPD1-FLAG and NRPD1 ASM-FLAG transgenic lines in *nrpd1-3* background as well as *nrpd1-3, rdr2-1* double mutant background. (B) *In vitro* activity assays with full tripartite oligo template analyzed by PAGE. (C) *In vitro* activity assays with a 76nt single-stranded DNA template analyzed by PAGE.

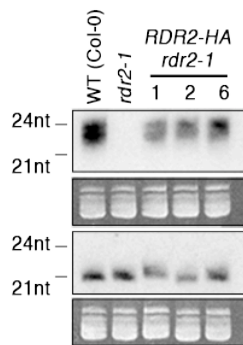
A. RNA-directed DNA methylation pathway



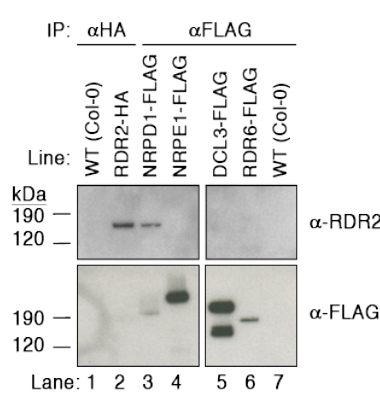
B. RDR2 peptides detected in affinity purified Pol IV

MVSETTNRSTVVKISNVQPTIVADELLRFLLEHLHGEDTVFALEIPIPTRDNWKPRDFARVQFTTLEVKSRQALLS
 SQSKLLFKTHNLRLEAYDDIIPRPVDPKRLDDIVLTVGFPESDEKRFCALEKWDGVRWCWILTEKRRVEFWW
 ESGDCYKIEVRFEDIETLSCCVNGDASEIDAFLLKLYGPKVFKRVTVHIATKFKSDRYRCKEDPDMWIRT
 TDFSGSKSIGTSTFCLEVHNGSTMLDIFSGLPYYREDTSLTYVDGKTFASAAQIVPLLNAAILGLEFPYEIL
 FQLNALVHAQKISLFAASDMELIKILRGMSLETALVILKLLHQSSICYDPVFFVKTQMOSVVKMKHSPASAY
 KRLTEQNIMSCQRAYVTPSKIYLLGPELETANYVVKNF~~AEHVSDFMR~~VTFVEEDWSKLPANALSVNSKEGYFVK
 PSRTNIYNRVLSILGEGITVGPKRFEFLAFSASQLRGNVVMFASNEKVKAEIDIREWMGCFRKIRSISKCAARM
 GQLFSASRQTLIVRAQDVEQIPDIEVTTDGADYCFSDGIGKISLAPAKQVAQKCGLSHVPSAFQIRYGGYKGI
 AVDRSSFRKLSLRDSMLKFDNSNRMLNVTWTESMPCFLNREIICLLSTLGIEDAMFEAMQAVHLSMLGNMLED
 RDAALNVLQKLSGENSKNLLVKKMLLQGYAPSSSEPYLSMMLRVHHESQLSELKSRICILVPKGRILIGCMDEMI
 LEYGQVYVRVTLTKAELKSRDQSYFRKIDETS~~SVVIGK~~VVVVTKNPNCLHPGDIRVLDALYEVHFEEKGYLDCIIF
 PQKGRPHPNESGGDLGDQFPVSWDEKIIPSEMDPPMDYAGSRPRLMDHDVTLBEIHKFFVDYMI~~SDTLGVI~~
 STAHLVHADRDPEKARSQKCLELANLHSAVDFAKTGAPAE~~MPYALKPREFPDFLER~~FEKPTYISESVFGKLYR
 AVKSSLAQRKPEAESD~~TVAYDVTLEEAGFESFIETAKAHRD~~MYEK~~LTSLMIYYGA~~NEEBEILTGLKTKEMY
 LARDNRRYGDMDRITLSVKDLHKEAMGWFEKSC~~EDEQQKKLASAWY~~YVYTPNHRDEKLTFLSFPPIVGDV
 LD~~IKAENAQRQ~~SVBEKTSGLVSI

C. HA-tagged RDR2 is fully functional



D. RDR2 co-IPs with FLAG-tagged Pol IV



E. Pol IV co-IPs with HA-tagged RDR2

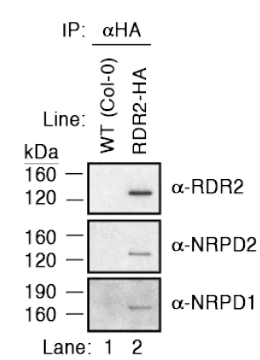
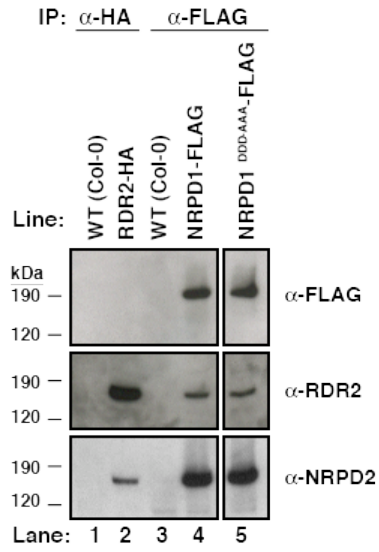
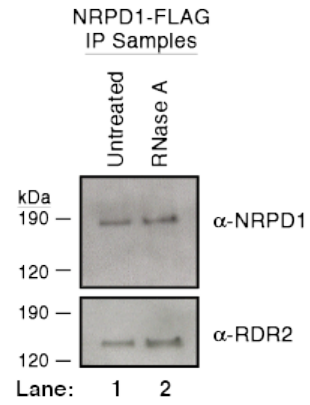


Figure 1

A. Pol IV active site mutant association with RDR2



B. RNaseA treatment



C. Pol IV and RDR2 immunolocalization

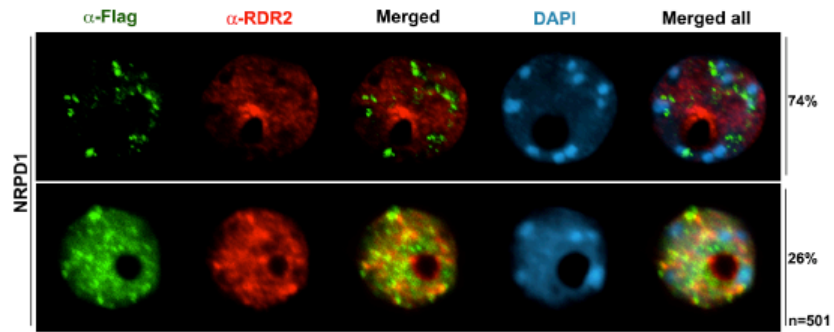
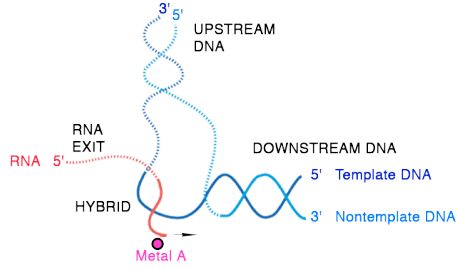
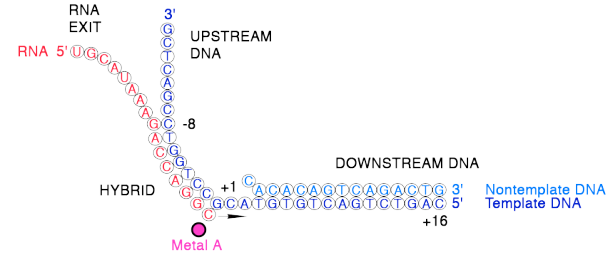


Figure 2

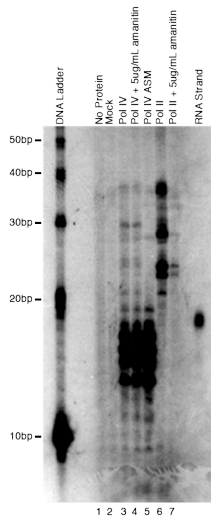
A. Model of Pol II Open Transcription Bubble



B. RNA Extension Template



C. RNA Extension Assay



D. Dissection of RNA Extension Assay Template

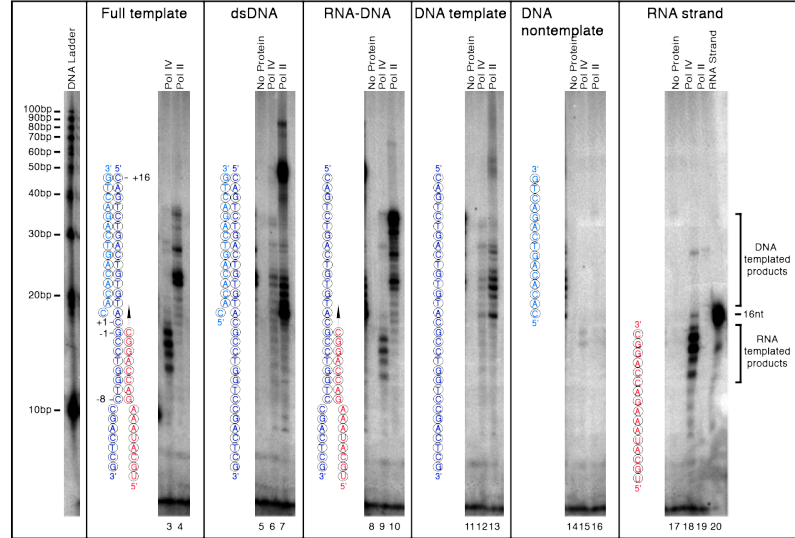
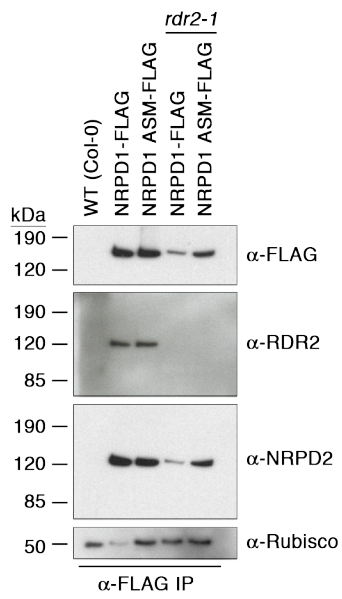
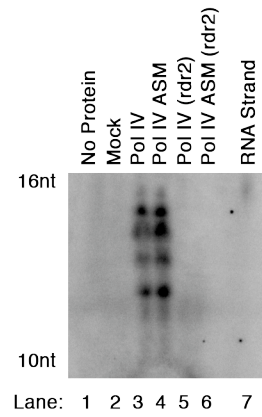


Figure 3

A. Co-IP and Western Analysis



B. RNA Extension Assay



C. Transcription with ssDNA Template

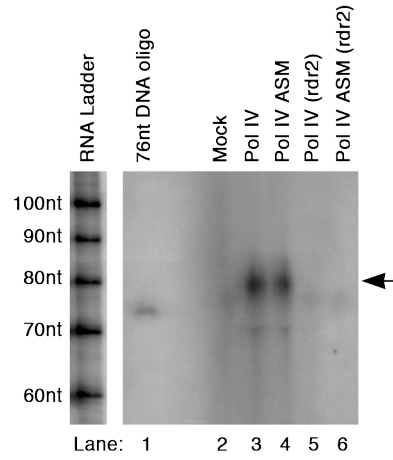


Figure 4

Supplemental Figures

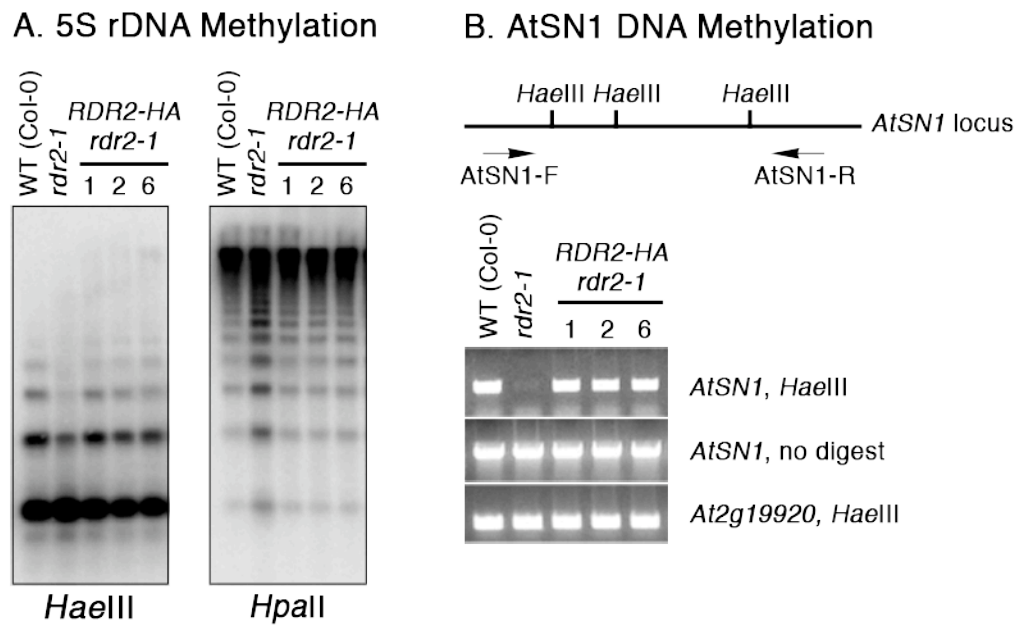
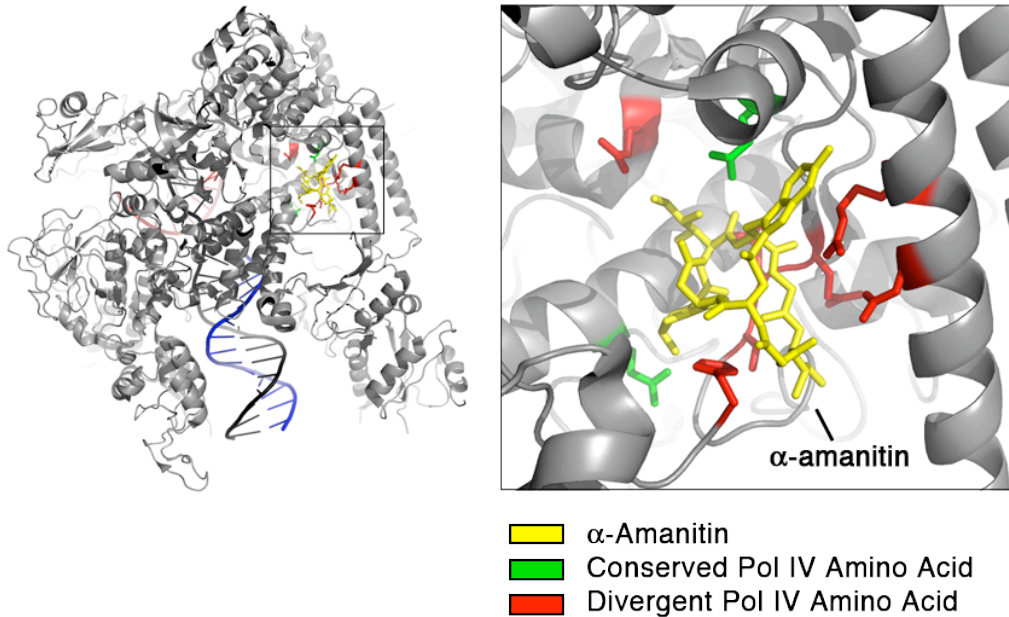


Figure S1. *RDR2* HA-tagged genomic transgene rescues *rdr2-1* mutant. (A) Southern blot analysis of 5S rDNA methylation with *Hae*III and *Hpa*II digested genomic DNA. (B) DNA methylation analysis at the *AtSN1* retrotransposon by chop-PCR with *Hae*III digested genomic DNA.

A. Pol II Amino Acids Bound by α -Amanitin



B. Conservation of H-Bond Contacts

Yeast Pol II	Pol IV	Pol V
His1085*	Ser898	Pro849
Glu822*	Asp818	Glu763
Ser 769	Asn751	Leu706
Gln763* (Rpb2)	His711	His711
Gln760*	Gln742	Gln697
Gln768*	Leu750	Leu705
Arg726†	Ala703	-
Gln767*	Leu741	Leu696
Asn723*	Ala700	-

C. Conservation of Hydrophobic Contacts

Yeast Pol II	Pol IV	Pol V
Gly819†	Tyr815	Ala760
Gly820*	Ser816	Ala761
Gly772*	Val754	Lys709
Val719	Leu696	-
Ile756†	Gly738	Thr693
Ala759	Val741	Val696

* = invariant in Pol I, II and III in Arabidopsis
 † = conserved in Pol II and III in Arabidopsis

Figure S2. Pol IV and Pol V are predicted to be alpha-amanitin insensitive. (A)

Crystal structure of alpha-amanitin bound to yeast Pol II largest and second-largest subunits as determined by Bruekner *et al* (2008) modeled in PyMOL. Enlargement focuses on Pol II alpha-amanitin binding pocket with alpha-amanitin in yellow. Pol II amino acids that form hydrogen bonds with alpha-amanitin are colored green if conserved in Pol IV and red if divergent in Pol IV. (B) Summary of hydrogen bond (C) and hydrophobic contacts between yeast Pol II and alpha-amanitin as determined by

Bruekner et al (2008) with corresponding Pol IV and Pol V amino acids from multiple protein sequence alignments (Haag et al, 2009). Conserved amino acids are highlighted green for hydrogen bond contacts and yellow for hydrophobic contacts. The “*” symbol denotes amino acids that are invariant in Arabidopsis Pol I, II and III, whereas the “‡” symbol denotes amino acids that are conserved in Arabidopsis Pol II and III.

CHAPTER 6

FUNCTIONAL ANALYSIS OF NRPD1 AND NRPE1 C-TERMINAL DOMAINS
REQUIRED FOR RNA-DIRECTED DNA METHYLATION

A manuscript in preparation

My contributions to this work:

I designed and performed all experiments unless otherwise noted. In particular, I cloned and generated a series of NRPD1 and NRPE1 C-terminal truncation and internal CTD deletion constructs to identify domains required for *in vivo* complementation. This work not only identified the DeCL-like domain as being required for Pol IV and Pol V function, but also revealed that the QS-rich, ten 16 aa repeats and majority of WG motifs are dispensable. Transgenic lines generated in this work and re-analysis of a previously generated NRPE1 CTD deletion line by another lab provide a more nuanced appraisal of the WG motif requirements as being important but not essential for Pol V function. To complement the loss-of-function analyses, NRPD1 and NRPE1 C-terminal domains were over-expressed in the wild type background and found to dominantly suppress RNA-directed DNA methylation. Ek Han Tan cloned the pENTR-NRPE1 aa 1243-1842 cDNA and Junchen Gu performed Western blot, DNA methylation and transcript analysis of the FLAG-NRPE1 and NRPD1 CTD over-expression domains under my supervision as a rotation student. Olga Pontes performed the localization analysis. I wrote the manuscript, with the assistance of Craig Pikaard.

Functional analysis of NRPD1 and NRPE1 C-terminal domains required for RNA-directed DNA methylation

Jeremy R. Haag, Junchen Gu, Olga Pontes, Ek Han Tan and Craig S. Pikaard¹

Biology Department, Washington University, St. Louis, MO, 63130

¹ Corresponding author: pikaard@biology2.wustl.edu
phone 314-935-7569, fax 314-935-4432

Running title: NRPD1 and NRPE1 C-terminal domains required for RdDM

Abstract

Plant-specific RNA Polymerases IV and V are specialized forms of RNA Polymerase II and are involved in the RNA-directed DNA methylation (RdDM) pathway. The Pol IV and Pol V largest subunits, NRPD1 and NRPE1, respectively, retain the conserved DNA-dependent RNA polymerase domains A to H present in all multisubunit RNA polymerases, but lack the C-terminal heptad repeats of the Pol II largest subunit. Instead, Arabidopsis NRPD1 and NRPE1 contain unique C-terminal extensions with domains that are conserved to varying degrees among diverse plant species. Complementation assays indicate that the Defective Chloroplast and Leaves-like (DeCL-like) domain is required for full function of both NRPD1 and NRPE1. The QS-rich domain and the ten 16 aa repeats present in the NRPE1 CTD are dispensable for function, as are the majority of WG motifs implicated in AGO4 interactions. Over-expression of the NRPE1 CTD domains in wild type plants has a gain-of-function phenotype resulting in dominant suppression of RdDM.

Introduction

DNA-dependent RNA Polymerases (DdRPs) catalyze the production of RNA from a DNA template. Bacterial DdRP complexes have 5 core subunits, whereas eukaryotic DdRP complexes are more complex, with 12 to 17 core subunits. Pol I transcribes 45S rRNA, Pol II transcribes mRNA as well as most micro RNA precursors, and Pol III transcribes 5S rRNA and tRNAs (Grummt, 2003; Schramm and Hernandez, 2002; Woychik and Hampsey, 2002). Plants are unique in that they encode two additional DdRP complexes named Pol IV and Pol V that produce noncoding RNAs (Matzke et al., 2009).

Pol IV and Pol V are members of the RNA-directed DNA methylation (RdDM) pathway, which is important for the silencing of retrotransposons and endogenous repeats. Pol IV transcripts are precursors for small RNA biogenesis in a process that requires RNA-DEPENDENT RNA POLYMERASE2 (RDR2) and DICER-LIKE3 (DCL3) (Herr et al., 2005; Onodera et al., 2005; Pontes et al., 2006) (Chapter 5). The siRNAs associate with ARGONAUTE4 (AGO4) in a RNA-induced silencing complex (RISC) that is required for DNA methylation and the generation of secondary siRNAs at some loci (Qi et al., 2006). Pol V transcripts are hypothesized to help recruit the silencing machinery to specific chromosomal loci for DNA methylation and chromatin modifications by serving as siRNA interaction scaffolds (Wierzbicki et al., 2008; Wierzbicki et al., 2009).

The Pol II largest subunit, Rpb1, or NRPB1 in plants, contains the DdRP conserved domains A-H that are conserved in all multisubunit RNA polymerase largest subunits from bacteria to eukaryotes followed by a unique C-terminal domain (CTD)

extension (Jokerst et al., 1989). The Rpb1 CTD is composed of a heptad repeat whose consensus sequence is YSPTSPS (Allison et al., 1985). This sequence is conserved among the Pol II largest subunits of animals, plants and fungi (Stiller and Hall, 2002). The heptad repeats are a target of post-transcriptional modifications and protein-protein interactions that control Pol II initiation, elongation, termination and pre-mRNA splicing events (Cho et al., 1997; Cramer et al., 1997; Ho et al., 1998; Liao et al., 1991; McCracken et al., 1997; Nonet and Young, 1989; Otero et al., 1999; Riedl and Egly, 2000; Yamamoto et al., 2001). The total number of heptad repeats varies by species, as does the minimum number of heptad repeats required for viability (Corden, 1990). The plant-specific Pol IV and Pol V largest subunits, NRPD1 and NRPE1, respectively, are evolved from Pol II NRPB1 (Luo and Hall, 2007). They contain the core DdRP conserved domains but lack the Pol II heptad repeats at their C-termini. *Arabidopsis thaliana* NRPD1 has a CTD of 179 amino acids (aa) whereas the NRPE1 is ~370 aa, twice the length of the CTD of the Arabidopsis Pol II largest subunit, NRPB1.

The DeCL-like domain is plant-specific and has no known function. The *Arabidopsis thaliana* genome encodes five Defective Chloroplast and Leaves-like (DeCL-like) domain-containing proteins, including NRPD1 and NRPE1. AtDCL (At1g45230) is required for chloroplast rRNA processing and correct ribosome assembly (Bellaoui and Gruissem, 2004; Bellaoui et al., 2003; Keddie et al., 1996). DOMINO1 (At5g62440) is an embryo-defective mutant that is nuclear localized and proposed to be involved in a process essential for nuclear and nucleolar functions (Lahmy et al., 2004). At3g46630 remains uncharacterized but is predicted to localize to the mitochondria (Lahmy et al., 2004).

N-terminal of the NRPE1 DeCL domain is a region consisting of ten imperfect 16 amino acid repeats (aa 1451-1651) rich in WG motifs that also occur flanking the repeats (El-Shami et al., 2007; Pontier et al., 2005). WG motifs have been implicated in the binding of Argonaute proteins (El-Shami et al., 2007; Takimoto et al., 2009; Till et al., 2007) and *in vitro* and *in vivo* experiments suggest that AGO4 can interact with the NRPE1 CTD via these WG motifs (El-Shami et al., 2007; He et al., 2009; Li et al., 2006).

At its extreme C-terminus, Arabidopsis NRPE1 contains a glutamine-serine rich (QS-rich) domain (aa 1851-1976). *Spinacia oleracea* has a short proline-serine rich (PS-rich) domain at this location rather than a QS-rich domain (Pontier et al., 2005).

To address the requirements of the NRPD1 and NRPE1 C-terminal domains for Pol IV and Pol V *in vivo* function, we generated a series of deletion constructs and assayed whether or not they were capable of complementing *nRPD1* and *nrpe1* mutants defective for DNA methylation, small RNA accumulation or transcriptional silencing. My analysis reveals that the DeCL-like domains of NRPD1 and NRPE1 are required for full activity. The NRPE1 QS-rich domain is dispensable, as is the domain consisting of the ten 16 aa repeats. Contrary to a previously published report, the NRPE1 WG motifs are not fully required for Pol V activity, as deletion mutants are capable of partial complementation. Over-expression of the NRPE1 CTD leads to dominant suppression of the RdDM pathway in transformed wild type plants. Collectively, these genetic studies show that the NRPD1 and NRPE1 CTDs play an important role in Pol IV and Pol V function.

Results

NRPD1 and NRPE1 CTDs have conserved domains among diverse plant species

Predicted full-length NRPD1 and NRPE1 sequences from diverse plant species were analyzed to determine the extent of CTD conservation. The DeCL-like domain is detected by the presence of the DFSYRK consensus sequence (Bellaoui and Gruissem, 2004; Bellaoui et al., 2003) and is present in all NRPD1 and NRPE1 proteins, with the exception of the NRPD1 and one of two NRPE1 proteins in *Physcomitrella patens* (Figure S1, S2 and S3). In the context of NRPE1, the DeCL-like domain is typically C-terminal of the 16 aa repeats and WG motifs. The NRPE1 16 aa repeats are imperfect and vary in number and length in different species (Figures S1 and S2). While the WG motifs are often embedded in the repeat sequence, exceptions do occur such as the *Physcomitrella patens*, *Vitis vinifera*, *Oryza sativa* and *Zea mays* NRPE1 proteins (Figures S1 and S2). The number of WG motifs and whether they are predominantly present as WG, GW, GWG or WGW motifs varies by species (Figures S1 and S2). The QS- and PS-rich domains appear unique to Arabidopsis and spinach, respectively, as no equivalent domains were detected in NRPE1 of other plants (Figures S1 and S2).

NRPE1 C-terminal domain deletions

The Arabidopsis NRPE1 CTD can be divided into four domains: a linker region that connects the CTD to the DdRP core, the 16 aa repeat and WG motif-containing domain, the DeCL-like domain and the QS-rich domain. To test for NRPE1 CTD functions, a series of six C-terminal deletion constructs and a full-length control construct were transformed into the *nrpe1* mutant to assay for complementation (Figure 1A). Each of the HA-tagged transgenes is expressed and encodes a protein of the predicted

molecular mass (Figure 1B). NRPE2 co-immunoprecipitates with all of the NRPE1 CTD deletion constructs, even when the entire CTD is deleted, suggesting that the CTD is not required for Pol V subunit assembly (Figure 1B). NRPE1 is typically detected on immunoblots as a doublet regardless of whether the native protein or C-terminal FLAG or HA epitope tagged proteins are detected (Pontes et al., 2006; Pontier et al., 2005; Ream et al., 2009). This banding pattern is observed in each of the C-terminal deletion constructs except for the full CTD deletion construct.

The NRPE1 DeCL-like domain is required for *in vivo* complementation

It has previously been determined that Pol IV and Pol V are required for DNA methylation and silencing of the *AtSN1* retrotransposon locus (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). DNA methylation at the *AtSN1* locus was analyzed by chop-PCR using the methylation sensitive *HaeIII* restriction enzyme (Figure 1C). If the *HaeIII* restriction sites in the *AtSN1* locus are methylated, DNA digestion will not occur and a PCR product will be obtained. If any of the *HaeIII* restriction sites are unmethylated, the DNA will be digested and PCR amplification of the region will fail. PCR amplification of the region was successful in the NRPE1 full-length and NRPE1 Δ 1851-1976 (QS-rich deletion) lines indicating these constructs successfully complement the *nrpe1* mutant and facilitate the methylation of the *HaeIII* sites. The NRPE1 Δ 1736-1976 protein (DeCL-like and QS-rich domain deletions) and remaining CTD deletions in the series fail to rescue *AtSN1* DNA methylation; a PCR product was not obtained, indicating that one or more *HaeIII* sites was susceptible to digestion. RT-PCR analysis demonstrates *AtSN1* transcript repression in the NRPE1 full-length and

NRPE1 Δ 1851-1976 lines and a failure to repress in the NRPE1 Δ 1736-1976 and remaining CTD deletions (Figure 1C). DNA methylation analysis at the 5S rDNA loci supports these results as Southern blot analysis of *Hae*III and *Hpa*II genomic DNA reveals that only the NRPE1 full-length and NRPE1 Δ 1851-1976 lines complement the DNA methylation defect of the *nrpe1* mutant (Figure 1D).

While NRPE1 is not absolutely required for the biogenesis of all siRNAs, *nrpe1* mutants do affect the accumulation of some siRNAs (Mosher et al., 2008). Small RNA Northern blot analysis of AtCopia, 45S rRNA and AtSN1 sequences demonstrates the QS-rich domain is dispensable for complementation but that the DeCL-like domain is required for wild-type levels of siRNA accumulation to occur (Figure 1E).

NRPD1 DeCL-like domain deletion

The Arabidopsis NRPD1 CTD is composed of a DeCL-like domain and a small linker region that connects it to the DdRP core structure. A NRPD1 DeCL-like deletion construct, NRPD1 Δ 1337-1453, as well as the previously published NRPD1 full-length control were transformed into the *nrpd1* mutant to determine if the NRPD1 DeCL-like domain is required for *in vivo* complementation (Figure 2A). The two FLAG-tagged NRPD1 constructs are both expressed at the protein level, and NRPD2 and RDR2 both co-immunoprecipitate with WT or Δ CTD proteins at equivalent levels (Figure 2B). These results suggest the NRPD1 DeCL-like domain is not required for Pol IV complex assembly or for mediation of the Pol IV-RDR2 interaction (Chapter 5).

The NRPD1 DeCL-like domain is required for siRNA biogenesis and transcript silencing but not DNA methylation

At *AtSN1*, the NRPD1 DeCL deletion mutant, NRPD1 Δ 1337-1453, restores DNA methylation to the same levels as the NRPD1 full-length transgene (Figure 2C). Similar results were observed at the 5S rDNA loci by Southern blot analysis of *Hae*III and *Hpa*II digested DNA (Figure 2D).

In contrast to the NRPD1 DeCL domain being dispensable for the restoration of DNA methylation, small RNA Northern blot analysis reveals that the NRPD1 DeCL-like domain is required for the wild-type accumulation of AtCopia, 45S and AtSN1 siRNAs (Figure 2E). Consistent with the failure to produce Pol IV-dependent siRNAs, it is found that the NRPD1 DeCL-like domain is required for suppression of *AtSN1* and *solo LTR* transcripts (Figure 2F).

NRPE1 CTD repeats are dispensable for *in vivo* complementation

Given the functional requirement for the NRPE1 DeCL-like domain, we were unable to conclude the significance of domains N-terminal to this domain using the C-terminal deletion series studied in Figure 1. To address the requirement for sequence elements between the NRPE1 DdRP core and the DeCL-like domain, three additional transgene deletion constructs were engineered and transformed into the *nrpe1* mutant for *in vivo* complementation assays (Figure 3A). NRPE1 Δ 1251-1426 contains a deletion in the linker region and deletes 3 of 18 WG motifs; NRPE1 Δ 1426-1651 deletes the ten 16 aa repeats and 13 of the 18 WG motifs, and NRPE1 Δ 1251-1651 deletes both regions and 16 of the 18 WG motifs.

The three NRPE1 internal CTD deletion lines were analyzed for rescue of DNA methylation at the 5S rDNA loci by Southern blot analysis of *Hae*III and *Hpa*II digested genomic DNA (Figure 3B). Deletion of the linker region (NRPE1 Δ 1251-1426) or the ten 16 aa repeats (NRPE1 Δ 1426-1651) resulted in full rescue of the *nrpe1* mutant. Only when these two regions were deleted together (NRPE1 Δ 1251-1651) was there a failure to fully complement, although DNA methylation levels are still increased relative to the *nrpe1* mutant. DNA methylation at *AtSNI* was also assayed by chop-PCR and similar results were observed with DNA methylation fully restored with the NRPE1 Δ 1251-1426 and NRPE1 Δ 1426-1651 transgenes and only partially with the NRPE1 Δ 1251-1651 transgene (Figure 3C).

In agreement with the *AtSNI* DNA methylation status, *AtSNI* transcription detected by RT-PCR demonstrates that only the NRPE1 Δ 1251-1651 transgenic line continues to express *AtSNI* transcripts, though below *nrpe1* mutant levels (Figure 3D). Unexpectedly, there are no observable defects in siRNA accumulation in any of the three deletion lines (Figure 3E).

The NRPE1 WG motifs are important but not required for NRPE1 function

It has previously been published that the NRPE1 WG motifs are required for *in vivo* complementation of 5S rDNA and *AtSNI* DNA methylation states in the *nrpe1-11* background (El-Shami et al., 2007). The NRPE1 transgene used in the study, NRPE1 Δ SD, had two deletions spanning aa 1411 to 1707 and aa 1875 to 1976. The transgene therefore deleted all ten 16 aa repeats, 16 of the 18 WG motifs and the QS-rich domain (Figure 4A).

Three independent NRPE1 Δ SD lines were compared side-by-side with the NRPE1 Δ 1251-1426, NRPE1 Δ 1426-1651, NRPE1 Δ 1251-1651 and NRPE1 Δ 1251-1976 deletion lines. Contrary to the published results (El-Shami et al., 2007), the NRPE1 Δ SD line does partially rescue DNA methylation at the *AtSN1* (Figure 4B) and 5S rDNA loci (Figure 4C). NRPE1 Δ SD DNA methylation levels are roughly equivalent to the NRPE1 Δ 1251-1651 transgenic line. The two do not display full complementation but they do facilitate significantly more DNA methylation than the *nrpe1* mutant. Transcription from the *AtSN1* and *solo LTR* loci in NRPE1 Δ SD and NRPE1 Δ 1251-1651 lines is partially suppressed (Figure 4D) in agreement with the DNA methylation results, showing increased methylation at these loci. Thus, the WG motifs may be important, but they are not required for NRPE1 to complement an *nrpe1* mutant.

Over-expression of the NRPE1 C-terminal domains dominantly suppresses the RdDM pathway

Having analyzed loss-of-function phenotypes with CTD deletions in the NRPE1 and NRPE1 proteins, we next tested for gain-of-function phenotypes. If the CTDs are a platform for protein-protein interactions, over-expression may titrate away silencing factors required for RdDM function. A YFP over-expression vector encoding NRPE1 aa 1234-1842, referred to as YFP-CTD (Figure 5A), was transformed into wild type Arabidopsis plants. In whole mounted Arabidopsis roots, the protein signal is detected throughout the nucleoplasm, with little to no cytoplasmic localization detected (Figure 5B). *AtSN1* DNA methylation, in ten of twelve independent transgenic lines, is reduced compared to wild type plants (Figure 5C) demonstrating that the transgene is capable of

dominant suppression of RdDM. *AtSN1* transcription is correspondingly activated in the lines that have reduced DNA methylation (Figure 5D). Lack of transgene RNA expression in line 182 (Figure 5D) explains why there is no dominant suppression phenotype in this plant. Because the transgene is expressed in line 172, a post-transcriptional gene silencing mechanism or mutation that prevents the protein from being translated or functioning properly may explain the lack of a dominant negative phenotype in this plant. Similar to *nrpe1* mutants, *AtCopia*, 45S and *AtSN1* siRNA accumulation is reduced in the YFP-CTD transgenic lines (Figure 5E) and these plants also display delayed flowering (Figure S4) similar to *nrpe1* mutants.

In an attempt to narrow down the region(s) capable of inducing dominant suppression of RdDM, three additional NRPE1 constructs were cloned, spanning aa 1426-1651, aa 1426-1851 and aa 1851-1977, in addition to the NRPD1 DeCL domain, aa 1337-1453 (Figure 5A). These cDNAs were recombined into over-expression vectors that add an N-terminal FLAG tag and transformed into wild type *Arabidopsis* plants. Protein blot analysis of immunoprecipitated protein samples confirmed expression of all the transgenes (Figure 5F).

Six independent lines for each transgene were analyzed for dominant suppression of the RdDM pathway. DNA methylation at the *AtSN1* locus was only marginally affected in three of the NRPD1 aa1337-1453 lines (Figure 5G). In contrast, multiple individuals for each of the three NRPE1 CTD over-expression constructs demonstrated significantly reduced *AtSN1* DNA methylation (Figure 5G). Corresponding with the DNA methylation results, transcription of *AtSN1* and *solo LTR* retroelements was activated in the NRPE1 CTD over-expression lines (Figure 5H). Weak expression of

AtSN1 is detected in several of the NRPD1 aa1337-1453 transgenic lines, although *solo LTR* expression does not appear to be activated (Figure 5H).

Discussion

Our results show that the DeCL-like domain is required *in vivo* for both Pol IV and Pol V function. NRPE1 is completely dependent upon this domain for function in the RdDM pathway, while NRPD1 requires the domain for complementation of siRNA biogenesis and suppression of retroelement transcription. Interestingly, DNA methylation is rescued despite deletion of the NRPD1 DeCL-like domain. Over-expression of the NRPD1 DeCL-like domain led to only subtle dominant negative DNA methylation defects, although release of transcriptional silencing was more pronounced, in agreement with the complementation assay results. In addition, the NRPD1 aa 1337-1453 lines displayed leaf curling and smaller plant size (Figure S5) similar to some of the reported phenotypes of plants over-expressing a plastid DeCL-like domain-containing protein, AtDCL (Bellaoui and Gruissem, 2004). The RdDM-defective phenotypes observed in the NRPD1 DeCL-like domain over-expression lines might be due to dominant-negative crosstalk with the three other DeCL-like domain containing proteins in Arabidopsis since *nrpd1* and *nrpe1* mutants lack these morphological phenotypes.

The QS-rich domain and ten 16 aa repeats in the NRPE1 CTD are not required for complementation of an *nrpe1* mutant, but each domain is sufficient to trigger dominant suppression of RdDM when over-expressed. The plants have no apparent morphological defects (data not shown). We suggest that the over-expressed domains either titrate away interacting proteins from the endogenous NRPE1 protein or in some other way interfere

with the function of the RdDM pathway. In agreement with this idea is the observation the YFP-tagged NRPE1 CTD localizes to the nucleus where other members of the RdDM pathway localize (Pontes et al., 2006). Interestingly, YFP-CTD was never observed in the nucleolus-associated Cajal body where siRNA biogenesis and processing are believed to occur (Li et al., 2006; Pontes et al., 2006), unlike the full-length NRPE1, suggesting the DdRP core is required for NRPE1 to localize here.

The NRPD1 Δ 1337-1453 and NRPE1 Δ 1251-1651 phenotypes are noteworthy since there is a breakdown in correlation between DNA methylation and siRNA production. In the case of NRPD1 Δ 1337-1453, DNA methylation is rescued despite the failure to restore siRNA production, and in the case of NRPE1 Δ 1251-1651, siRNA production is rescued despite the failure to restore DNA methylation. Neither restores retroelement transcript suppression. These results suggest siRNA production and DNA methylation are unable to establish a transcriptionally silenced state independent of one another. Building upon this idea, there may be two parallel pathways in plants that converge on the same target that are both required for the establishment of silencing. Perhaps DNA methylation provides an independent check on the siRNA-mediated silencing pathway in plants, and vice versa. At the very least, the results imply that Pol V-directed DNA methylation is important for transcriptional silencing but not Pol V-derived siRNAs and that Pol IV-derived siRNAs are important for transcriptional silencing but not Pol IV-directed DNA methylation.

In disagreement with a previously published report (El-Shami et al., 2007), the majority of NRPE1 WG motifs can be deleted and still largely complement the *nrpe1* mutant (Figure 4). This suggests that the WG motifs are important but not required for

Pol V function. Reports of *in vitro* interaction between bacterially expressed NRPE1 CTD protein and AGO4 in plant extracts (El-Shami et al., 2007; He et al., 2009; Li et al., 2006) have been confirmed (Figure S6) and demonstrate that AGO4 is capable of binding NRPE1 aa 1426-1651 but not a NRPE1 CTD construct that lacks this region. However, if NRPE1 and AGO4 do directly interact via the WG motifs *in vivo*, this interaction is not required for the RdDM pathway to function because the NRPE1 Δ 1426-1651 line fully complements the *nrpe1* mutant. It must be stated that despite repeated efforts, the reported *in vivo* interaction between NRPE1 and AGO4 (Li et al., 2006) cannot be confirmed despite numerous co-IP approaches (Figure S7) and mass spec analysis of both NRPE1 and AGO4 purified samples (Haag, Ream, Pikaard, EMSL, unpublished). Thus, if NRPE1 and AGO4 do interact *in vivo*, it is possibly a weak or transient interaction mediated by AGO4 binding of Pol V transcripts (Wierzbicki et al., 2009) with the WG motifs acting to help stabilize the interaction.

While the NRPD1 and NRPE1 CTDs have little resemblance to the CTD of NRPB1, the Pol IV and Pol V complexes are evolutionarily derived from Pol II (Luo and Hall, 2007; Ream et al., 2009) and like Pol II, Pol IV and Pol V require distinct C-terminal domains for proper function. It is likely that the unique roles of these related polymerases arise from differential use of Pol II-derived small subunits (Ream et al., 2009) and their unique CTD architectures. Whether the CTDs play a role in regulating Pol IV and Pol V transcription or post-transcriptionally process Pol IV and Pol V transcripts is still an open question. The NRPD1 and NRPE1 CTDs are likely to be involved in protein-protein interactions and may be the target of post-translational modifications, like the NRPB1 CTD. Evidence for alternative splicing or post-

translational modification of the NRPE1 CTD is hinted at by the observation that the NRPE1 doublet pattern is lost when the full CTD is deleted (Figure 1B) and the over-expressed NRPE1 QS-rich domain migrates much larger than the predicted 14kD size (Figure 5F). Proteomic analyses to identify protein-protein interactions and post-translational modifications in the NRPD1 and NRPE1 CTDs are currently underway.

Materials and Methods

Plant materials. *Arabidopsis thaliana* mutant lines *nrdp1-3*, *nrdp2* (*nrdp2a-2*, *nrdp2b-1*) and *nrpe1-11* have been described previously (Onodera et al., 2005; Pontier et al., 2005), as have transgenic lines *NRPD1-FLAG* (*nrdp1-3*) and *NRPD1^{DDD-AAA}-FLAG* (*nrdp1-3*) (Haag et al., 2009; Pontes et al., 2006). The NRPE1 Δ SD-FLAG (*nrpe1-11*) transgenic line was kindly provided by Thierry Lagrange.

Cloning, vectors and transgenic lines. The pENTR-NRPE1 full-length genomic sequence with its endogenous promoter (Pontes et al., 2006) was recombined into pEarleyGate301 (Earley et al., 2006) using LR Clonase (Invitrogen) in order to add a C-terminal HA epitope tag in lieu of the normal stop codon. C-terminal domain deletions were obtained by using pENTR-NRPD1 and pENTR-NRPE1 full-length genomic clones with endogenous promoters (Pontes et al., 2006) as the DNA template and reverse primers that truncated the 3' end (Table S1). Pfu Ultra (Stratagene) was used to amplify the sequences. The PCR products were gel purified and cloned into pENTR-TOPO S/D (Invitrogen) before being recombined into pEarleyGate 301 (NRPE1 C-terminal truncations with HA epitope) or pEarleyGate302 (NRPD1 C-terminal truncation with

FLAG epitope). Internal C-terminal domain deletions were obtained by the SLIM method (Chiu et al., 2004) using the pENTR-NRPE1 full-length genomic clone as the DNA template and the appropriate primers (Table S1). Constructs were recombined into pEarleyGate301. CTD over-expression lines were generated by cloning *NRPD1* and *NRPE1* cDNA sequences (Table S1) and recombining into pEarleyGate104 (35S promoter with N-terminal YFP fusion) or pEarleyGate202 (35S promoter with N-terminal FLAG epitope). pEarleyGate plasmids in *Agrobacterium tumefaciens* strain GV3101 were used to transform *Arabidopsis thaliana* (Col-0) plants by the floral dip method (Bechtold and Pelletier, 1998) as modified by Clough and Bent (Clough and Bent, 1998). The *NRPD1* and *NRPE1* genomic clones were transformed into *nrpd1-3* and *nrpe1-11*, respectively, while the over-expressed cDNA clones were transformed into wild type plants. T1 seeds were sown on soil and transformants were selected by spraying 2-week old seedlings with BASTA herbicide. *NRPE1* Δ SD-FLAG transformants were selected as described previously (El-Shami et al., 2007).

DNA methylation analysis. Southern blot analysis of *Hae*III and *Hpa*II digested DNA at the 5S rDNA locus was performed as in (Haag et al., 2009). The AtSN1 DNA methylation assay involving PCR amplification of undigested or *Hae*III-digested genomic DNA was performed as previously described (Herr et al., 2005).

RNA analysis. Small RNA was isolated and analyzed as previously described (Haag et al., 2009). RT-PCR was performed as previously described (Haag et al., 2009) using primers in Table S1.

Antibodies. Affinity purified anti-NRPD2 and anti-RDR2 have been described previously (Haag et al., 2009; Onodera et al., 2005). Anti-FLAG M2-HRP and anti-HA are commercially available (Sigma).

Immunoprecipitation and immunoblotting. Frozen leaf tissue (4.0g) was ground in mortar and pestle and protein extracted as in (Pontes et al., 2006). Supernatant was incubated with 35uL anti-FLAG-M2 or anti-HA resin (Sigma) for 3 hours at 4 °C on a rotating mixer. Resin was washed two times with extraction buffer supplemented with 0.5% NP-40. Washed immunoprecipitates were eluted from the resin with two bed volumes of 2x SDS sample buffer and boiled 5 min. Protein samples were run on Tris-glycine gels by SDS-PAGE and transferred to nitrocellulose or PVDF membrane.

Antibodies were diluted in TBST + 5% (w/v) nonfat dried milk (Schnucks) as follows: 1:500 NRPD2, 1:250 anti-RDR2, 1:3,000 anti-HA and 1:2,000 anti-FLAG-HRP. 1:5,000 to 1:10,000 anti-rabbit-HRP (Amersham) was used as secondary antibody. ECL Plus (GE Healthcare) was used for chemiluminescent detection of proteins. Membranes were stripped with 1% SDS, 25 mM glycine, pH 2.0 and re-equilibrated with TBST prior to subsequent blocking and immunoblotting.

Whole mount localization. Whole roots were fixed in 4% formaldehyde in PBS, pH 7.4 for 20 min at room temperature and washed in 1X PBS, pH 7.4 at room temperature. Nuclei were stained with 2.5 ug/ml propidium iodide (Invitrogen) and observed with Leica SP2 confocal microscope using 488 nm and 561 nm laser lines.

Acknowledgements

JRH and CSP designed the study and wrote the paper. EHT cloned the pENTR-NRPE1 aa 1234-1842 cDNA. OP performed the microscopy for Figure 5B. JG helped screen the CTD over-expression lines and prepared Figures 5F, G and H. JRH performed all remaining experiments. We thank Mike Dyer and the greenhouse staff for expert plant care, members of the Pikaard lab for helpful discussion and Tom Ream for proofing the manuscript. The National Institutes of Health grant GM077590 supported this research. The Pontes laboratory is supported by a grant from the Edward Mallinckrodt Foundation. The authors have declared that no competing interests exist.

References

- Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell*, 42(2), 599-610.
- Bechtold, N., and Pelletier, G. (1998) In planta *Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Methods Mol Biol*, 82, 259-266.
- Bellaoui, M., and Gruissem, W. (2004) Altered expression of the *Arabidopsis* ortholog of DCL affects normal plant development. *Planta*, 219(5), 819-826.
- Bellaoui, M., Keddie, J.S., and Gruissem, W. (2003) DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. *Plant Mol Biol*, 53(4), 531-543.
- Chiu, J., March, P.E., Lee, R., and Tillett, D. (2004) Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res*, 32(21), e174.
- Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev*, 11(24), 3319-3326.
- Clough, S.J., and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J*, 16(6), 735-743.
- Corden, J.L. (1990) Tails of RNA polymerase II. *Trends Biochem Sci*, 15(10), 383-387.

- Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997) Functional association between promoter structure and transcript alternative splicing. *Proc Natl Acad Sci U S A*, 94(21), 11456-11460.
- Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J*, 45(4), 616-629.
- El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev*, 21(20), 2539-2544.
- Grummt, I. (2003) Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev*, 17(14), 1691-1702.
- Haag, J.R., Pontes, O., and Pikaard, C.S. (2009) Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS One*, 4(1), e4110.
- He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H., and Zhu, J.K. (2009) An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell*, 137(3), 498-508.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science*, 308(5718), 118-120.
- Ho, C.K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B., and Shuman, S. (1998) The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*, 273(16), 9577-9585.
- Jokerst, R.S., Weeks, J.R., Zehring, W.A., and Greenleaf, A.L. (1989) Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*. *Mol Gen Genet*, 215(2), 266-275.
- Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet*, 37(7), 761-765.
- Keddie, J.S., Carroll, B., Jones, J.D., and Grissem, W. (1996) The DCL gene of tomato is required for chloroplast development and palisade cell morphogenesis in leaves. *EMBO J*, 15(16), 4208-4217.
- Lahmy, S., Guilleminot, J., Cheng, C.M., Bechtold, N., Albert, S., Pelletier, G., Delseny, M., and Devic, M. (2004) DOMINO1, a member of a small plant-specific gene family, encodes a protein essential for nuclear and nucleolar functions. *Plant J*, 39(6), 809-820.
- Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell*, 126(1), 93-106.
- Liao, S.M., Taylor, I.C., Kingston, R.E., and Young, R.A. (1991) RNA polymerase II carboxy-terminal domain contributes to the response to multiple acidic activators in vitro. *Genes Dev*, 5(12B), 2431-2440.

- Luo, J., and Hall, B.D. (2007) A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol*, 64(1), 101-112.
- Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009) RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol*, 21(3), 367-376.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, 385(6614), 357-361.
- Mosher, R.A., Schwach, F., Studholme, D., and Baulcombe, D.C. (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci U S A*, 105(8), 3145-3150.
- Nonet, M.L., and Young, R.A. (1989) Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of *Saccharomyces cerevisiae* RNA polymerase II. *Genetics*, 123(4), 715-724.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, 120(5), 613-622.
- Otero, G., Fellows, J., Li, Y., de Bizemont, T., Dirac, A.M., Gustafsson, C.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (1999) Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol Cell*, 3(1), 109-118.
- Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell*, 126(1), 79-92.
- Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev*, 19(17), 2030-2040.
- Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*, 443(7114), 1008-1012.
- Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. (2009) Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell*, 33(2), 192-203.
- Riedl, T., and Egly, J.M. (2000) Phosphorylation in transcription: the CTD and more. *Gene Expr*, 9(1-2), 3-13.
- Schramm, L., and Hernandez, N. (2002) Recruitment of RNA polymerase III to its target promoters. *Genes Dev*, 16(20), 2593-2620.
- Stiller, J.W., and Hall, B.D. (2002) Evolution of the RNA polymerase II C-terminal domain. *Proc Natl Acad Sci U S A*, 99(9), 6091-6096.
- Takimoto, K., Wakiyama, M., and Yokoyama, S. (2009) Mammalian GW182 contains multiple Argonaute-binding sites and functions in microRNA-mediated translational repression. *RNA*, 15(6), 1078-1089.

- Till, S., Lejeune, E., Thermann, R., Bortfeld, M., Hothorn, M., Enderle, D., Heinrich, C., Hentze, M.W., and Ladurner, A.G. (2007) A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nat Struct Mol Biol*, 14(10), 897-903.
- Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*, 135(4), 635-648.
- Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet*, 41(5), 630-634.
- Woychik, N.A., and Hampsey, M. (2002) The RNA polymerase II machinery: structure illuminates function. *Cell*, 108(4), 453-463.
- Yamamoto, S., Watanabe, Y., van der Spek, P.J., Watanabe, T., Fujimoto, H., Hanaoka, F., and Ohkuma, Y. (2001) Studies of nematode TFIIE function reveal a link between Ser-5 phosphorylation of RNA polymerase II and the transition from transcription initiation to elongation. *Mol Cell Biol*, 21(1), 1-15.

Figure Legends

Figure 1. The NRPE1 DeCL-like domain is required for *nrpe1* in vivo

complementation. (A) Genomic HA-epitope tagged *NRPE1* C-terminal domain deletion series transformed into *nrpe1-11* mutant background. Black colored regions denoted with a “Δ” represent deletions. (B) Western blot analysis of HA-immunoprecipitated NRPE1 proteins from whole plant extracts and co-immunoprecipitated NRPE2. (C) Agarose gel results of chop-PCR DNA methylation assay and transcript expression at the *AtSN1* retroelement. (D) 5S rDNA methylation analysis by Southern blot of *HaeIII* and *HpaII* digested genomic DNA. (E) Northern blot analysis of AtCopia, 45S rRNA, miR171 and AtSN1 small RNAs with images of ethidium bromide (EtBr) stained gels below.

Figure 2. The NRPD1 DeCL-like domain is required for *nrpd1* in vivo

complementation. (A) Genomic FLAG-epitope tagged *NRPD1* C-terminal domain deletion transformed into *nrpd1-3* mutant background. Black colored regions denoted with a “Δ” represent deletions. (B) Western blot analysis of FLAG-immunoprecipitated NRPD1 proteins from whole plant extracts with co-immunoprecipitated RDR2 and NRPD2. (C) *AtSN1* chop-PCR DNA methylation assay. (D) 5S rDNA methylation analysis by Southern blot of *HaeIII* and *HpaII* digested genomic DNA. (E) Northern blot analysis of AtCopia, 45S rRNA, miR171 and AtSN1 small RNAs with images of ethidium bromide (EtBr) stained gels below. (F) RT-PCR analysis of *AtSN1* and *solo LTR* transcription with *GAPA* and no RT controls.

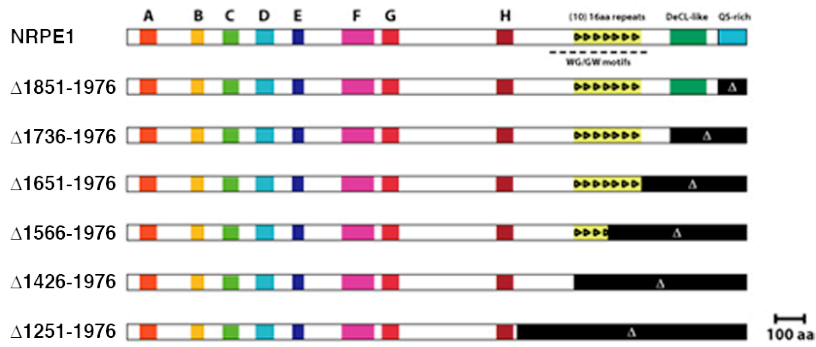
Figure 3. The NRPE1 repetitive elements and majority of WG motifs are not required for *nrpe1* complementation. (A) Genomic HA-epitope tagged *NRPE1* CTD internal deletion series transformed into *nrpe1-11* mutant background. Black colored regions denoted with a “Δ” represent deletions. (B) 5S rDNA methylation analysis by Southern blot of *HaeIII* and *HpaII* digested genomic DNA. (C) *AtSN1* chop-PCR DNA methylation assay. (D) RT-PCR analysis of *AtSN1* transcription with *actin* and no RT controls. (E) Northern blot analysis of 5S rRNA, AtCopia, 45S rRNA and miR163 small RNAs with image of ethidium bromide (EtBr) stained gel below.

Figure 4. The NRPE1 WG motifs are important but not required for *nrpe1* in vivo complementation. (A) Genomic HA-epitope tagged *NRPE1* CTD internal deletion series transformed into *nrpe1-11* mutant background. Black colored regions denoted with a “Δ” represent deletions. (B) *AtSN1* chop-PCR DNA methylation assay. (C) 5S rDNA methylation analysis by Southern blot of *HaeIII* and *HpaII* digested genomic DNA. (D) RT-PCR analysis of *AtSN1* and *solo LTR* transcription with *GAPA* and no RT controls.

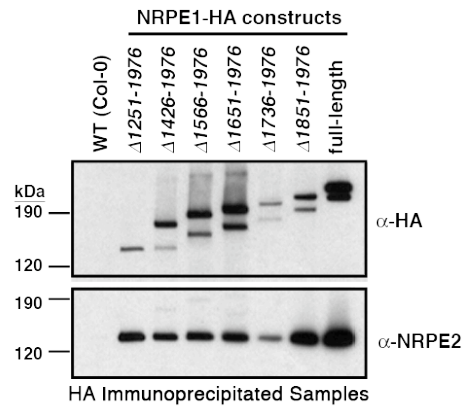
Figure 5. Over-expression of the NRPE1 CTD dominantly suppresses the RddM pathway. (A) 35S promoter driven N-terminally tagged cDNA constructs transformed into wild type *Arabidopsis thaliana*. (B) Whole mount localization of YFP-CTD in *Arabidopsis* root with enlargements of a single nucleus showing YFP signal, propidium iodide (PI) signal for stained DNA, and overlaid images. (C) *AtSN1* chop-PCR DNA methylation assay with YFP-CTD transformants. (D) RT-PCR analysis of *YFP-CTD*

transgene and *AtSN1* transcription with *actin* and no RT controls. (E) Northern blot analysis of *AtCopia*, 45S rRNA, miR171 and *AtSN1* small RNAs with images of ethidium bromide (EtBr) stained gels below. (F) Western blot analysis of immunoprecipitated over-expressed FLAG epitope tagged NRPE1 and NRPD1 CTD protein domains. An arrow denotes predicted full-length proteins. (G) *AtSN1* chop-PCR DNA methylation assay of over-expressed CTD domains. (H) RT-PCR analysis of *AtSN1* and *solo LTR* transcription with *GAPA* and no RT controls in over-expressed CTD transformants.

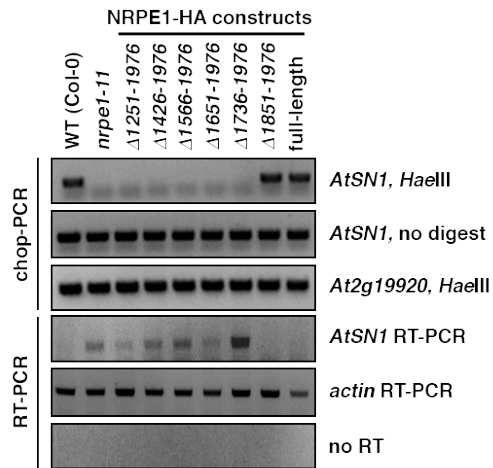
A. NRPE1 C-Terminal Domain Deletion Constructs Transformed into the *nrpe1-11* Mutant



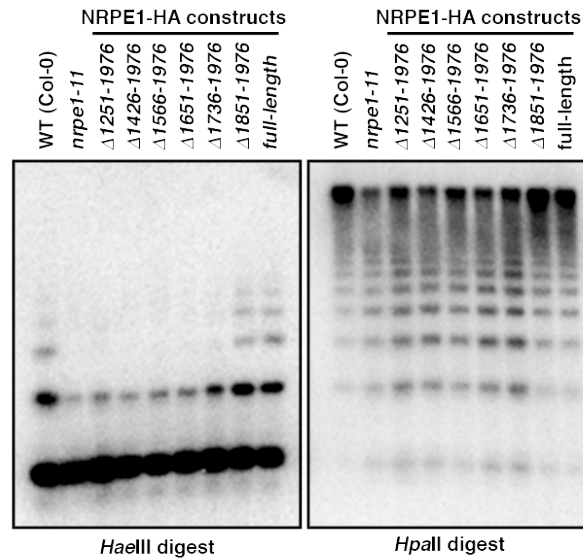
B. Co-IP Analysis in Transgenic Lines



C. AtSN1 DNA Methylation and Expression



D. DNA Methylation at 5S rDNA



E. Small RNA Accumulation

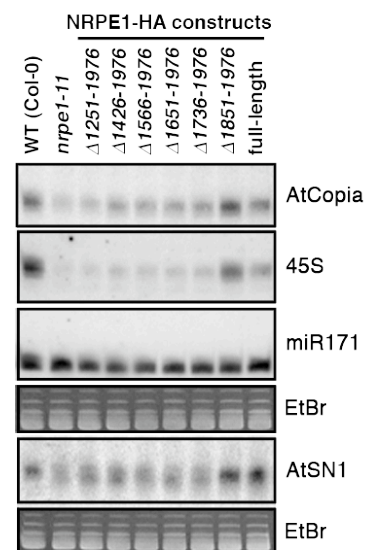
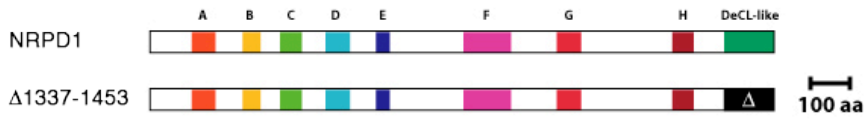
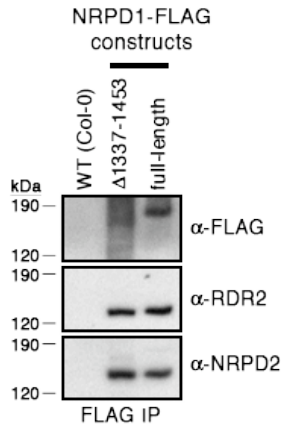


Figure 1

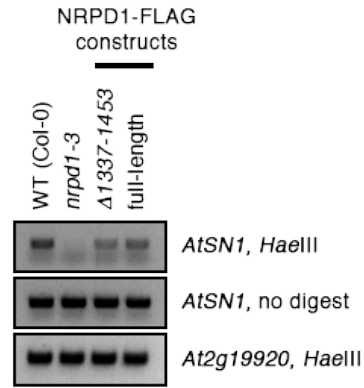
A. NRPD1 DeCL-like Domain Deletion Construct Transformed into the *nripd1-3* Mutant



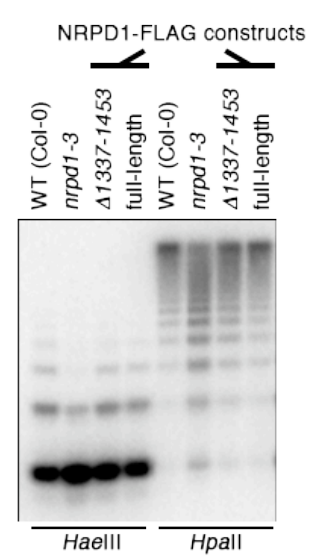
B. Co-IP Analysis



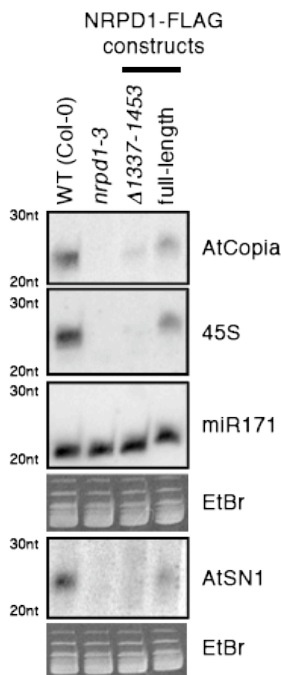
C. AtSN1 DNA Methylation



D. 5S rDNA Methylation



E. Small RNAs



F. Transcription Analysis

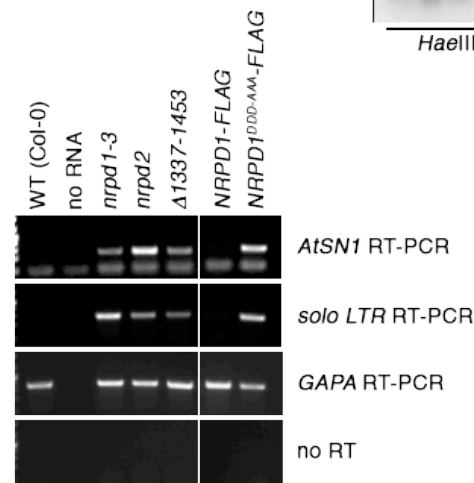
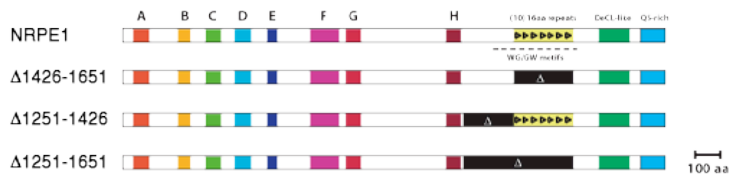
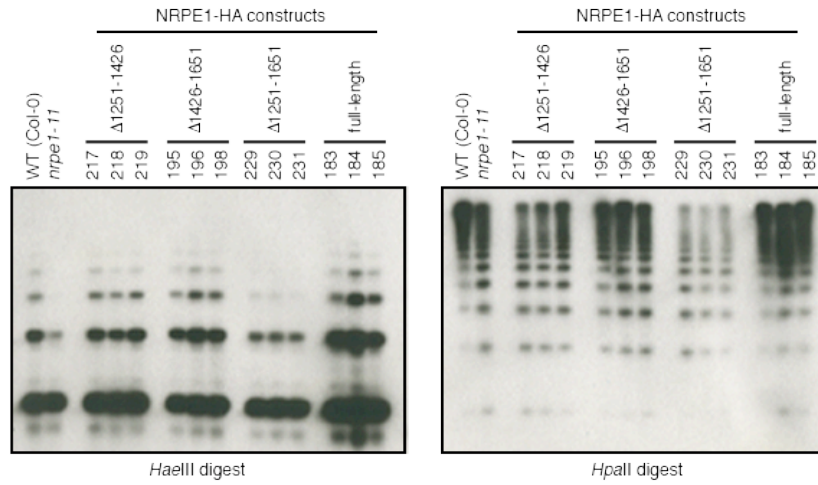


Figure 2

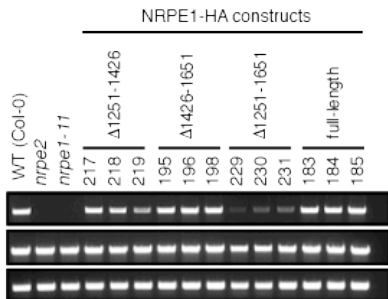
A. NRPE1 Internal Deletion Constructs Transformed into the *nrpe1-11* Mutant



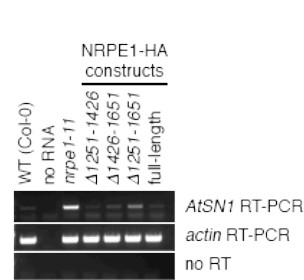
B. 5S rDNA Methylation



C. *AtSN1* DNA Methylation



D. *AtSN1* Transcripts



E. Small RNA Accumulation

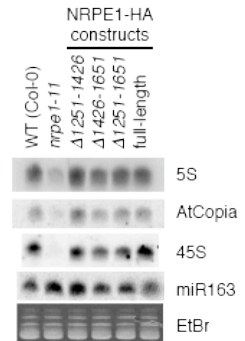
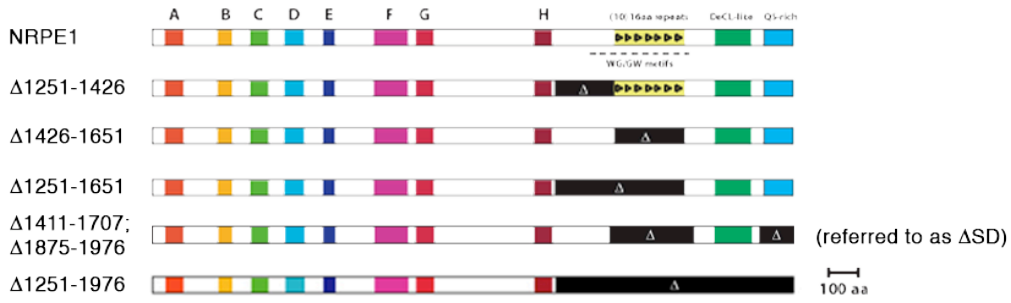
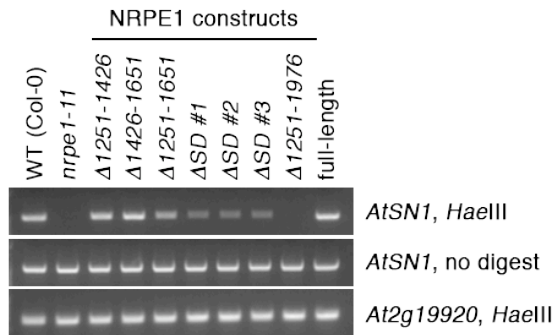


Figure 3

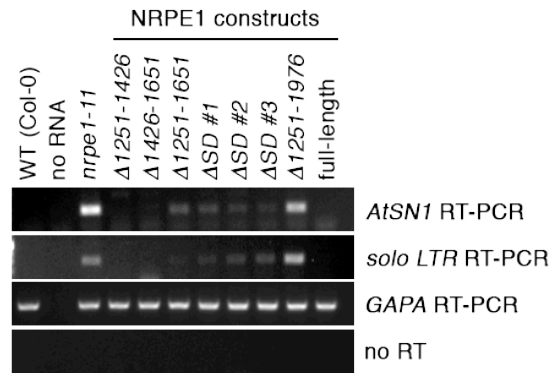
A. NRPE1 Internal Deletion Series in *nrpe1-11* Mutant Background



B. *AtSN1* DNA Methylation



D. Retroelement Transcription



C. DNA Methylation at 5S rDNA

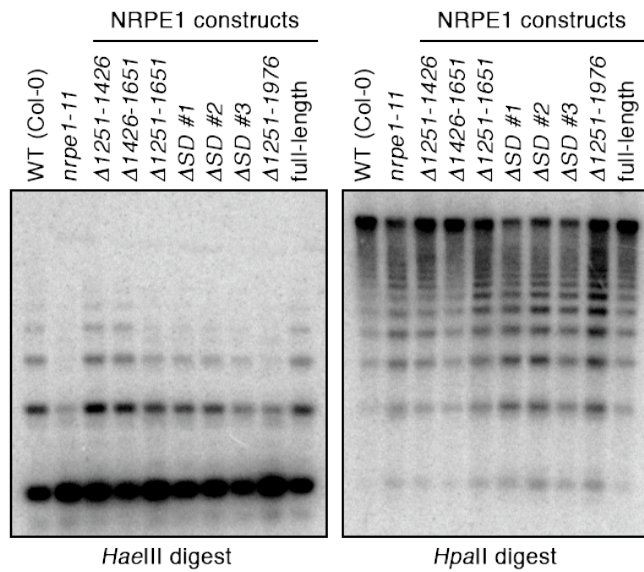


Figure 4

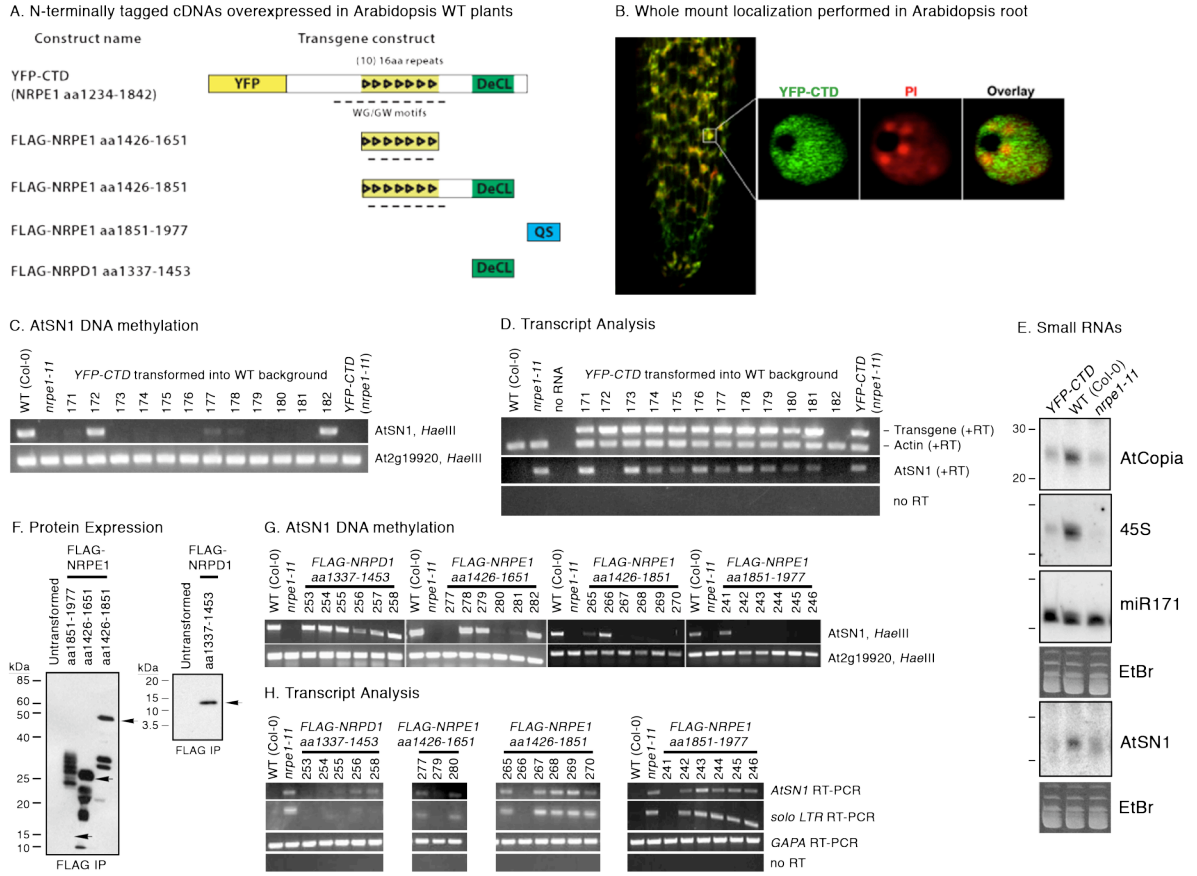


Figure 5

Table S1. Primers used in this study.

Target	Primer	Sequence (5' to 3')	Application
NRPDI Δ1337-1453	NRPDI-F NRPDI d1337-R	CAC CGG TGT CTC ACA TTC CAA AGT CCC C CCA TGT AAA GAT CGT TCT AAG CAG TGA CAT AGG AAT	Generate genomic NRPDI Δ1337-1453 clone; deletes DeCL domain
NRPEI Δ1251-1976	NRPEI-F NRPEI d1251-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A GAT AAA GAA GAA ACA GAT GTG TAC AGC TTC CTT	Generate genomic NRPEI Δ1251-1976 clone; deletes entire CTD
NRPEI Δ1426-1976	NRPEI-F NRPEI d1426-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A CCA CGA TTT GTC TGA AAC AGA TTT GTG TCC	Generate genomic NRPEI Δ1426-1976 clone; deletes all repeats, DeCL and QS-rich domains
NRPEI Δ1566-1976	NRPEI-F NRPEI d1566-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A CCC CAT ACC CCA ACC AGC AGG	Generate genomic NRPEI Δ1566-1976 clone; deletes 4 repeats, DeCL and QS-rich domains
NRPEI Δ1651-1976	NRPEI-F NRPEI d1651-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A GTC TTC TGC AGT GGG ACT TGG C	Generate genomic NRPEI Δ1651-1976 clone; last repeat at C-terminus; deletes DeCL and QS-rich domains
NRPEI Δ1736-1976	NRPEI-F NRPEI d1736-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A CTC AGA GGT GAA TGA GTC CAA GCG	Generate genomic NRPEI Δ1736-1976 clone; deletes DeCL and QS-rich domains
NRPEI Δ1851-1976	NRPEI-F NRPEI d1851-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A GAA TTC ATT GAC AAG TAC TTT ACG AAA CCT	Generate genomic NRPEI Δ1851-1976 clone; deletes QS-rich domain
NRPEI Δ1251-1426	d1251-1426 mut-F d1251-1426-F d1251-1426 mut-R d1251-1426-R	GTG TAC AGC TTC CTT GAC AAA AAG AAC TGG GGA ACT GAA TCA GC GAC AAA AAG AAC TGG GGA ACT GAA TCA GC AAG GAA GCT GAA CAC ATC TGT TTC TTT ATC ATC TAG ACC AGT CTG C ATC TGT TTC TTC TTT ATC ATC TAG ACC AGT CTG C	Generate genomic NRPEI Δ1251-1426 clone using SLIM strategy (Chiu et al., 2004); deletes linker between domain H and CTD internal repeats
NRPEI Δ1251-1651	d1251-1651 mut-F d1251-1651-F d1251-1651 mut-R d1251-1651-R	GTG TAC AGC TTC CTT AAG GAT ACC AAT GAG GAT GAT AGA AAT CCG TG AAG GAT ACC AAT GAG GAT GAT AGA AAT CCG TG AAG GAA GCT GAA CAC ATC TGT TTC TTT ATC ATC TAG ACC AGT CTG C ATC TGT TTC TTC TTT ATC ATC TAG ACC AGT CTG C	Generate genomic NRPEI Δ1251-1651 clone using SLIM strategy (Chiu et al., 2004); deletes linker and CTD internal repeats
NRPEI Δ1426-1651	d1426-1651-F d1426-1651-R	GTT TCA GAC AAA TCG TGG AAG GAT ACC AAT GAG CTC ATT GGT ATC CTT CCA CGA TTT GTC TGA AAC	Generate genomic NRPEI Δ1426-1651 clone using Stratagene strategy; deletes CTD repeats
NRPEI aa1234-1842	NRPEI 1234-F NRPEI 1842-R	CAC CAA AGA GAC TGG TCT AGA TGA TAA AGA AGA AAC AGA TG TTA GAA TTC TTC AGC ACG GTC AGG GT	cDNA clone of NRPEI CTD (-QS domain) used for bacterial expression and transgenics
NRPEI aa1426-1651	NRPEI 1426-F NRPEI 1651-R	CAC CAT GTG GGA CAA AAA GAA CTG GGG AAC TG TCA GTC TTC TGC AGT GGG ACT TGG C	cDNA clone of NRPEI repeats used for bacterial expression and transgenics
NRPEI aa1426-1851	NRPEI 1426-F NRPEI 1851-R	CAC CAT GTG GGA CAA AAA GAA CTG GGG AAC TG TCA AGG TTT CGT AAA GTA CTT GTC AAT GAA TTC	cDNA clone of NRPEI repeats and DeCL used for transgenics
NRPEI aa1851-1977	NRPEI 1851-F NRPEI 1977-R	CAC CAT GCC TCG GCC TAG CCG AAA CAG TTA TGT CTG CGT CTG GGA CCG	cDNA clone of NRPE QS-rich domain used for bacterial expression and transgenics
NRPDI aa1337-1453	NRPDI 1337-F NRPDI 1453-R	CAC CAA AAA CAT CGA GTT GCT TTC CCA GTC ATT G TCA CGG GTT TTC GGA GAA ACC AC	cDNA clone of NRPDI DeCL domain used for transgenics
NRPEI aa1251-1425 , 1652-1977	NRPEI 1251-F NRPEI 1977-R	CAC CCT TCA AAT GGT CAT ATC CAC GAC AAA CGC TTA TGT CTG CGT CTG GGA CCG	cDNA clone of NRPEI repeat internal deletion used for bacterial expression; cloned from NRPEI Δ1426-1651-HA total RNA

Table S1. Primers used in this study (continued).

AtSN1	AtSN1-F AtSN1-R	AGG ATT TAT TTC AAT CCA CGA ACC T CGA CTC CCA TAA GTA ACG AGT TG	Chop-PCR (Herr et al., 2005)
At2g19920	AtSN1 control-F AtSN1 control-R	CTC TGG GTT ACC TTT CAG GAA TCA G CTA AAT TGA AGA GCT TAC CTG CTT G	Chop-PCR control (Herr et al., 2005)
AtSN1	AtSN1 RT-F AtSN1 RT-R	ACC AAC GTG CTG TTG GCC CAG TGG TAA ATC AAA ATA AGT GGT TGT TGT ACA AGC	RT-PCR (Herr et al., 2005)
solo LTR	solo LTR-F solo LTR-R	ATC AAT TAT TAT GTC ATG TTA AAA CCG ATT G TGT TTC GAG TTT TAT ICT CTC TAG ICT TCA TT	RT-PCR (Wierzbicki et al., 2008)
Actin	Actin-F Actin-R	TCA TAC TAG TCT CGA GAG ATG ACT CAG ATC ATG TTT GAG TCA TTC TAG AGG CGC GCC ACA ATT TCC CGT TCT GCG GTA G	RT-PCR (Herr et al., 2005)
GAPA	GAPA-F GAPA-R	GGT AGG ATC GGG AGG AAC GAT AAC CTT CTT GGC ACC AG	RT-PCR, glyceraldehyde 3-phosphate dehydrogenase A (Kanno et al., 2005)

Supplemental Data

Supplemental Methods

Sequence analysis. Full-length NRPD1 and NRPE1 protein sequences were obtained from NCBI GenBank and the publicly available genome sequencing efforts of JGI (<http://www.jgi.doe.gov/>). When necessary, cDNA predictions were made using FGENESH+ (<http://www.softberry.com>). Repeat elements were identified with XSTREAM (<http://jimcooperlab.mcdb.ucsb.edu/xstream/>) and by manual analysis.

***In vitro* co-immunoprecipitation.** NRPE1 cDNA constructs were recombined into pDEST17 (N-terminal GST fusion construct for bacterial expression) and expressed in the BL21.AI strain. A single colony of each construct was inoculated in 5 mL 1xLB (50 ug/mL Carb) and incubated overnight at 37 degrees C. Overnight culture was then used to inoculate fresh 1xLB (50 ug/mL Carb) and samples were incubated at 37 degrees C to an OD₆₀₀ of 0.4. Expression was induced with the addition of L-Arabinose to 0.2% final concentration and incubated another 3 hours at 37 degrees C. Bacteria were pelleted and washed once with 1x Binding Buffer. The pellet was resuspended in 1x Binding Buffer and lysed by sonicating a total of 1 min at Duty Cycle 40% and Output 1.5 in a Branson Sonifier. Samples were centrifuged at 10,000 x g for 15 min at 4 degrees C. The soluble fraction was retained and GST-tagged recombinant protein purified with glutathione resin (Amersham).

MYC-AGO4 protein extract was isolated from 4.0 g of inflorescence tissue by grinding under liquid nitrogen in a mortar and pestle and resuspending in 14 mL Baumberger buffer. Extract was filtered through two layers of Miracloth and centrifuged

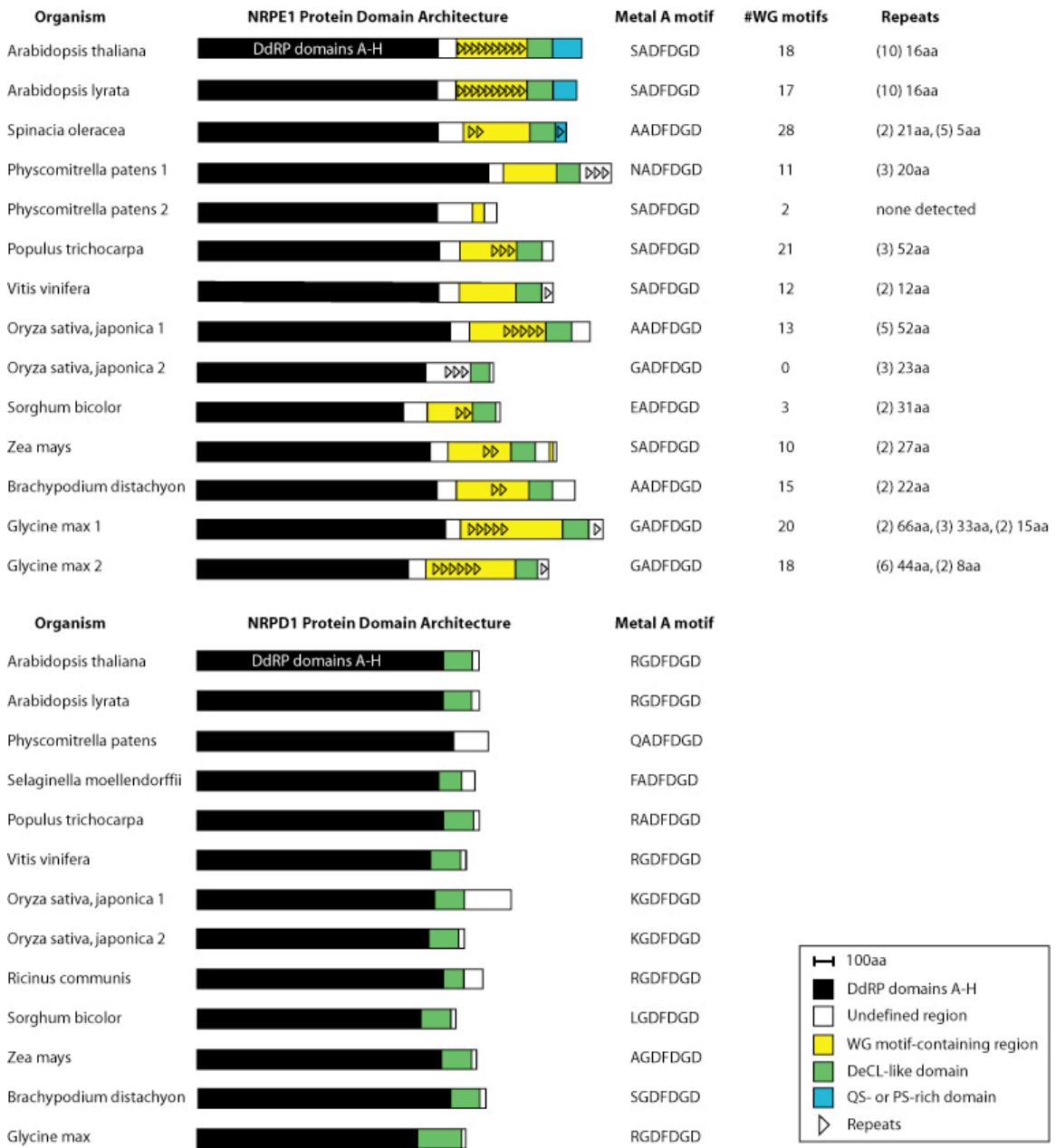
15 min at 11,500 rpm. Supernatant (300 μ L) was added to the washed glutathione resin with bound GST recombinant proteins and the volume was brought up to 1 mL with Baumberger buffer and incubated for 3 hrs at 4 degrees C. The glutathione resin was washed 5 times for 2 min each with 1 mL Baumberger Wash Buffer and pelleted by centrifugation at 200 rpm for 2 min. Protein was eluted from the resin by adding 50 μ L 2x SDS loading buffer and incubating at 95 degrees C for 5 min.

Samples were split and run on 4-12% Novex gels. One sample set was Coomassie stained while the other was transferred to PVDF membrane for Western blot analysis.

***In vivo* co-immunoprecipitation using native antibodies.** All steps were performed at 4 degrees C unless otherwise stated. Frozen inflorescence tissue (0.7 g) was ground in liquid nitrogen and homogenized with 2 mL extraction buffer (50mM Tris-HCl [pH 7.5], 150 mM NaCl, 5 mM MgCl₂, 10% glycerol, 0.1% NP-40) containing 2 mM DTT, 1 mM PMSF, and 1/100 plant protease inhibitor cocktail (Sigma) [Li et al, 2006]. Sample was transferred to a 2.0 mL microcentrifuge tube and centrifuged twice at 13,000 rpm for 5 min. Samples were precleared with 20 μ L Protein A agarose beads (Pierce) for 30 min. The samples were then incubated with 1:250 anti-NRPE1 or 1:250 anti-AGO4 for 3 hrs. Protein complexes were captured with 60 μ L Protein A agarose beads (Pierce) for 2 hrs and then washed five times with extraction buffer. Samples were boiled in SDS loading buffer and run on a 7.5% Tris-glycine gel followed by transfer to PVDF membrane. Western blot was performed with 1:5000 anti-Myc monoclonal antibody (Upstate) O/N at 4C followed by anti-mouse-HRP and ECL Plus detection.

In vivo co-immunoprecipitation analysis comparing the extraction buffers from [Li et al, 2006] and [Baumberger et al, 2005] was performed as above, except one set of samples was incubated with anti-FLAG agarose beads and the other with anti-cMyc agarose beads (Sigma) for 4 hrs at 4 degrees C. The Protein A preclearing step was skipped.

Figure S1.



Comparison of NRPD1 and NRPE1 C-terminal domain architectures among diverse plant species. Domain features of illustrated full-length protein predictions are based on sequence analysis presented in Figures S2 and S3. The *Arabidopsis lyrata*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Populus trichocarpa*, *Vitis vinifera*, *Sorghum bicolor*, *Brachypodium distachyon* and *Glycine max* NRPD1 and NRPE1 sequences were produced by the US Department of Energy Joint Genome Institute, <http://www.jgi.doe.gov/> and are provided for use in this publication only. *Zea mays* NRPE1 was kindly provided by Lyudmila Sidorenko (Chandler lab). The *Brachypodium distachyon* sequences were identified by Tom Ream in the Pikaard lab. Remaining sequences have previously been published or are available from NCBI GenBank.

Figure S2. Predicted NRPE1 protein sequences among diverse plant species with key domain features denoted to the right-hand side. The Metal A motif is in black bold type; the conserved DdRP H domain is underlined in bold; WG/GW/WGW/GWG motifs are in bold; repeat elements are underlined with solid and dotted lines; the DeCL signature motif is in bold blue type.

```

>Arabidopsis_thaliana_NRPE1 (At2g40030)
MEEESTSEILDGEIVGITFALASHHEICIQSISESAINHPSQLTNAFLGLPLEFGKCESCGATEPDKCEGH
FGYIQLPVPIYHPAHVNELKQMLSLCLCKLKIKKAKGTSGGLADRLLGVCCEEASQISIKDRASDGASYL
ELKLPSSRRLQPGCWNFLERYGYRYGSDYTRPLLAREVKEILRRIPEESRKKLTAKGHIPOEGYILEYLPV
PPNCLSVPEASDGFSTMSVDPSSRIELKDVLKKVIAIKSSRSGETNFESHKAEASEMFRVVDTYLQVRGTAK
AARNIDMRYGVSKIISDSSSSKAWTEKMRTLFIKRGSGFSSRSVITGDAYRHVNEVGIIPIEIAQRITFEERV
SVHNRGYLQKLVDDKLCLSYTGSTTYSLRDGSKGHTELKPGQVVHRRVMDGDVVFINRPPTTHKHSGLQAL
RVYVHEDNTVKINPLMCSPLSADFDGDCVHLFPQSLSAKAEMELFSVEKQLLSHTGQLILQMGSDSL
SLRVMLERVFLDKATAQQLAMYGSLSLPPPALKSSKSGPAWTVFQILQLAFPERLSCKGDRFLVDGSDLL
KFDFGVDAMGSIINEIVTISIFLEKGPKETLGFDFSLQPLLMESLFAEGFSLSLLEDLSMSRADMDVIHNLII
REISPMVSRRLRSYRDELQLENSIHKVKEVAANFMLKSYSIRNLIDIKSNSAITKLQQGTGLGLQLSDKK
KFYTKTLVEDMAIFCKRKYGRISSSGDFGIVKGCFFHGLDPYEEMAHSIAAREVIVRSSRGLAEPGTLFKN
LMAVLRDIVITNDGTVRNTCSNSVIQFKYGVDSERGHQGLFEAGEPVGVLAATAMSNPAYKAVLDSSPNSN
SSWELMKEVLLCKVNFQNTTNDRRVILYLNECHCGKRFCQENAACTVRNKLKLVSLKDTAVEFLVEYRKQP
TISEIFGIDSLHGH IHLNKTLLQDWNISMQDIHQKCEDVINSLGQKKKKATDDFKRTSLSVSECCSFRD
PCGSKGSDMPCLTFSYNATDPDLERTLDVLCNTVYPVLEIVIKGDSRICESANI IWNSSDMTTWIRNRHAS
RRGEWVLDVTVEKSAVKQSGDAWRVVIDSCLSVLHLIDTKRSIPYSVKQVQELLGLSCAFEQAVQRLSASV
RMVSKGVLKEHI ILLANNMTCSGTMLGFNSGGYKALTRSLNIKAPFTEATLIAPRKCFEKAEEKCHTDSLS
TVVGSCSWGKRVDVGTGSOFELLWNQKETGLDDKEETDVYSFLQMVISTTNADAFVSSPGFDVTEEEMAEW
AESPERDSALGEPKFEADSADFQNLHDEGKPSGANWEKSSSWDNGCSGGSEWGVSKSTGGEANPESNWEKTT
NVEKEDAWSSWNTRKDAQESSKSDSGGAWGIKTKDADADTTPNWETSPAPKDSIVPENNEPTSDVWGHKSV
SDKSWDKKNWGTESAPAAWGSTDAAVWGSSDKKNSETESDAAWGSSRDKNNSDVGSGAGVLGPWNKKSSET
ESNGATWGSSDKTKSGAAAWNSWDKKNIETDSEPAAWGSQGGKKNSETESGPAAWGAWDKKKSETEPGPAGW
GMGDKKNSETELGPAAMGNWDKKKSDTKSGPAAWGSTDAAAWGSSDKNNSETESDAAWGSSRNKKTSEIES
GAGAWGSWGQPSPTAEDKDTNEDDRNPWVSLKETKSREKDDKERSQWGNPAKKFPSSGGWSNGGGADWKGN
RNHTPRPPRSEDNLAPMFTATRQLDSFTSEEQELLSDVPEVMRTLRLKIMHPSAYPDGDPISDDDKTFVLE
KILNFHPQKETKLGSGVDFITVDKHTIFSDSRCFFVVDGAKQDFSYRKSLNNYLMKKYPDRAEEFIDKY
FTKPRPSGNRDRNNQDATPPGEEQSQPPNQSIGNGGDDFQTQTQSQSPSQTTRAQSPSQAQAQSPSQTQSQ
QSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQS

```

Metal A

H

(10) 16 aa repeats & (18) WG motifs

DeCL

QS-rich

>Phycomitrella_patens_NRPE1-1
 MQVMEAAAWRQPSQAPTADLVGLQIGLATTSEILGHSVIESRSKDTLISLVDPRGLPAEDERCATCGGTN
 YDECTGHFAHVKLTQPIFHPNYIRCVQRVLQKICLACGVPKVKMKSFSEEAANLKQNFDRIDSEDVGGNG
 EHPVLLLEADAIEKDADDVVILLSSDEEEYPRDILRVVPSGPMDFLIRSTNESAIADLPQLKSYKSKSKAHA
 NGFSHVDVTRKSTRKSSSKSSSTQNPVKIYKGTAGLDVLNADTLRTAEPLDTNTCPYCSPGYPDYRHIL
 VKILPVKGRKKNVDSQIILLEVQGSCKGKFLLPDFWFSFIKGAAYPENEEVPKSHVLSPLEALSILKKS
 DTAIGKLGMLNGLVARPEGLIMKCVPIPPNCTRTTDDYKYVSNTTAVRFGTDRVTRTLQNLVNEIGRIQRTRT
 GKIMKRGQRDEVKVLQVLTAEYLREKGAPKAVPGKEPLKKDRNGRFTKQDDHRWTKDWISQNYLKGKGGNYT
 ARAVVAGDPSLAIEETIGVPLEIAQKLTVPERATKWNRSKLOEYVDRTQMLQOQSGKPGATRIVRNEEAFQV
 WANSTHTVQIGDVIHRNIQDGFVYVNRPPSVHKHSLMALKVQVHYGLVLTINPLVCPFF**NADFDGD**IFHV
 FIPQSLQAI AELEHLMVAPQOIISDHGGQPLLGLTQDITLLAAYLLTSSKLLVDKAGMDQLCLWALKQPPDA
 AIVKSPKGGPFWTGEQIFGLTLPTDLQVGAPEHEVFI EGGEVIRWSNGAKSLRKDSEGIAAALCVQLGPVA
 LVNYLNTATGLLHAWLQMHGFSTGLADFQVTSNSADRQKMLKSI FEDYYQKSIQESCDSVRILDAKVQAMG
 QEVISSPDHLTRNINFLEQAAQQTFRNRESEVESIVMKYAARDNGLLMMVRSKSGSRGKLLQOIAGMGLQ
 LYKQHLPLPFGSRRSSMSNSSELDWEDKGLVRSLLVDGLNPSSELFNHVIADRTVILRKHVEVVQPGTLF
 KSLMLFLRDLHVMYDGSVRNQCCKNIVQFCYGAIGVLKRSIPKERLSRSQFEVVNPATPIVTTWEEDDLKR
 WPLSILAGEPVGVLAAATAISQPAYELMLDAPCLNGPFKPRPLELVQETLYPRAKSVLKPIDRTAIIRLVNC
 PCTQPLCLERRVLAVQAHLKKSILKATAEASCAVEFWNMENFEVAGPSGEALRMGSPWLGHIKLSLNLKMQQL
 QVDVELMVERLRQRFSGI IKNPKKHPMGQIFFCVSYNCGISNGLCLHFSKLPKNKMQNQRNDEIYNTALLA
 LLLKIRGTIISGLLDCTVKGDERIESVIIVSEDPSTTWHRGLTCNQELEELVLEVVSPTKSKSKRGDA
 WASVKQACLPLMHMVDWNRSMYPYSIQEIRHALGVEASYQMSQRLGLVLDKTAPHTRSVHVKLVDMMTFS
 GDANGFNFSGFQDMNKSTGI SAPFTEASFQKPIKTLMDAAGRATDSVESVLASCV**WG**KEAPLGTGSNFEL
FWQPSKQSRLAASRKAEKDVHMIWKDLHEKCSIDKVLPPSPPSLPGLPTLPDGDVLDGAGFSPHAS
 NDAADDT**WGS**PHRNNGGDGV**WGD**SPVVRDDDG**WGW**AVGKGNDSNEVDGYDQDNSTGASKELSGW**SK**PASE
 RS**GWGS**MSDKEGSSRNAWDDDFGKEDRHE**GWG**DGATEPINEGG**WGS**LNNEEGTTSKAKCSSD**WG**TNAVQEIIG
 DGG**W**DAVSIIEVPEGD**GW**DSLKVPQTENAIEVGSSEHADRSYGPAGDVSQEQFRARGEESRRGGRPWTSRD
 RRRWRGRGSFGKDRGSSGRMSPGNRQNSGTISRQEQTPWVQGSTKADAWAKHAWASFGSSQGEVQAGGD**GW**
 DAVLPDNCGASNRAHSTYPIAGSMPPTSRQDEVEPECKDIDDLVKSMRRIILFNPRNELGGRLSDEDDDELVQ
 TVLAYHPKLSEKAGCGTAYIKVDRSAGFVNNRCFWLVRTDGSEI **DFS****FK**CLKEKVAREFPFSLDRYDDVY
 QAHRPFPTANFEENKSAAQGNIDAGPSAAHLLLEDMPIDHEDLDARPAAAHLPEGIPIDQEDLDAQPAVAH
 LSEDTPIDQENLDAQPAANSISVDTHFDQQEDIDTQTGQESAPSIGVSSATKLI CKKLTPEVHEHQDTSGP
 H

Metal A
 H
 (11) WG motifs
 DeCL
 (3) 20aa repeats

>Phycomitrella_patens_NRPE1-2 (phya_79970)
 MQIKSEDWTWTPGNVPIPPPSAEIVGLQFGLTTANEINRARDTLSSSIDPRLGLPAENERCATCSGTNIN
 ECTGHFGHLKLTQPIFHPHHVRLQLQVLSKICLACGSLKGGKALAILKKIPEGAIGKLGMLNRLVARPEGL
 IMKCVLI PPNCTRTTDDYKHVNNTTAVRFGTDNVTRTLQKLVAEIVHIRKTRAGKATNRTQRDESTKLQILT
 AEYLREKGAPKAVPGKEPLKRDRNGRVTKQDYHRWTKEWLSQNVLGKSGNFTAKAVLAGDPFLGIEQIGIP
 WLIAQKLTLPERASQWNHTKLQEYVNVSQKLOQESSENTAHATRVERNEVVYQVLSKTSLVQVIGDIVHRHI
 QDGDYVYVNRPPSVHRHSLVALKVHIHQPTITVNPLICPPFF**SADFDGD**IFHIFAPQSLQAI AELDQLMAV
 KQQVI SEHGGQPLLELTQSQSIAFNVLNQNNTLLAAHLLTSKKLFLDKATMDQLCLWASKKPPEAAIILKS
 PKGGPFWTGEQVFALTLPEDFELGAPQEEVFIQGGEEIRWRNGTKLLRKGNDSVAAALCVQLGPVALVDYL
 NTATGVLHTWLQVQGFSTGLTDFQVTPNRTKRQEMLSILEESFLKSIQESCDFVRILDAKVQALDSDENP
 SPESLTKNIRFLEQVAREIFQKRRSEAGRIVAKYAEQRNSLLMMVESGSKGSMEKLLQOIAGMGLQLYKQ
 HLLSYSSSRPAMTYSSQLDWWEDMGLVRSLLVDGLKANELFRHVIADRTGILRKHVEVVQPGTLFKALMF
 FLRDLHIMYDGSVRSQCSKNLIQFCYGGARGSLIPRKPTEETLAWEEEDHRRWPLSVLAGEPVGVLAAAAI
 SQPAYELMLDAPSLNGPFKPRPLNLIQRLSTTWRFAHETLYPREKSSLKPTDRCVVLRVLVHCECTESLCL
 RRVLEVQAHLKRINLRMMAESVAVEYWNMEDSRAAGPSGDLVRLGSPWLGHINLSQDAMKQCEVNVEDIVK
 RLCQKFSQTAGYVLKKNKMGQIFFCHRIQETIIPGLLDCTMKGDERIETVRVVCCEGPASTTWHRRFAHCTG
 NLDEELVLEVYVSPSSSKSRGMASVQACVSLKDLVDWNRSMYPYSIQEIRCSLIGIEVAYQIVVQTA
 PTHFVHVKLVAEMMTFSGDAIGFTFSGFKDMNRSISVAPFSEASFOASAOPIRTLLGAAGRATDSVEGVM
NCI**WG**KEAPLGTGGNFGLFWQPKAIKSFLLCCVVKQRFNTNICLLIGSHLQKFI VFYALMVLVFLDLKQVPL
IFQGIQRFGASKEAVKDVHTILKDLEDECIPDRFISSMPTLLPPLHLILPEGNLEFDDGAGFSPQRVSDCN
 EGLDDRNHGNSVDDQRGVSDTAVDGNVPIDWIKEEIIYQNSDIKPEDELGAWQPTSQYGGG**W**DDIDTVPGL
 RSLDNVSSDATGFKCYDTSKNSKNEEVVMVETTMGFGSINWGTNCIQDIGSD**GW**DVPSSEVATGGSWDFL
 DKKCQNDSGCGSKHLDHKHGSSGKSILLQERQFTAHEALDQDPAK

Metal A
 H
 (2) WG motifs

>Spinacia_oleracea_NRPE1

RYPVPVPPNCLSVDPDISDGVSVMSDDLCSAMLLKVLRLQIEVIRSSRSRSGEPNFESHEVEANDLQVAVSQYLQV
RGTGKAARAADNRYGVSKEGNNSKAWVEKMRTLFI SKGSGFSSRSVITGDAYRAVNEVGPCEIAQKMTF
EERVNVHNIQYLQGLVDKNLCLTFRDGLSTYSLREGSKGHTFLRLGQMVHRRIMDGDIVFINRPPTTHKHS
LQALRVYIHDDHVVKINPLMCGPLAADFDGDCVHLFYQSLSARAEVLELFSVEKQLLSSHSGNLNLQLST
DSLSSLKTMFEVYFLDRASANQLAMYASSLLPSPALWKACSSNAKKAHSSGPRWTAQQVLQATALPSHFE
CHGDRLLIHDSEILKLDNFNRDIVASVISDVLTSLFFNKSPKDALDFDLSLQPLLMLNLFSEGFVSVLHDF
FPKSELQNIQRNIQDLSPLLLQLRSSFNELVQVQFENHIREFKSPVGNFILISSALGSMIDSRSDSAIDKI
VQQIGFLGLQLSDRRKFYSRGLVEDVASLFHQKYPFADVYPSEEFQFVSRFCFFHGLDPYEEIVHSIATREV
IVRSSKGLAEPGTLFKNLMAVLRDVIICYDGTVRNISSNSVIQFEYGVGGMQSQNLFPAGDPVGVLAATAM
SNPAYKAVLDSSPNSNSWDMMEKILFCRANFRNDINDRRVILYLNDCCCGRKYCQENASCLVKNHLKVKV
LRDAEIELAIEYKRPKLEPESCEIDAGLVGHIHLNSGLLKASGIGMHDILQKCEEQVNLRLKVKKYGYHFK
RILLSVSDCCFFNHSDSKWTDMPCLKFFWQDMDTDTDLERTKHMADMICPVLLDTIIKGDPRISTVNI IWI
NPGTTTTWVQSPCSSTKGELEVALEKEAVRLTGDAWRIVLDCCLPVFHLIDTRRSIPYAIKQIQDLFGIS
CAFDQAVQRLSTSVTMVTKGVLKEHLLLLASSMTCAGNLVGFNTSGIKALCRALNVQVPFTEATLYTPRK
FERASEKCHVDTLASIVGSCSWGKRVSIGTGAKFDLLWETKEIEMADKPTDVYNFLHLVSSANEEVDSGG
LGEDIESFEKDVYMEPALSPEQENKAVFEETLEIGVDSITGAESSWDAFPSSGTGWANKIDTGSLSAE
GGSSSWGSKKDQANPEDSSKTGGWSSGGSKOKQPEDSSKSGGWDAKSWGGSNQGDPSPVWGQPVKATND
ISIENDHSGSAEGGWANSGMKKDLKQENSSTAGGWDAKSWSGSKPKDPSSAWGAGKKTDDNNGWKK
DSKKDLASGSVEDGGCSGWGPKKDLLQPEDSAGENGWGAASKSKSKEPSSAWGKPAQETDNI GWKKNPQRD
SENLEGTSGWNDKLQENKSFQKQSPASSKDWSTGNITAGSTGFVGEKGNKPDVAVSNVSVKKSTWGQ
TGGNSWKKNEQDEKQDGLPWGKSHKSSDSWTSGQGNQHPVSVQGVSEKQGTLSWGQPRDSSQKNNNEN
GVSSNFRQAGKSWDSKKKESNVQSSWAQQGDSTWKSKEARSSVKANNSTNSGWSTGKALVDGVSSSW
GSQKEDRPQPKSNDRSVGDGNFDKDAKEEGLSSWDAKKVERKTQSSWGQPSSEKNSAQSSADHWGSDKSNQ
PGKSSWGSEDTNAGKDEKQDSSWGSNVSTWKKESGEKLGSDDSQSPWGQPGGSGWNNKQPEGGRGWGS
SNTGEWKSARKNQNNQNNQNNRPPRPNDDSPRVALTATRKRMEFPTTEKDLVSEVESLMQSI RIRIMHQS
GCVDGEPLLPDDQTYLIDNILNYHPDKAAKIGAGVDFITVKKHSNFQESRCFYVSTDGKDTDFSYIKCIE
TFVKGKYPVSAESFTSKYFRRSQRPOPASPSPASPSPTSPSPASPSAPPNPPTPT

Metal A

H

(28) WG motifs & (2) 21aa repeats

DeCL

PS-rich repeats

>Populus_trichocarpa_NRPE1

CTASISDCPISSHSQLTNPFLGLPLEFGKCESCGTSEPGKCEGHFGFIHLPIPIYHPSHISELKRMLSLIC
LKCLKLRNKIQIKSNGVAERLLSCCEECAQISIREVKNNTDGACFLELKLPSRSRLRDGCWNFLERYGFRY
GDDFTRPLLPCEVMQILKRI PAETRKKLSGKGYFPQDGYILQQLPVPPNCLSVVSDGITVMSSDLSISM
LKKVLKQAEVIRSSRSGAPNFDAAHKDEATSLQSMVDQYLQVRGTTKTSRDVDTRYGVKKESESTTKAWLE
KMRTLFIKRGSGFSSRSVITGDAYTLVNQVGIPEYIAQRITFEERVSVHNMRYLQELVDNKLCLTYKDGSS
TYSLREGSKGHTFLRPGQVHRRIMDGDIVFINRPPTTHKHSQALSIVYVHDDHAVKINPLICGPLSADFD
GDCVHLFYQSLAAKAEVLELFSVEKQLLSSHSGNLNLQLTTDSLSSLKMMFKACFLGKSAQQLAMFISP
YLPQALLKVNCFPPHHTAHQILQMALPACFNCSEGERFLIINSNFLKVDNFNRDVASVINEILISMFPEKG
SGAVLKFNSLQPLMLNLFSEGFVSLEDFSI SRAVKQRIPEFKAISPLLCNLRSTFNELVELQVENHI
RDVKQPVREFILTSSALGYLIDSKSDAAVTKVQVQIGFLGLQVSDRGKLYSKTLVEDLASHFLSKYPANLF
DYPSAQYGLIQNSFFHGLDAYEEMAHSISTREVIVRSSRGLSEPGTLFKNLMAILRDVVIICYDGTVRNVSS
NSIIQFEYGVKVGTESQSLFPAGEPVGVLAATAMSNPAYKAVLDSTPSSNCSWDMMEKILLCKVGFKNDLA
DRRVILYLNDGCGGRNYCQERAAAYLVKNHLEKVSCLKDIACFMIEYKSQQIPESEFGSDAGLVGHVHLDRK
LQDLNITAQVILEKQETVNTFRKKKKVGNLFKKTILLVSESCSFQQCIDESPCLMFFWQGADDVHLERTS
NILADMICPVLLETIIKGDHRI SCANI I WATPETNTWIRNPSRTQKGEALDIVLEKSVVKKSGDAWRIVL
DSCLPVHLINTTRSI PYAIKQVQELLGVSCAFDTAVQRLSKSVTMVAKGVLKEHLILLGNSMTCAGSLIG
FYTGGYKTLRSLDIQVPFTEATLFTPRKCFEKAEEKCHTDSLSSIVASCAGWKHVTGTSHFVLDLWDTK
EACLNPEGSMDVYSFLNMVRSSTAGGEESVTAACLAGAEVDDLMEDEDDWNLSPHNSSSDKPTFEDSAEFQDF
LGNQPAESNWEKISSLKDRSRSNGNWDVDKNDGAVKEKPSWLGMTAEANDVASSGWDTAAARTTNNNSWNS
ENNVAQNSNSFGWATKKPEPHNGFATKVQEEPTTSDNDWDAGAAGWRKDRDNKFAETNASKSWGKVTGDGE
SGQNKSKNRPEDQDVGTHGWDDKMSQDQSI SGWASKTTQEATTESLWDSKGNNSNPGDAACGWKAASTWG
AENTDGDGLWGEKVSNSQADTASGWGKPKSPEISLWGSTKESVKS DRGWGVSSSGGRDKKTENQSLAGQ
GKESGGWGNKVTNSQADTASGWGKPKSSENSQGWLSKESGKEVHEWGVNSAGGNGSETNNNNENQSLVE
QGKESGWDNKASSNQEGTASGWGKPKSPALSEGWGS PREPVKAVHWGVNSGGGNDWKNKRNRPSKPHED
LNASGIFTTRQRLDVFTSQQEIDILSDIEPLMLSIRIRIMHQTGYNDGDPLSADDQSYVLDNVFHYHPDKAV
KMGAGIDHVTVSRHSNFQESRCFYIVSTDGCKQDFSYRKCLENFIKGYPLADEFIA

Metal A

H

(21) WG motifs (3) 52aa repeats

DeCL

>Vitis_vinifera_NRPE1

MEEDSSTILDGEISGIRFGLATRQEICIASVSDCPI SHASQLTNPFLGLPLEFGKCESCGTAEPGQCEGHF
GYIELPIPIYHPGHVSELKRMLSLCLCKLKRKSKVTNNGITEQLLAPCCQDSPQVSVREFRPTGACFL
ELKIPSRSRPKDGFWDLARYGYRYGHNL SRILLPSEVMEILRRI PEDTRKKLVRKGYFPQDGYILQYLPV
PPNCLSVPI SDGVS IMSSDL SVSMLK KVLKQIEVIKGSRS GEPNFESHKIEANNLQSSIEQYLEVRGTAK
TSRSLDTRFRGSSKEPNESSTKAWLEKMRTLFI RKGSGFSSRSVITGDAYKRVNEI GLPFEIAQRITFEERV
NVHNMKHLQNLVDEKLCLTYRDGLSTYSLREGSKGHTFLRPGQVVHRRIMDGDIVFINRPPTTHKHSLOAL
SVYVHDDHTVKINPLICGPL **SADFDGD**CVHLFY PQSLGAKAEVLELFSVEKQLLSSHSGNLLQLATDSSL
SLKVL FERYFLNKAQQQLVMFVMSLPRPALLKSPCSGPCWTALQILQ TALPSYFDCIGERHWISK SAIL
KVDYNRDVLQSLVNEIVTSIFSEKGPNEVLKFFDSLQPLLMENLFSEGFSVSLEDFSI PSEVTQNIQKNVE
DISSLLYNLRSMYNELLQLQAENHLRLTKVPVANFILNSSALGNLIDSKSDSAINKVVQQIGFLGQQLSEK
GKFYSRLTVEGMAYLFKSKYPFHGADYPSGEFGLIRSCFFHGLDPYEEMVHSISTREIIVRSSRGLSEPGT
LFKNLMAILRVDVICYDGTVRNVCSNSIIQFEYGVKARTKPKQHFFPAGEPVGVLAAATAMSNPAYKAVLDSS
PSSNSSWELMKEILLCQVNFKNDLIDRRVILYLNDCCGRKYCRENAAYLVKNQLKKASLKDTAVEFMIEY
VKQHAVSGSSEPGTGLVGH IHLNKKLLQLDLNVSMQEVCKCEETINSFRKKNVGPFFKKIILSFRECCTF
QHSCQSKGSDMPCLLFFWQGNRDDNLEQILHILAHKICPVLLQTI IKGDSRVCTVNI IWISPDTTTTWIRNP
CKSRK GELALDIVLEKAAVKQRGD AWRIVLDA CLPVLHLIDTRRSIPYAIKQVQELLGISCAFDQAVQRLS
KSVTMVAKGVLKEHLI LANSMT CAGNLIGFNSGGYKALS RALNLQVPFTEATLFTPRKCFEKASEKCHTD
SLSSIVASC SWGKHVTVGTSRFDVLDWDTKEIGPAQDGGIDIYSFLHLVRS GSYGKEPDTACLGA EVEDLI
LEDENLELGMSPEHSSNF EKPVFEDSAEFQNTWENHVP GSGGDWAVNQNKETTASTLKPSAWSS **SWG**TDKVT
MKDTFSTREPDESSRSAGWDDKGT **TWGT**DKAQNTAFRRTHEDSPRSSGRDETFRDGRPQFASSAWGKKIDEA
DKT **GW**NKNDGKPKQMDKLR ESYDWDCKVAQEKTQSTYGGISSTTG DWKKNELQMEVVQHDES PVNEH SWDA
NLPEDPLAQATT SV **GW**DSSTGKDWTKRKLQSPSEQQRDPAIKSWSSSHNMKEQSNQPASTH **GW**DS PGAKG
WNDVEEQSQWNQRGSAVKNDQSESSH **GW**GPSNEQNQLPSSQ **GW**GPSNAGAGHESETQS **QW**QPSGKKSRPE
GSR **GW**GSNNTTEWKNKKNRPNKPKQGPLNDDYSAGGI FTATRQRVDI FTSEEQDILLDVEPIMQSI RIRIMHQA
GYNDGDPLSADDQSYILDKVFNNHPDKAVKMG TGIDYVMVSRHSSFLESRCFYV VSTDGHKED **DFS**YRKCLE
NFIKEKYPDNAETFIGKYFRRPRAGGNRERSVIPED **GGN**REOSVVP EETGSENRO

Metal A

H

(12) WG motifs

DeCL

(2) 12aa repeats

>Oryza_sativa_J_NRPE1-1 (OsJ_05410)

MEEDQSAIPVAEGAIAKSIKLSLSTED EIRTY SINDCPVTHPSQLGNPFLGLPLETGKCESCGASENGKCEG
HFGYIELPVPIYHPCHVTELRQILNVVCLKCLRVKKGKVKQTEGKDNTSALS CYCRDLPALSLKEIKTAD
GAFRLELKMPPRKFMT EGSWNFLDKYGFHGGTSHCRTLLPEEALNILKKIPEETKRKLAARGYIAQSGYV
MKYLPVPPNCLYIPEFTDQSIMSYDISISLLK KVLQKIEQIKKSRAGSPNFESHEVESCDLQLSIAQYIH
LRGTTTRGPQDNTKRFAISTDPSALSTKQWLEKMRTLFI SKGSGFSSRSVLTGDPYIGVDVIGLPSEVAKRI
TFEEQVTDINLNRLQEI VDKGLCLTYRDGQATYAITVSGKGTTLKVGQTI SRRIVDGDVVFLNRPPSTHK
HSLQAFRVYVHEDHTVKINPLICAPF **AADFDGD**CVHIYYPQSLAAKAEALELFSVEKQLTSSHSGKVNLLQ
VSDSLLALKHMSRTMLSKEAANQLAMLVTC SLDPAVIKSKPYWTISQIVQ GALPKALTSQGDKHVVRDS
TIIKLDLDKESVQTSFSDLVYSTLSVKGPGEALQFLNVLQPLLMELILLDGF SVSLQDFNVPKVLL EEAQK
NIEKQSLILEQSRFAENQV VEMRVDNNDKDIKQI SDFVVKRSHLGLLIDPKSDSSVSKVVQQLG FVGLQL
YREGKFYSRRLVEDCYT FVNKH PAVREEHSPEAYGLVRSSYFHGLN PYEELVHAISTREAI VRSRGLTE
PGTLFKNLMAILRVDVICYDGTVRNVCSKSI IQLNYTEDDALDFPSAIGPGE PVGVLAAATAISNPAYKAVL
DASQSNNTSWERMKEILQTTSRYKNDMKDRK VILFLNDCSCAKKFCKEKAAIAVQGC LRRITLED CATDIC
IEDGNWAAPAGFQHPVPPPQCKILPVPIPIPAHGSVKFPPVPIPAPEHLKYNIHVVRYQKQIGLDGTSEAA
PALVGH IHLDR AHLERINISTEDILQKCQEVSGKYGKKKGHL SNLFKNITFSTCDCLFTQKLVDGKLPKLP
CLQFFVSDNMIVSESV ERAVSVLADSLCGVLLNTIIKGDPR IQEAKIVVWGS DATSWVKNTQKASKGEP AV
EIIVEEEALHIGDAWR TTM DACIPVLN LIDIRRSIPYGIQQVRELLGISCAFDQVVQRLSTTVRMVAKDV
LKDHLVLVANSMTFTGNLNGFN NAGYKATFRSLKVOVPFTESTLITPMKCFEKA AEKCHSDSLGCVVSSCS
WGKHAASGTGSSFQILWNESQLKSNKEYGDGLYDYLALVRTDEEKARYTFFDDVDYLA EENEADVCLSP EL
DGTIGQPIFDDNLEEQDVQNNSSWDNGTTTNASWEQNGSAGNDSDK **WGW**NDAAAGADTGVTKPANQGN SC
WDVPATVEKSSSD **WGW**GTEKAKEKEKISEEPAQHDAWSVQGPKRATDGGASWKKQSSSTQNDGNSWKENKG
RGSNGGSWEKDNAQKGS **WGR**GNDEAENNDVQNKSWETVAADAHASTEKS **WGN**VTASPSDNAWSAAPVSQ
NGSSDTKQSDSWD **GW**KSAGVDKAINKDKE SLGNVPASPSFSAWNASPVSOGNERSDAKQSDSWD **GW**KSAGV
DKAINKDKE SLGNVPASPSFSAWNAAPVSQGNERLDAKQSDSWD **GW**KSAGVDDSVKDKE **WGN**VPASPSDS
AWNAAPVSOGNESSDAKQSDSWD **GW**KSAGVDASTNKDKE **WGN**VPASPSDSAWNAAPVSQDDVWNSAEAN
ESRNKDWKSD **WGW**ARGGNWRGQRNNPGRPPRKPDPGRGLPRRPDERGPPRRHFDLTAE EEEKILGEI EPTVLS
IRKIFRESIDS IKLSPEDEKFIKENVLEHHPEKQSKVSGEIDHIMVDKHQVFQDSRCLFV VSSDGT RS **DFS**
YLKCMENFVRKTYPEHGDSFCKKYFKRRRDQPPAADGGTAPGTPAGATQSTAVDTQEGTSQQTQDPDIATAP

Metal A

H

(4.3) 52 aa repeats & (13) WG motifs

DeCL

AATQQETLQDTPAPPADDGLLGKGPSPSD

>Oryza_sativa_J_NRPE1-2 (OsJ_04874)

MEGHPDPTSAATAMIPEASIRRINLSITSNEEILKAQPVNELEKPIPIITHQSOLLNNPYLGLPLQVGCQS
CGSNAIEECEGHFRFIELPMPIFHPSHVTELSQILNLIICLRCLKIKNRKELPPLCVAEVKKSNGARGLELR
APIKKELEEGFWSFLDQFGSCTRGTSHCRPLLPEEVQNI IKKIPEETRRWLSVRGYIPQDGFILSYLCVPP
NCLRVS NVLDGNTFSCSGTSTNLLRKALRKIQQIRGSRIGSSNIQVDQVADDLQVDVANYINLGGTTKGGH
DDTFTSQPTAMQWKQKMKTLFISKSSSFSSRGVITGDPYIIGLNVVGVPEEVAKRMSVEEKVTDHNI AQLQD
MMNKGLCLTYTDANSITYSLDAGKDNPNKKHTILKVG EIVNRRVFDGDIVFLNRPSTDKHSVEAFYVQVH
NDHTIKINPLICDPL**GADFDG**DCVQIFYPRLSARAEAKELYTVDKQLVSSHNGKLNQFQKND FSLALKIM
CGREYSEREANQITNAMFSSGMYPQKPLIGGPYWTFPQILETTKSNAITLADHLDRSEV GALATGTTISSI
LSTKGPREATEFLNLLQPLLMESLLIDCF SINLGDFTVPSPILEAIQNNPLELNKYREPIMDFITHSSAIG
LLVDPKSDSNMKNVVEQLGFLGPQLQHNGRLYSSRLVEDCLSKSLHRCCGSTNCCNPLEEYGTVRSSIIYHG
LNPYEALLHSICEREKIMRASKGLVEPGSLFKNMMSRLRDVTACYDGSIRTSSGNLVLQFGSRDASNCVTP
GDPVGILAAATAVANAAYKAVLAPNQNNII SWDSMKEVLLTRASTKADANHRKVILYLNQCSCENECMERAL
TIRACLRRIKLEDCTTEISIKYQQQATQAAHHLVGH IHLDKQLNQIETIMDSVLHKCQETFRNNIKKKS
MREILKTVTFISSTSLCDQHTDDD KKFQVSCLOFFLPGSITKNI SESTERVIDFMTNAIFPIILD TVIKGD
PRVEEANLVRIEPESTFWVQSSGAEQKGEAALEITVEEAAA AESGNAWGVAMNACIPVMDLIDTTRSM PYD
IQQVRQYLSKSVGMITKSVLQEH LTTVASSMCTGDLHGFNNSGYKATCQSLKVOAPFMEATLSRSIQCFE
KAAAKAYS DQ L GNVVSACS **WG**NNAEIGTGSAFEILWNDENMSSSKSILGGYGLYDFLEAVETT GATKDKAI
VPHNYCLYDVDCI PEDKVCLEENNQITWTDKPKAEFLMESEGRRAGMHSTGQKHPRKPNWHEGNTKSSPNS
TAVEFTGQVFQRRQLKTKSNWNSDATQQDDKPSWYSSNSAGTQNF T IAGSSRPGEWNRKNNNRGQGGGREGV
WKSEGP H RGGSSSNRNQGGG RAVWKSEASHRGSGNNRNRGGG RAVWKSEASRRGGSMRQVASC AFTPV EQQ
IFEQIEPITKNVKRI IRESRDGIKLPDDEKFI VTNVLMYHPERKKKIAGNGNYITVDRHQVFHGSRC L YV
MSSDGSRK**DFSYKK**CLENYIRAQYPDAADSF CRKYFK

Metal A

H
(3) 22aa repeats
DeCL

>Oryza_sativa_I_NRPE1-1 (OsI_05888)

MEEDQSAIPVAEGA IKS IKLSLSTED EIRTY SINDCPVTHPSQLGNPFLGLPLETGKCESCGASENGKCEG
HFGYIELPVPIYHPCHVTELRQILNVVCLKCLRVKKGVKQTEGKDNTSALSICYCRDL PALSLKEIKTAD
GAFRLELKMPPRKFMTEGSWNFLDKYGFHHGGTSHCRTLLPEEALN I LKKIPEETKRKLAARGYIAQSGYV
MKYLPVPPNCLYIPEFTDGQSIMSYDISISLLKVLQKIEQIKKSRAGSPNFESHEVESCDLQLSIAQYIH
LRGTTTRGPQDNTKRFAISTDPSALSTKQWLEKMR TLFISKSGSFSSRSVLTGDPYIGVDVIGLPSEVAKRI
TFEEQVTDINLNLQEI VDKGLCLTYRDGQATYAITVSGSKGHTTLKVGQTI SRRIVDGDVVFLNRPPSTHK
HSLQAFRVYVHEDHTVKINPLICAPF**AADFDG**DCVHIYYPQSLAAKAEALELFSVEKQLTSSHSKVN LQL
VSDSLLALKHMSSRTMLSKEAANQLAMLVTCSLPDAVIKSKPYWTISQIVQ GALPKALTSQGDKHVVRDS
TIIKLDL DKE SVQTSFSDLVYSTLSVKGPGEALQFLNVLQPLLMELILLDGF SVSLQDFNVPKVLL EEAQK
NIEKQSLILEQSRFAENQV VEMRVDNNLKD IKQI SDFVVKRSHLGLLIDPKSDSSVSKVVQQLG FVGLQL
YREGKFYSRRLVEDCY YTFVNKHPAVREEHSP EAYGLVRSSYFHGLN PYEELVHAISTREAI VRSSRGLTE
PGTLFKNL MALLRDVVICYDGTVRNVCSKSI IQLNYTEDDALDFPSAIGPGE PVGVL AATAISNPAYKAVL
DASQSNNTSWERMKEILQTT SRYKNDMKDRKVI LFLNDCSCAKKFC KEKAAIAVQGC LRRITLED CATDIC
IEYQKQIGLDGTSEAAPALVGH IHLDR AHLERINISTEDILQKCQEVSGKYGKKKGHLSDPRIQEAKI VVW
GSDATSWVKNTQKASKGEPAVEIIVEEEEEALHIGDAWRTTMDACIPVLN LIDIRRSIPYGIQQVRELLGIS
CAFQDQVQRLSTTVRMVAKDVLKDHLV LVANSMFTFTGNLNGFN NAGYKATFRSLKVQVPFTESTLITPMKC
FEKAAEKCHSDSLGCVVSSCS **WG**KHAASGTGSSFOILWNESQLKSNKEYGDGLYDYLALVRTDEEKARYTF
FDDVDYLAEE NEADVCLSP ELDGTIGQPI FDDNLEE QDVQNNSSWDNGTTTNASWEQNGSAGNDS DK**WGGW**
NDAAAGADTGVTKPANQGN SCWDVPATVEKSSSD**WGGWG**TEKAKEKEKISEEPAQHD AWSVQGPKRATDGG
ASWKKQSTQNDGNSWKENKGRGSSNGGSWEKDNAQKGS**WGR**GNDEAENNDVQNKSWETVAADAHASTEKS
WGNVTASPSDNAWSAAPVSQGNSSDTKQSDSWD**GW**KSAGVDKAI NKDKESLGNVPASPSFSAWNASPV SQ
GNERSDAKQSDSWD**GW**KSAGVDKAI NKDKESLGNVPASPSFSAWNAAPVSQGNERLDAKQSDSWD**GW**KSAG
VDDSVKDKES**WGN**VPASPSDSAWNAAPVSQGNESSDAKQSDSWD**GW**KSAGVDASTNKDKES**WGN**VPASPSD
SAWNAAPVSQGDVWNSAEANESRNKDWKSD**GW**GARGGNWRGQRNNPAEEEEKILGEIETT VLSIRKIFRES
IDSIKLSPEDEKFIKENVLEHHPEKQSKVSGEIDHIMVDKHQVFQDSRCLFVVSSDGT RS**DFSYLK**CMENF
VRKTYPEHGDSFCKKYFKRRRDQPPAADGGTAPGTPAGATQSTAVDTQEGTSQQTQPD IATAPAAATQQETL
QDTPAPPADDGLLGKGPSPSD

Metal A

H
(4.3) 52aa repeats &
(13) WG motifs
DeCL

>Oryza_sativa_I_NRPE1-2 (OsI_05331)

MGFSPAISLRNLSMVALRIWESGTTVYIAAAAVPGAKLVVLVILITAGRPFLLTHLYMRYSRTEMEGHPDPT
SAATAMIPEASIRRINLSITSNEEILKAQPVNELEKPIPIPTHQSOLLNNPYLGLPLQVGDIAIEECEGHFGF
IELPMPFIHPSHVTELSQILNLI CLRCLKIKNRKVQNI IKKIPEETRRWLSVRGYIPQDGFILSYLCVPPN
CLRVS NVLDGNTFSCSGTSTNLLRKALRKIQQIRGSRIGSSNIQVDQVADDLQVDVANYINLGGTTKGHGD
DTFTSQPTAMQWKQKMKTLFISKSSSFSSRGGVITGDPYI GLNVVGVPEEVAKRMSVEEKVTDHNI AQLQDM
MNKGLCLTYTDANSITYSLDAGKDNPNKHTILKVGEIVNRRVFDGDIVFLNRPPSTDKHSVEAFYVQVHN
DHTIKINPLICDPL**GADFDDDC**VQIFYPRLSARAEAKELYTVDKQLVSSHNGKLNQFKNDFSLALKIMC
GREYSEREANQITNAMFSSGMPYQKPLIGGPYWTFPQILETTKSNAITLADHLDRESV GALATGTTISSIL
STKGPREATEFLNLLQPLLMESLLIDGFSINLGDFTVPSPILEAIQNNPLELNKYREPIMDFITHSSAIGL
LVDPKSDSNMNKVVEQLGFLGPQLQHNGRLYSRRLVEDCLS KSLHRCCGSTNCCNPLEEHGTVRSSIYHGL
NPYEALLHSICEREKIMRASKGLVEPGSLFKNMMSRLRDVTACYDGSIRTSSGNLVLQFGSRDASNCVTPG
DPVGILAATAVANAAYKAVLAPNQNNII SWDSMKEVLLTRASTKADANHRKVILYLNQCSCEENECMERALT
IRACLRRIKLEDCTTEISINTSLCDQHTDDDQEFRVSLQFFLPASITKNISESTERVIDFMTNAIFPIIL
DTVIKGDPRVEEANLVRIEPESTFWVQSSGAEQKGEVALEITVEKAAAESGNAWGVMADACIPVMDLIDT
TRSMYPDIQQVRQYLSKSVGMITKSVLQEHLLTVASSMTCTGDLHG FNNSGYKATCQSLKVQ**APFMEATLS**
RSIQCFEKAAAKAYS DQLGNVVSACS WGNNTEIGTGS**AFEILW**NDENMSSSKSILGGYGLYDFLEAVETT
ATKDKAIVPHNYCLYDVDCIPEDKVCLEENQITWTDKPKAEFLMESEGRRAGMHSTGQKHPRKPNWHEGN
TKSSPNSTAVEFTGQVFQRRQLKTKSNWNSDATQODDKKPSWYSSNSAGTQNFYIAGSSRPGEWNRKNNNR
GQGGGRAVWKSEGPHRGGSSSNRNOGGGRAVWKSEASHRGSNNRNRGGGRAVWKSEASRRGGSMRQVASC
AFTPVEQQIFEQIEPITKNVKRI IRESRDG IKLPDGEKFIVTNVLMYHPERKKKIAGNGNYITVDRHQVF
HGSRCLYVMSSDGSRK**DFSYKK**CLENYIRAQY PDAADSFCRKYFK

Metal A

H

(3) 22a repeats

DeCL

>Zea_mays_NRPE1

MEEDHSVILISEGAIKSIKLSLSTGEEICTYSINECPVTHPSQLGNPFLGLPLEAGKCESCGASENDKCEG
HFGYIELPVPIYHPCHVTELRQLLSLICLCLRIKKGKDI PALSLKEIKTTDGAIRLELRAPHNKHMTERS
WNFLDKYGFHHGGC SHHRTLLPEEALNLIKVPDDTRRKLAAARGYIVQTGYVMKYLPVPPNCLYIPEFTDG
QSIMSYDISIALKKVLQKIEQIKRSRSGSPNFESHDAESCDLQLAIGQYIRLRGTTTRGPQDNTKRFTVGS
ADSAALSTQWLEKMYTLFISKSGSFSRSGVLTGDPYI GLGVVGLPSEVAKRMTFEEQVTDININRLQDVV
DKGLCLTYRDQATYAITVGSKGYYTTLKVGQTI SRRIYDQDVVFLNRPPSTHKHSLQAFYAYVHDDHTVKI
NPLMCGP**SADF**GD CVHIYYPQSLAAKAEALELFSVERQLISSHSGKVNQLGNDSL VAMKAMSHTTMLH
KELANQLAMFVFPFSLAPAVIKPVPSWTISQIVQGAFFANLTCQGDTHLVRDSTIIRLDLGKESVQDSFPD
LVSSILREKGPKEALQFLNVLEPLLMEFLLLDGLSISLRDFNVPKALLEEAQKDIRNQSLILEQSRCSTSQ
FVEFRVENNLKKNVQQISDSVKGKFDLGLLIDPKKEASMSKVQVQVGFVGLQLYREGKLYSRRLVEDCFTN
FVNKHLAIGDEYPPEAYGLVQSSYFHGLNPYEELIHAISTREAMIRSSRGLSEPGTLFKNLMAILRDVVIC
YDGTVRNICSNSIIQLKYGEDDETSSSVVPPGEPVGVLAATAISNPAYKAVLDSSQSNNASWESMKEILQ
TRTSYKNDVKDRKVVFLNDCSCAKKFKERAALAVQ SCLKRVTLGDCATDICI EHQKQINLDGTSEAAPT
LVGHIHLDKGHLERINISTQDILQKCQEMPIDGKLHKVPCVQFAFSDDIVLSESIERAVNVIADSVCSVLL
DTIIKGDPRIQAAKVIWVESDAASVWKHTRKVS KGESALEIIVEKDDAVSNGDAWRTAIDA CLPVLNLI DT
RRSIPYGIQQVRELIGISCAFDQVVQRLSTTVKMVNKGVLKDHLILVANSMTCTGNLIGFNIAGYKATFRS
LKVQVPFTESTLFTPMKCFEKAAEKCDSDSLGCVVSSA**WGK**HAAVGTGSSFOILWNEINQVCLSYQPELIA
YISLYQTDYMFLLDDVDYLVEENAADDMCLSPEDGTLGKPTFEDNFEEQNIQKGS SWEIGITTNSSWEQNA
SVANDSGD**WGGW**SSGGGAAAKPADQDNSWEVHAKVQDNSTTD**WGGW**SVEKPTGEATVSGEPAETDTWADKG
AKMESDAGDGNWEKSSSTPEASKKNDSSSENTWDKRKGDDGAWGNRSDDGHGNWEHPSNWNQSLDQDQDT
WGNARGKKKADGNYCQWEEQPSNYKQKKTNADHDSYNNVMPSSEIAWNAGDGTGRPNAKSNAESS**WGEED**
KMESDDHPKVPKESDTWNTGRSNE SPWDNTDALQDSWVKSAAARNNTQDGSWDKVVSMKDLDSLQDSWSKA
TIQTNDAQNDSDWNVAKNAPDSAAEDS**WGA**ATPAETDTSGNKEWKS**DGWG**AKSGNWSSQRNNPGRPPRRPD
ERGP PPRQRFELTVAEKNILLEVEPIKLRVRSI FREACDGVRLNPEDEKFI LEKVLEHHPEKQSKVSGEI
DYLTVNKHQTFQDTRCFFVSTDGSQ**DFSYLK**CLENFVRKSYTEDADTFCKMYLRPETEQGTPPAPQAE
VPQETWGS PAVPLEGGTHIAGPDSTGDAVILGEQHDLTPASPAVAPQVASEPDTTDTGTLGKAPQAD**WGP**
RFDAD

Metal A

H

(2) 27aa repeats & (10) WG motifs

DeCL

>Solanum_lycopersicum_NRPE1 (DQ020653) - incomplete N- and C-termini
DFDGDCVHLFYPQSLSAKAEVLELFAVGKQLLSSHTGNFNSQLATDSLSSLLKLMFSSHFFDKAAAQQLAMF
 LPMALPDSAVVDVRKSGAMWTTLQILGAALPDGDFDSCGETHTIGKSQFLGIDYHRDLISSILNDVITSIYF
 MKGPNVDLKFNFSLQPLLMENLCTEGFSISLRDFYMTKAVRDGIQERIQCMSKLLHHLRSSYNESVEVQLE
 HHLRNEKLPVIDFVLKSSGMGVLIDSKSESFAFNKVVQQIGFLGLQISDRGKFYXXTLVHDMALFQKKYPS
 VGTNPSEEFGLVRSCLFYGLDPYQGMIIHSSSREVIVRSTRGLTEPGTLFXNLMAILRDVVICYDGTVRNV
 SSNSIIQFEYGSSSGSNLPSEFCAGDPVGVLAATAMSNPAYKAXLDSSPSSNSSWEMMKEILLCGVSFKND
 VSDRRVILYLNDCCGRGRCREKAAAYVVKHLSKVCLKDAADFLIEYAGRQAGYENSETGTGLIGHIRLN
 QGQLENLGISVLEVHERCQENISSFRXKKKIGNLFRKRVLSVSEFCFCHNSGSKCLNAPCLRFVSWPDASD
 DHLERVSHILADMIXPILLDTVIKGDPRVSSANIAWISPDMTSWIRSPSKSQRGELALDIVLEKEAVKXRG
 DAWRXLMDNCLPVIHLIDTTRSIPYAIKQVQELIGISCAFEQAVXRLSTSVTMVTKGVLLKDHVLLANSMT
 CAGNLVGFNAGGIKALSRSLNVQIPFTEATLFTPRKCFERAAEKCHVDLSLSSIVASC**SWGKHVAVGTGSRF**
EVLLNTRNVEWNIPDTRDVYSFLHLVRNTSAQEVEGTSCLGAEIIDELEEDEDMGLYLSPNRDSGSEMPTFE
 DRAEFDYNENLDEGKPSGSAWEEASSGSVKSGSWDMAGKTQNGAEEGVNQSDSWSS**WG**KKVDEPENNRQQ
 SGSGEQSGSWSPWGRRWKMMVVLGDEPKQLNSESS**WG**KAPNGGGLGSATAEGNRRLDQSVNDWSSSVSRDG
 QYKKWWLEFFKRWWLELS**GW**QWKNNRPARSADDSNRGGHFTATRQKIDLFTAEEQEIIISDVPIMLKVKS
 DPLSADDQSYIIDTVLNYHPDKAVKMGAGLDYITVSKHTNFQDTRCFYVVSTDGAKQELAAV

Metal A

H

(3) WG motifs

>Glycine_max_NRPE1-1 (Glyma15g37710)
 MEDNPPSSVLDGTVVGIKFGMATRQEICTASISDSSISHASQLSNPFLGLPLEFGRCESCGTSEVGKCEGH
 FGYIELPIPIYHPSHISDLKRMLSMVCLNCLKLRKTKLPASSSGLAQRLISPCQEDKAALVSIREVKTSD
 GACYLALKVSKSMQNGFWSFLEKYGYRYGGDHTRALLPCEAMEIIKRIPIETKKKLAGKGYFPQDGYVLK
 YLPVPPNCLSVPEVSDGVSVMSSDPSITILRKLRLKVEIIKSSRSRGEPNFESHVVEANDLQSVVDQYFQIR
 GTSKPARDIETHFGVNKELTASSTKAWLEKMRTLFIRKSGSFRSRVITGDCYKRINEVGIPEVEAQRITF
 EERVNIHNIYRLQKLVDEHLCLTYKEGGSTYSLREGSKGHIYLPKQIVHRRIMDGDIVFINRPPTTHKHS
 LQALYVYIHEDHTVKINPLICGPL**GADFDGD**CVHLFYPQSLAAKAEVLELFSVENQLLSSHSNGLNLQST
 DSLSSLLKMLVKRCFFDRAANQLAMFILLPLPRPALLKASSGDACWTSIQILQCALPLGFDCTGGRYLIRQ
 SEILEFEFSRDVLPATVNEIAASVFFGKGPKEALNFFDVLQPFMLMESLFAEGFSVSLEEFISIIRAIR
 KSIDGVSSLLYQLRSLYNELVAQOLEKHIRDVELPIINFALKSTKLGDLIDSKSKSAIDKVVQQIGFLGQQ
 LFDGRGFYKGLVLDVASHFHAKCCYDGDGYPSEAYGLLKGCFNGLDPYEEMVHSISTREIMVRSRGLS
 EPGLTFKNLMAILRDVVICYDGTVRNICSNSIIQFEYGIQAGDKSEHLFPAGEPVGVLAAATAMSNPAYKAV
 LDASPSNSSWELMKEILLCKVNFRELVDRRVILYLNDCCGGSYCRENAAYSVKDQLRKVSLKDAAVEF
 IIEYQQQRTQKENSETDVGLVGHIIYLDDEMMLEELKISMAVYFDKCHERLKSFSQKKKKMTLFLSYLIVRG
 TVKCSIFVVSRIQDLYFIDHEYCTWKTVMVFLSVSETIKNEIFPGLFMTISYLLFFTIPTESCSSHPAAPC
 LTFWLKNYDSDLNNAVKVLAEKICPVLFKTIIQGDPRISSASIIWVSPDTNTWVRNPKSSNGELALDIIIL
 EKEAVKQSGDAWRVVLDAQLPVLHLIDTTRSIPYAIKQIQELLGISCTFDQAIQVAASVKMVAKGVLRH
 LILLASSMTCGGLVGFNIGGYKALSRQLNIQVPFTDATLFT**PKKCFERAAEKCHTDSLSSIVASCWGKH**
VAVGTGSKFDVVDANEIKSNEIEGMDVYSFLHMVKSFTNGEEETDAQLGEDIDDLLEEEYMDLGMSPQHN
 SGFEAVFEENPEVLNGSTSN**GW**DVSSNQGESKTNEWS**GW**ASSNKAEIKDGRSEIAPKNS**WG**KTVNQEDSSK
 SNPWSTSTIADQTKTKSNEWS**AWG**SNKSEIPV**GW**ASSNKTEIKDGRSETAQENS**WG**KTVNQEDSSKSNAWN
 TSTTVDHANTKSNEWS**AWG**SNQSEIPAGGSKAVQEDS**WG**SSKWKADVAQEDNSRLGAWDANAADQTKSSEW
SGWGKKKDVTQEDNSRLGAWDANAADQTKSRDWS**GWG**KKKDIQEDNSRLGAWDANAADQTKSSEWS**GWG**
KKDQIRQNLNMNGQVGERRKKLPKKTIPGLVLGMQIQQIRQNLNMNEDQTKSNEWS**GWG**KKKDVQEDNSRLG
 AWDANAADQTKSNEWS**DWG**KKKEVTQEDNVQDS**SWG**SGKRKDKVTQEDNSGSG**GW**GANRTDLAKSKSSEWSS
WGKNKSEIPAGGSENVQND**SWG**SGKLEDDTQKENSASVVRNKAETIDGGSEKQEDAWNNGNKAESKVG
 NAS**WG**KPKSSESQAWDSHNQSNQNSSSQ**GW**ESHIASANSESEKGF**QWG**KQGRDSFKKRFEGSQGRGSNAG
 DWKNNRPRAPRQRLDIYSSGEQDVLKDIPIQSIIRIMQSQGYNDGDLAAEDQLFVLENVFEHHPDK
 ETKMGTGIDYVMVNKHSFQESRCFYVVKDGE**DFSYR**KLANYISKYPDLAESFLGKYFRKPRARGD
 QTATPGRDEAATPGEQTATPGRDEAATPAEQI STPTMETNE*

Metal A

H

(2) 66aa repeats, (3) 33aa repeats & (19) WC motifs

DeCL

(2) 15aa repeats

>Glycine_max_NRPE1-2 (Glyma13g26690)
 MEI I K R I P I E T K K K L A G K G F F P Q D G Y V L K Y L P V P P N C L S V P E V S D G A S V M S S D P S M T I L R K L L R K V E I I K S
 S R S G E P N F E S H H V E A N D L Q S V V D Q Y F Q I R G T S K P A R D I E T H F G V N K E L T A S S T K A W L E K M R T L F I R K G S G F
 S S R N V I T G D C Y K R I N E V G I P V E V A Q R I T F E E R V N I H N I R Y L Q K L V D E H L C L T Y K E G V S T Y S L R E G S K G H I Y
 L K P G Q I V H R R I M D G D I V F I N R P P T T H K H S L Q A L Y V Y I H E D H T V K I N P L I C G P L **G A D F D G D C V H L F Y P Q S L A**
 A K A E V V E L F A V E N Q L L S S H S G N L N L Q L S T D S L L A L K M L V K R C F L G R A A A N Q L A M F L L L P L P R P A L L K A S S D
 D A C W T S I Q I L Q G A L P M G F D C T G G R Y L I R Q S E I L E F D F S R D A L P A T I N E I A A S I F F G K G P M E A L K F F D V L Q P
 F L M E S L F A E G F S V S L E E F S I S R A I K R I I R R S I G K A S S L Y Q L R S L Y N E L V A Q Q L E K H I Q D V E L P I I N F A L K
 S T K L G D L I D S K S K S T I D K V V Q Q V G F L G Q Q L F D R G R F Y S K G L V D D V A S H F H A K C C Y D G D G Y P S A E Y G L L K G C
 F F N G L D P Y E E M V H S I S T R E I M V R S S R G L S E P G T L F K N L M A I L R D V V I C Y D G T V R N I C S N S I I Q F E Y G I Q A G
 D K T E H L P A G E P V G V L A A T A M S N P A Y K A V L D A S P N S N S S W E L M K E I L L C K V N F R N E P V D R R V I L Y L N D C D C
 G G S C C R E N A A Y S V K N Q L R K V S L K N A A V E F I I E Y Q Q R T Q K E N S E T D A G L V G H I Y L D E M M L E E L K I S M A N V F
 E K C L E R L K S F S R K K A R Q S F L I I R G T V N E S C S S S H P A A P C L T F W L K N H S D L D N A V K V L S E N I C P V L F E T I
 I K G D P R I S S A S I I W V S P D T N T W V R N P Y K S S N G E L A L D I V L E E E A V K Q S G D A W R I V L D S C L P V L H L I D T R R S
 I P Y A I K Q I Q E L L G I S C T F D Q A I Q R V A A S V K M V A K G V L R E H L I L L A S S M T C G G N L V G F N T G G Y K A L S R Q L N I
Q V P F T D A T L F T P K K C F E R A A E K C H T D S L S S I V A S C S W G K H V A V G T G S K F D I V W D S S E V F D N T D L I L D L I R I
 G I K S N E I E G M D V Y S F L H M V K S V T N G E E E T D A C L G E D I D D L L E E E Y M D L G M S P Q H N S G F E A V F E E N P E V L N G
 S T S N G W D V S S N Q T Q S K T N E W S G W A S S N K D G R S E T A Q E N S W G K T V N Q E D S S K S N A W N T S T T A D Q T K T K S N E W
S D W G S N K S E I P A G G S K A V O E D S S K S N A W N T S T T S N O T K T K S K E W S A W G S N K S E I P A C G S K A V O E D S S K S N T
W N T S T T A D Q T K T K S N E W S A W G S N K S E I P A G G S K A V O E D S S K S N A W N R S T T A D Q T K T K S N E W S A W G S N K S E I
P A G G S K A V O E D S S K S N A W N T S T T A D Q T K T K S N E W S A W G S N K S E I P A G G S K A V O E D S S K A W N T S T T A D Q T K T
K S N E W S A R V S N K S E I P A G G S K A V O E D S W G S S K W K A D V A Q E D N S R L G A W D A N A A D Q T K S N E W S G W G K K K D V T
 Q E D N V Q H S W G S G K R K D K V T Q E D N S G S G D W G A N R T D L A I T K S S E W S S W G K N K T E I P A G G S A N V Q N D S W G L G K
 L N D T Q K D N S G C G A W G E N S G S A W P Q E D A W N S G N W K A E S K V G N T T W G K P K S S E S H A W D S H N Q S N Q N S S S Q G W E
 S H I A S A N S E N E K G F Q W G K G R D S N R P P R A P G Q R L D I Y S S E E Q D V L K D I E P I M Q S I R R I M Q Q Q G Y S D G D P L A A
 E D Q L F V L E N V F E H H P D K E T K M G A G I D Y V M V N K H S S F Q E S R C F Y V V C K D G Q S K **D F S Y R K** C L A N Y I S K K Y P D L
 A E S F L G K Y F R K P R A R G D Q T A T L G G D Q T A T P A Q D E A A T S G P G Q R Q E *

Metal A

H

(6) 44aa repeats & (18) WG motifs

DeCL

(2) 8aa repeats

>Brachypodium_distachyon_NRPE1 (Bradi4g45070 and Bradi4g45060)
 M E E D Q S A V L V A E G A I K S I K L S L S T E D E I L T Y S I N D C P V T H P S Q L G N P F L G L P L E T G K C E S C G A S E N G K C E G
 H F G Y I E L P V P I Y H P C H V S E L R Q L L S L V C L K C L R I K K G K A K Q S N G K E N V S V T A C S Y C R D V P A L S L K E V K T A D
 G A F R L E L R A P P R R L M K D S S W N F L D K Y G F H H G G A S H F R T L L P E E A L N I L K K I P D D T R K K L A A R G Y I A Q S G Y V
 M K Y L P V P P N C L Y I P E F T D G Q S I M S Y D I S I S L L K K I L H R I E Q I K K S R A G T P N F E S H E A E S S D L Q I S I A Q Y I H
 L R G T T K G P Q D T K R F T I S T D S S H L S T K Q W L E K M R T L F I S K G S G F S S R S V L T G D P Y I G V D V V G L P S E V A K R I T
 F E E Q V T D I N I K R L Q E V V D K G L C L T Y R D G Q T T Y A I T V G S K G Y T T L K V G Q T I S R R I V D G D V V F L N R P P S T H K H
 S L Q A F Y V Y I H D D H T V K I N P L I C S P L A A D F D G D C V H I Y Y P Q S L A A K A E A L E L F S V E K Q L T N S H N G K V N L Q L S
 N D S L L A L K H M S S R T V L S K E S A N Q L A M L L S F S L P D A V V K L K P C W T I T Q I I Q G A L P A A L T C E G G R F L V K D S T
 V I K L D L A K E S V Q A S F S D L V S S I L C V K G P G G A L Q F L N A L Q P L L M E Y L L L D G F S V S L Q D F N V P K V L L E E V H K S
 I Q E Q S L V L E Q S R C S K S Q F V E M R V D N N L K D V K Q Q I S D F V V E S S H L G L L I D P K S E P S M S K V V Q Q L G F V G L Q L Y
 R E G K F Y S S R L V E D C F S S F V D K H P P I V G N Q H P P E A Y G L V Q N S Y F H G L N P Y E E L V H S I S T R E A I V R S S R G L T E
 P G T L F K N L M A I L R D V V I C Y D G T V R N I C S N S I M Q L K Y N E D D A T D I P S A L T P G E P V G V L A A T A I S N P A Y K A V L
 D A S Q S N N T S W A S M K E I L Q T K V S Y K N D T N D R K V I L F L N D C S C P K K F C K E K A A I A V Q N R L K R V T L E D C A T D I C
 I E Y H K Q I L D G S S E A T P A L V G H I H L E K A R L D M I N V S T E D I L Q K C Q E V S L K H G K K K G H L G H L F K K I T F S T C D C
 S F T Q K P M I D G K L P K V P C L Q F S F S E D I P M L S E S V E R A V S V L A N S L C D V L L D T I I K G D P R I Q E A K I M W V G S D A
 Q S W V K N T R K V S K G E P T V E I V V E K N E A S K Q G D A W R I A M D A C I P V I D L I D T R R S I P Y G I Q Q V R E L L G I S C S F D
 Q I V Q R L S T T M K T V A K G I L K D H L I L V A N S M T C T G N L Y G F N T G G Y R A T F R A L K V Q V P F T E S T L F T P M K C F E K A
A E K C H S D A L G C V V S S C S W G K H A A L G T G S S F Q I L W N E N Q L K S N K E Y G D G L Y D F L A M V R T D Q E K A R Y T F L D D V
 D Y L V E D N A M D D I C L S P E L N G T H G V P T F E D N F E H Q D T Q N G N S W E N G T K A N A S W E Q N A S A G N D S D N W G G W S N A
 A A A A D T G A A K P A D Q G N S S W D V P A T A E N D S T D W G G W G N E K A K D N R T V S T E P A E L D T W S D R G A K K G T D G G G G S
W G K Q T N T C E D S G T N L E R N S W A K R P S S P S L S T W A K K N S D G G D G T W D K Q A N S C K K N V E Q D S W K N M P V S P A R N A
 W N K K E S R G D A T W E M R A S T L E E K K T S E S N E G S W E K S N A Q K D S W G N T Q H G S S D K M A V K D N D M Q Q D P W G H I A T
 Q N I N A Q D D L W G S V A A K A Q T S T A E N T D A Q D D S W G A V A A K A Q T S T A Q E S W G N V A A S P S D N A W K A P P I S Q T S A A
 E H T D A H N D S W G I V A A K A Q T S T A Q Q E S W G N A T A S P S D N A W N A A P M D L D A K Q P G S W D G W S S A L A E D S N K A D D S
 S N K N K G W K S D G W G A K G N R R D Q R D N P S M P P M R P D E R P P R P R F E V P A E A K K I L R E I E P I V S M V R K I F R E S C D G
 V R L P L E D E K F I K E S I L E H H P E K E R K V P G E I D H I M V N K H H I F Q E S R C F Y V V L A D G T H T **D F S Y N K** C M D N Y V R K
 T Y T D A A E H A D L V S Q M Y F K K R D R D R A A A V D G G S T P A N A S Q S T Q V M E T S Q D E A P Q E A Q P E T C V A T Q E E T R V S P
 Q E T P A A T T Q Q E E T E N N P D S A S E A D Y H S A S E A G L P E G V

Metal A

H

(2) 22 aa repeats & (15) WG motifs

DeCL

>Sorghum_bicolor_NRPE1 (Sb03g046922)

MEDDDPAAAGLTVPEAFIRRVKLSVTSNQEIVSTSPLFSPQDPIPIITHCSQLQDNPSLGLPLQDGSTCESC
GATQLDKCDGHFGFIKLEPIYHPSHIAELGKILNLVCLRCLRLKPKKVTGKESRFTSCSYCQELSPLCV
SQVKKSNARSLELKLPLKQEVADGFWSFLDQFGFHTSGTSHRRPLHPKEVQDIMKKITEKTRARLAARGY
NLQDGFVMDNMSIPPNCLOISNMLDENTEMCPPTSKGLLHKVLRITIEQIESLNI SHPNIEARELGADDLQV
AVADYMMMGAAKVSQHVTFTTRQPAPKQWHKMKTLFSLKSSSYTCRAVITGDPYIIGLDVVGVPDEIARM
SVQECVTNYNIARLQDMMNKGLCLTYTDLNTNTYDLGKKGKNCIMLRVGETVDRRLVDGLVFLNKPPS
TDMHSIQALYVHVHDDHTIKINPLICGPLEADFDGDCVHIFFPSSVLARVEAAELFAVEKQLLNSHNAKLN
FQIKNDYLLALRIMCDRSYSKEKANQIAMFSSGMI PPCNPWTICDRWTIPQILQTTDALRIVPSHPNTVGA
SVTAIITSTLSEKGPRAEIKLINLLQPLLMESLLMDGFSISLKDLDGQSAMQKANQSSISLEIDKFSKSIDV
FIANSALGLLVDPKNDLSALMNLVEQVGLGYLQESTDRLYSNNLVEDCYNFLEKRSRGSTKCYDPPKGHDF
VTSSFYNGLNPIYEELLHSISVREKIERSSSKGLAEAGNLFKNMMAMLRDVTVCYDGMTRTSYNNSIVQFDS
TNVSSSLTPGDSIGILAATVFANAAYKAVLVPNQKNMTSWDSMKEVLLTNACSKTGTIDQKAILYLKNCFC
GLKFCSELAHRVQSCLKRIKLEYCAIEVSIKYQQEATQAAQCLVGHIIHLDKQLNWMEITMGNILQTCQK
NVNKHVMKNRQLMQILKTTEIISSEYCLCGQDIGDERALQVSCLQCFIHASTTTVQPESNVIQMMTNTIFP
ILLDVTIKGDPQVQEAkliwvePKLTRWVKNSAEQKGElaveITVEKIAAAENGGTWGVVMDACVPVMDL
IDTTRSAPCNIQEVQKVFGISSVFDRVVQFLMFCPLGFFQHLKAVGMVTKSVLMEHLITVASSMTCTG
SLHGfNRSgSKATFQSLKvQAPFTEATLSRPMQCfRksAEKVDSdQLdSVVSTCSWGNHAAIGTGSafKIH
WNDENQsasNEILREYNLYDFLEAVGRIGATEQKTDAPHSLCLYDVGLPEDEVQEDVVCFGGTSPISWT
DKPKGDSLLHDFMGRAGMWS TVQKHQEMQNKTkwnSASTRGQNKRFQTVQVYARKQPKHSWSQAATHQNNK
LSWCGENVAGAQDFANAESSKGGWNRKNSGfGRGGHRRGGGRGMAFANAESSSSGGWNRKNSGfGRGGRRGG
GRGMWkSEGSHRGGSNSTNWRAQNNSARQCGISYSFTPVEQQIYTQVEPIIKNVKRIIRESRDGMKLSQD
DEMFI MNKILMYHPEKEKKMAGQGNyIMVNKHQTFPSSRCLYVASSDGSSSDFSYKKCLENFIRIHYPHAA
ESFCRKYFK

Metal A
H
(2) 31aa repeats &
(3) WG
DeCL

>Arabidopsis_lyrata_NRPE1 (483042)

MEESSSEILEGEIVGIKfALATHHEICIASISGSAINHPSQLTNSFLGLPLEFGKCESCGATEPDKCEGH
FGYIQLPVPIYHPAHVNELKQMLSLCLCKLKIKKAKSTSGGLADRLLGVCCEEASQISIRDRASDGASYL
ELKLPsRSLQAGCWNFLERYGYRYSdyTRPllAREVKEILRRIPEETRKKLTAKGHIPOEGYILEYLPV
PPNCLSVPDVSDGYSSMSVDPsRIELKDVlKkVIAIKSSRSGETNFESHKAEANDMFRVVDTYLQVRGTAK
AARNIDMRYGVSKI SDSSSKAWTQKMRTLfIRKGSgfSSRSVITGDAYRHVNEVGPIEIAQRITFEERV
SVHNIGYLQKLVDKLCLSYtQGStTYSLRDgSKGHTVlKPGQVVHRRVIDGDVVFINRPPTTHKHSLQAL
RVYVHEDNTVKINPLMCSPLSADFDGDCVHLfYpQSLsAKAEVMELfSVEKQLLSSHTGQLILQMGCDsLL
SLRVMLEGVFLDKATAQQLAMYGSLTLPPPALRKSSKSGPAWTVfQILQLAFPERLSCKGDRFMVDGSDLL
KFDFGVDAMASINEIVTSIFLEKGPkETLGFfDSLQPLLMESLFAEGFSVSLLEDLSMSRADMDVIHNLII
REISPMVSRRLRSYRDELQLENSLHKVKEVAANFMLKSYSMRNLIDIKNSAITKLvQQTGfLGLQLSDKK
KFYTKTLVEDMALFCKRKYGRISSSGDFGIVKGCFFHGLDPYEEMAHSIAAREVIVRSSRGLAEPGTLfKN
LMAVLRDIVITNDGTVRNTCSNSVvQFTYGVdSERGHQGLFEAGEPVGvLAATAMSNPAYKAVLDSTANSN
SSWEQMKEVLLCKVNFQNTTNDRRVILYLNECHCGKRfCQENAAyTVRNKLKkVSLKDTAVEFLVEYRKQQ
TISEIFGIDSLHGHIHLDKTLQDWNISMQDILQKCEDVINSLGQKKKKATDDFKRTSLSVSECCSFQD
PCGRKDSMPCLMfSYSATDPDLERTLDVLCNTIYPVlLETVIKGDPRICSANI IWNSSDMTTWIRNCHAS
RRGEWLDVTVEKSAVQSGDAWRVVIDACLsvLHLIDTKRSIPYSIKQVQELLGLScafeQAVQRlsASV
RMVSKGVLKEHIILLANMTCsgNMLGFNSGGYKALTRSLNIKAPFTEATLITPRRCFEKAAEKCHTDsLS
TVVGCSWgKRVDVGTGSQfELLWNQKETGLDDKEETDVYSfLQmVRSTTNADAYVSSPGFDVTEEEMAEW
AESPERDSALGEPKFEDSAEFQNLHDEGKpSESNWEKSSSWDNGCSGGSEWgVSKNTGGEANPEsnWEKTT
NVEKEDAWSSWNTKKDAQESSKSDSGVAWGLKTKDDADTTPNWETRPAQTDsIVPENNEPTSDVWGHKSG
SDKSWDKKNGGTESAPAAWgSTDAAVWgSSDKKNSETESDAAAWgSRDKKNSEVGSgAGVLGPWNKSSKT
ESDgATWgSSDKTKSGAAAWSSWDKKNMETDSEPAAWgSQSKNKPETESGPSTWgAWDTKKSETESGPAGW
GIVDKKNSETESGPAAMGNWDKKNSTESGPAAWgSTDAAVWgSDKKNSETESDAAAWgSRDKKtSETES
GAAAWgSWGQPTPTAANEDANEDDENPWVSLKETKSRDKDDKERIQWGNPAKkFPSSGGSNGGGADWKGK
RNHTPRPPRSEDNLAPMFTATRQLDSFTSEEQELLSdVEPVMRTLRKIMHPSAYPDGDPIsDDDKTFVLE
KILNFHPQKETKLGSGVDFITVDKHTIFSDSRCFFVvSTDGAKQDFSYRkSLNNYLMMKYPDRAEEFIDKY
FTKPRPSGNRDRNNQDATPPGEEQSPPTQSIGNGGDDFNtQTQSPSQTQAQAQAQAQAQSPSQTQTQSPS
PSQTQTQSPSQTQAQAQSPSQTQTYs

Metal A
H
(10) 16 aa repeats &
(17) WG motifs
DeCL
QS-rich

Figure S3. Predicted NRPD1 protein sequences among diverse plant species with key domain features denoted to the right-hand side. The Metal A motif is in black bold type; the NRPD1 signature motif (Erhard et al, 2009) in the DdRP G domain is underlined; the conserved DdRP H domain is underlined in bold; the DeCL signature motif is in blue bold type.

```

>Arabidopsis_thaliana_NRPD1 (Atlg63020)
MEDDCHEELQVPVGTLTLSIGFSISNNNDRDKMSVLEVEAPNQVTD SRLGLPNPDSVCRTCGSKDRKVCEGHF
GVINFAYSIIINPYFLKEVAALLNKICPGCKYIRKKQFQITEDQPERCRYCTLNTGYPLMKFRVTTKEVFRR
SGIVVEVNEESLMKLLKRGVLTLPDPYWSFLPQDSNIDESCLKPTRRIITHAQVYALLLGGIDQRLIKKDIIP
MFNSLGLTSFVPVTPNGYRVTEIVHQFN GARLIFDERTRIYKKLVGFEGNTLELSSRVMECMQYSRFLFSETV
SSSKDSANPYQKKS DTPKLCGLRFMKDVLLGKRS DHTFRTVVVGDPSLKLNEIGIPESIAKRLQVSEHLNQ
CNKERLVT SFVPTLLDNKEMHVRRGDRLVAIQVNDLQTGDKIFRSLMDGDTVLMNRPPSIHQHSLIAMTVR
ILPTTSVVS LNPICCLPFRRGDFDGDCLHGYV PQSIQAKVELDELVALDKQLINRQNGRNLSSLGQDSLTA
YLNVNVEKNCYLNR AQMQQLQMYC PFLPPPAIIKASPSSTEPQWTGMQLFGMLFPPGF DYTYP LNNVVVSN
GELLSFSEGS AWLRDGE GNFIERLLKHDKGKVL DIIYSAQEMLSQWLLMRGLSVSLADLYLSSDLQSRKNL
TEEISYGLREAEQVCNKQQLMVESWRD FLAVNGEDKEEDSVSDLARFCYERQKSATLSELAVSAFKDAYRD
VQALAYRYGDQSN SFLIMSKAGSKGNIGKLVQHSMCIGLQNSAVSLSFGFPRELTCAA WNDPNSPLRGAKG
KDSTTTESYV PYPYGV IENSFLTGLNPLESFVH SVTSRDSSFSGNADLPGLTSRRLMFFMRDIYAA YDGTVRN
SFGNQLVQFTYETDGPVEDITGEALGSL S ACALSEAAYSALDQPI S LLETSP LLLN LKNVLECGSKKQREQ
TMSLYLSEYLSK KKHGFEYGSLEIKNHLEKLSFSEIVSTSMIIFSPSSNTKVPLSPWVCHFHISEKVLKRK
QLSAESVVS LNEQYKSRNRELKLDIVLDL IQNTNHCSDDQAMKDDNVCITVTVVEASKHSVLELDAIRL
VLI PFLLDSPVKG DQGIKKVNILWTD RPKAPKRNGNHLAGELYLVKVTMYGDRGKRNCWTALLETCLPIMDM
IDWGRSHPDNIRQCCSVY GIDAGRSIFVANLES AVSDTGKEILREHLLL VADSLSVTGEFVALNAKGWSKQ
RQVESTPAPFTQAC FSSPSQCFLKAAKEGVRDDLOGS IDALAWGKVPGFGTGDQFEIIISPKVHGFTT PVD
VYDLLSSTKTM RRTNSAPKSDKATVQPFGLLHSAFLKDIKVL DGGKIPMSLLRTIFTWKNI ELLSQSLKRI
LHSYEINELLNERDEGLVKMVLQLHPNSVEKIGPGVKGIRVAKSKHGDSCCFEVVRIDGTFEDFSYHKCVL
GATKIIAPKMMNFYKSKYLKNGTLES GGFSEN

```

Metal A
NRPD1 sig
H
DeCL

```

>Phycomitrella_patens_NRPD1 (phya_90112) (complete?)
MELQDPEAGEAPLAEVMGIQFGILSAKDIVTLSVFEREHSIIITAKDLWDSRLGIYNLPGNNHCQTCGARK
ASDCDGHFGHITLPMPIYHPLHIYFLKLLNQICLVCKRFKEKVFTLTSYFN SPLYQSSSESSDDGKACKWC
GVNNSYETIEMKASVKEGKLP LDYWNFVCGNPERAYNILQSLSKKVIQKLGMD EYVARPEALILHFVPVPP
SGSRITEVDFGSSLP RTHMVGGRRFRFDKQHKLLQRLSFEVKRLQSLRTGMPDWATTKNEVMELQLLASSY
LTGSKWEHGLNPKAYDAVVKSDVQKSDRYMKGHILAKTNNSSARMIVVDPSIKIEEILLPVFLVEQLTIP
EKVTA FNIERLQRYVDNGPYADLPGRDRVRLHSRLKRMVVEIGD TVHRHIKDGDLVIVNRPPSLTKHAIMA
MEVRLHHSCSLAINPLICAPFQADFDGDCMHLFVPQTSEAHAEAEHELLKVS NQLINPQGGQSN SALTEDSR
LGAYLMTSSCIFLNKMEVSQ LSTSSLVSLPIPAILKSPNKREPLWTGQQLYSTILPEGICYKVTDKKFSTD
VERGILISNGELLVCNGNSNWLGD AFDALTA VIHTSQGPAAALVYLNRAQELANFLRDRGFSVGLQDFQL
SRDRS QLLRRRLEEVSIGNREALFR TLLMDEHVQREELNKNPASKRGLTAETECIKSKGLYL GATGIVKQV
EALDKVAVDRFQTKFRE STKRLAKDYCKRMNPLLVMINAGSKGSM SKLVQQTISVGLQLFKGEHLLPLNVP
DFCQKQLTDVSTLRATDFLQFERRVPSANLSGYWESRGIITSSYLDGLSPLQFFIHTLSSRYGIMRSKVEE
PNLLLLKRLLLFLRNLYVEYDGSVRSLEGQQIVQFKYGRYIEGQRGAIITTEGPKIWCEAGEPVGILAATAI
TEPAYQLKLDSPHNVGAKAIGPLDLIN ETLSPSNPLKLD RRVLLRFPLALKSRRHGQENGAMRILQH LKP
VLSMVATTTMIEYRKAQTVVGEHGRSSPWVGHIRLGVVKKIYQLLVADLVGSLETQYT NCKFASSHSCQ
FGSSGVTQE QPNPCHFFVDDSTLVATLDDKEYDEVLSNSLEVMKNVILPILLRTPIKGDARIESVNLLWE
DMEWNPRCTKYLSKKPKNGT GELVLEVTVKKECKSRGKAWKIVTESCLPIMQLLDWQRCTPYSIQELN
HVFGLEAAKGVLLQRLELAIAGMGKPVNLEHLELIADTMVTS GKVSGASLSGYKDLCKTISRSAPFSTAAF
LNPKNSFVVAGR HGIS E TMEGALSSSVWGKAPSLGTGSNFEFFWQAKARAREVCNIREGFDIHEYLA KLNS
SALKPCEGVPVPQH HNESQCVSTTMIQGHCDMVMS PDDFKLKQTND ELEIHLRSKEDFPQVGNHNGVLKQQ
ASSPTHISHPPVTDPIRTEGAVTSRSEACEDSSSFHTPNETLELTRQDSSNSSPCSSFRKDLFPPTVLHDD
SEGDETS GIV

```

Metal A
H

>Populus_trichocarpa_NRPD1

CSTCGSRDLKSCEGHFGVINFPYITIVHPYFLSEVVQILNKICPGCKSIRLAKATELITKENPQRKGCKYCA
GNSLGLWYPPMKFKVSSKEIFRKTAAIAEIRETLSSKPKQKGFKKILAADYWDIFPKDEQEEEEETNAKPNRR
VLSHSQVRHMLKDVDPNFIKLSILKTDITIFLNCFPVTPNSHRVTEVTHAFSNGQRLIFDERTRAYKKMVD
RGVANTLSFHVMDCLKTSKLNPKSGNIDPWTAPPKSNDYVNNASGLRWIKDVVLGKRNDHSFRMVIVGD
PHLQLHEIGIPCHIAERLQISESLTAWNWEKLNACFEKSRFEKGMHVRREGNLVVRVHMKELRLGDIYR
PLNDGDTVLINRPPSIHQHSLIALSVKVLVPSVLA INPLCCPPFRADFDGDC LHGYVPQSV DTRVELTEL
VSLDKQLTNWQSGRNLLSLSQDSLTA AHLVLEDDVFLSSFELQQLQMFRRPERFLLPAVKAPSANALVWTGK
QLISMLLPVGFDDHDFPSCNVCIRDGDLVSSEGSFWLWDTDGNLFQSLVKHCHGQVLDFLYAAQRVLCEWLS
MRGLSVLSLDLYLCPDSSNRKNNMDEIWYGLQDADYACNLKHLMVDS CRDFLTGNNEEDQCNVERLRFSLG
CSEEDYCVMAFDGERLICYEKQSAALSQSSVDAFRLVFRDIQSLVYKYASQDNSFLAMFKAGSKGNLLKLV
QHSACLGLQHALASLSFRIPHQLSCAGWNKQKADDATE SAKRYIPHAVVEGSFSLGNPIECFVHSVTSRD
SSFSDNADLPGTLFRRMMFFMRDLHGAYDGTVRNAYGNQLVQFSYNIDMDPSGSVDEINNSDGIAGR
PLAACAISEAAYSALDQPI S LLEKSP LLNLK NVLE CGLKRNSAHQTMSLFLSEKLGQRHGFYAALEVQN
HLERLLFSDIVSFVRIIFSPQSDGRMHFSPWVCHFHVKYIILHKVFFSFQEIVKRSRSLKVHYIIDALEKQ
CKSKTRFPKVQITSRYALWFLNTHQIRDWRTIYADTWKEKKEFCITVTIVETSNEFIELETIQDLMP
FLLETVIKGFMEIQKVDILWNDKPKIPKSHNRLRGELFLRVHMSRGS DKTRLWNQLMDDCLSIMDLIDWAR
SHPDNIHECCLAYGIDAGWKFFLNNLQSAMSDVGKTVLPEHLLL VANCLSVTGEFVGLNAKGLKRQREHAS
VSTPFVQACFSNPGDCFIRAAGVDDLOGSIDALAWGKVP AIGTGQFDIVYSGKGLEFSKPV DVYNLLG
SQMISTEQNTEFGVLDQAQIYKSDKCGAQFLHKFGGCGPKGFKVKEGIPRSFLRRLTTYDDIQRMSYTVRKI
LNKYSVDQQLNESDKSVLMMTLYFHPRRDEKIGIGAKDIKVINHPEYQDTRCFSLVRTDGTIEDFSYRKCL
HNALEIIAPQRAKRYCEKYLT SKVSATDNSG

Metal A

NRPD1 sig

H

DeCL

>Vitis_vinifera_NRPD1

MDNDFLEEQQVPSGLLIGIKFDVSTEEDMGADSGSRRLRSKGCYCAANSNDWYPTMKFKVSSKDLFRKTA
IIVEMNEKLPKKLQKKSFRPVLPLDYWDFIPKDPQOEENCLNPNRRVLSHAQVHYLLKIDIPGFIKEFVSR
MDSFFLNCLPVT PNNHRVTEITHALSNGQTLIFDQHSRAYKKLVDFRGTANELSCRVLDC LKTSKLRSEKS
TSKDSASKMSGLKWIKEVLLGKRTNHSFRMIVVGD PKLRLSEIGIPCHIAEELLI SEHLNSWNWEKV TNGC
NRLLEKQTYVRRKGT LAPVRRMND FQAGDIYRPLTDGDIVLINRPPSIHQHSVIALSVKVLPLNSVVS
INPLCCSPFRGDFDGDCLHGYIPQSVDSRVELSELVALNRQLINRQSGRNLLSLSQDSL SAHLVME DGV
LNL FQMQLLEMFCPYQLQSPAIIKAPLLDTQVWTGKQLFSMLLPPGFNYVFP LNVGRISD GELISSSDGSA
WLRDIDGNLFS SLVKDCQ GKALDFLYAAQEV LCEWLSMRGLSVLSLDIYLSSDSISRKNMIDEVFCGLLVA
EQTCHFQQLLDSSQNFLIGSGENNGVVPDVQSLWYERQGSAA LCQSSVCAFQKFRDIQNLVYQYANK
DNSLLAMLKAGSKGNLLKLVQQGLCLGLQHS LVP LSFKIPHQLSCAAWNKQKVPGLIQNDTSEYAESYIPY
AVVENSFLMGLNPLECFVHSVTSRDSSFS DNADLPGTLTRRLMFFMRDLYIAYDGTVRNAYGNQLVQFSYN
IEHTSTPSDGINEDTCAYDMGGQPVGSI SACAISEAAYSALDQPI S LLEPSPLLNLKRVLE CGLRKSTADR
TVSLFSLK KLEKRKHGFEYGALEVKNHLEKLLFSDIVSTVMIVFSPQNGSKTHFSPWVCHFHVCEEIAKRR
SLKPHSII DALYMKCN SARAESKINLPDLQITSNGRDCFVDMEKEDSDCFCITVSI VNSKKSCIQLD TVRD
LVIPFLLGAVVVI PSSIKDAILSWHGLLDVKKVDILWNDNPDSVLKSSSGLYLRVYVSGDCGKKNFWGV
LMDACLQIMDMIDWERSHPDNIHDI FVVYGIDAGWKYFLNSLKS AISDIGKTVLPEHLLL VASCLSATGEF
VGLNAKGMARQKELTSI SSPFMQGC FSSPGSCFIKAGKRAVDNLHGSLDALAWGKIPSVGSGGHFDILYS
AKGHELARPEDIYKLLGSQTSCH EQNLKVKVPITCYQTTTKCGAQLVYANGDSASKGCKSLEKISKSVLRS
FLSLNDIQKLSRRLKFI LQKYPINHQLSEIDKTTLMALYFHPRRDEKIGPGAQNIKVRYH SKYHNTRCFS
LVRTDGT EEDFSYHKCVHGALEIIDPRRARSYQSRWLPYSEV

Metal A

NRPD1 sig

H

DeCL

>Oryza_sativa_J_NRPD1-1 (OsJ_15844)

MLLEPELSPGSLGTRTRGEGWMEEPSLEVNNPVAELNAIKFSLMTSSDMEKLSSATIIEMCDVTNAKLG
NGAPQCATCGSRSIRDCDGGKLLTGKLLGHFGVIKLAATVHNSYFIEEVVQLLNQICPGCLTLKQNGDTKK
ADGTTIQGTCKYCSKDGSKLYPSIIFKMLTSPRVTLRSRKLHRNTSVMDKMSIIAEVAGGVAHKS
KNAKPHETLPQDFWDFIPDDNQPIFNVTKKILSPYQVFHMLKLLDPELINQDDRTKAYKRMVDLY
SKKSDDESSASTDTYGTWKWKDIILSKRSDNAFRSIMVGDPKINLNEIGIPMGLALNLV
VSEQVSSYNFETINLKC�LHLLTKEVLLVRRNGNLI
FVRKANQLEIGDIAYRLLQDGDVLVLRPPSVH
QHSLIALSAKLLSTQSAVSINPLCCDPFKGDFDGD
CLHGYIPQCLQSRIELEELVGLSGQLLNQDGRSLVSL
THDSLAAAHQLTNADVFLKAEFQQQLQMLSSSI
SLTPMPSVFKSTNSQGPLWTGKQLFGMLLPYGMNI
SFDQKLHIKDSEVLTCSSGFSWLQNTSSLSFV
MFKYCGKALEFLSSTQDVLCEFLTMWGLSVLS
SDLYLFSHDYSRKLEEVHLALDEAEAFQIKQI
LLNSVSI
PNLKYYDGGDDRSNTDEQSGFTQVSLPI
IRSSMTSFKSVFNDLLKMVQQYVSKDN
SMMTIMNSGSKGSVLKQVQATACVGLQLPASK
FPFRI
PSQLSCVSWNRHKS
LNCEITDGTSECVGGQDMYAVVRNSFLDGLN
PLECLLHAISGRANFFSENADVPGTLTRKLMY
HLRDTYVAYDGTVRSSYQQIVRFSYDTADGMY
SDHDLEGE
PGAPVGSWAACSI
SEAAYGALDHPVNSLEDSPLMNLQEV
LKCHKGTNSLDHTGLLFLSKHLRKYRYGFEYAS
LEVKDHLERVDFSDMVDTVILIYGGSDMQKTK
GNPWITHFHLNQETMKIKRGLGFEIVREI
IDQYNTLRKQLNNAIPSVSISNSETLHLKME
NKSGKLGKNLGTGNECVKNQTCVMTMVVQVE
INSMSQLDVIKERVIPSILATLLKGFLEFK
NVKVQCQEDNELVLKVMSEHCKSGKFWATLQ
NACIPI
MELIDWERSRPERVYDNFCSYGIDS
AWKFFVESLRSTTDAIGRNIHQHLLV
VADCLSRPAHSFINAAKRDSDVNL
SGTLDAIAWGKEPCAGSSGPFKILYSGKSHET
KQNEHIYDFLHNPEVQALEKNVMDTYRKR
TEKTSKRRSALNSEGNATINGGAI
SFNQKFLNAKVGIWENIIDMRTSLQ
NMLREYTLNEVVTEQDKSCLMEALKFH
PRGYDKIGVGI
REIKIGVNP
GHPSSRCFIVLRNDDTTADFSYNNRFP
CRYLHSELPEAPPERLRPSHRPSAAACG
GGGGGNCVVSSTREKPKFFLSGDCRYG
DECRCYLHAGSINDGFSLLTPLRGHQ
KEPLL
FVGIPDAVKI
WDTGAEMSLSEPTGEYMH
WRLAMGCSSLQCNYTSLG
CYGKLETGSLAVTYTHNEDHGALALAGM
QDAQLNPILLWSTNYNIVHLYELPSMEE
QVRKAVFLNRET
FGSQFALAISRI
PYSVVEEY
TSTGLEELFADVGTWKKQN

Metal A

NRPD1 sig

H

DeCL

>Oryza_sativa_J_NRPD1-2 (OsJ_30285)

MAGGVREGREIEMAPRRATILLGRIGMEEPSLEVMP
EADLKAVKFS
LMTSSDMEKLSSASIIEMCDVTNA
KLGLPN
GAPCATCGS
QSVRDCDGHFGVIKLAATVHNPCIEEVVQLLNQICPGCLTLKQNGDTKKTDGTTI
QTTCKYCSKDGAKLYPSVIFKMLTSPRVTLRSRKLHRNTSVMDKISIIAEVAGGVT
HNSKNKAPHETLPQDFWDFVPDDNQPPQSNVAKKILSPYQVFHMLKLLDPELINQ
LQYSRKSDGEDPTSPDTYGTWKWKDIILSKRSDNAFRSIMVGDPKINLNEIGIP
TDLALNLV
VSEQVSYNFETINLKC�LHLLTKEVLLVRRNGKLI
FVRKANQLEIGDIAYRLLQDGDVLVLRPPSVH
QHSLIALSAKLLPIQSAVA
INPLCCDPFKGDFDGDCLHGYVPQTLQSRV
ELDGLVSLSGQMLNAQDGRSLVSLTHDSLAAAHQLTSADVFLQKAEFQQLQLL
CSSISPTPEPSVVKANFQGS
LWTGKQLFGMLLP
SGMNI
SFDQKLHIKDSEVLTCSSGFSWLQNTSSVFSVMFKEYGSKALEFLSSTQDVL
CEFLTMKGLSVLSDFYLFSDHYSRKKLSEIHLALDEAEAFQIKQILLNTVSI
PNLKHYDGP
DNLSNSHGQSDFTQVSLPIIKSSITGFKSVFNDLLKMVLQHVSKDN
SMMAMINS
GSKGSVLKQVQATACVGLQLPASTFPFRI
PSELSCVSWNRQKSLNCEITNNTSECMAGQ
NMYAVIRNSFLDGLNPLECLLHAISGRANFFSENADVPGTLTRKLMYHLRDTYVAY
DGTVRSSYGRQIVQFSYDTADGMNNDHDLEGE
PGAPVGSWAACSISEAAYGALDHPVNALEDSPLMNLQEV
LKCHKGTSAVHTGLLFLSKYLKRYRYGFEYAS
LEVKDHLERVDFSDLVDTETMKIKRRLGFI
VRELIDQYNALRKKLNMI
PSVICISYSKCSV
GNECVKNRSCCVTMAQV
ESNSTSQLDI
IKERVIPSILATLLKGFLEFENVKVE
CQQDSELVVKVMSEHCKTGKFWATLQ
NACIPI
MELIDWERSRPERVYDIFCSYGIDS
AWKYFVESLRSTTDAIGRNIHQHLLV
VADCLISGQFHGLSSQGLKQQR
AWLSISSPFSEACFSRPAYSFINAAKRDSDVNL
SGALDAIAWGKEPCAGTSGPFKVLYSGKSQKTKQNK
NIYDFLHNPEVQALEKNFM
DQTEKPSKQSAFSSKGNATINGGTISVNQKFL
DSKVGIWENIIDMRTCLQ
NMLREYTLNEVVTEQDKSCLIEALKFH
PRGYDKIGVGI
REIKIGVNP
GHPNSRCFIVQ
RSDDT
SADFSYNNKCVLGAANSISPELGSYIEKILSNRAIRPHQL

Metal A

NRPD1 sig

H

DeCL

>Oryza_sativa_I_NRPD1 (OSIGBa0147H17.3)

MEEPSLEVNNPVAELNAIKFSLMTSSDMEKLSSTATIEMCDVTNAKLGLPNGAPQCATCGSRSIRDCDGHF
GVIKLAATVHNSYFIEEVVQLLNQICPGCLTLKQNGDTKKADGTTIQGTCKYCSKDGSKLYPSIIFKMLTS
PRVTLRSRKLHRNTSVMDKMSIIAEVAGGVVAHKSKNKAPHETLPQDFWDFIPDDNQPPIFNVTKKILSPYQ
VFHMLKKLDPELINQVTRRRELLFLSCLPVTNCHRVAAEMPYGHS DGPRLAFDDRTKAYKRMVDLYSKKSD
DESSASTDITYGIKWLKDIILSKRSDNAFRSIMVGDPKINLNEIGIPMGLALNLVSEVQVSSYNFETINLKC
NLHLLTKEVLLVRRNGNLI FVVRKANQLEIGDIAYRLLQDGDVLVLRNPPSVHQHSLIALSAKLLSTQSAVS
INPLCCDPFKGDFDGDCLHGYPQCLQSRIELEELVSLSGQLLNQDGRSLVSLTHDSLAAAHQLTNADV
LEKAEFQQLQMLSSSISLTPMPVSVFKSTNSQGPLWTGKQLFGMLLPYGMNISFDQKLHIKDSEVLTCSGGS
FWLQNNNTSSLFVSMFKEYGCKALEFLSSTQDVLCEFLTMWGLSVLSLDLYLFS DHYSRRKLSEEVHLALDE
AEEAFQIKQILLNSVSI PNLYYDGGDDRSNTDEQSGFTQVSLPIIRSSMSTFKSVFNDDLKMKVQQYVSKD
NSMMTMINS GSKGVLKQVQQTACVGLQLPASKFPFRIPSQLSCVSWNRHKS LNCEITDGTSECVGGQDMY
AVIRNSFLDGLNPLECLLHAISGRANFFSENADVPGLTRKLMYHLRDTYVAYDGTVRSSYQQIVRFSYD
TADGMYSDHDLEGE PVAPVGSWAACSI SEAAYGALDHPVNSLEDSPLMNLQEV LKCHKGTNSLDHTGLLFL
SKHLRKYRYGFEYASLEVKDHLERVD FSDMVDTEMKIKRGLGFEFIVREIIDQYNTLRKQLNNAIPSVSIS
NSKCSVGNECVKNTCCVSMVVQVEINSMSQLDVIKERVIPSILATLLKGFLEFKNVKVCQEDNELVLKV
GMSEHCKSGKFWATLQACIPIMELIDWERSRPERVYDNFCSY GIDSAWKFFVESLRSTTDAIGRNIHRQH
LLVVADCLSVSGQFHGLSSQGLKQORTWLSISSPFSEACFSRPAHSFINAAKRDSDVNLSGTLDAIASDMV
DKEPCTGSSGPFKILYSGKSHETKQNEHIYDFLHNPEVQALEKNVMDTYRKRTEKTSKRRSALNSEGNATI
NGGAI SFNQKFLNSKVG I WENIIDMRTSLQNM LREYTLNEVVTEQDKSCLIEALKFHPRGYDKIGVGIRES
KIGVNP GHPSSRCFIVLRNDDTTADFSYNKCVLGAANSISP ELGSIENRRSNRAVRPHQL

Metal A

NRPD1 sig

H

DeCL

>Solanum_lycopersicum_NRPD1 (DQ020654) - incomplete N-terminus

FRTVVVGDPNIELGEIGIPCXXAENLHMAETLSLRNWERMTDLC DLMILQRGGILVRRNGVLVRI SVMDGL
QKGDIIHRPLVDGDVVMINRPPSIHQHSLIALSVRILPINSVLSINPLVCSPPFRGDFDGDCLHGYPQSID
STIELSELVALKQQLLDGQNGQNL LSLSHDSLTA AHLILEPGVFLDRFQMQLQMF CPRQLGMTAIVKAPP
GNICYWTGKQLFSLLLPSDLEYVFP SNGVCISEGEIVTSSGGSSWLRDASDNLFYSLVKHNGGDTLDLLYA
AQTVLC EWLSMRGLSVLSLDLYISADSY SRENMIDEVCSGLQEAERLSYIQLLMIKYNKDFLSGNLEESKI
SMGFDFEFMSIMQK SASLSQASASAFKVF RDIQNLVYNYASNDNSLLAMLKAGSKGNLLKLVQHNMC LG
LQQSLVPVSFRMPRQLSCDAWNNHKSHLVIEKPKVPECPGSYIPSAVVKSSFLAGLNPLECFVHSLTTRD
SSFSGHADVSGTLNRKLMFFMRDLYVGYDGTVRNAYGNQIVQFSY YEAEQIASTKVTGEALESHNHAIGGH
PVGSLAACAI SEAA YCALDQPVSALESSPLLN LK KILESGAGSRTGEKTASMFLSKRLGRWAHGF EYGALE
VKGHLERLLLSEVVSTVMICFSPETRKSTHNC PWVCHFHDKENVKTRRLKLR SVLDALNMRYRAATTKAG
NDLPNLHITCKDCSVAEVQKEKSEICITVSVVETS KDPSSLLDTLRDVVI PFLLETVIKGFSAFKKVDILW
KELPSPSKSSRGPTGELYLQVFMSESCDRIKF WNALVDSCLQIRDLIDWERSYPDDVHDLTVAYGIDVAWE
YFLCKLHSAVSETGKKILPEHLVLAADSLTTTGEFVPLSAKGLTLQRKAAGVVSPFMOACFTNPGDSFVRA
AKMGLSDDLQGSLESLAWGKTPSIGTGSSFDIMYSGGYELAEQIN VYTLLRNLVTVDTPNVKVTLGK DGG
MDGMSLVRRDLRLDDLKKSCKSELSFTKLSYFSFN DIKKSLSQLKQMLSKYDIGRELNEADKCLAMMAL
QFHPRNEKIGKGAPKEIKIGYHQEFEGSR CFMVVRSDDTVE DFSYRKC MQHALELIAPQKAKTSRWLNGA
SA

Metal A

NRPD1 sig

H

DeCL

>Ricinus_communis_NRPD1 (RCOM_1683300)

MEADLFEEERQQLPSALLTAITFGVSTEAEEKLSVLTIDTVSEVTDSKLGLPNPTNQ CSTCGSKDLKSCEG
HFGVIKFPFTILHPYYLSEVVRILNQVCPKCKSIRKESKVRCLNHLNPKLPVLLILLCWYPAMKFSVSSEE
IFRKNVIIAKFSERPTNKSQKRGFKKLAADYWDII PKDEQQEENITRPNQRVLSHAQVIHLLLENIDPNFI
RKFVLRKDSIFLNCFSVTPNCHRVTEVTHAFSNGQRLVDDRTRAYKKMVDFRGI AKELSFRVLDCLKTSK
INPKSVNNDYMALQRKMNDSSSSSSGLRWIKDVVLGKRNDNSFRMVVVGDPNIKFSEIGIPCPPIAERLQ
ISEHLTWNWDKLNNTCCEVRLLEKGMHVRREGKLVVRRTKELRIGDIIYRPLNDGDTVLINRPPSIHQH
SLIALSVKVL PATSVLAINPLICAPF**RGDFDGD**CLHGYVVPQSV DTRVELRELVALDKQLIN VQNGRNL SF
SQDSLVA AHLVMEDGVLLSLQQMQQLQMF CPHQLFSPA VRKAPSLNGCAWTGKQLI SMLLPRGFDHECPSS
DVYIRDGELISSEGSFWLRD TDGNLFQSLIKQCQDQVLDFLYIAQEVLC EWLSMRGLSVLSLDLYLCPDSD
SRENMMDEVLFGLQDAKGT CNMKQFMVDS CRDFLASI DEDEQYSVNFDVEHLCEKQSAALSQASVDAFK
HVFRDIQTLGYKYASKDNALMAMFKSGSKGNLLKVVQHS MCLGLQHSLVPLSFRMPLQLSCDAWNKQKAEN
AVECARSYI PSAVVEGCFLTGLNPLECFVHSVTSRESSFS DNADLPGTLTRRLMFFMRDVHAAYDGSVRS
YGNQLIQFSYNI DEGRSAETYGTAKIVDNYDGMAGKPVGSLAACSI SEAAYSALDQPI SLLEKSPLLNLKN
VLECGLKKSNAHKSM SLFLSEKLGRRRHGF EYGALKVQDHLERLLFS DIVSVSRIIFSSQSESKTCFSPWV
CHFHVYKEIMKKRN LNVD SIINILNGRCKSNTNL PNVQISCKSCSIADNHREKEETLCITVTIVERSKNSS
TRLATI QDLMIPF LLETVLKGLMEINKVDILWKDWPRI SKTHNQPYGELYLRV SMSADSEKTRLWNLLMDY
CLPIMDMIDWTC SRPDNVRDFSLAYGIDAGWKF FLQRLES AISDVGKSVLPEHMLLVANCLSVTGEFVGLN
AKGWKRQREDASV SSPFVQACFSSPGNCFIKAAGVKDDLOGSLDALAWGKVPVSGTGQFDIVYSGVKVL
LLFLLV KRVKLKT PPSFVVLTVFLETPLINLLVWYSVDQQLNEADKCTLTMALYFHPRKEEKIGSGFKDIK
VVKHPEYQDSRCFSLVRS DGTIE**DFSYRK**CVYGALEII IAPHKARSQIEFFQNSDVVAII GRITYKLFVVGQS
EVKELPWEVHVACGLGKHSNRVI SMLCYVQGSCKVDLALC NGLGRRLALVTANRA

Metal A

NRPD1 sig

H

DeCL

>Zea_mays_NRPD1

MELHREPPEAILNAIKFDLMTSTDMEKLSMSII EVSDVTS PKLGLPNGSLQCETCGSQRGRDCDGHFGVT
KLAATVHNPFYFID DVVHFLNRICPGCLSPREGIDTKRLEREKVQATCKYCSKDGSKLYPSIVFKTLSSPRV
LLFKSKLHRNASVMERISIVAEAADRMPNRSKGGKSL EGLPLDFWDFVPS ENKQVQSNMTKIILSPYQV FY
MLKSDPELIKQFVSRRELLFLSCLPVT PNCHRVVEIGYGLPDGRLTFDDRTKAYKRMVDVSRRIDDYRQH
PHFSVLASSLVSSRVSECLKSSKLYSKADGETSTDTYGMKWLKDVVLSKRSDNVFRS IMVGDPKIKLWEI
GIPEDLSSSLVSEHVSSYNFQSTNLKCNLHLLAKQELF IRRNGKLMFLRKADQLEIGDIA YRPLQDGDII
LINRPPSVHQHSLIALSAKILPIHSVVSINPLCCTPF**AGDFDGD**CLHGYIPQSIRSRVELEELVSLHNQLL
NMQDGRNLVSLTHDSLAAHLLTSTDVFLK KSELQQLQMLC LSVSTPAPAVIKSMNFQGS LWTGKQLFSML
LPSGMNFSCDTELHIMDSEVLTCSLGSSWLQNNTSGLFSVMFKQY GCKALDFLSSAQEVLCEFLTMRGLSV
SLSDLYMFS DHYSRRKLAEGVKLALYEAEEAFRVK KILLDPINIPVLKCHDETEDV TYRQSDCIQSNPSVI
RSSIMAFKDVFRDLLKMVQQHVSNDNSMMVMINAGSKG SMLKYAQQTACIGLQLPASKFPFRIPSQLSCIS
WNGQKSLNYEAESTSERVGGQNL YAVIKNSFIEGLNPLECLLHAISGRANFFSENADVP GTLTRKLMYHLR
DIHVAYDGTVRSSYGQQIVQFSYDSVDDLVDKLGAPVGCRAACSI SEAA YGALEHPVNGLEDSPLMNLQEV
FKCHKATNSGDHIGLLFLSRHLK KYRYGLE YASLEVKNHLERVNFSDLVETIMI IYDGHDKIRNEGMWTT
FHINKAMMKKRLGLRFVVD ELAKEYDTRDQLNNAI PSIRI SRRKCLVGD EGVKSSSCC IAVVAHAERNS
ISQLDTIKTRVIPSILD TLLKGFLEFKDVEIQCPHDGELLV KVCMS EHCCKGRFWPTLQ NACIPVMELIDW
ELSQPSNVSDIFCSY GIDSAWKYFVESLKSATTD TGRNIRREHLLVIADSLSVTGQFHALSSQGLKQQRTR
LSI SSPFSEACFSRPAQSFINA AKQCSVDNL CGSLDAVW**WGKEPFNGTSGPFEIMHSGKPHEPEQNESIYD**
FLCSSKVRNFEKNHLDTRRQSTENASICRLACKSSKGSTTVNGVAITIDQDFLHAKVSIW DNIIDMRTSLQ
NMLREYPLNGYVAEPDKS QLIEALKFH SRGAEKIGVGVREIKIGLNP SHPGTRCFILLRNDTTE**DFSYHK**
CVQGAADSI SPQLG SYLKKLYRA

Metal A

NRPD1 sig

H

DeCL

>Glycine_max_NRPD1 (Glyma11g02920)

MENIAVLEINAAGQVTGSSSLGFPNASDECATCGSKDKRFCEGHFGVVKFPTPI LHYPYFMSEIAHILNKICP
VCKSIRHKSKVIYLLLPNTGILSFYELASMDFIITCFLPPIYSSIVFLQGVRLIYGTKRNSDCNYCSAYP
SMKFRVSSNDLFRRTAIIVEVKASKKTLGTEIPADYWNFI PCDAQQEENYVNRRLVSPAQVLNLLNGVDPD
FIEKYIPRKNLLYLNCFPVTPNCHRVTEVPYAI SIFNIIIFINCHMGTPNELSSRVLDCLRISKARCSAVL
AFRLCFSFDEMQLNPKTPNSIFADIQQRKIGENACNSSGLRWIKDVVLGKRNDSSLRTVVVGVDPDLELSE
VGIPCHIAESLQVSEYVNRQNREKLLYCCELRLLEKGGKIDVCRNGSKVHLYKKEDLQIGDKIYRPLADGDK
VLINRPPSIHQHSMIALTVRVLPISSVVCINPLCCSPL**RGDFDGD**CLHGYIPQSVTARIELNELVALDRQL
INGQSGRNLLSLSQDSLTAAYLLMEDGVLLNVYQMQQLQMLSISDKRLIPPVAVVKAPSSNSSLSWSGKQIFS
MLLPYDFDYSPSDGTVVSDGELVSSSEASGWL RDS DYNVVFQSLVEHYQGKTLNFLYTAQKVLCEWLSMTG
FSVLSLDLYLSSDSYARKNMIEEIFYGLQDAEQAYKYLLLSVKRQLMMLLGKFFAIFKAGSKGNLLKLVQHS
MCLGMQNSLVRLSYRLPRHLSYVFCFLTGLNPLECFVH SVTNRDS SFSDHADLPGLTLRRLMFFMRDLHD
AYDGTVRNLYGNQLIQFSYDIEEDSSCDKGFQEYAIIGGEPVGAISACAISEAAYSALGQPVSLLETSPLLN
LKNVLECGSRKRNGDQTVSLFLSEKLGKQRHGFYAAAEVKNYLERLLFSNIVSTVMIIFTPHDGSSQEKY
SPWVCHFHL DKEIVTRRKLKVHSIIDSLYQRYYSQRKDSKVCFTNLKISSNILRFSHHHEFLYCSLGLFDV
KKVDVLWNNQSKVKNSCNGFSGELYLRVTLSSSEGRGRFWGVLLNLCHKIMHIIDWTRSHPDNINHFSAY
GIDAGWQYFFNVCMIKNFPSFNPGSCFIKAAKSGVTDNLOGSLDAL**AWGNCLSMGTSGMFDIIYSEKYFSP**
CNAHDKCYTGLFLTIDTTSFPYLLIYRKEVDKNSISCYSKNHETTFCPRYKVAKSGNVYELLEASFDPKPN
KAGTHLHKYSSDKCGSEFRHKNGYALKEGKQWKTILRNFTYCWVVFVIMPCNEFMLLCLLGKYYSQLGS
RVVNFVLRMDFSRKYSIDELLSESDRSTMLRVLNFHPRKSEKFGIGPQDIKVGWHPKYKDSRCFHIVRIDG
TVE**DFSYRKC**CILGALDIVDPKSKIQEKKWSGHGNT*

Metal A

NRPD1 sig

H

DeCL

>Selaginella_moellendorffii_NRPD1 (Smo:441655)

MASSKRSSHRDRAL EEA TGT LIALDFRPLTSEEIIRASVYEVKTVRALQNNRFGLPNLSDCCTSCGAKRT
DASNSACPGHSGHIELPVLVYHWDRI SALEAILNRVCLHCYSFKHKGRKKELRTLSSLEQV ASGVDAHQAD
IGAVPNGARAPEAEENPGKCTGPAAAVKKI FKKVGTANVPALLEIDGKVRREDIPPGFQSLILKDEMTPO
WRSKMLDPNQVLRILKCLPQETIDKLRDEKLPSIPAEDYFIKSLPVPNWMRYSTNEFYFQDKTTKNLKH
LTKIKSIVYTRDEDKISLLTEQKVMIEIQAAATQCI RANPLYGNVSDDEDPRYGNVSDSKPLSGLHFLRSLT
GKYCGSSARAVVIGDPALKLEEIGISARIAAGLVVLETVTSSNII FLQSYAYNNPGLKVVVRGGEVCTARSC
KKLQVGDV IHRSLKDGQVFNRPPTFHKHALIGLKS KVI RNNVFAVNPLICPPL**FADFDGDTL**LALYLPQS
LQVRAEVAELVALPKQLVSSQGGQSI IGLTQDALLGAHLMTRKNVFLDKLDMDQLRMWCPSAEVVPVAIVK
SPRKSPLWTGQQLFQMTLPTTFDWESDDGGLIIRQGEILRTSDKSSAWLGKDGLMTTICRRYGPDRAL EHL
DIAQGI AVDWISERGF SVGLCDFYMAADAVSRKLEEETLCAVEEAKISSLAHQIVSDPRFQVNSVSRPRC
NSWNERVQPVTSVNEATQQAASAFQSTMKAFERTIEEHVRENSRENSLLRMVEANSKGSFSKMMQGGCL
GLQLRQGEFVYHRVKS LFPRAVENESRGYLTSSSELWKS MGLV ESSFLDGLDPREFFIHSLSRKGNDGSQQ
RCASFFRFLMSYMKDIRVEYDNTIRSTHGGHIFQFSYGATAEPGEPVGLLAGTAVIEPVYDQVMSSSPQAS
TMLKTLQNILFSNSFKDIDRCVTLKLQKLPVQPEWIALQVQDFLKPVTIGMLASKIWI EYSPCSEVGGQKK
RVPWIGCFQLRAEAMERCSLNIDTIVCHLRKLLPTSLDDPDAFIQGLHFFSRDVEVLCFFPITSSVSNYDS
KQIHKHMIGTMFGNLLQVVVKGCPRGIEFVN VKWEDELCEVAFLSRTRGVPWTHALEACGSI SHLVDWQK
STPLSIQEVHVAFGIEAAYQYLLEKLKEFTKSGSVLRKPWKNI DANESGYEAFVKNLSGCSPLAFAMGKSP
GGVFEEAAMNREVDYLAGANELAFCGKSPSLGTGANIELFFKEDKGPVSRFPDFESLVFSRRVDDTVSAT
LSAKDREIVWARIDQRSQKLHDILRKS LTTGTPVSAANEAVILDTLKYHPMMDSKVGCGVRHIRVDNHH SFG
GRCFHIVRLDGSVE**DFSYHK**CLLERIKGNTVLVQRYKKKFMGGKNGRKEEVPVEIFSQKNDTGRMYDKKTH
GFLLENHFVPVKTLKKT*

Metal A

no NRPD1 signature motif

H

DeCL

>Sorghum_bicolor_NRPD1 (Sb06g025933)

MELHRELPEATLNAIKFDLMTSTDMEKLSMSVIEVSDVTS PKLGLPNASPQCETCGSKSGRDCDGHFGVT
KLAATVHNPFYFIDDVVHFLNQICPGCLSPREGINMKKDGSKLYPSVIFKTLSSPRVLLSKSKLHRSPSVME
RISIVAAEAERVSNRSGKGLLEGLPQDYWDFVPS ENKQVQSNMTKIILSPYQVFHMLKKS DP ELIKQFVS
RRELLFLSCLPVT PNCHRVEIGYGLSDGRVTFLYSKKTYGETSTDPSGMKWLKDAVLSKRSDNAFRSTMV
GDPKIKLWEIGIPEDLASNLVSDHVNSYFENINLKC NLHLLTKEELFIRRNGKLMFLRKADQLEIGDIA
YRPLQDGD LILINRPPSVHQHSLIAFSAKILPIHSVVSINPLCCTPF**LGDFDGD**YGRSLVSLTHDSLAAA
LLTSTDVFLKKSEFQQLQMLC LSVLTPVPAVIKSMNFQGS RWTGKQLFSMLLPSGMKFS CDRMLHILNGEV
LTCSLGSSWLQNNTSGLFSVMFKQYGCKALDFLSSAQEVLCEFLTMRGLSVLSLSDMFSDHYSRRKLT
LALDEAAEFRIKQIILLDPINIPVLKQDETEDVTYRQSDCIQNNPSVIRSSIMAFKDFVSDLLKMQHV
SNDNSMMVMINAGSKGSMKYAQQTACVGLQLPASKFPFRVPSQLSCIRWNRQKVALDYEAGTNERVGGQ
LYAVIRNSFIEGLNPLECLLHAISGRANFFSENADVPGTLTRKLMYHLRDIHVAYDGTVRSSYQQIVQFS
YDSADDPVDKLGAPVGCWAACSISEAAYGALEHPVNGLEDSPLMNLQEVFKCHKATNSGDHIGLLFLSRHL
KKYRYGLEAYASLEVKNHLEQVNFSDLVETIMIMLEM MKKRLGLRFVIEELTKEYNATRDQLKNAIP SICI
SRRKCVVGDEGVKISACCI AVVALAEPNSMSQLDTIKKRVIPIILD TLLKGFLEFKDVEIQ CQH DGELLVK
VCM SHHCKGGRFWATLQACIPVMELIDWELSRPSNVADIFCSYGIDSAWKYFVESLKSATTDIGRNIRRE
HLLVIADSMSVTGQFHAIS SHGLKQQRTRLSISSPFSEACFSRPAQSFIDAAKQCSVDNL CGSLDAI**AWGK**
EPFNGTSGPFEIMHSGKPHEPEQDES IYDFLRS PKVQNV EKNHLDTRRQSTENASICRLACKSKGSATVNG
VAITSDQDFLHAKVSIWDNIIDMRASLQNM LREYPLNGYVMEPDKSKLIEALKFHPRGAEKIGVGVREIKV
GLNPNHPGTRCFILLRNDTTE**DFSYHK**CVHGAANSISPQLG SYLKKLYHRA

Metal A

NRPD1 sig

H

DeCL

>Brachypodium_distachyon_NRPD1 (Bradi2g34870 and Bradi2g34880)

MVRSLLSVIREVTQGSEHSPTKEVQNTGELEKGGVSLPRPAVHLPLL VQGVRAPRRSSDMS EWT DGP NNE
MDVPM AELKALKFDLLSSADIETLSSANIIEASDVTS AKLGLPNAA PQCVT CGSQNV RDCDGHSGVIKLP
TVYSPYFLEQLVQFLNQICPGCWTPKQNRDTRKSDAATIQEPCKYCSKDGLYPSVIFKVLTS PRITLSKSK
LQRNTSVM D KVS VTAEVINMSKNKSSLEVLPHDYWNFVPHNQPPQNTTKILLSPYQVFHILKQVDLELIT
KFAPRRELLFLSCLPVT PNRHRVAEMPYRFSDGPSLAYICMLYSKKT D KESSTDSYGT SVKKNDSYGT KWL
KDAILSKRSDYAFRSIMVGD PKIRLHEIGIPMDLADLFVPEHVS IYNFKS INLKC NLHLLAKELLIARRNG
KLIYVRKENQLEIGDIVYRPLQDGD LILVNRPPSVHQHSLIALSAKLLPVQSVVA INPLNCAPL**SGDFDGD**
CLHGYPVQSIGSRVELGELVSLSHQLLNMQDGRSLVSLTHDSLAAA HLLTSSGVLLNKTEFQQLQMLCVSL
SPTPVPSVIK SINPQG PLWTGKQLFGMLLPSGMNFSPDKLHIKDSEVLACSGGSFWLQNNTSGLFSVLFK
QYGGEALEFLSSAQDMLCEFLTMRGLSVLSLSDIYLFSDHYSRRKFAEEVNLALDEAAEFRTQILLSPNF
IPHLKCYDDCDDLSDSYEQSDFVQSNLPIIKSSIMAFKSVFSDLLKMQVQHTPKDNSMMAMINAGSKGSM
KFVQQAACVGLQLPAGKFPFRIPSELTCASWNRHKS LDCDI SEGARKRLGGQNSHAVIRNSFIEGLNPLEC
LLHSISGRANFFSENADVPGTLTKNLMYHLRDIYVAYDGTVRSSYQQIVQFTYD TAEDIYTD CGQEGEFG
APVGSWAACSISEAAYGALDHPVNVIEDSPLMNLQEV LK CQGTNSLDHFGLLFLSKNLKKYRYGF EYASL
YVQNYLEPMDFSELVNTVMIQYDGGGVQKTKGSPWITHFHISKEMMKRRLGLRLLVEDLTEHYNAKR DQL
NNVIPKVYISKCKCSDDDDCINNQTCCITVVAQDESNSTSTSQLDDLKRAIPVLLATPVKGFLEFKDVEI
QCQRDNELVVKVNM SKHCKSGIFWTTLKKACIGIMGLIDWERSRPGSVYDIFCPCGIDSAWKYFVESLRSK
TDDIGRN IHREHLLV VADT LSVSGQFHGLSSQGLKQQR TQLSTSSPFSEACFSR PADTFIKA AKQCSVDNL
CGNIDALAWGKEPPAGTSGPFKIMYAGKPHEPVQENIYGF LHNPEVWGPEKNHMETDSTRTKNASERWSS
GNATFNNGGTISVEQNYLGAKVGVWDSIIDMRTCLQNM LREYQLDEYVVELDKSRVIEALRFHPRGREKIGV
GIRDIKIGQHP SHPGTRCFILVRNDTTE**DVSYKK**CVQGAADSI SPQLGSHMEKILQTRSF CRDSWR

Metal A

NRPD1 sig

H

DeCL

>Arabidopsis_lyrata_NRPD1 (924683)

MEDDCEELQVPVGTLTLSIGFSISNNTDRDTMSVIKVEAPNQVTD SRLGLPNPDSICKTCGSKDRKVCEGHF
GVINFQYSIINPYFLKEIAALLNKICPGCKYIRKKQFQITEDQPERCRYCTSNTGYPLMKFRVTTKEVFRR
SGIVVEVNEESLMKLLKRGVLALPPDYWSFVPQDSNIDESCLKPTRRILTHAQVYALLSGIDQRLIKKDIP
MFDSLALTSFPVTPNGYRVTEIVHQFNARLVFDERTRIYRKLVGFEENTLELSSRVIECMQYSRLFSENV
SSSQDSANPYQKKS DTPKLCGLRFMKDVLLGKRS DHTFRTVVVGDPSLKLHEIGIPERIAKRLQVSEHLNN
WNNERLVTFCSPNLF DNKEVHVRRGDRLVAIRVSDLQTGDKIFRNLMGDGTVLMNRPPSIHQHSLIAMTVR
VLPTTSVVSLNPICCLPFR**RGDFDGD**CLHGYVPQSIQAKVELDELVALDKQLINRQNGRNLLSLGQDSLTA
YLVNVEKNCYLNRAQMQLQMYCPFLPPPAAIKASPSSTEPQWTGMQLFGMLFPPGFYDITYPLNDVVVSN
GELLSFSEGSAWLRDGEENFIQGLIKHDKRKVLDIIYSAQEMLSQWLLMRGLSVSLADLYLSSDPQSRKNL
TEEISYGLREAEQVCNKQQLMVESWRDFLAVNGEDEGEDSVARDLARFCYERQKSATLSKIAVSAFKDAYR
DVQALAYRYGEQSNFSLIMSKAGSKGNIGKLVQHSMCIGLQNSAVLSYGFPRELTCASWNDPNSPLRGAK
GEDSTATESYVPGVIENSFLTGLNPLESFVHVS VTSRDSSFSGNADLPGTLSRRLMFFMRDIYAAYDGTVR
NSFGNQLVQFTYETDGPVEDITGEALGSL SACALSEAAYSALDQPI S LLETSPLLNLKNVLECGSKKGQRE
QTMTLYLSETLSKKKHGFEYGSLEIKNHLEKLSFSEIVSTSMIIFSPSTNTKVPLSPWVCHFHISEKVLKR
KQLNVE SVVSSLNEQYKSRNRELKLDIVDLDIQSTNHCSDDKAMKDDSDFCITVTVIEASKHSVLELDAIR
LVLI PFLLDSPVKGSQEIKKVDILWTD RPKAPKRNGDHLAGELYLRVTMYGDRGKRNCWTALLETCLPIMD
MIDWSRSHPDNIRQCCSVY GIDAGRSIFVANLES AVSDTGKTILKEHLLL VADSLSVTGEFVALNAKGWSK
QRQVESTPAPFTQACFSSPSQCFLKAAKEGVRDDLOGSIDAL**AWGKVP**PGFGTGDQFEIIISPKVHGFTTPV
NVYDLLSSTPPKTN SAPKSDKVTVQPFDLLGTAF LKGIKVLDGKGISMSRLRTIFTWENIEKLSQSLKRIL
TSYEINDPLNGRDEELVMMVLHLHPNSADKIGPGLK GIRVAKSKHGDSRCFEVVRIDGTFE**DFSYHK**CVLG
ATKI IAPKKVNLYKSKYLKNGTHQPGRLSEN PQTVK

Metal A

NRPD1 sig

H

DeCL

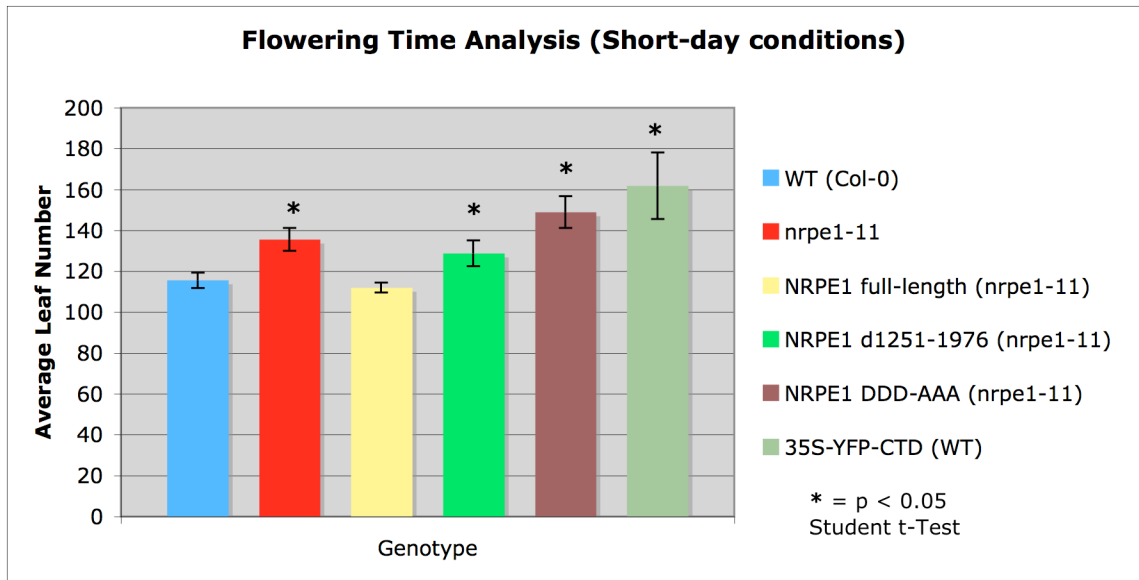


Figure S4. Flowering time experiment with Arabidopsis plants grown under short-day conditions (8 hrs light/16 hrs dark) and randomly rotated every 4 to 6 days. Rosette leaf number was counted when the bolt reached 5 cm in height.



Untransformed FLAG-NRPD1 aa1337-1453 #258 (T2 generation)

Figure S5. Visible phenotypes observed among wild type *Arabidopsis* plants transformed with pEarleyGate202-NRPD1 aa1337-1453 (Line #258, T2 generation). Plants display a range of smaller statures and curled rosette leaves. The survival rate was lower than that of other CTD over-expressed domains transformed and planted side-by-side. This rate was not quantified but it took three flats of planted seed to obtain (9) T1 individuals after BASTA selection (~0.5 to 1.0 mL seed planted per flat) compared to the typical single flat that results in at least (30) BASTA survivors.

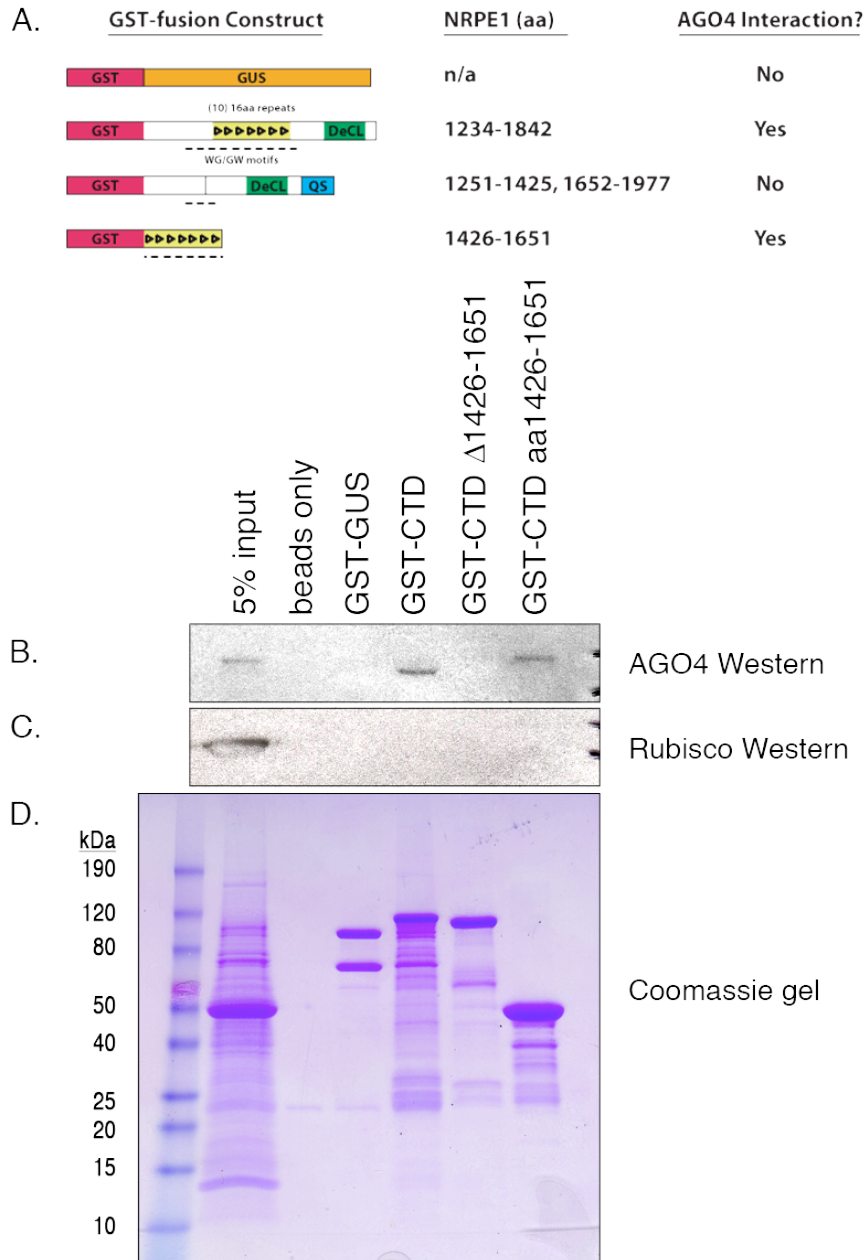
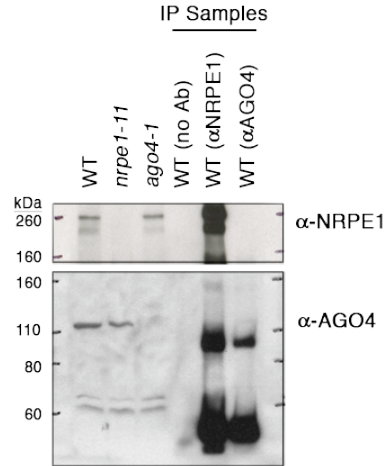
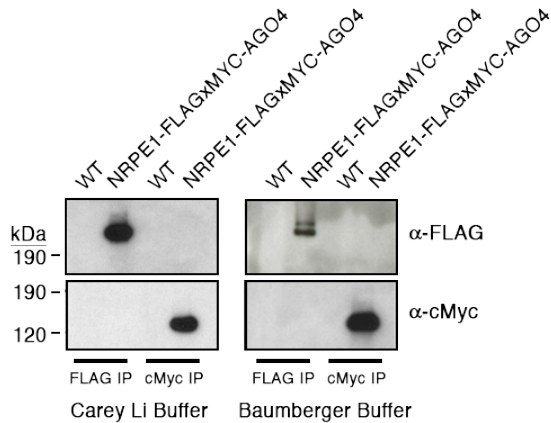


Figure S6. AGO4 *in vitro* interaction with the NRPE1 CTD. (A) Bacterially expressed N-terminal GST tagged constructs used for the *in vitro* protein-protein interaction experiment. Total protein extract from MYC-AGO4 expressing plants was incubated with GST-tagged proteins bound to glutathione resin. The resin was washed and bound proteins analyzed by Western blot. (B) AGO4 Western was performed using the anti-cMyc, clone 9E10. (C) Rubisco Western to demonstrate adequate resin washing. (D) Coomassie stained gel of the eluted bound protein fractions demonstrating roughly equal protein inputs.

A. Co-IP Trial with Native Antibodies



B. Extraction Buffer Co-IP Comparison



C. Co-IP Trial with Dual Tagged Line

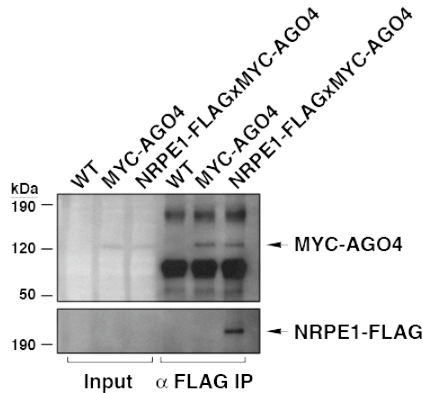


Figure S7. Failure to verify reported NRPE1-AGO4 interaction *in vivo*. (A) Western blot analysis showing lack of co-immunoprecipitation between NRPE1 and AGO4 using native antibodies. Wild type, *nrpe1-11* and *ago4-1* total protein extract controls demonstrate the specificity of these antibodies. (B) A transgenic line bearing both MYC-AGO4 and NRPE1-FLAG genomic constructs was generated by crossing lines from Li et al (2006) and Pontes et al (2006). The possibility exists that the NRPE1-AGO4 interaction is sensitive to buffer conditions so a side-by-side comparison was performed with the extraction buffer and techniques used in the originating report (Li et al, 2006) and the buffer and techniques typically used in the Pikaard lab (Baumberger et al, 2005 with modifications in this manuscript). Reciprocal co-IPs were performed with FLAG and cMyc resin under both conditions. Interaction between NRPE1 and AGO4 was not observed in either immunoprecipitate with either buffer. (C) Western blot analysis showing non-specific IP of MYC-AGO4 with anti-FLAG resin from whole plant extract. This is the only case where an apparent interaction was observed between NRPE1 and AGO4. The result cannot be trusted, though, since the control sample showed immunoprecipitation of MYC-AGO4 with the anti-FLAG resin.

CHAPTER 7
CONCLUSIONS AND FUTURE DIRECTIONS

Prologue

The activity of Arabidopsis RNA Polymerases IV and V has been difficult to assess with a lack of biochemical evidence. The discovery of Pol IV DNA-dependent RNA polymerase activity *in vitro* and Pol V transcripts *in vivo* that are dependent upon the Metal A and Metal B sites demonstrates that these are functional polymerases and opens the door to additional avenues of exploration. The continuing effort to define the biochemistry of these two polymerases will likely shed new light on not only the functions of Pol IV and Pol V, but potentially offer new perspectives on Pol II mechanisms as well. Determining protein-protein interaction networks and potential post-translational modifications of the largest subunit CTDs will provide an additional level of insight into the regulation of these enzymes and the RNA-directed DNA methylation (RdDM) pathway as a whole. Finally, structural analysis of Pol IV and Pol V will allow a greater understanding of the sequence divergence these enzymes have undergone and hopefully provide clues about their mechanistic significance.

i.

BIOCHEMICAL ELUCIDATION OF THE RNA-DIRECTED DNA METHYLATION PATHWAY

Introduction

The RNA-directed DNA methylation (RdDM) pathway in plants is an exciting and challenging pathway given the growing number of proteins involved: its roles in genome defense, development and stress response and its diverse biochemical processes,

many of which still remain unknown or only inferred by homology and/or genetic evidence. While siRNA biogenesis and incorporation into RNA-induced silencing complexes (RISC) with Argonaute proteins is rather well understood based on experiments in yeast, fly and plants, the stages immediately before and after are largely black boxes, especially in plants. Research on RNA Polymerases IV and V has begun to shed light on these stages of the pathway while at the same time revealing how much we have yet to understand.

RNA Polymerase IV

RNA Polymerase IV is a DNA-dependent RNA polymerase as demonstrated by *in vitro* experiments using a tripartite template imitating an open transcription bubble (Chapter 5). To aid in future analysis, it would be most useful to develop a large-scale column purification of Pol IV and Pol V complexes. Currently each *in vitro* activity assay sample requires an individual FLAG immunoprecipitation from freshly prepared whole plant extract. Being able to work off of frozen purified protein stocks would help speed up and standardize experiments. A column-based purification strategy would have the added advantage of potentially containing cofactors lost during immunoprecipitation. This could have important implications for both obtaining transcriptional activity (discussed below with regard to Pol V) and for identifying protein-protein interactions.

Arabidopsis thaliana is the most natural source of material to use given the large number of epitope-tagged genomic and mutant lines generated in the Pikaard lab. Alternatively, maize could be used since the immature cob is rich in protein. The Chandler lab (University of Arizona) is developing epitope-tagged NRPD2/NRPE2 lines

(personal communication), which will be important for sorting out the maize Pol IV and Pol V complexes in addition to performing biochemical analyses. Broccoli is another option since it is a relative of Arabidopsis and can be transformed (Chen et al., 2001). The broccoli head is nothing more than a huge mass of inflorescence tissue, a tissue source known to have the highest Pol IV/Pol V protein levels (Pontier et al., 2005).

Now that DNA-dependent RNA polymerase activity has been identified for Pol IV, follow-up analysis needs to be performed to determine the optimum reaction conditions. Thus far reaction conditions optimized for yeast Pol I *in vitro* activity have been used. Since the NRPD1 and NRPE1 transgenes are capable of complementing *nripd1* and *nripe1* mutants in the T1 generation, it stands to reason that Pol IV and Pol V do not require a methylated DNA template. However, whether they prefer a methylated DNA template to an unmethylated DNA template has yet to be determined.

While the Metal A motif has been conserved in NRPD1 and NRPE1 sequences across plants, it is curious why they have diverged in the larger context from the extended YNADFDGDEMN motif found among eukaryotic Pol I, II, III and archaeal polymerases (Haag et al., 2009) (Chapter 4). The NRPD1 and NRPE1 proteins also lack a region of sequence between the DdRP conserved domains F and G (Luo and Hall, 2007) effectively eliminating the trigger loop that is critical for bacterial and Pol II polymerase functions (Landick, 2009). The large degree of sequence divergence in NRPD1, NRPE1 and NRPD2/NRPE2 amino acids positioned in the active site region using yeast Pol II as a model (Haag et al., 2009) (Chapter 4) is hypothesized to either compensate for this sequence loss or to confer yet undiscovered properties. As mutations in the trigger loop can lead to decreased polymerization rates and higher nucleotide misincorporation rates

(Brueckner and Cramer, 2008; Kaplan et al., 2008), this may explain the apparent low *in vitro* activity levels of Pol IV relative to Pol II. Inserting the *NRPBI* genomic sequence encoding the region between domains F and G into *NRPDI* to see if activity levels increase would be one option to explore this possibility.

Identification of Pol IV transcriptional inhibitors would also be of value. Pol IV is resistant to α -amanitin up to at least the 250 $\mu\text{g/mL}$ tested (Haag and Pikaard, unpublished). Chemical inhibitors can be screened, though this would be a time intensive labor. Yasuyuki Onodera, Tom Ream and I have generated many different antibodies in the Pikaard lab against Pol IV and Pol V subunits using both peptide and recombinant protein antigens. These antibodies should be supplemented into *in vitro* reactions to test for Pol IV inhibition. This approach has shown past success in inhibiting the polymerase activity of hepatitis C virus RdRP (Moradpour et al., 2002) and the cleavage activities of DCL1 and DCL3 (Qi et al., 2005). The same can be done to test the inhibitory properties of antibodies raised against other proteins in the Pikaard lab that are involved in the RdDM pathway such as RDR2, DCL3, HEN1, Pol V and DRM2 (discussed in more depth below).

Whether Pol IV transcripts are long or short in nature, 5' triphosphorylated, 5' capped or 3' polyadenylated are all unknown. Pol IV transcripts have yet to be detected *in vivo* despite attempts by RT-PCR (Wierzbicki et al., 2008) and nuclear run-on transcription (Erhard et al., 2009). It is hypothesized that these attempts have failed because Pol IV transcripts are low in abundance and short-lived being made double-stranded by RDR2 and diced by DCL3. *NRPDI-FLAG (nrpd1-3)* and *NRPDI^{DDD-AAA}-FLAG (nrpd1-3)* transgenic lines have been crossed into the homozygous *rdr2-1* mutant

background for *in vitro* analysis of Pol IV activity (Chapter 5). It is believed these lines will provide a suitable background for identifying Pol IV transcripts by RT-PCR, RNA immunoprecipitation or nuclear run-on transcription assays. Candidate loci would be those categorized as Pol IV-dependent by Mosher et al (2008). Whole genome ChIP to identify Pol IV loci is another approach to this question.

Beyond RNA catalysis, DNA-dependent RNA polymerases also have backtracking, proofreading and cleavage activities either intrinsic or in complex with other proteins. Yeast Pol I has been demonstrated to have intrinsic RNA cleavage activity dependent on the Pol I-specific A12.2 subunit (Kuhn et al., 2007), while Pol II requires the TFIIS cleavage factor (Johnson and Chamberlin, 1994). Initial attempts to identify cleavage activity for Pol IV and Pol V have failed (Figure 1). If Pol IV and Pol V do cleave their transcripts for 3'-terminal trimming or proofreading, they too may require TFIIS. Experiments should be performed to test for Pol IV and Pol V interaction with TFIIS, TATA-BINDING PROTEIN (TBP) and other such Pol II associated complexes as well as *in vitro* stimulation of Pol IV and Pol V activities when supplemented to reactions. Transgenic FLAG-tagged TFIIS (At2G38560) and TBP1 (At3g13445) Arabidopsis plants have been generated and await *in vivo* and *in vitro* testing (Haag and Pikaard, unpublished).

Characterization of the Pol IV-RDR2 relationship

The discovery that Pol IV and RDR2 are physically coupled for the production of siRNA precursors (Chapter 5) helps resolve why RDR2 is the favored RNA-dependent RNA polymerase (RdRP) for the RdDM pathway. Since the Pol IV-RDR2 interaction is

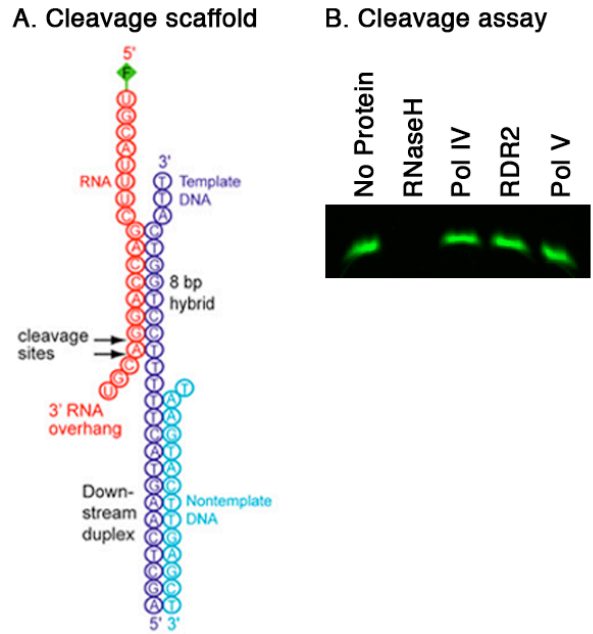


Figure 1. Test for intrinsic RNA cleavage activity. (A) 5'-FAM-RNA cleavage scaffold as used by Kuhn et al, 2007. (B) Denaturing polyacrylamide gel analysis of full-length 5'-FAM labeled RNA remaining after incubation of the cleavage scaffold with immunoprecipitated complexes from whole plant extract or RNaseH positive control.

resistant to RNase treatment and does not rely on active Pol IV transcription, it is hypothesized that the interaction is not via an RNA intermediate but is protein-protein mediated (Chapter 5). This raises the question of which Pol IV subunits are mediating the interaction, either directly or indirectly. Since RDR2 does not interact with Pol II or Pol V (Figure 2), it stands to reason that RDR2 is interacting with a Pol IV-specific subunit. NRPD1 and NRPD7 are the only two subunits that fit this criterion (Ream et al., 2009). The NRPD1 C-terminus containing the DeCL domain was tested as a candidate interaction domain. The NRPD1 Δ 1337-1453-FLAG transgenic line was tested for *in vivo* co-IP of RDR2 but the interaction was unaffected suggesting the NRPD1 CTD is not

required for RDR2 interaction (Chapter 5). Western blot analysis of the 35S::FLAG::NRPD1 aa 1337-1453 immunoprecipitated protein for co-IP of RDR2 failed suggesting the NRPD1 CTD is not sufficient for RDR2 interaction, either (Haag and Pikaard, unpublished). These results do not rule out the possibility that the NRPD1 DdRP core (aa 1-1336) and RDR2 interact, however.

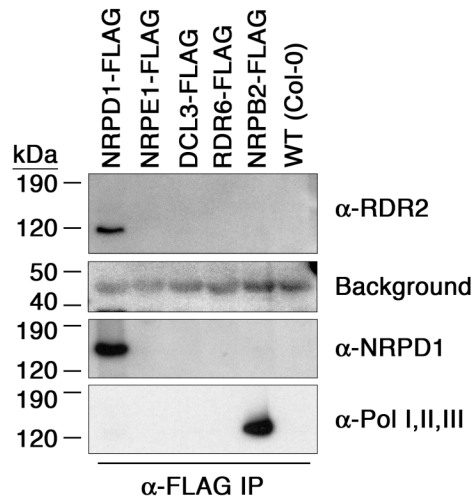


Figure 2. Western blot analysis of Pol IV-RDR2 interaction specificity by co-IP.

NRPD7 is a very interesting candidate for mediating the RDR2 interaction with Pol IV. NRPB7 is known to form a Pol II dissociable subcomplex with NRPB4 (Edwards et al., 1991; Larkin and Guilfoyle, 1998). Arabidopsis Pol IV and Pol V share a Rpb4 subunit paralog distinct from that used by Pol II, named NRPD4/NRPE4 (He et al., 2009a; Ream et al., 2009). The Rpb4/7 subcomplex is positioned near the RNA exit channel and adjacent to the CTD linker region (Armache et al., 2005) (Figure 3) and Rpb7 has a functional RNA binding domain (Mitsuzawa et al., 2003; Ujvari and Luse, 2006). Rpb4/7 interact with the RNA product co-transcriptionally in the nucleus and are able to dissociate from Pol II and chaperone the mRNA to the cytoplasm to stimulate

mRNA decay (Goler-Baron et al., 2008; Lotan et al., 2005; Lotan et al., 2007; Selitrennik et al., 2006). Additionally, the yeast Rpb7 subunit, along with Rpb2, is required for siRNA-dependent heterochromatin formation (Djupedal et al., 2005; Kato et al., 2005).

The properties of the Pol II Rpb4/7 subcomplex are consistent with what may be hypothesized as needed to mediate the transfer of Pol IV transcripts to RDR2. Close

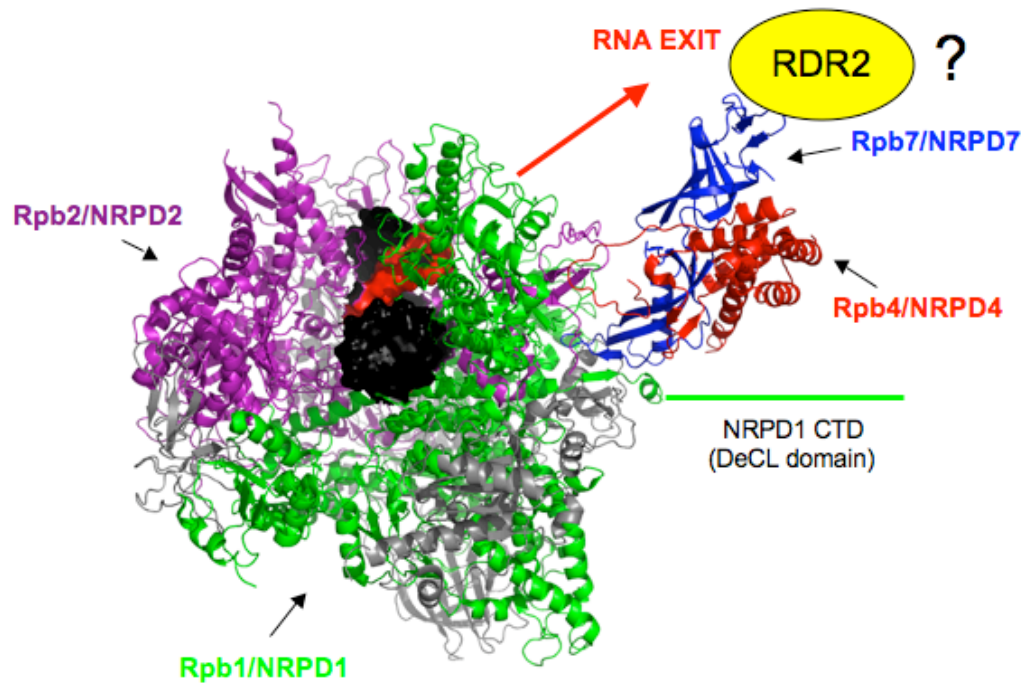


Figure 3. Model of the proposed Pol IV-RDR2 interaction interface via the NRPD4/7 subcomplex. Crystal structure of the complete yeast Pol II elongation complex (PDB1Y1W) modeled in PyMOL. Pol IV is hypothesized to have a homologous structural organization. Rpb1 is green, Rpb2 is purple, Rpb7 is blue, Rpb4 is red and other subunits are in gray. The dsDNA is colored black with the RNA exiting in red at 2:00. RDR2 is hypothesized to interact either directly or indirectly with the NRPD4/7 subcomplex at the RNA exit channel to make dsRNA from Pol IV transcripts.

proximity between the two enzymes would aid this transfer making it more efficient and specific. This may very well be the case as none of the other five RdRp's in *Arabidopsis thaliana* act redundantly with RDR2 (Dalmay et al., 2000; Mourrain et al., 2000; Xie et al., 2004). A Pol IV transcript could be transiently bound by the NRPD4/7 subcomplex, which in turn may either directly hand-off the transcript to RDR2 or may dissociate from Pol IV and chaperone the transcript to RDR2.

Many experiments are required to test these hypotheses. *In vitro* co-IP experiments need to be performed with the Arabidopsis Rpb4-like and Rpb7-like family members to determine their interaction specificities with one another as predicted by the LC-MS/MS analysis (Ream et al., 2009). *In vitro* interaction with RDR2 can also be tested using bacterially expressed NRPD4 and NRPD7 proteins as bait and whole plant protein extract to see if RDR2 binds. Tom Ream has cloned NRPB4, NRPD4, NRPB7, NRPD7 and NRPE7 cDNAs in the Pikaard lab. Preliminary bacterial expression trials have failed to isolate soluble protein (Haag, Ream and Pikaard, unpublished). Optimization of growth and induction procedures is still needed or alternatively, yeast could be used as an expression system.

In vivo analysis can be performed by crossing the *NRPD1-FLAG (nrpd1-3)* transgenic line into the *nrpd4* and *nrpd7* mutant backgrounds. Both of these mutant lines are viable and available for study. Western analysis of immunoprecipitated NRPD1-FLAG samples would reveal if RDR2 co-immunoprecipitates in the absence of either or both of these proteins. Co-localization of NRPD4/NRPE4 with NRPD1 and NRPE1 suggests that the NRPD4/NRPE4 is not always associated with Pol IV and Pol V (He et al., 2009b) lending credence to the hypothesis that the NRPD4/7 subcomplex may be

dissociable. It would be informative to test for NRPD4 co-localization with RDR2 to determine if they co-localize to a greater extent than NRPD1 and RDR2 co-localize (Chapter 5).

What are the implications of the Pol IV-RDR2 interaction with regard to RDR2 transcription initiation? In other words, does RDR2 require a free 3' end to initiate or is it capable of transcribing Pol IV products internally (Figure 4)? This may have important implications for understanding the role of NRPD4/7. If RDR2 requires free 3' ends to initiate dsRNA production from Pol IV products, then it is hypothesized that NRPD4/7 may play an important role in chaperoning the Pol IV transcript to RDR2, regardless of whether the NRPD4/7 subcomplex actually mediates the Pol IV-RDR2 interaction. Current evidence from Arabidopsis RDR6 supports the 3' end initiation model because *in*

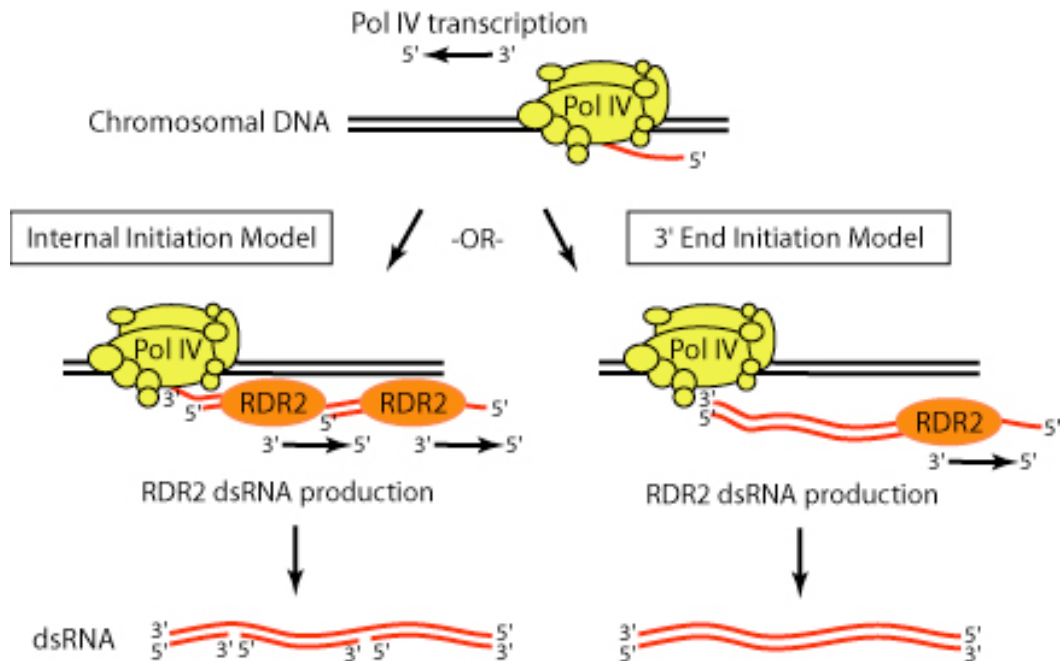


Figure 4. Two proposed models for RDR2 polymerase initiation using Pol IV transcript as a template.

in vitro assays demonstrate RDR6 initiates from the 3' end to form stable dsRNA products and that RDR6 does not use miRNA primers to initiate internal dsRNA production (Curaba and Chen, 2008). This is unlike the case of *Neurospora crassa* QDE-1 which has both primer-dependent and independent activities (Makeyev and Bamford, 2002).

Still to be addressed is whether Pol IV activity is dependent on RDR2. It has been observed in preliminary analysis that Pol IV *in vitro* activity is weaker in the *rdr2-1* mutant background (Chapter 5). Pol IV protein stability does not appear to be an issue in the *rdr2-1* mutant background since NRPD1 and NRPD2 are still detectable in Western blot analysis of immunoprecipitated NRPD1-FLAG protein (Chapter 5). Thus, it is hypothesized that RDR2 may actually stimulate Pol IV transcriptional activity.

Interestingly, the RdRPs of *S. pombe* and *Tetrahymena* have both been demonstrated to stimulate *in vitro* dicer cleavage activity when the two interact (Colmenares et al., 2007; Lee and Collins, 2007).

Under normal lab growth conditions, Pol IV interacts with RDR2 but not RDR6 *in vivo* (Figure 2). Interestingly, Pol IV plays a role in the natural-antisense siRNA (nat-siRNA) pathway with RDR6, not RDR2, and members of the trans-acting siRNA pathway (Borsani et al., 2005). It would be informative to test NRPD1-FLAG plants under salt-stressed conditions to determine if Pol IV preferentially interacts with RDR6 reflecting the requirements of the nat-siRNA pathway. The determination of factors that control such a switch could reveal a great amount about how substrates are channeled through the various Arabidopsis RNA silencing pathways.

RNA Polymerase V

Despite *in vivo* evidence that Pol V is a DNA-dependent RNA polymerase producing short RNA transcripts that are 5' triphosphorylated or capped and lack 3' polyadenylated ends (Wierzbicki et al., 2008), Pol V *in vitro* activity has remained elusive (Huang et al., 2009; Onodera et al., 2005) (Haag and Pikaard, unpublished) (Chapter 5). Experiments to date have focused on using Pol V affinity purified samples bound to anti-FLAG resin. It is possible that non-ideal reaction conditions and/or the wrong nucleic acid template are to blame. Efforts should be made to design a column-based purification approach in case the FLAG resin interferes with Pol V function or required cofactors are lost in the immunoprecipitation procedure. If Pol V-enriched column fractions still fail to display *in vitro* activity, individual column fractions can be added back to the Pol V-enriched fractions to determine if any stimulate Pol V activity. Subsequent purification steps can then be employed to identify the required factors.

In the meantime, much can still be learned from Pol V by studying its *in vivo* functional requirements. It is hypothesized that Pol V transcripts form RNA scaffolds that help recruit DNA methylation and chromatin modification machinery (Wierzbicki et al., 2008; Wierzbicki et al., 2009). An entire series of NRPE1 CTD deletions have been analyzed for their ability to complement *nrpe1* mutants (Chapter 6), but the effect of the NRPE1 CTD deletions on Pol V-dependent transcription *in vivo* remains uncharacterized. It is hypothesized that the NRPE1 CTD is required for Pol V-dependent transcription and Pol V association with Pol V-dependent loci. This can immediately be tested by RT-PCR and chromatin immunoprecipitation (ChIP) experiments, respectively.

The NRPD1 and NRPE1 CTD deletion analysis brought to light the interesting observation that restoration of siRNA production or DNA methylation independent of the

other is not capable of bringing about a silenced state (Chapter 6). The *AtSNI* locus and other potentially affected loci should be analyzed by ChIP to determine the chromatin marks present. It is hypothesized that affected loci will still have active marks (H3Ac) indicating both Pol IV-generated siRNAs and Pol V-directed DNA methylation are required for a switch to the silenced state (H3K27 and H3K9) at the chromatin level.

Steps towards in vitro reconstitution of the RdDM pathway

The Arabidopsis RdDM pathway has largely been elucidated via genetic screens and studies that infer biochemical activities based on molecular phenotypes. It is impressive that the field has managed to piece together as much of the pathway as it has based on this approach. This does not negate the value of confirming the biochemistry hypothesized by the genetic evidence, though. Work by many labs is taking us closer to the day when the entire RdDM pathway will be able to be reconstituted *in vitro*. Affinity purified Pol IV from whole plant extract has been demonstrated to be a DNA-dependent RNA polymerase physically coupled to RDR2 (Chapter 5). Affinity purified DCL3 from whole plant extract is capable of cleaving dsRNA to produce 24nt siRNAs (Qi et al., 2005). Bacterially expressed and affinity purified HEN1 protein samples methylate both 3' overhang strands of siRNA and miRNA duplexes (Yang et al., 2006; Yu et al., 2005). Affinity purified AGO4 protein has demonstrated siRNA loading and mRNA target cleavage activities (Qi et al., 2006).

Thus, only a few components of the pathway remain to be biochemically elucidated *in vitro*. RDR2 activity has yet to be convincingly demonstrated from affinity purified whole plant extracts (Haag and Pikaard, unpublished), though RDR6 *in vitro*

activity has been published (Curaba and Chen, 2008). Pol V *in vitro* activity has yet to be obtained, but work is ongoing. DRM2, a *de novo* cytosine DNA methyltransferase required for the RdDM pathway, has also yet to be biochemically defined *in vitro* as a bona fide DNA methyltransferase. Ek Han Tan in the Pikaard lab has attempted to obtain DRM2 activity from protein expressed in bacteria and immunoprecipitated from whole plant extract. These efforts may require alternative strategies utilizing column-purified protein samples or protein expressed in baculovirus, *in vitro* wheat germ transcription/translation systems or in Agro-infiltrated tobacco leaves. These approaches still offer the benefits of eukaryotic post-translational modifications and, in the plant-based systems, the ability to assemble required complexes. Arabidopsis NRPE2, the second-largest subunit of Pol II, has successfully been expressed and purified from Agro-infiltrated tobacco leaves and found to assemble with endogenous tobacco Pol II subunits to be transcriptionally active *in vitro* (Haag and Pikaard, unpublished). This approach was successful for obtaining large quantities of Arabidopsis RDR6 for *in vitro* studies (Curaba and Chen, 2008). Attempts to recapitulate the entire RdDM pathway *in vitro* will provide an important means of confirming what the genetic data suggests or point out gaps in our knowledge not made apparent by the genetic data.

ii.

ROLES OF THE NRPE1 AND NRPE2 C-TERMINAL DOMAINS

Introduction

The C-terminal requirements of NRPD1 and NRPE1 have only begun to be elucidated on a domain-by-domain basis (El-Shami et al., 2007) (Chapter 6). Future work should focus on dissecting the requirements of the Defective-Chloroplast and Leaves-like (DCL) domain, which is required for both Pol IV and Pol V *in vivo* function (Chapter 6). While the NRPE1 WG motifs have received widespread acceptance as a required platform for AGO4 interaction and Pol V function (El-Shami et al., 2007; Till and Ladurner, 2007), our findings temper this view calling into question both the prevalence of an *in vivo* Pol V-AGO4 interaction and the requirement for the majority of WG motifs for *in vivo* complementation (Chapter 6).

It is difficult not to make comparisons between the Pol IV and Pol V largest subunit CTD extensions and that of Pol II from a functional perspective. Now that we know which domains are required for *in vivo* function, the focus must now turn to why they are required. This will take us into the realm of protein-protein interactions, post-translational modifications and possible enzymatic or regulatory functions.

Defective Chloroplast and Leaves-like Domain

The Defective Chloroplast and Leaves-like (DeCL) domain is required for full complementation of both *nrpd1* and *nrpe1* mutants, but has no known function (Chapter 6). The ancestral NRPD1 largest subunit is believed to have arisen from a genomic DNA duplication of the Pol II largest subunit after the common ancestor of Charales and land plants diverged from other green algae (Luo and Hall, 2007). Sometime after this duplication event but before the duplication of NRPD1/NRPE1, it is the hypothesis of this author that a C-terminal duplication of one of the DeCL domain genes was integrated

at the 3' end of the ancestral *NRPD1* gene. This would therefore explain the presence of the DeCL domain in both NRPD1 and NRPE1 proteins and the conserved intron/exon structure with AtDCL and At3g46630 (Haag and Pikaard, unpublished).

Luo and Hall (2007) reported that the DeCL domain exists at the C-terminus of NRPD1 and NRPE1 proteins in angiosperms, but not bryophytes (non-vascular land plants that reproduce via spores, i.e. mosses). This assertion was based on *Spagnum* NRPD1 sequence analysis. Since the publication of their article, the genome of *Physcomitrella patens* has been released. The DeCL domain is present in one of the two *Physcomitrella* NRPE1-like proteins suggesting the DeCL domain insertion event occurred earlier than once thought. The *Physcomitrella* NRPD1 protein discovered by BLAST search lacks the DeCL domain but the entire contig is not available so it is not certain that the sequence represents the full-length gene. Thus, the jury is still out as to whether *Physcomitrella* NRPD1 contains the DeCL domain, as well as, whether any or all of the predicted *Physcomitrella* NRPD1 and NRPE1 subunits are functional. This is a question that Andrzej Wierzbicki, a postdoc in the Pikaard lab, will be pursuing as an assistant professor at the University of Michigan.

So what is the function of the DeCL domain? BLAST analysis (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) of the DeCL domain sequence results in predominantly plant-specific hits. There are three other proteins in Arabidopsis that contain this domain. AtDCL is a plastid protein required for the processing of ribosomal RNA and ribosome biogenesis (Bellaoui et al., 2003). DOMINO1 is a nuclear localized protein required for embryogenesis with phenotypes consistent with defects in ribosome biogenesis (Lahmy et al., 2004). The remaining DeCL domain containing protein in

Thus, while no specific function has been ascribed to the DeCL domain, it can be hypothesized that it directly or indirectly plays a role in RNA-related processes. Given that NRPD1 and NRPE1 are the largest subunits for RNA Polymerase IV and V complexes, respectively, it stands to reason that the DeCL domain may be directly binding or processing Pol IV and Pol V transcripts, interacting with siRNAs or interacting with proteins that in turn are related to RNA binding and/or processing events. The DeCL domain cDNA (encoding NRPD1 aa 1337-1453) has been cloned and expressed in *Arabidopsis thaliana* (Chapter 6). In addition, the protein has been successfully expressed in bacteria (Haag and Pikaard, unpublished). These two tools can serve as the starting point to determine if the DeCL domain binds RNA/siRNA *in vivo* and *in vitro*, or is capable of *in vitro* RNA cleavage. Additionally, the DeCL domain contains a highly conserved DFSYRKC motif with other invariant amino acids upstream and downstream (Figure 6). Site-directed mutagenesis of these amino acids in NRPD1 and NRPE1 could be performed to test if the mutants are capable of *in vivo* complementation. The site-directed mutagenesis constructs could also be tested for abrogation of any observed *in vitro* RNA binding/processing activities. Finally, a yeast two-hybrid screen could be performed to determine if there are any protein-protein interactions (discussed below).

Platform for protein-protein interactions

Like the Pol II CTD, the NRPD1 and NRPE1 CTDs are potential platforms for protein-protein interactions. These interactions may regulate Pol IV and Pol V activities, modify and/or process RNA transcripts, or recruit Pol IV and Pol V to specific loci. The Pikaard lab and others are using many complementary approaches to identify protein-



Figure 6. Conserved DeCL domain sequence block in NRPD1 and NRPE1 proteins from diverse plant species with

AtDCL, DOMINO1 and At3g46630 reference sequences. Generated by Jalview using ClustalW2 protein sequence alignment

with consensus sequence below.

protein interactions to elucidate the RdDM pathway and better understand Pol IV and Pol V regulation. Two genetic screens and mass-spec analysis of affinity purified Pol V have identified a putative transcription factor, named KTF1 or SPT5-like (Bies-Etheve et al., 2009; He et al., 2009b; Huang et al., 2009). The Pikaard lab has designed a genetic screen of its own to identify modifiers of the NRPE1 CTD (Haag, Tan and Pikaard; described in detail in a later section).

In collaboration with the Pacific Northwest National Laboratory, Tom Ream, Ek Han Tan, Todd Blevins, Alexa Vitins and I have been analyzing affinity purified protein samples from *Arabidopsis* by LC-MS/MS to identify protein-protein interactions. The focus has been on members of the RdDM pathway (NRPD1, RDR2, DCL3, HEN1, AGO4, NRPE1, NRPE5 and DRM2) in addition to members of related silencing pathways (RDR6, DCL2, DCL4, DRB4, SGS3, MBD6 and HDA6) with the relevant controls. Tom Ream has already had great success with this approach for the elucidation of the complete subunit compositions of Pol I, II, III, IV and V in *Arabidopsis thaliana* complemented by genetics and co-IP approaches (Ream et al., 2009) (Ream, Pontvianne, Haag, Nicora, Norbeck, Pasa-Tolic and Pikaard, unpublished).

One difficulty with the resultant data sets is the large number of candidate protein-protein interactions identified. Several different filters are being used. These include comparison to vector only and wild type controls, co-expression analysis (<http://www.arabidopsis.leeds.ac.uk/act/coexpanalyser.php#CO1>), predicted or known localization patterns, literature searches and a bit of common sense. In the end though, one is still left with a large number of candidates that must be screened for RdDM defects by isolating homozygous mutants and performing reciprocal co-IP analysis. This

involves the generation of transgenic lines that complement the mutant and/or the production of antibodies.

In an attempt to streamline these efforts, the NRPD1 and NRPE1 epitope-tagged CTD deletion lines and individual over-expressed CTD domains have been sent off for LC-MS/MS analysis. It is hoped that by comparing the NRPD1 and NRPE1 full-length data sets with these and the proper controls, the number of false-positives will be reduced and also allow identification of the protein-protein interaction domains. Results are still being analyzed, but some interesting trends are emerging.

A glutamine-rich protein, GRP23, was detected with both the NRPE1 full-length and NRPE1 Δ 1251-1976 affinity purified proteins, suggesting it interacts with the Pol V core and not the NRPE1 CTD (Haag, Norbeck, Nicora, Pasa-Tolic and Pikaard, unpublished). GRP23 is a pentatricopeptide repeat (PPR) protein that has been published to interact with the RNA Polymerase II Rpb3-like subunit in Arabidopsis via a yeast two-hybrid (Y2H) screen and bimolecular fluorescence complementation (BiFC) (Ding et al., 2006). Interestingly, this Rpb3-like subunit is NRPE3b, which favors association with Pol V, though it is found to a low degree with Pol II and Pol IV (Ream et al., 2009). GRP23 was not found to interact with NRPB3/NRPD3/NRPE3a by Y2H. GRP23-YFP transgenic plants have been obtained from Dr. Wei-Cai Yang for immunoprecipitation to confirm association with the Pol V complex by Western blot and to also test for association with the Pol II and Pol IV complexes. Since *grp23* mutants are embryo lethal (Ding et al, 2006), GRP23 likely does have a required role with Pol II transcription as Pol IV and Pol V mutants are viable. This suggests that the two NRPB3 paralogs in Arabidopsis have distinct functions and may preferentially associate with a given

polymerase in a tissue-, stress- or developmentally-specific manner. Such a role for subunit variants has been hypothesized previously (He et al., 2009a; Ream et al., 2009) and future work should focus on exploring this potential additional layer of RNAP subunit composition complexity among the Arabidopsis multi-gene subunit families.

The NRPE1 QS-rich domain (aa 1851-1977) has yielded a number of interesting candidate interactions. Among them are seven subunits of the Arabidopsis Mediator complex (Haag, Norbeck, Nicora, Pasa-Tolic and Pikaard, unpublished) (Table 1). Mediator is a eukaryotic, multi-subunit complex that interacts with yeast Pol II subunits Rpb1, Rpb2, Rpb3, Rpb6, Rpb11 and Rpb12, including contacts with the Rpb1 CTD (Chadick and Asturias, 2005). T-DNA insertion mutant lines have been ordered for MED4, MED8 and MED14. Mediator mutants have been isolated previously in Arabidopsis so are known to be viable (Autran et al., 2002; Backstrom et al., 2007). Once homozygous mutants are isolated they will be screened for defects in RdDM. Antibodies are also commercially available for MED6 and MED7 (www.agrisera.com)

Protein	AGI	NRPE1 QS
MED4	At5g02850	5/6
MED8	At2g03070	4/4
MED9	At1g55080	2/2
MED14/SWP	At3g04740	2/2
MED15	At1g15780	2/3
MED21	At4g04780	1/1
MED27	At3g09180	1/1

Table 1. Mediator subunits found in LC-MS/MS analysis of FLAG-NRPE1 aa1851-1977 (NRPE1 QS). The first numeral in the NRPE1 QS column represents the number of unique peptides identified and the second number represents the total scan count.

and should be used to test for MED6 and MED7 co-IP with immunopurified Pol II, Pol IV and Pol V complexes.

Continuing with this theme, candidate proteins have been identified that are putative Pol II transcription repressors, transcription factors, DNA-binding proteins, RNA-binding proteins, an exoribonuclease and a TFIID interactor (Table 2). Given Pol IV and Pol V evolution from Pol II (Luo and Hall, 2007; Ream et al., 2009), the shared

Table 2. Selection of candidate proteins identified by LC-MS/MS that may interact with the NRPE1 CTD. The first number in the NRPE1 QS column refers to the number of unique peptides identified and the second number refers to the total peptide scan count.

AGI	NRPE1 QS	Annotation
At4g27740	37/85	Zn-finger, nuclear localization (Yippee-like)
At3g22380	22/39	nuclear reg in A.t. circadian clock (TIC)
At1g72010	10/15	TCP family txp factor, plant specific
At4g32551	11/22	LUG, forms a co-repressor complex w/ SEU, HDA19 and Mediator (similar to yeast Tup1)
At1g43850	9/11	SEU, forms a co-repressor complex w/ LUG, HDA19 and Mediator (similar to yeast Ssn6)
At1g17440	9/18	nuclear localized, interacts w/TFIID (similar to yeast Taf61)
At1g14580	9/16	Zn-finger (C2H2 type)
At3g04590	7/17	DNA-binding family protein
At2g44710	7/8	RNA recognition motif (RRM)
At2g31370	6/6	bZIP txp factor (POSF21)
At3g04590	5/14	DNA-binding family protein
At3g54230	5/6	nucleic acid binding
At5g16840	4/8	RNA recognition motif (RRM)
At1g07920	4/6	elongation factor 1-alpha, EF-1-alpha
At3g47620	4/6	ATTCP14, TCP family txp factor
At5g60390	4/6	elongation factor 1-alpha, EF-1-alpha
At1g06070	4/5	bZIP txp factor (bZIP69)
At5g08330	3/5	TCP family txp factor
At2g02080	3/5	ATIDD4, txp factor
At5g23280	3/4	TCP family txp factor
At5g51660	3/4	CPSF160, txp factor

Table 2 (continued)

At5g52040	3/4	ATRSP41, Arg/Ser-rich splicing factor
At4g09000	3/4	GF14, GRF1 (GENERAL REGULATOR FACTOR 1)
At1g07930	3/4	elongation factor 1-alpha, EF-1-alpha
At1g13960	3/4	WRKY4, DNA binding, txp factor
At2g03340	3/4	WRKY3, DNA binding, txp factor
At5g20730	3/3	MSG1, ARF7, TIR5, BIP, NPH4, txp factor
At1g58220	2/4	myb family txp factor
At1g78300	2/3	GF14 OMEGA, GRF2 (GENERAL REGULATORY FACTOR 2)
At1g15780	2/3	protein binding, txp cofactor
At1g49600	2/3	ATRBP47A, RNA binding
At1g35160	2/3	GF14 PHI, GRF4 (GENERAL REGULATORY FACTOR 4)
At3g02520	2/3	GF14 NU, GRF7 (GENERAL REGULATORY FACTOR 7)
At3g01210	2/3	nucleic acid binding
At3g27010	2/3	PCF1, AT-TCP20, txp factor
At4g17950	2/3	DNA binding family protein
At5g38480	2/3	RCI1, GRF3 (GENERAL REGULATORY FACTOR 3)
At2g21660	2/2	GR-RBP7, GRP7, CCR2, ATGRP7
At1g75660	2/2	XRN3 (5'-3' exoribonuclease)
At1g58100	2/2	TCP family txp factor
At1g19220	2/2	IAA22, ARF11, ARF19, txp factor
At3g06590	2/2	transcription factor
At4g25500	2/2	ATRSP40, ATRSP35 (Arg/Ser-rich splicing factor)
At5g19790	1/2	RAP2.11 (related to AP2 11) DNA binding/txp factor
At4g00830	1/2	RNA recognition motif (RRM)
At4g27000	1/2	ATRBP45C, RNA binding

use of Pol II regulatory machinery will likely be an emerging theme in the field during the coming years. Genetic screens are likely to miss Pol IV and Pol V regulatory machinery that is shared with Pol II because such mutants will probably be lethal or display a weak or no phenotype due to being members of multi-gene families.

Y2H analysis is a complementary approach and will be performed by Todd Blevins in the Pikaard lab at Indiana University. The NRPD1 DeCL domain and the NRPE1 CTD will be used as bait sequences. Interestingly, a Y2H screen was conducted previously using the AtDCL full-length protein as bait (Mohammed Bellaoui, personal

communication). The results have not been published, nor have they been confirmed, but Dr. Bellaoui has kindly shared them since this project is no longer being pursued. As expected, the majority of candidates are predicted chloroplast proteins since the AtDCL protein is plastid-localized, but a number of interesting candidates were identified that are predicted to be nuclear localized (Table 3). These likely are not true partners with the chloroplast AtDCL protein, but may be recognizing the conserved DeCL domain and be true interacting partners with the nuclear localized NRPD1, NRPE1 or DOMINO1. T-DNA mutants should be obtained from the ABRC and tested for defects in RdDM. The PRH75 protein is especially interesting given its experimentally determined localization pattern in the nucleus and nucleolus and the fact that it is associated with a 500kD complex of unknown composition (Lorkovic et al., 1997; Lorkovic et al., 2004).

AGI	Annotation	Comment
At1g21200	Transcription factor	
At2g22430	Homeodomain leucine zipper class I protein, ATHB6	regulates hormone responses in Arabidopsis (Himmelbach et al, 2002)
At2g32030	GCN5-related N-acetyltransferase	
At3g11100	Transcription factor	Similar to At1g21200
At5g03180	Zinc finger (RING/FYVE/PHD-type) family protein	
At5g13920	Zinc knuckle (CCHC/GRF-type) family protein	
At5g18650	Zinc finger (RING/FYVE/PHD-type) family protein	
At5g62190	DEAD/DEAH box RNA helicase, PRH75	nuclear localized, present in a 500kD complex (Lorkovic et al, 1997)

Table 3. Predicted nuclear-localized proteins that interact with AtDCL by Y2H

(Bellaoui and Gruissman, unpublished).

Target of post-translational modifications

The NRPE1 C-terminal domain extension is a potential target for post-translational modifications based on parallels with the NRPB1 CTD and NRPE1 Western blot analysis (Pontes et al., 2006; Pontier et al., 2005). Western blot analysis of individually over-expressed NRPE1 CTD domains demonstrates that only the QS-rich domain, NRPE1 aa 1851-1977, migrates at larger than predicted molecular weights (Chapter 6). Interestingly, it is this region of the NRPE1 CTD that is most highly predicted to be a target of phosphorylation and glycosylation modifications using NetPhos2.0 (<http://www.cbs.dtu.dk/services/NetPhos/>) and Yin-O-Yang (<http://www.cbs.dtu.dk/services/YinOYang/>) predictive models. LC-MS/MS analysis of large-scale affinity purified FLAG-NRPE1 aa 1851-1977 has identified several candidate amino acid positions with detectable phosphorylation (Table 4) (Haag, Ream, Nicora, Norbeck, Pasa-Tolic and Pikaard, unpublished).

These results are currently in the midst of being replicated with both the FLAG-NRPE1 aa 1851-1977 and NRPE1-FLAG full-length affinity purified proteins from Arabidopsis. If the results are confirmed, it will be the first experimental evidence that the NRPE1 CTD is post-translationally modified. Follow-up experiments may include the development of phospho-specific peptide antibodies, site-directed mutagenesis of candidate amino acids and *in vivo* ³²P labeling. The significance of these post-translational modifications is still uncertain as NRPE1 aa 1851-1977 can be deleted with no detectable impact on *nrpe1* complementation. This may suggest that the right experimental assay has not yet been performed to identify a mutant phenotype or these

NRPE1 Amino Acid	Context v	PTM	Score (NetPhos2.0)	Peptides	ScanCount
S1540	SETESGPAA	P	0.034	1	1
K1551	AWDKKKSET	U	n/a	1	1
S1553	DKKKSETEP	P	0.985*	1	1
T1864	NQDATPPGE	P	0.509*	2	2
S1877	PPNQSIGNG	P	0.049	1	1
T1887	DDFQTQTQS	P	0.350	1	1
T1889	FQTQTQSQS	P	0.291	1	1
S1891	TQTQSQSPS	P	0.036	1	1
S1893	TQSQSPSQT	P	0.978*	5	18
S1895	SQSPSQTRA	P	0.680*	2	2
T1897	SPSQTRAQS	P	0.228	1	1
S1901	TRAQSPSQA	P	0.988*	9	30
S1903	AQSPSQAQA	P	0.115	3	3
S1909	AQAQSPSQT	P	0.631*	6	11
S1911	AQSPSQTQS	P	0.054	1	1

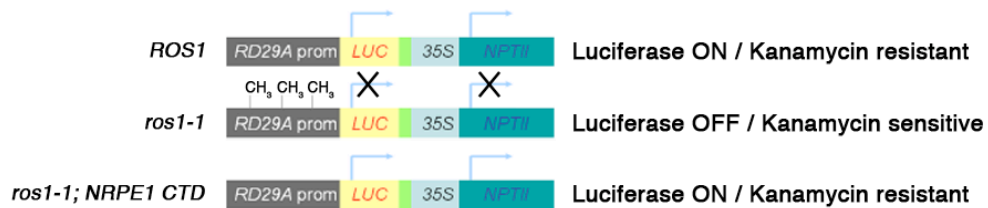
Table 4. Predicted and experimentally observed NRPE1 amino acids that are phosphorylated or ubiquitinated. Amino acids highlighted in yellow have a high predictive NetPhos2.0 score (*) and were identified by multiple peptides and scan counts in the LC-MS/MS analysis. PTM = Post-Translational Modification; P = Phosphorylated; U = Ubiquitinated; Peptides = the number of unique peptides identified with the amino acid bearing a particular PTM; ScanCount = the total number of peptides identified with the amino acid bearing a particular PTM.

post-translational modifications are not functionally significant. LC-MS/MS analysis of the NRPE1 CTD in the context of the full-length protein will hopefully resolve this.

Applications for dominant suppression of RdDM

The C-terminal domains of NRPE1 are capable of dominantly suppressing the RdDM pathway when over-expressed in wild type Arabidopsis plants (Chapter 6). The transgenic plants still have a functional endogenous *NRPE1* gene but they behave as *nrpe1* mutants. To test if the over-expressed NRPE1 CTD (NRPE1 aa 1234-1842; referred to as 35S::YFP::CTD) is capable of releasing a transgene from the silenced state, the 35S::YFP::CTD transgene was transformed into a *ros1-1* mutant background that contained a silenced luciferase reporter (Tan, Haag and Pikaard, unpublished; the reporter line was provided by Jian-Kang Zhu) (Figure 7A). In the wild type background, the stress-inducible RD29A promoter is activated by cold, ABA or salt stress and the plants

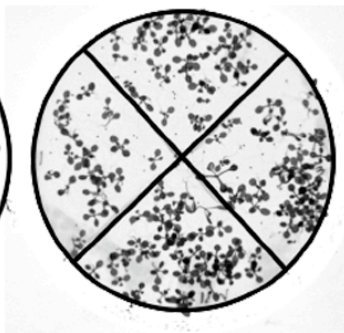
A. Basis of suppressor screen



B. Genotypes planted



C. Visible light



D. Luciferase

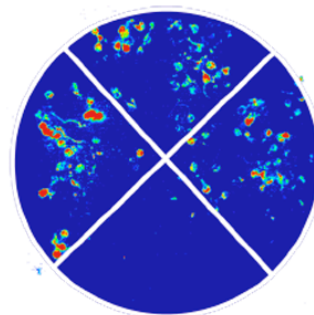


Figure 7. Luciferase reporter screen to detect defects in the RdDM pathway. Over-expression of the NRPE1 CTD (35S::YFP::CTD) dominantly suppresses silencing of the RD29A promoter and leads to activation of the luciferase reporter under stress-inducible conditions. Figure modified from Tan and Pikaard, unpublished.

express the luciferase reporter and NPTII selectable marker conferring kanamycin resistance. In the *ros1-1* mutant background, a DNA demethylase enzyme is no longer functional and the RD29A promoter becomes hypermethylated. Luciferase and kanamycin resistance both fail to be activated in these plants after stress treatment (Gong et al., 2002) (Figure 7B-D). Transformation of this genetic background with the 35S::YFP::CTD transgene dominantly suppresses silencing of the transgene promoter and reactivates luciferase expression under stress-inducible conditions (Figure 7B-D) (Tan and Pikaard, unpublished). Thus, over-expression of the NRPE1 CTD is capable of dominantly suppressing the silencing of both endogenous and transgene targets.

The *RD29A::LUC::35S::NPTII; ros1-1; 35S::YFP::CTD* genetic background could serve as the basis for an EMS mutagenesis screen to identify suppressors of the over-expressed NRPE1 CTD. In such mutants, the *RD29A* promoter would be hypermethylated and luciferase activity silenced. It is believed this screen would identify interactors and/or modifiers of the NRPE1 CTD in addition to mutants in the RdDM pathway already discovered with the *ros1-1* suppressor screen (He et al., 2009a).

There are also biotechnology applications for the ability to dominantly suppress RdDM. Transformation of agriculturally significant crops such as soybean, maize, cotton and rice has been a major investment made by seed companies in the previous decades with the goal of increasing yield, stress tolerance, and conferring insect and pesticide resistance (Shewry et al., 2008). This process can be very time-consuming as transformation strategies are not always efficient, crop generation times can be lengthy and transgenes may be silenced. One way to minimize the chances of transgene silencing is to select single insertion events but even this is not always effective. Transformation

of crops first with the over-expressed NRPE1 CTD could dominantly suppress RdDM and possibly provide a genetic background more amenable to the testing of new transgenes. This is hypothesized to reduce the chances of transgene silencing. Once a transgene has been determined to have the desired effects and is ready to go on to later stages of development, the over-expressed NRPE1 CTD transgene could either be crossed out of the genetic background or the transgene transformed into a more suitable genetic background for production and marketing. An added benefit of this approach is that plant genomes already encode the NRPE1 sequence and therefore would not be harboring “foreign” genes.

Towards this end, one must determine if the *Arabidopsis thaliana* NRPE1 CTD is capable of dominantly suppressing RdDM in distantly related plants or if it is only effective in close relatives due to the divergence of the NRPE1 CTD across plant species (Chapter 6). Because transformation of maize is a time consuming process, *Arabidopsis thaliana* plants were transformed with a portion of the *Zea mays* NRPE1 CTD, aa 1287-1612, and tested for dominant suppression of RdDM (Haag and Pikaard, unpublished). The clone contains the maize NRPE1 WG motifs and two 27 aa repeat elements but lacks the DeCL domain (maize *NRPE1* genomic sequence data for primer design was kindly provided by Vicki Chandler). Over-expression of the maize NRPE1 CTD in wild type *Arabidopsis thaliana* plants failed to dominantly suppress DNA methylation at the *AtSN1* locus, a marker of RdDM (Figure 8). This suggests that the *Arabidopsis thaliana* NRPE1 CTD would not be effective at dominantly suppressing RdDM in distantly related plants either. Experiments to test the maize NRPE1 CTD in *Zea mays* and other monocots such as *Oryza sativa* (rice) and *Brachypodium distachyon* (a model for grasses and cereals)

would be required. Given the higher degree of CTD conservation among these close plant relatives (Chapter 6), it is hypothesized that such a strategy could be effective.

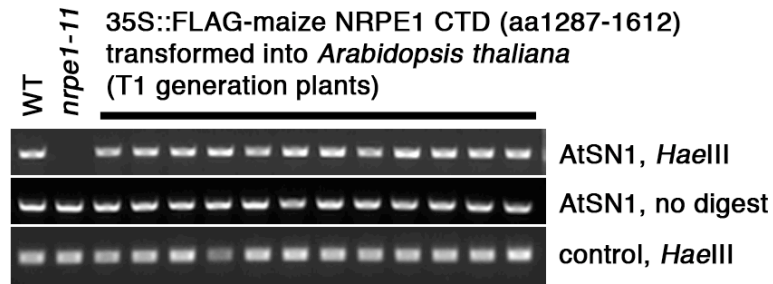


Figure 8. Chop-PCR experiment to assay DNA methylation at the *AtSN1* locus.

iii.

STRUCTURE-FUNCTION ANALYSIS

Introduction

Primary sequence analysis of the known Pol IV and Pol V subunits is able to identify regions of sequence variance that may be important in determining what makes these two polymerases functionally distinct from each other and Pol II (Haag et al., 2009; He et al., 2009a; Herr et al., 2005; Lahmy et al., 2004; Landick, 2009; Ream et al., 2009). The challenge lies in determining if and how the primary sequence divergence translates to divergence from Pol II at the tertiary level and if and how it affects the function of Pol IV and Pol V. To get at these questions, the elucidation of Pol IV and Pol V structures is required.

Determination of Pol IV and Pol V structures

The 10-subunit core and the complete 12-subunit atomic structures of yeast Pol II have been resolved by x-ray crystallography (Armache et al., 2005; Cramer et al., 2001) and have offered a greater understanding of Pol II transcription (Gnatt et al., 2001; Westover et al., 2004a; Westover et al., 2004b). By extension, the structures have been used to help interpret the 12 Å cryo-electron microscopic (cryo-EM) structure for the complete 14-subunit yeast Pol I (Kuhn et al., 2007). To explain the differences between the EM map and the shared Pol II core structure, a homology model was constructed for the Pol I core. This analysis identified conserved folds between Pol I and Pol II despite divergent primary sequences as well as helped define Pol I-specific surfaces. This Pol II homology modeling approach was also successfully used for modeling the 9-subunit core of yeast Pol III (Jasiak et al., 2006). Both studies also incorporated x-ray structures of Pol I and Pol III-specific subcomplexes to obtain a complete 14-subunit yeast Pol I structure and 11-subunit yeast Pol III structure (Jasiak et al., 2006; Kuhn et al., 2007).

A similar strategy could be used to determine the complete subunit structures of Pol IV and Pol V via cryo-EM. The conserved subunit composition of these complexes with yeast Pol II would allow a direct comparison. This would require large-scale affinity purification of Arabidopsis Pol II, IV and V complexes, a technique already worked out by the Pikaard lab (Ream et al., 2009). If the protein quantity obtained from Arabidopsis is still not great enough for EM analysis after scaling up, one could turn to alternative tissue sources such as broccoli, cauliflower or maize. As discussed previously, the Pikaard lab is working to transform broccoli with epitope tagged *NRPB2*, *NRPD1* and *NRPE1* transgenes; cauliflower has been used for Pol V affinity purification and subunit composition analysis (Huang et al., 2009), and transgenic maize has been

generated with two of the three *NRPD2/NRPE2* genes epitope tagged (Vicki Chandler, personal communication).

As with the yeast Pol I analysis, a homology model could be built between yeast Pol II and each of the Arabidopsis complexes being studied—Pol II, Pol IV and Pol V. Regions of conservation and divergence could be identified for all three. Most interesting would be the active site centers of Pol IV and Pol V which have undergone a great degree of primary sequence divergence from Pol II as well as the region corresponding to the Pol II bridge helix which is predicted missing in *NRPD1* and *NRPE1* proteins.

It is possible that in addition to the Pol IV subunit structure, the structure of RDR2 and its contacts with Pol IV will be revealed. RDR2 co-IPs with Pol IV using the Arabidopsis large-scale affinity purification protocol (Chapter 5) and should be detectable as a unique electron density not present in the Pol II or Pol V structures. This can be verified by comparing the Pol IV EM structure with that of Pol IV purified from an *rdr2* mutant background, *NRPD1-FLAG (nrpd1a-3; rdr2-1)* (Chapter 5).

The *NRPB4/7* (Pol II), *NRPD4/7* (Pol IV) and *NRPE4/7* (Pol V) subcomplexes would be good candidates for performing x-ray structure analysis. Each polymerase has a unique Rpb7-like subunit (Ream et al., 2009). Pol II has a unique Rpb4-like subunit while Pol IV and V share a Rpb4-like paralog (He et al., 2009a; Ream et al., 2009). Efforts should continue in the Pikaard lab to express these subunits in bacteria, or alternatively in yeast. The human and yeast *Rpb4/7* subcomplexes have previously been crystallized in addition to the archaeal RNAP E/F and yeast Pol III C17/25 counterparts (Armache et al., 2005; Jasiak et al., 2006; Meka et al., 2005; Todone et al., 2001) and can

thus be used as guides for the crystallization conditions in addition to structural comparison.

The Rpb5-like five-member gene family would be another candidate for x-ray structure analysis as the NRPE5 subunit is distinct to Pol V and has a unique N-terminal extension and the absence of a C-terminal motif present in the NRPB5/NRPD5 subunit shared by Pol II and Pol IV (Lahmy et al., 2009; Larkin et al., 1999; Ream et al., 2009). The x-ray structure of yeast Rpb5 (Todone et al., 2000) has previously been solved and could offer guidance.

Discovery of Pol IV-nucleic acid contacts

The divergent active site regions of Pol IV and Pol V (Haag et al., 2009) may have novel surfaces and therefore make novel contacts with the DNA template and/or RNA transcript. To assess this possibility, a strategy that identifies protein-DNA contacts by photocrosslinking and mass spectrometry can be utilized (Geyer et al., 2004). Briefly, a photoactivatable DNA oligo template is fed to Pol IV *in vitro* and photocrosslinked. The crosslinked sample is protease digested, DNA-peptide conjugates purified, and the sample hydrolyzed to remove DNA. Peptides are then identified by MALDI-TOF-MS/MS to define DNA-protein contacts. A similar experimental procedure could be performed using a photoactivatable RNA oligo being extended by Pol IV as in the tripartite dsDNA-RNA template (Chapter 5). This would complement the structural analysis and help define the Pol IV template entry channel, active site region and both the DNA and RNA exit channels.

Elucidation of the eukaryotic DdRP subunit assembly pathway

To date eukaryotic polymerase subunits for Pol I, II or III have not successfully been reconstituted *in vitro* to form a transcriptionally active complex (Acker et al., 1997; Kimura and Ishihama, 2000), though the feat has been accomplished with the similarly complex archaeal RNAP (Werner and Weinzierl, 2002). The study of the subunit assembly pathway *in vivo* is limited by the fact that Pol I, II and III are essential for viability. Pol IV and Pol V offer the unique opportunity to assess the contribution of individual eukaryotic RNAP subunits in the assembly of a core RNAP complex. Pol IV and Pol V are not essential to plant viability and mutants have successfully been isolated in four of the five the subunits not shared with Pol I, II or III - NRPD1, NRPE1, NRPD2/NRPE2, NRPE3b, NRPE5, NRPD7, and NRPE7 (He et al., 2009a; Herr et al., 2005; Huang et al., 2009; Kanno et al., 2005; Lahmy et al., 2009; Onodera et al., 2005; Ream et al., 2009). Purified Pol IV or Pol V complexes can be isolated from these homozygous mutant lines and analyzed by cryo EM. Losses in electron density should be able to be compared with the complete core electron density to determine the presence/absence of RNAP subunits and thus infer at least some of the requirements for *in vivo* RNAP assembly.

References

- Acker, J., de Graaff, M., Cheynel, I., Khazak, V., Keding, C., and Vigneron, M. (1997) Interactions between the human RNA polymerase II subunits. *J Biol Chem*, 272(27), 16815-16821.
- Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem*, 280(8), 7131-7134.
- Autran, D., Jonak, C., Belcram, K., Beemster, G.T., Kronenberger, J., Grandjean, O., Inze, D., and Traas, J. (2002) Cell numbers and leaf development in Arabidopsis: a functional analysis of the STRUWWELPETER gene. *EMBO J*, 21(22), 6036-6049.
- Backstrom, S., Elfving, N., Nilsson, R., Wingsle, G., and Bjorklund, S. (2007) Purification of a plant mediator from Arabidopsis thaliana identifies PFT1 as the Med25 subunit. *Mol Cell*, 26(5), 717-729.
- Bellaoui, M., Keddie, J.S., and Grissem, W. (2003) DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. *Plant Mol Biol*, 53(4), 531-543.
- Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R., and Lagrange, T. (2009) RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep*, 10(6), 649-654.
- Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell*, 123(7), 1279-1291.
- Brueckner, F., and Cramer, P. (2008) Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat Struct Mol Biol*, 15(8), 811-818.
- Chadick, J.Z., and Asturias, F.J. (2005) Structure of eukaryotic Mediator complexes. *Trends Biochem Sci*, 30(5), 264-271.
- Chen, L.F.O., Hwang, J.Y., Charng, Y.Y., Sun, C.W., and Yang, S.F. (2001) Transformation of broccoli (*Brassica oleracea* var. *italica*) with isopentenyltransferase gene via *Agrobacterium tumefaciens* for post-harvest yellowing retardation. *Molecular Breeding*, 7(3), 243-257.
- Colmenares, S.U., Buker, S.M., Buhler, M., Dlakic, M., and Moazed, D. (2007) Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Mol Cell*, 27(3), 449-461.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, 292(5523), 1863-1876.
- Curaba, J., and Chen, X. (2008) Biochemical activities of Arabidopsis RNA-dependent RNA polymerase 6. *J Biol Chem*, 283(6), 3059-3066.
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*, 101(5), 543-553.
- Ding, Y.H., Liu, N.Y., Tang, Z.S., Liu, J., and Yang, W.C. (2006) Arabidopsis GLUTAMINE-RICH PROTEIN23 is essential for early embryogenesis and

- encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. *Plant Cell*, 18(4), 815-830.
- Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005) RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev*, 19(19), 2301-2306.
- Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J*, 45(4), 616-629.
- Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem*, 266(1), 71-75.
- El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev*, 21(20), 2539-2544.
- Erhard, K.F., Jr., Stonaker, J.L., Parkinson, S.E., Lim, J.P., Hale, C.J., and Hollick, J.B. (2009) RNA polymerase IV functions in paramutation in *Zea mays*. *Science*, 323(5918), 1201-1205.
- Geyer, H., Geyer, R., and Pingoud, V. (2004) A novel strategy for the identification of protein-DNA contacts by photocrosslinking and mass spectrometry. *Nucleic Acids Res*, 32(16), e132.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*, 292(5523), 1876-1882.
- Goler-Baron, V., Selitrennik, M., Barkai, O., Haimovich, G., Lotan, R., and Choder, M. (2008) Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes Dev*, 22(15), 2022-2027.
- Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L., and Zhu, J.K. (2002) ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell*, 111(6), 803-814.
- Haag, J.R., Pontes, O., and Pikaard, C.S. (2009) Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS One*, 4(1), e4110.
- He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S., and Zhu, J.K. (2009a) NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes Dev*, 23(3), 318-330.
- He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H., and Zhu, J.K. (2009b) An effector of RNA-directed DNA methylation in *Arabidopsis* is an ARGONAUTE 4- and RNA-binding protein. *Cell*, 137(3), 498-508.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science*, 308(5718), 118-120.
- Huang, L., Jones, A.M., Searle, I., Patel, K., Vogler, H., Hubner, N.C., and Baulcombe, D.C. (2009) An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol*, 16(1), 91-93.

- Jasiak, A.J., Armache, K.J., Martens, B., Jansen, R.P., and Cramer, P. (2006) Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell*, 23(1), 71-81.
- Johnson, T.L., and Chamberlin, M.J. (1994) Complexes of yeast RNA polymerase II and RNA are substrates for TFIIS-induced RNA cleavage. *Cell*, 77(2), 217-224.
- Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet*, 37(7), 761-765.
- Kaplan, C.D., Larsson, K.M., and Kornberg, R.D. (2008) The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol Cell*, 30(5), 547-556.
- Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005) RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science*, 309(5733), 467-469.
- Kimura, M., and Ishihama, A. (2000) Involvement of multiple subunit-subunit contacts in the assembly of RNA polymerase II. *Nucleic Acids Res*, 28(4), 952-959.
- Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007) Functional architecture of RNA polymerase I. *Cell*, 131(7), 1260-1272.
- Lahmy, S., Guilleminot, J., Cheng, C.M., Bechtold, N., Albert, S., Pelletier, G., Delseny, M., and Devic, M. (2004) DOMINO1, a member of a small plant-specific gene family, encodes a protein essential for nuclear and nucleolar functions. *Plant J*, 39(6), 809-820.
- Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T., and Lagrange, T. (2009) PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. *Proc Natl Acad Sci U S A*, 106(3), 941-946.
- Landick, R. (2009) Functional divergence in the growing family of RNA polymerases. *Structure*, 17(3), 323-325.
- Larkin, R.M., and Guilfoyle, T.J. (1998) Two small subunits in Arabidopsis RNA polymerase II are related to yeast RPB4 and RPB7 and interact with one another. *J Biol Chem*, 273(10), 5631-5637.
- Larkin, R.M., Hagen, G., and Guilfoyle, T.J. (1999) Arabidopsis thaliana RNA polymerase II subunits related to yeast and human RPB5. *Gene*, 231(1-2), 41-47.
- Lee, S.R., and Collins, K. (2007) Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat Struct Mol Biol*, 14(7), 604-610.
- Lorkovic, Z.J., Herrmann, R.G., and Oelmuller, R. (1997) PRH75, a new nucleus-localized member of the DEAD-box protein family from higher plants. *Mol Cell Biol*, 17(4), 2257-2265.
- Lorkovic, Z.J., Hilscher, J., and Barta, A. (2004) Use of fluorescent protein tags to study nuclear organization of the spliceosomal machinery in transiently transformed living plant cells. *Mol Biol Cell*, 15(7), 3233-3243.
- Lotan, R., Bar-On, V.G., Harel-Sharvit, L., Duek, L., Melamed, D., and Choder, M. (2005) The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes Dev*, 19(24), 3004-3016.

- Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G., and Choder, M. (2007) The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *J Cell Biol*, 178(7), 1133-1143.
- Luo, J., and Hall, B.D. (2007) A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol*, 64(1), 101-112.
- Makeyev, E.V., and Bamford, D.H. (2002) Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol Cell*, 10(6), 1417-1427.
- Meka, H., Werner, F., Cordell, S.C., Onesti, S., and Brick, P. (2005) Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Res*, 33(19), 6435-6444.
- Mitsuzawa, H., Kanda, E., and Ishihama, A. (2003) Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res*, 31(16), 4696-4701.
- Moradpour, D., Bieck, E., Hugle, T., Wels, W., Wu, J.Z., Hong, Z., Blum, H.E., and Bartenschlager, R. (2002) Functional properties of a monoclonal antibody inhibiting the hepatitis C virus RNA-dependent RNA polymerase. *J Biol Chem*, 277(1), 593-601.
- Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jouette, D., Lacombe, A.M., Nikic, S., Picault, N., Remoue, K., Sanial, M., Vo, T.A., and Vaucheret, H. (2000) Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*, 101(5), 533-542.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, 120(5), 613-622.
- Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell*, 126(1), 79-92.
- Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev*, 19(17), 2030-2040.
- Qi, Y., Denli, A.M., and Hannon, G.J. (2005) Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol Cell*, 19(3), 421-428.
- Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*, 443(7114), 1008-1012.
- Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. (2009) Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell*, 33(2), 192-203.
- Selitrennik, M., Duek, L., Lotan, R., and Choder, M. (2006) Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of *Saccharomyces cerevisiae* RNA polymerase II by two pathways. *Eukaryot Cell*, 5(12), 2092-2103.

- Shewry, P.R., Jones, H.D., and Halford, N.G. (2008) Plant biotechnology: transgenic crops. *Adv Biochem Eng Biotechnol*, 111, 149-186.
- Teixeira, M.T., and Gilson, E. (2007) La sets the tone for telomerase assembly. *Nat Struct Mol Biol*, 14(4), 261-262.
- Till, S., and Ladurner, A.G. (2007) RNA Pol IV plays catch with Argonaute 4. *Cell*, 131(4), 643-645.
- Todone, F., Brick, P., Werner, F., Weinzierl, R.O., and Onesti, S. (2001) Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol Cell*, 8(5), 1137-1143.
- Todone, F., Weinzierl, R.O., Brick, P., and Onesti, S. (2000) Crystal structure of RPB5, a universal eukaryotic RNA polymerase subunit and transcription factor interaction target. *Proc Natl Acad Sci U S A*, 97(12), 6306-6310.
- Ujvari, A., and Luse, D.S. (2006) RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol*, 13(1), 49-54.
- Werner, F., and Weinzierl, R.O. (2002) A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell*, 10(3), 635-646.
- Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004a) Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*, 119(4), 481-489.
- Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004b) Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*, 303(5660), 1014-1016.
- Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*, 135(4), 635-648.
- Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet*, 41(5), 630-634.
- Wolin, S.L., and Cedervall, T. (2002) The La protein. *Annu Rev Biochem*, 71, 375-403.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol*, 2(5), E104.
- Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. (2006) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res*, 34(2), 667-675.
- Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science*, 307(5711), 932-935.

APPENDIX A

ROLES OF RNA POLYMERASE IV IN GENE SILENCING

A review published in *Trends in Plant Science* (2008), 13 (7): 390-397.

My contributions to this work:

For this review article I was responsible for writing the “Roles of Pol IV in the spread of silencing” section and generating Figure 1. I participated in discussions concerning the scope of this work and also aided in the editing process.

Special Issue: Noncoding and small RNAs

Roles of RNA polymerase IV in gene silencing

Craig S. Pikaard, Jeremy R. Haag, Thomas Ream and Andrzej T. Wierzbicki

Department of Biology, Washington University, 1 Brookings Drive, St. Louis, MO 63130, USA

Eukaryotes typically have three multi-subunit enzymes that decode the nuclear genome into RNA: DNA-dependent RNA polymerases I, II and III (Pol I, II and III). Remarkably, higher plants have five multi-subunit nuclear RNA polymerases: the ubiquitous Pol I, II and III, which are essential for viability; plus two non-essential polymerases, Pol IVa and Pol IVb, which specialize in small RNA-mediated gene silencing pathways. There are numerous examples of phenomena that require Pol IVa and/or Pol IVb, including RNA-directed DNA methylation of endogenous repetitive elements, silencing of transgenes, regulation of flowering-time genes, inducible regulation of adjacent gene pairs, and spreading of mobile silencing signals. Although biochemical details concerning Pol IV enzymatic activities are lacking, genetic evidence suggests several alternative models for how Pol IV might function.

RNA polymerases IVa and IVb: non-essential polymerases devoted to gene silencing

In all eukaryotes, DNA-dependent RNA polymerases (Pol I, II and III) transcribe essential genes, including rRNAs, mRNAs and tRNAs (see Glossary for abbreviations used in the article). Pol I, II and III are complicated enzymes with 12–17 subunits, which include structural and functional homologs of the five bacterial RNAP subunits [1]. The largest and second-largest Pol subunits, the homologs of bacterial β' and β , interact to form the DNA entry and RNA exit channels in addition to the catalytic center of RNA synthesis (Figure 1a) [2].

At present, the catalytic subunits homologous to those depicted in Figure 1a are the only known Pol IVa and Pol IVb subunits in *Arabidopsis*, a species discussed throughout this review. These subunits were initially identified by C.S. Pikaard, who examined the newly sequenced *Arabidopsis* genome and found two genes comprising an atypical fourth class of polymerase largest subunits, and two genes for an atypical class of second-largest subunits. His collaborator J. Eisen (Institute for Genomic Research, Rockville, MD) confirmed that these putative subunits are founding members of novel plant-specific clades [3] (see also [4–6]). As with the Pol I, II and III subunits, the atypical subunits are nuclear proteins [4,7,8], representing a new class of polymerase that has been designated nuclear RNA polymerase IV (Pol IV) [4,5].

Glossary

AGO: ARGONAUTE, proteins in this family bind to small RNAs, including siRNAs and miRNAs, and are capable of cleaving RNAs complementary to the small RNAs, a process known as slicing.

CLSY1: CLASSY1, a putative chromatin remodeling protein involved in RNA-directed DNA methylation.

CTD: C-terminal domain.

DCL1: *Arabidopsis* DICER-LIKE 1, involved primarily in miRNA biogenesis.

DCL2: *Arabidopsis* DICER-LIKE 2, generates 22-nt siRNAs.

DCL3: *Arabidopsis* DICER-LIKE 3, involved in 24-nt siRNA biogenesis.

DCL4: *Arabidopsis* DICER-LIKE 4, generates 21-nt siRNAs.

DeCL: Defective chloroplasts and leaves. Also known as DCL in the literature, which can cause confusion with Dicer-like proteins.

DRD1: DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1, a putative chromatin remodeling protein involved in RNA-directed DNA methylation.

DRM2: DOMAINS REARRANGED METHYLTRANSFERASE 2, the primary *Arabidopsis de novo* DNA methyltransferase.

dsRNA: double-stranded RNA.

GFP: Green fluorescent protein, initially derived from jellyfish.

HEN1: HUA ENHANCER 1; methylates the 2' hydroxyl groups of siRNA and miRNA 3'-terminal nucleotides.

HST1: HASTY1, an exportin 5 homolog implicated in nuclear export of miRNAs.

HYL1: HYPNASTIC LEAVES 1, a dsRNA-binding protein that interacts with DCL1.

I-siRNA: long siRNA of ~40 nt, as opposed to the predominant 21–24-nt size range.

miRNA: microRNA, small RNAs transcribed from dedicated genes, mediate mRNA cleavage or translational arrest.

nat-siRNA: siRNA derived from natural antisense transcripts derived from adjacent genes.

Pol I: DNA-DEPENDENT RNA POLYMERASE I, synthesizes the precursor for the three largest rRNAs.

Pol II: DNA-DEPENDENT RNA POLYMERASE II, transcribes most genes, including mRNAs and miRNAs.

Pol III: DNA-DEPENDENT RNA POLYMERASE III, mostly transcribes 5S rRNA genes and tRNA genes.

Pol IVa: nuclear RNA polymerase IVa, includes the NRPD1a and NRPD2a subunits.

Pol IVb: nuclear RNA polymerase IVb, includes the NRPD1b and NRPD2a subunits.

RdDM: RNA-directed DNA methylation, one of several gene silencing pathways in the nucleus.

RDR2: RNA-DEPENDENT RNA POLYMERASE 2, required for the biogenesis of 24-nt siRNAs in *Arabidopsis* in the RNA-directed DNA methylation pathway.

RDR6: RNA-DEPENDENT RNA POLYMERASE 6, involved in the ta-siRNA, nat-siRNA, I-siRNA, transgene and viral silencing, and long-distance silencing pathways.

RISC: RNA-induced silencing complex, includes an ARGONAUTE protein and siRNA (siRISC) or miRNA (miRISC).

RNA: Ribonucleic acid.

RNA-FISH: RNA fluorescent *in situ* hybridization, a means for locating specific RNAs.

RNAP: DNA-dependent RNA polymerase.

RNP: ribonucleoprotein, a complex of RNA and proteins.

rRNA: ribosomal RNA, four rRNAs are present in ribosomes.

SDE3: SILENCING DEFECTIVE 3, a putative RNA helicase.

SGS3: SUPPRESSOR OF GENE SILENCING 3, a putative coiled-coil protein.

siRNA: small interfering RNA.

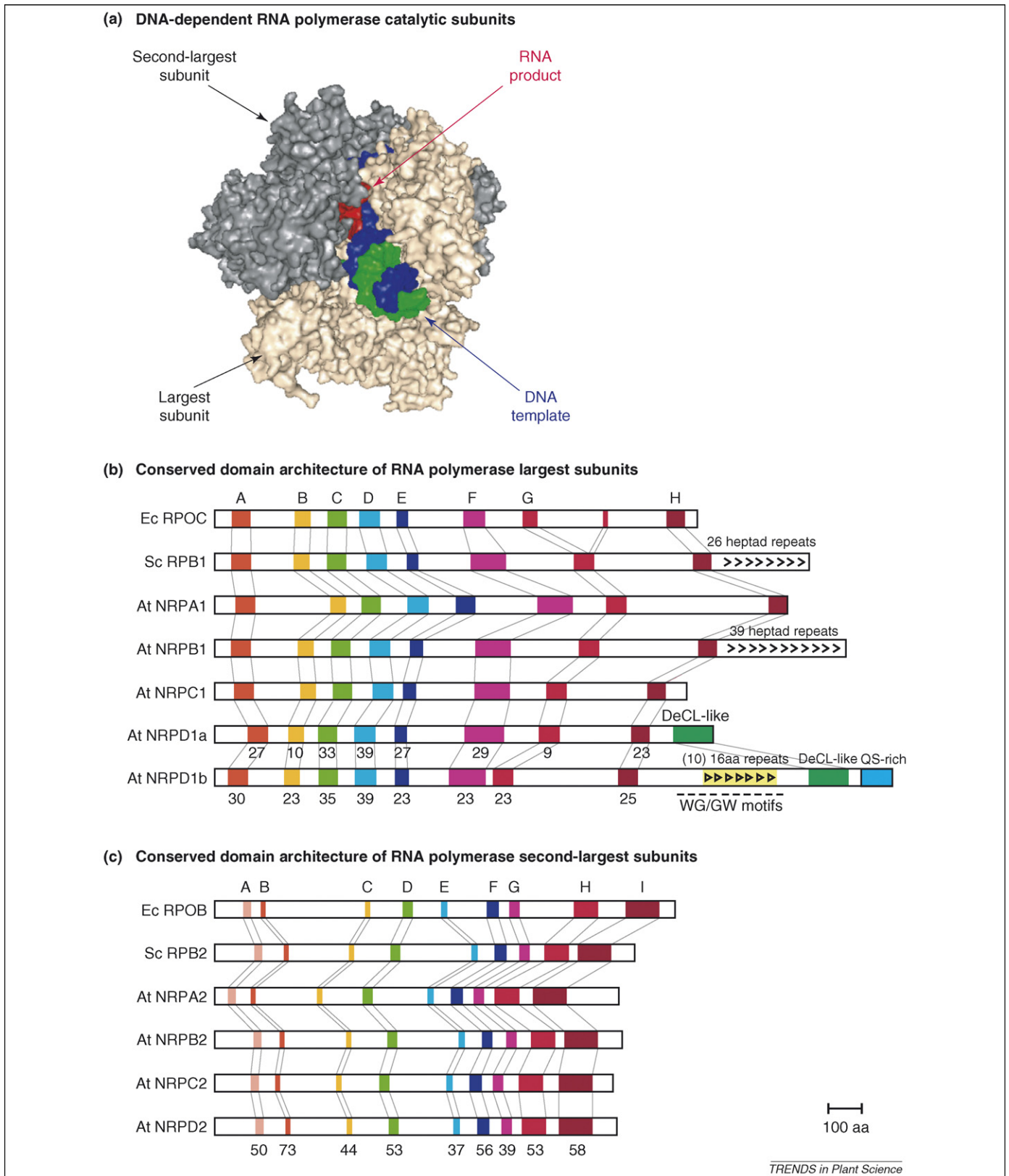


Figure 1. Catalytic subunits of DNA-dependent RNA polymerases. **(a)** The largest and second-largest subunits form the catalytic center. The image is a surface rendering generated using the crystal coordinates for a yeast Pol II elongation complex determined by K. Westover, D. Bushnell and R. Kornberg [PDB: (Protein Data Bank [http://www.rcsb.org/pdb/home/home.do] 1R9T)]. Only the two largest Pol II subunits are shown. The DNA template strand is shown in blue, the non-template strand in green, and the nascent RNA in red. **(b)** Domain structures of the largest subunits of RNAP. The largest subunits of *E. coli* (*Ec* RPOC) and yeast Pol II (*Sc* RPB1) are compared with the largest subunits of *Arabidopsis* Pol I (*At* NRPA1), Pol II (*At* NRPB1), Pol III (*At* NRPC1), Pol IVa (*At* NRPD1a) and Pol IVb (*At* NRPD1b). Positions of conserved domains A–H are highlighted. Numbers below Pol IV domains indicate the percentage identities to corresponding *Arabidopsis* Pol II subunit domains. CTDs of the largest subunits of yeast and *Arabidopsis* Pol II have 26 or 39 copies, respectively, of a seven amino acid (heptad) repeat. The domain with similarity to the DEFECTIVE CHLOROPLASTS AND LEAVES protein (DeCL domain), present in the CTDs of the largest subunits of Pol IVa and Pol IVb, is shown in green. The CTD of NRPD1b also includes a region rich in WG–GW motifs, overlapping ten, imperfect, 16-amino-acid repeats, and a domain composed of alternating glutamines and serines (QS-rich domain). **(c)** Domain structures of the second-largest subunits of RNAP. *E. coli* (*Ec* RPOB) and yeast Pol II subunits (*Sc* RPB2) are compared with the second-largest subunits of *Arabidopsis* Pol I (*At* NRPA2), Pol II (*At* NRPB2), Pol III (*At* NRPC2) and Pol IV (*At* NRPD2). Positions of conserved domains A–I are highlighted. Numbers below Pol IV domains are percentage identities to the corresponding *Arabidopsis* Pol II subunit domains.

Box 1. Pol IV subunit nomenclature

Nomenclature for Pol IV subunit genes derives from naming systems used in other eukaryotic model systems (e.g. budding yeast [*Saccharomyces cerevisiae*], in which RNA polymerase I, II and III are designated *RPA*, *RPB* and *RPC*, respectively). In *Arabidopsis*, an N, for 'nuclear', was added (e.g. *NRPA*, *NRPB* etc.) to polymerase subunit gene names to circumvent nomenclature conflicts with unrelated genes. The resulting gene names were registered with The *Arabidopsis* Information Resource by joint request of the David Baulcombe and Craig Pikaard laboratories. Largest subunits that are homologs of bacterial β' are designated, by convention, with the number 1, such that the unique *Arabidopsis* genes *NRPA1*, *NRPB1* and *NRPC1* encode the largest subunits of Pol I, II and III, respectively. Likewise, the genes encoding the second-largest subunits of *Arabidopsis* Pol I, II and III are designated *NRPA2*, *NRPB2* and *NRPC2*, respectively. On the basis of this naming scheme, the two related, but distinct, Pol IV largest subunits were designated *NRPD1a* and *NRPD1b*. Likewise, the two Pol IV second-largest subunit genes are designated *NRPD2a* and *NRPD2b*. Only *NRPD2a* is functional in the Col-O ecotype of *Arabidopsis* that has been studied to date [4,5,9,10]. Therefore, *NRPD2a* can be referred to simply as *NRPD2*. In other plant species, there are numerous functional genes for both the largest and second-largest subunits of Pol IV.

NRPD1a is the largest subunit of Pol IVa [4,5], whereas *NRPD1b* is the largest subunit of Pol IVb [9,10] (subunit nomenclature is discussed in Box 1). The largest subunits in both Pol IVa and Pol IVb have C-terminal domains (CTDs) that share similarity with the DEFECTIVE CHLOROPLASTS AND LEAVES protein (abbreviated DeCL in this article), which is required for 4.5S rRNA processing in chloroplasts (Figure 1b) [11]. The CTD of *NRPD1b* also includes ten imperfect 16-amino-acid repeats within a tryptophan and glycine (WG–GW)-rich region. A glutamine and serine (Q–S)-rich domain is present at the distal end of the CTD (Figure 1b). The WG–GW motifs are proposed to mediate Argonaute protein interactions [8,12], but the significance of the DeCL and Q–S domains is unknown. However, the DeCL and Q–S domains might facilitate additional molecular interactions in a manner analogous to the function of the CTD of the largest subunit of Pol II. This CTD mediates numerous interactions that govern processes such as transcriptional activation by enhancers, transcription elongation, and several mRNA processing steps [13–15]. Both Pol IVa and Pol IVb have an *NRPD2* subunit that is encoded by the same gene, *NRPD2a* [4,5,9,10]. *NRPD1a* and *NRPD1b* each co-immunoprecipitate and co-localize with *NRPD2* [7], but the alternative largest subunits do not immunoprecipitate with one another, indicating that Pol IVa and Pol IVb are distinct physical entities.

The full subunit compositions of Pol IVa and Pol IVb are not known, nor are their templates or enzymatic products. However, a flurry of studies in the past three years has shown that Pol IVa and, to a lesser extent, Pol IVb are crucial for several RNA-mediated gene silencing phenomena. These pathways, and the roles of Pol IV in them, are the focus of our review.

Roles of Pol IVa and Pol IVb in the RNA-directed DNA methylation pathway

Arabidopsis has four Dicer endonucleases (DCLs), six single-subunit RNA-dependent RNA polymerases (RDRs)

and ten Argonaute proteins (AGOs) that participate in microRNA (miRNA)- and small interfering (siRNA)-mediated transcriptional or post-transcriptional silencing [16–19]. In the RNA-directed DNA methylation (RdDM) pathway of transcriptional gene silencing [20–23], double-stranded RNAs generated with the involvement of RDR2 are cleaved by DCL3, and the resulting siRNAs are loaded into AGO4–RISC and/or AGO6–RISC complexes that mediate the *de novo* methylation of cytosines within DNA sequences complementary to the siRNAs [22,24–28]. The realization that Pol IVa and Pol IVb are players in the RdDM pathway came from a combination of genetic screens [5,10] and reverse-genetic analyses [4,9]. Silencing-defective (*sde*) mutants were identified in screens for the de-repression of a silenced transgene locus, and analysis of these mutants led to the identification of *sde4* as an allele of *NRPD1a* [5]. A subsequent test to determine if one of the atypical second-largest subunit (*NRPD2*) genes might partner with *NRPD1a* revealed that insertional mutants of *NRPD2a* also disrupted the silencing pathway. Coinciding with this disruption was the disappearance of 24-nt siRNAs and the loss of cytosine methylation at corresponding loci [5]. Our laboratory initially focused on *NRPD2*, showing that its activity was not redundant with that of the equivalent Pol I, II or III subunits and that it did not co-purify with Pol I, II or III [4]. However, *NRPD2* was found to localize within the nucleus and to affect the coalescence of heterochromatic sequences into chromocenters [4]. Heterochromatic DNA is typically heavily methylated, and loss of cytosine methylation occurred at a subset of heterochromatic loci in *nrdp2* mutants as well as in *nrdp1a* mutants [4]. Collectively, the initial studies of *NRPD1a* and *NRPD2* pointed to the existence of Pol IVa.

Kanno *et al.* [29] carried out a genetic screen for mutations causing the de-repression of a reporter gene silenced by RdDM. This led to the identification of *DRD1*, a member of the SWI2–SNF2 chromatin remodeling protein family, in addition to *DRD2* and *DRD3*, which turned out to be *NRPD2a* and *NRPD1b*, respectively [10]. The realization that the *NRPD1b* gene had been mistakenly annotated as two genes [4,5,10] also led to a reverse-genetic examination of cytosine methylation and siRNA phenotypes in *nrdp1b* insertional mutants [9]. Collectively, these independent studies revealed the existence of Pol IVb and showed that siRNAs eliminated in Pol IVa mutants [4,5] are not abolished in Pol IVb mutants [9,10], despite similar losses of cytosine methylation [9,10]. These observations, based on a small number of loci, indicated that Pol IVa and Pol IVb act at different steps in the RdDM pathway, with Pol IVa acting upstream of siRNA production, and Pol IVb functioning at a later step in the pathway, mostly downstream of siRNA production [10]. Recent genome-wide analyses of small RNA populations have shown that there are at least 4600 *Arabidopsis* loci that give rise to small RNAs, with 94% of them being dependent on Pol IVa [30]. Pol IVb plays little, if any, role in siRNA abundance at approximately one-third of these loci; it has intermediate effects at another one-third of the loci; and it is absolutely required for siRNA production at one-third of the Pol IVa-dependent loci [30]. However, there are no definitive examples of siRNAs that

are dependent on Pol IVb only, and which do not require Pol IVa. These results are consistent with the hypothesis that Pol IVa acts upstream of siRNA production. The role of Pol IVb in siRNA production is less clear, and it could be indirect. A positive feedback relationship exists between the formation of heterochromatin and the continued production of siRNA. As such, the role of Pol IVb in facilitating RdDM might explain the influence of Pol IVb on siRNA abundance, as has been depicted in circular models for the RdDM pathway [7,8].

The localization of proteins involved in RdDM has provided insight into the RdDM pathway [7,8,31,32]. Pol IVa, Pol IVb and DRD1 co-localize with chromosomal loci that are both sources and targets of abundant siRNAs, suggesting that these proteins are involved in the generation of siRNA precursors or the targeting of siRNA-directed chromatin modifications [7]. AGO4 and DRM2, the primary *de novo* DNA methyltransferase, also co-localize at source/target loci in some nuclei [32]. RNA-FISH combined with protein immunolocalization has shown that siRNAs co-localize with RDR2, DCL3, AGO4 and NRPD1b within a nucleolar compartment interpreted to be an siRNA processing center [7]. This processing center includes several molecular markers of Cajal bodies [8], which are dynamic compartments important for assembling ribonucleoprotein complexes involved in pre-mRNA splicing, pre-rRNA processing, RNA methylation and pseudo-uridylation, telomerase assembly and histone mRNA 3' end formation [33,34]. Formation of siRNA-RISC complexes is consistent with the overall theme of assembling ribonucleoprotein complexes within Cajal bodies [8,33–35]. Recent evidence suggests that miRNA processing in plants also occurs within nucleolus-associated Cajal body-like entities that include the spliceosomal proteins SmB and SmD3 – both found in Cajal bodies and spliceosomes – but which lack the canonical Cajal body protein coilin [36]. Other groups have suggested that these miRNA processing centers are not Cajal bodies, because they lack coilin [37,38]. However, *Drosophila* lacks coilin yet has functional Cajal bodies [39]. These observations can be reconciled by the hypothesis that there are numerous sub-classes of Cajal bodies, some of which have coilin and some of which do not [34,35,39].

Because Pol IVa co-localizes with loci that give rise to abundant 24-nt siRNAs and because loss of *NRPD1a* function causes all other known components of the RdDM pathway to mislocalize, Pol IV is thought to act at an initial step of the pathway, upstream of RDR2 [7]. CLSY1, which like DRD1 is an SWI–SNF family protein, co-localizes with RDR2 at the inner perimeter of the nucleolus; and, in *clsy1* mutants, RDR2 localization is severely disrupted [40]. Pol IVa localization is also affected, albeit to a lesser degree [40], suggesting that CLSY1 functions at the interface between Pol IVa and RDR2, presumably facilitating the generation of dsRNAs that are diced by DCL3 and loaded into AGO4 effector complexes [16,17,26,41] within the nucleolar siRNA processing center [7,8]. NRPD1b co-localizes with AGO4 both within the processing center [7,8] and at target loci [32], interacting with AGO4 through the CTD [8,12]. Current models suggest that siRNA-AGO4–Pol IVb effector complexes then locate their targets by

virtue of siRNA-target base-pairing interactions [7,8]. Pol IVb, DRD1 and DRM2 are then thought to collaborate in the siRISC-directed DNA methylation process through an as yet unknown mechanism [21]. DNA methylation then appears to feed back on the production of siRNAs, such that siRNAs are depleted in *drm* mutants at some loci [4,7,41] and in *ddm1* (*decrease in DNA methylation 1*) or *met1* (*cytosine methyltransferase 1*) mutants that are required for maintaining DNA methylation patterns at other loci [42]. Therefore, it is possible that Pol IVa preferentially transcribes methylated DNA [4] or aberrant RNAs generated from methylated loci [7,43,44] as a means of perpetuating the repression cycle.

A role for Pol IV in flowering

Although they are non-essential in terms of viability, Pol IVa and Pol IVb nonetheless play roles in development, affecting flowering time in the context of the RdDM pathway. Under short-day conditions, flowering in *nprp1a* and *nprp1b* mutants is significantly delayed, as is also the case in *rdr2*, *dcl3*, *ago4* and *drm* mutants [9,45]. The flowering-time regulators *FCA* and *FPA* were identified in screens for mutants that disrupt RNA-directed gene silencing, and they appear to be players in the RdDM pathway, wherein they act at some, but not all, loci [46]. At least two flowering genes, *FWA* and *FLC*, appear to be targets of silencing through Pol IV-dependent siRNA pathways [45,47,48].

The role of Pol IV in abiotic and biotic stress-inducible siRNA production

Pol IV plays an important role in the production of natural antisense transcript siRNAs (nat-siRNAs) [49–53]. These siRNAs are generated from dsRNAs derived from the overlapping 3' ends of convergently transcribed gene pairs. Expression of one member of the gene pair is constitutive, but expression of the other is inducible, as in the case of the *P5CDH* and *SRO5* gene pair, respectively. Salt stress induces *SRO5* expression such that its transcript can anneal with the *P5CDH* mRNA to form a region of dsRNA. In a process involving Pol IVa, RDR6, SGS3 and DCL2, a 24-nt nat-siRNA is produced, and this is thought to guide the cleavage of *P5CDH* transcripts, setting the stage for generation of additional DCL1-dependent 21-nt siRNAs [49]. The resulting downregulation of *P5CDH* results in increased proline synthesis, a physiological response that helps to confer salt tolerance.

Pathogen-inducible siRNAs provide two examples of additional means for generating nat-siRNAs [54,55]. In the first, infection of *Arabidopsis* with *Pseudomonas syringae* generates a 22-nt nat-siRNA in a pathway that requires Pol IVa, RDR6 and SGS3. This pathway is similar to that which generates the salt stress-induced nat-siRNA, except that DCL2 is not involved; instead, DCL1, HYL1 and HEN1 – which are typically involved in miRNA biogenesis – are required for siRNA production in the pathogen response. The end result is the downregulation of *PPRL*, a negative regulator of pathogen resistance. More recently, investigators demonstrated that *Pseudomonas syringae* infection induces expression of a 39–41-nt RNA [54]. This so-called long siRNA (l-siRNA) matches the overlapping region of the *SRRLK* and *AtRAP* gene pair,

and it specifically downregulates *AtRAP*, another negative regulator of the pathogen defense response, in a pathway requiring Pol IVa and Pol IVb, DCL1, HYL1, HEN1, HST1 (HASTY1), RDR6, DCL4, AGO7 and SDE3. Most of these proteins (i.e. DCL1, HYL1, HEN1, HST1, RDR6, DCL4 and AGO7) are also players in the so-called *trans*-acting siRNA (ta-siRNA) pathway, in which miRNA-mediated cleavage of a specific target mRNA initiates the subsequent production of siRNAs from the cleaved mRNA [56–59]. Resulting siRNAs then target additional mRNAs for cleavage, thereby amplifying the signal in a regulatory cascade. It is not yet clear whether a similar regulatory cascade occurs upon bacterial infection and, if so, where Pol IVa and Pol IVb fit within such a pathway.

Roles of Pol IV in the spreading of silencing

Pol IVa is required for both short-range spreading of RNA silencing cell-to-cell through plasmadesmata and long-range silencing through the phloem [60,61]. Two independent screens revealed a requirement for Pol IVa and RDR2 in the short-range spreading of silencing [40,62], and DCL4 [40,63], DCL1, HEN1 and AGO1 [62] are also required. By contrast, HYL1, DCL3, AGO4, RDR6 [40,62,63], Pol IVb (NRPD1b) and DRD1 [40] are all dispensable. Although both 24-nt and 21-nt transgene-specific siRNAs are produced, the DCL4-dependent 21-nt siRNAs are believed to be the primary short-range mobile signals [40,62,63]. However, longer siRNAs can suffice when overproduced in mutants of *DRB4*, which encodes a dsRNA binding protein that partners with DCL4 in the production of 21-nt siRNAs [62].

In Pol IVa mutants, silencing is impaired even in the phloem cells where the silencing signal is initiated, suggesting that Pol IVa acts at an initiating step in the process that ultimately gives rise to the mobile silencing signal(s). Interestingly, the spreading of silencing can be dramatically enhanced in *dcl3* and *ago4* mutants [40], coincident with increased 21-nt siRNA production and loss of 24-nt siRNAs. A possibility is that Pol IVa/RDR2-dependent dsRNA substrates can be channeled into either 24-nt or 21-nt siRNA production, with the 21-nt siRNAs acting as the primary short-range mobile signals.

An ability to distinguish between production and perception of silencing signals has come from a study in which wild type or mutant rootstocks or scions (shoots) were grafted onto one another and monitored for long-distance silencing of a green fluorescent protein (GFP) transgene [64]. Pol IVa (NRPD1a), RDR2, DCL3, AGO4 and RDR6 are all required for the scion to respond to a silencing signal derived from a dsRNA hairpin expressed in the rootstock [65]. However, none of these proteins are required to generate the mobile signal. Interestingly, RDR6 is required for the perception of the long-distance signal [65] but is dispensable for short-range silencing [40,62]. Pol IVb (NRPD1b) is dispensable for both short and long-distance silencing, consistent with the hypothesis that Pol IVb functions in chromatin modification rather than in RNA production.

The nature of the long-distance silencing signal is unknown, but *dcl1–8* hypomorphs and *dcl2;dcl3;dcl4* triple mutants defective for miRNA or siRNA production, respectively, continue to produce the mobile signal in roots,

as do mutants for Pol IVa, Pol IVb, *RDR2* and *RDR6* [65]. Therefore, it seems unlikely that Dicer-generated small RNAs are the long-distance signaling molecules. Instead, larger RNAs might serve as the mobile signal(s). An intriguing observation is that siRNAs produced in the scion upon reception of the silencing signal do not correspond to the approximately two-thirds of the GFP gene that was used as the hairpin trigger sequence; instead, the siRNAs neatly correspond to the third of the GFP transgene located downstream (3') of the trigger sequences [65]. It is not clear why this should be the case if siRNAs are the mobile signal. Antisense siRNAs could anneal anywhere throughout the first two-thirds of the target mRNA and might be expected to prime RDR activity in the upstream direction. Likewise, siRNA-directed cleavage of target mRNAs, which would render the 3' target fragment uncapped, 'aberrant' and a potential substrate for RDR6 [66], would generate a diverse set of cleaved fragments throughout the first two-thirds of the GFP target. Therefore, a possibility is that the dsRNA trigger molecule itself, or its component strands, is the mobile signal(s), which is plausible given the evidence that intact mRNAs can traffic through phloem [67]. If the antisense strand of the dsRNA trigger were to anneal to the intact mRNA in the shoot such that only the 3' portion of the GFP mRNA were to remain single-stranded, the resulting structure might somehow direct RDR6- and Pol IVa-dependent amplification of the single-stranded sequences 3' of the trigger sequence.

Unsolved mysteries and future directions

Pol IVa is integral to numerous RNA silencing pathways, including the RdDM pathway, the nat-siRNA and l-siRNA pathways, the short-range spreading of silencing pathway, and the pathway for the perception of long-distance silencing signals (Figure 2). Pol IVb is apparently less gregarious, acting primarily in the RdDM pathway [30], but also playing an undefined role in the l-siRNA pathway [54]. It seems probable that both Pol IVa and Pol IVb possess enzymatic activity, given that the NRPD1a, NRPD1b and

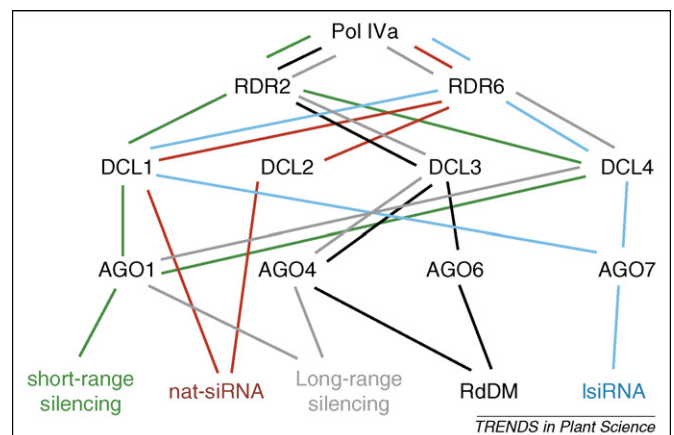


Figure 2. A variety of proteins participate in Pol IVa-dependent silencing pathways. The figure shows a subset of the proteins that are involved in RdDM, nat-siRNA, l-siRNA, short-range silencing, and long-distance silencing pathways. Proteins involved in the various pathways are linked by color-coded lines. The diagram does not imply the order of events, but illustrates the diversity of functional collaborations that are possible. Not all mutants have been tested in every pathway; therefore, other potential connections might exist. However, the figure reflects the models provided by the authors of the studies discussed in the text.

NRPD2 subunits possess the key conserved amino acids of the metal A and metal B sites found within the catalytic centers of other multi-subunit RNA polymerases [68,69]. But what do Pol IVa and Pol IVb transcribe, and what are their products? At present, we have no answer. In fact, our only biochemical clue is a negative result: a conventional, promoter-independent transcription assay [70] using sheared double-stranded template DNA revealed that chromatographic fractions enriched for Pol IV lack DNA-dependent RNA polymerase activity, unlike fractions enriched for Pol I, II and III [4]. Based on this result, it seems likely that Pol IVa and Pol IVb use very specific templates.

A distinct possibility is that Pol IVa transcribes RNA [7,43,44]. Pol IVa is mislocalized by RNase treatment of nuclei, but not by DNase treatment, whereas Pol II shows the opposite nuclease sensitivities [7]. Moreover, there is precedent for DNA-dependent RNA polymerases transcribing RNA. Hepatitis Delta Virus (HDV) and plant viroid RNAs are replicated by Pol II transcription [71,72]. Likewise, *Escherichia coli* RNAP is regulated by binding to 6S RNA, which is transcribed in order to be released [73].

Previous models for the RdDM pathway have suggested that Pol IVa transcribes methylated DNA or transcripts of methylated loci, with resulting Pol IVa transcripts being amplified or made double-stranded by RDR2 (Figures 3ab). However, in the nat-siRNA and l-siRNA pathways, regions of dsRNA are apparently generated by Pol II transcription of overlapping gene pairs, and these transcripts persist in *nrd1a* mutants, suggesting that there is no need for Pol IVa in the initial formation of dsRNA. Likewise, Pol IVa plays roles in short-range spreading of silencing triggered by dsRNA hairpin trigger sequences, and in long-distance silencing likely to involve annealing of a mobile RNA to target mRNAs, thereby forming dsRNA. In each of these cases, there is no obvious need for Pol IVa in the initial generation of dsRNAs.

Pol IVa might use initial dsRNAs as templates, generating transcripts that are then made double-stranded by RDR2 or RDR6, one or both of which are involved in all known Pol IVa-dependent pathways (Figure 3c). Subsequent dicing, siRNA-mediated target slicing *in trans*, and RDR transcription of sliced templates might then amplify the initial signal and generate small RNAs beyond the region of initial transcript overlap. Alternatively,

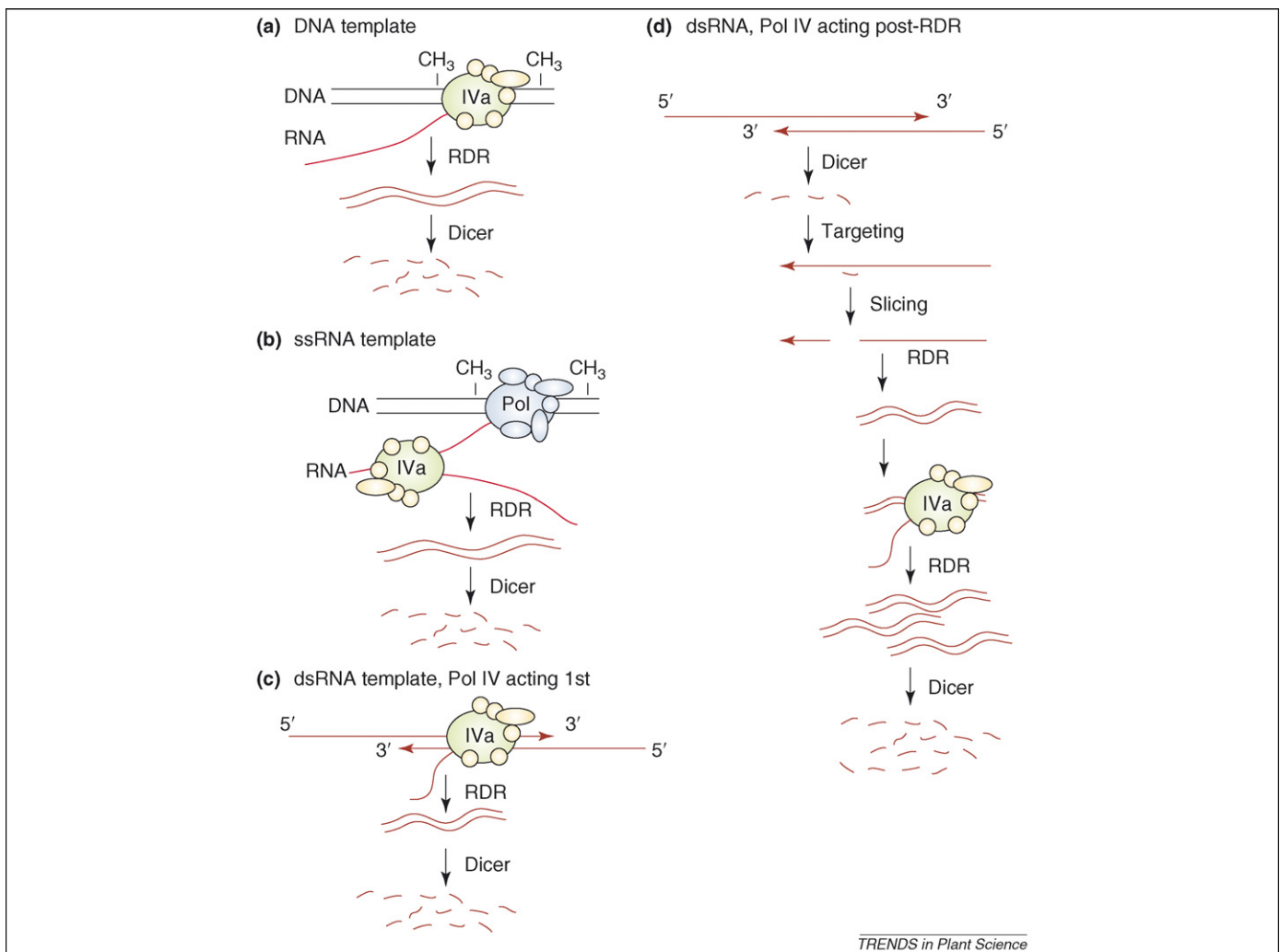


Figure 3. Possible modes of Pol IVa function. Pol IVa might transcribe a specialized DNA template, such as methylated DNA (a) or single-stranded RNA transcripts derived from methylated DNA loci (b). Alternatively, Pol IVa might transcribe dsRNA generated from bidirectional transcripts, including transcripts of natural antisense gene pairs, or dsRNAs resulting from the annealing of long-distance mobile RNAs with target mRNAs (c) and (d). The model shown in (d) might account for the involvement of numerous Dicer proteins and numerous RDR inputs in the nat-siRNA and long-distance silencing pathways.

licing of initial dsRNA regions might lead to the production of siRNAs that prime RDR on sliced or unsliced target RNAs, resulting in secondary dsRNAs that are then transcribed by Pol IVa and amplified by further RDR activity (Figure 3d). The model in Figure 3d would account for the involvement of more than one Dicer and more than one RDR-requiring step in the nat-siRNA and long-distance silencing pathways.

Pol IVa appears to be dispensable in some dsRNA-initiated phenomena. For instance, one group [10] screened for methylation-defective mutants by using a dsRNA hairpin to trigger RNA-directed DNA methylation. They recovered nine alleles of *NRPD1b*, and twelve alleles of *NRPD2a*, but no alleles of *NRPD1a* or *RDR2* were identified [10], suggesting that the production of dsRNA hairpins had bypassed a need for Pol IVa or RDR2. Similarly, deep sequencing of small RNA libraries has shown that more than 90% of all siRNAs are Pol IVa-dependent and are mostly derived from transposable elements and tandem repeats [30,74]. Inverted repeats, however, can contribute to the siRNA pool by a Pol IVa-independent mechanism [74]. Because transcription of inverted repeats can produce hairpin dsRNAs on their own, their Pol IVa-independence fits with the idea that Pol IVa functions at other loci in the production of dsRNAs that then feed into siRNA production. Why some dsRNA hairpin-initiated silencing phenomena require Pol IVa, but others do not, is not clear. The strength of the promoters driving hairpin formation might be an important variable.

Pol IVb is even more of a mystery than Pol IVa. *NRPD1b* mostly appears to reinforce Pol IVa-dependent siRNA production [9,30] yet is required, in addition to Pol IVa, for RdDM [9,10,75]. One possibility is that Pol IVb binds to DNA and interacts with AGO4 through its CTD [8,12], facilitating siRNA–DNA base-pairing, which in turn enables the recruitment of DRM2. Alternatively, siRNA–AGO4 complexes might anneal to Pol IVb transcripts, thereby recruiting DRM2 and/or histone modifying enzymes to the vicinity of the corresponding DNA, as in models for siRNA-mediated silencing in fission yeast (*Schizosaccharomyces pombe*) [76,77]. AGO4 can slice RNAs in an siRNA-guided process, providing evidence that AGO4–siRNA–RISC complexes can interact with RNA transcripts [78]. Nonetheless, direct siRNA interactions with DNA cannot be ruled out.

Clearly, there is much that needs to be learned concerning the templates, products, subunit structures, and interacting partners of Pol IVa and Pol IVb. Development of *in vitro* assays will be invaluable for deciphering the functions of these enigmatic polymerases and is a major challenge for the future.

Acknowledgements

Our work is supported by National Institutes of Health grants R01GM60380 and R01GM077590, and by the Monsanto Company–Washington University Plant Biology Research Agreement.

References

- 1 Werner, F. (2007) Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* 65, 1395–1404
- 2 Kettenberger, H. *et al.* (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* 16, 955–965
- 3 The-Arabidopsis-Genome-Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
- 4 Onodera, Y. *et al.* (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120, 613–622
- 5 Herr, A.J. *et al.* (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* 308, 118–120
- 6 Luo, J. and Hall, B.D. (2007) A multistep process gave rise to RNA polymerase IV of land plants. *J. Mol. Evol.* 64, 101–112
- 7 Pontes, O. *et al.* (2006) The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* 126, 79–92
- 8 Li, C.F. *et al.* (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* 126, 93–106
- 9 Pontier, D. *et al.* (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* 19, 2030–2040
- 10 Kanno, T. *et al.* (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* 37, 761–765
- 11 Bellaoui, M. *et al.* (2003) DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. *Plant Mol. Biol.* 53, 531–543
- 12 El-Shami, M. *et al.* (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* 21, 2539–2544
- 13 Phatnani, H.P. and Greenleaf, A.L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* 20, 2922–2936
- 14 Hahn, S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* 11, 394–403
- 15 Shilatifard, A. (2004) Transcriptional elongation control by RNA polymerase II: a new frontier. *Biochim. Biophys. Acta* 1677, 79–86
- 16 Brodersen, P. and Voinnet, O. (2006) The diversity of RNA silencing pathways in plants. *Trends Genet.* 22, 268–280
- 17 Baulcombe, D. (2004) RNA silencing in plants. *Nature* 431, 356–363
- 18 Herr, A.J. and Baulcombe, D.C. (2004) RNA silencing pathways in plants. *Cold Spring Harb. Symp. Quant. Biol.* 69, 363–370
- 19 Chapman, E.J. and Carrington, J.C. (2007) Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* 8, 884–896
- 20 Wassenegger, M. (2000) RNA-directed DNA methylation. *Plant Mol. Biol.* 43, 203–220
- 21 Huettel, B. *et al.* (2007) RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. *Biochim. Biophys. Acta* 1769, 358–374
- 22 Wassenegger, M. (2005) The role of the RNAi machinery in heterochromatin formation. *Cell* 122, 13–16
- 23 Bender, J. (2004) DNA methylation and epigenetics. *Annu. Rev. Plant Biol.* 55, 41–68
- 24 Cao, X. *et al.* (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* 13, 2212–2217
- 25 Zilberman, D. *et al.* (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299, 716–719
- 26 Xie, Z. *et al.* (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* 2, E104
- 27 Matzke, M. *et al.* (2006) RNA-directed DNA methylation and Pol IVb in *Arabidopsis*. *Cold Spring Harb. Symp. Quant. Biol.* 71, 449–459
- 28 Zheng, X. *et al.* (2007) Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J.* 26, 1691–1701
- 29 Kanno, T. *et al.* (2004) Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr. Biol.* 14, 801–805
- 30 Mosher, R.A. *et al.* (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3145–3150
- 31 Pikaard, C.S. (2006) Cell biology of the *Arabidopsis* nuclear siRNA pathway for RNA-directed chromatin modification. *Cold Spring Harb. Symp. Quant. Biol.* 71, 473–480
- 32 Li, C.F. *et al.* (2008) Dynamic regulation of ARGONAUTE4 within multiple nuclear bodies in *Arabidopsis thaliana*. *PLoS Genet.* 4, e27

- 33 Cioce, M. and Lamond, A.I. (2005) Cajal bodies: a long history of discovery. *Annu. Rev. Cell Dev. Biol.* 21, 105–131
- 34 Handwerger, K.E. and Gall, J.G. (2006) Subnuclear organelles: new insights into form and function. *Trends Cell Biol.* 16, 19–26
- 35 Pontes, O. and Pikaard, C.S. siRNA and miRNA processing: new roles for Cajal bodies. *Curr. Opin. Gen. Dev.* (in press)
- 36 Fujioka, Y. *et al.* (2007) Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in *Arabidopsis*. *Plant Cell Physiol.* 48, 1243–1253
- 37 Fang, Y. and Spector, D.L. (2007) Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. *Curr. Biol.* 17, 818–823
- 38 Song, L. *et al.* (2007) *Arabidopsis* primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5437–5442
- 39 Liu, J.L. *et al.* (2006) The *Drosophila melanogaster* Cajal body. *J. Cell Biol.* 172, 875–884
- 40 Smith, L.M. *et al.* (2007) An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in *Arabidopsis*. *Plant Cell* 19, 1507–1521
- 41 Zilberman, D. *et al.* (2004) Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* 14, 1214–1220
- 42 Lippman, Z. *et al.* (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* 1, E67
- 43 Vaughn, M.W. and Martienssen, R.A. (2005) Finding the right template: RNA Pol IV, a plant-specific RNA polymerase. *Mol. Cell* 17, 754–756
- 44 Vaucheret, H. (2005) RNA polymerase IV and transcriptional silencing. *Nat. Genet.* 37, 659–660
- 45 Chan, S.W. *et al.* (2004) RNA silencing genes control *de novo* DNA methylation. *Science* 303, 1336
- 46 Baurle, I. *et al.* (2007) Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. *Science* 318, 109–112
- 47 Swiezewski, S. *et al.* (2007) Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, FLC. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3633–3638
- 48 Liu, J. *et al.* (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev.* 18, 2873–2878
- 49 Borsani, O. *et al.* (2005) Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 123, 1279–1291
- 50 Henz, S.R. *et al.* (2007) Distinct expression patterns of natural antisense transcripts in *Arabidopsis*. *Plant Physiol.* 144, 1247–1255
- 51 Jen, C.H. *et al.* (2005) Natural antisense transcripts with coding capacity in *Arabidopsis* may have a regulatory role that is not linked to double-stranded RNA degradation. *Genome Biol.* 6, R51
- 52 Wang, X.J. *et al.* (2005) Genome-wide prediction and identification of *cis*-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol.* 6, R30
- 53 Jin, H. *et al.* (2008) Small RNAs and the regulation of *cis*-natural antisense transcripts in *Arabidopsis*. *BMC Mol. Biol.* 9, 6
- 54 Katiyar-Agarwal, S. *et al.* (2007) A novel class of bacteria-induced small RNAs in *Arabidopsis*. *Genes Dev.* 21, 3123–3134
- 55 Katiyar-Agarwal, S. *et al.* (2006) A pathogen-inducible endogenous siRNA in plant immunity. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18002–18007
- 56 Vazquez, F. *et al.* (2004) Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* 16, 69–79
- 57 Allen, E. *et al.* (2005) microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* 121, 207–221
- 58 Peragine, A. *et al.* (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* 18, 2368–2379
- 59 Yoshikawa, M. *et al.* (2005) A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* 19, 2164–2175
- 60 Voinnet, O. (2005) Non-cell autonomous RNA silencing. *FEBS Lett.* 579, 5858–5871
- 61 Himber, C. *et al.* (2003) Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *EMBO J.* 22, 4523–4533
- 62 Dunoyer, P. *et al.* (2007) Intra- and intercellular RNA interference in *Arabidopsis thaliana* requires components of the microRNA and heterochromatic silencing pathways. *Nat. Genet.* 39, 848–856
- 63 Dunoyer, P. *et al.* (2005) DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nat. Genet.* 37, 1356–1360
- 64 Palauqui, J.C. *et al.* (1997) Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *EMBO J.* 16, 4738–4745
- 65 Brosnan, C.A. *et al.* (2007) Nuclear gene silencing directs reception of long-distance mRNA silencing in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 14741–14746
- 66 Axtell, M.J. *et al.* (2006) A two-hit trigger for siRNA biogenesis in plants. *Cell* 127, 565–577
- 67 Kim, M. *et al.* (2001) Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* 293, 287–289
- 68 Lehmann, E. *et al.* (2007) Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* 450, 445–449
- 69 Cramer, P. (2006) Recent structural studies of RNA polymerases II and III. *Biochem. Soc. Trans.* 34, 1058–1061
- 70 Schwartz, L.B. *et al.* (1974) Isolation and partial characterization of the multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in the mouse myeloma, MOPC 315. *J. Biol. Chem.* 249, 5889–5897
- 71 Greco-Stewart, V.S. *et al.* (2007) The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. *Virology* 357, 68–78
- 72 Ding, B. and Itaya, A. (2007) Viroid: a useful model for studying the basic principles of infection and RNA biology. *Mol. Plant Microbe Interact.* 20, 7–20
- 73 Wassarman, K.M. and Saecker, R.M. (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* 314, 1601–1603
- 74 Zhang, X. *et al.* (2007) Role of RNA polymerase IV in plant small RNA metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4536–4541
- 75 Huettel, B. *et al.* (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J.* 25, 2828–2836
- 76 Buhler, M. *et al.* (2007) RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* 129, 707–721
- 77 Buhler, M. *et al.* (2006) Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* 125, 873–886
- 78 Qi, Y. *et al.* (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* 443, 1008–1012

APPENDIX B

THE ARABIDOPSIS CHROMATIN-MODIFYING NUCLEAR siRNA PATHWAY
INVOLVES A NUCLEOLAR RNA PROCESSING CENTER

Published in *Cell* (2006), 126 (1): 79-92.

My contributions to this work:

I cloned the genomic sequences for NRPD1a/NRPD1 and NRPD1b/NRPE1 and generated transgenic plants that complemented the *nrpd1a-3/nrpd1-3* and *nrpd1b-11/nrpe1-11* mutations, respectively. With these lines I was able to demonstrate rescue of DNA methylation at the 5S rDNA (Figure 1E) and that both NRPD1a/NRPD1 and NRPD1b/NRPE1 interact with NRPD2/NRPE2 by co-IP and Western blot experiments (Figure 1F and G).

The *Arabidopsis* Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center

Olga Pontes,¹ Carey Fei Li,² Pedro Costa Nunes,^{1,4} Jeremy Haag,¹ Thomas Ream,¹ Alexa Vitins,¹ Steven E. Jacobsen,^{2,3} and Craig S. Pikaard^{1,*}

¹Biology Department, Washington University, 1 Brookings Drive, St. Louis, MO 63130, USA

²Department of Molecular, Cell and Developmental Biology

³Howard Hughes Medical Institute

University of California, Los Angeles, Los Angeles, CA 90095, USA

⁴Secção de Genética, Centro de Botânica Aplicada à Agricultura, Instituto Superior de Agronomia, Tapada da Ajuda, 1349-017 Lisboa, Portugal

*Contact: pikaard@biology.wustl.edu

DOI 10.1016/j.cell.2006.05.031

SUMMARY

In *Arabidopsis thaliana*, small interfering RNAs (siRNAs) direct cytosine methylation at endogenous DNA repeats in a pathway involving two forms of nuclear RNA polymerase IV (Pol IVa and Pol IVb), RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), DICER-LIKE 3 (DCL3), ARGONAUTE4 (AGO4), the chromatin remodeler DRD1, and the de novo cytosine methyltransferase DRM2. We show that RDR2, DCL3, AGO4, and NRPD1b (the largest subunit of Pol IVb) colocalize with siRNAs within the nucleolus. By contrast, Pol IVa and DRD1 are external to the nucleolus and colocalize with endogenous repeat loci. Mutation-induced loss of pathway proteins causes downstream proteins to mislocalize, revealing their order of action. Pol IVa acts first, and its localization is RNA dependent, suggesting an RNA template. We hypothesize that maintenance of the heterochromatic state involves locus-specific Pol IVa transcription followed by siRNA production and assembly of AGO4- and NRPD1b-containing silencing complexes within nucleolar processing centers.

INTRODUCTION

In diverse eukaryotes, small interfering RNAs (siRNAs) regulate processes that include mRNA degradation, viral suppression, centromere function, and silencing of retrotransposons and endogenous DNA repeats (Almeida and Allshire, 2005; Baulcombe, 2004; Grewal and Rice, 2004; Tomari and Zamore, 2005). siRNAs are generated by Dicer endonuclease cleavage of double-stranded

RNAs (dsRNAs), whose production in *Neurospora*, *C. elegans*, *S. pombe*, and plants involves one or more RNA-dependent RNA polymerases (RdRPs) (Baulcombe, 2004; Wassenegger and Krczal, 2006). Following dicing of dsRNAs into ~20–25 bp duplexes (Bernstein et al., 2001; Hannon, 2002), one RNA strand is loaded into effector complexes that carry out the silencing functions. A defining feature of these effector complexes is the inclusion of an Argonaute (AGO) family protein (Carmell et al., 2002; Sontheimer and Carthew, 2004). In RNA-slicing effector complexes, the AGO-associated siRNA base pairs with its target, thereby positioning the target RNA for endonucleolytic cleavage (Song et al., 2004). Within effector complexes that direct chromatin modifications (Grewal and Rice, 2004; Verdell et al., 2004; Volpe et al., 2002; Wassenegger, 2005), the mechanisms by which siRNAs guide target modifications are not yet understood.

In *Arabidopsis thaliana*, silencing at endogenous repeat loci involves histone H3K9 methylation and RNA-directed DNA methylation that is correlated with the production of homologous siRNAs (Cao et al., 2003; Lippman et al., 2003; Xie et al., 2004; Zilberman et al., 2004). Key players in this chromatin-modifying nuclear siRNA pathway include DICER-LIKE 3 (DCL3), ARGONAUTE4 (AGO4), RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), and two forms of nuclear RNA polymerase IV (Pol IV). The largest and second largest subunits of Pol IV are similar to the catalytic β and β' subunits of *E. coli* DNA-dependent RNA polymerase and to the corresponding subunits of eukaryotic nuclear RNA polymerases I, II, and III (see Onodera et al., 2005 and references therein). Two genes encode distinct Pol IV largest subunits, and two genes encode Pol IV second largest subunits. Both of the largest-subunit genes (*NRPD1a* and *NRPD1b*) are expressed, but only one of the second-largest-subunit genes (*NRPD2a*) is functional (Herr et al., 2005; Onodera et al., 2005; Pontier et al., 2005). As a result, there are two genetically nonredundant forms of Pol IV, namely Pol IVa and Pol IVb,

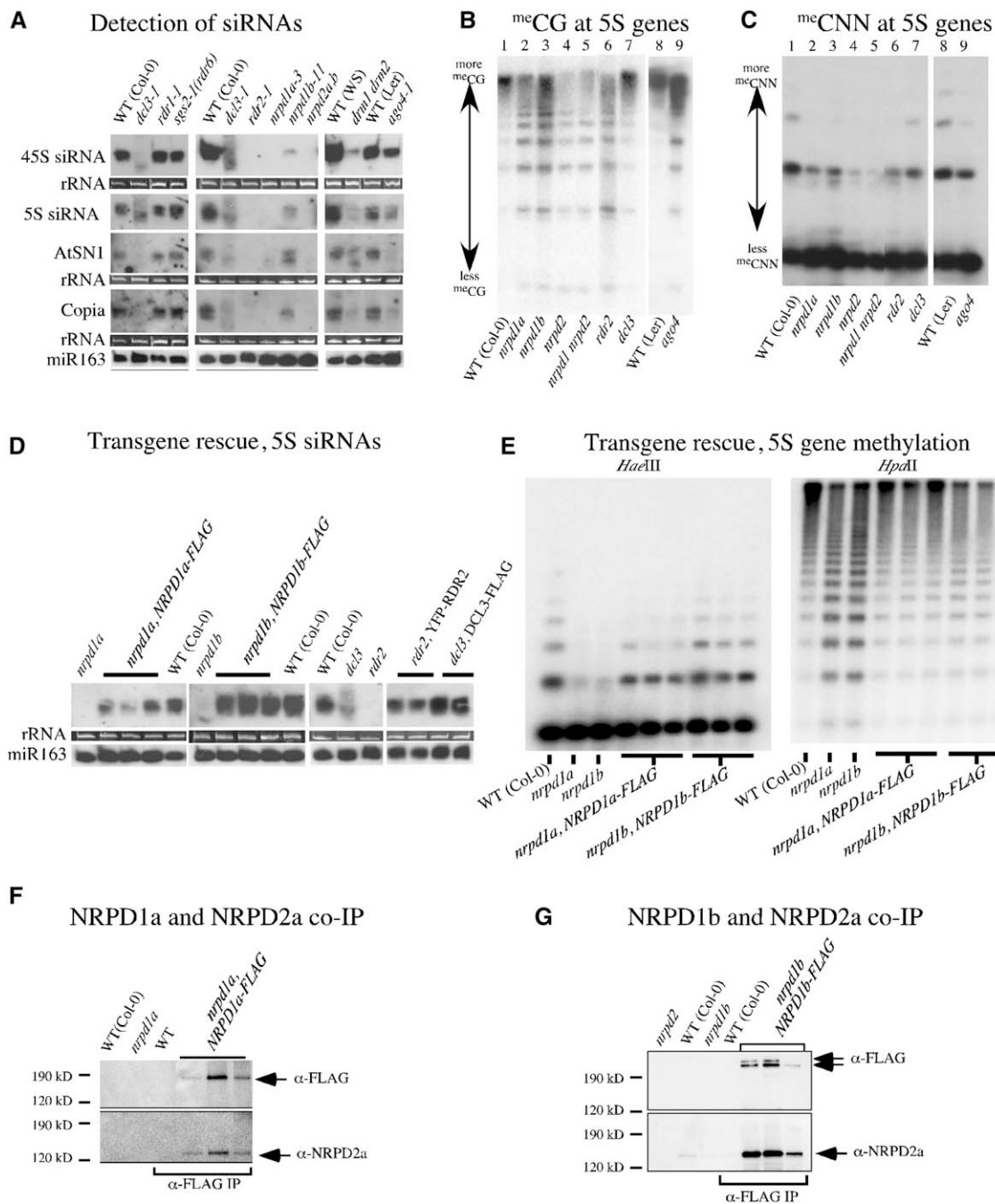


Figure 1. Loss of siRNAs and Cytosine Methylation at Repeated DNA Sequences in Mutants of the Nuclear siRNA Pathway

(A) siRNAs of wild-type (WT) and mutant plants. RNA blots were hybridized to probes corresponding to the 45S rRNA gene intergenic spacer (45S siRNA), the 5S rRNA gene siRNA siR1003, the *AtSN1* family of retroelements, the *Copia* transposable element family, or the microRNA miR163.

(B and C) Loss of CG or CNN methylation at 5S gene repeats. Genomic DNA digested with *HpaII* or *HaeIII* was hybridized to a 5S gene probe. *nprpd1a*, *nprpd1b*, *nprpd2*, *rdr2*, and *dcl3* mutants are in the Col-0 genetic background. *ago4* is in the Ler background.

(D) siRNA production in *nprpd1a*, *nprpd1b*, *rdr2*, and *dcl3* mutants is rescued by corresponding transgenes. Genomic clones under the control of their own promoters and encoding C-terminal FLAG-tagged proteins rescued the *nprpd1a*, *nprpd1b*, and *dcl3* mutants (three, three, and two independent transformants, respectively), whereas a YFP-RDR2 cDNA fusion under the control of the cauliflower mosaic virus 35S promoter rescued *rdr2* (two independent transformants shown).

(E) Transgene rescue of 5S rDNA methylation in *nprpd1a* and *nprpd1b* mutants. Southern blot analysis of *HaeIII*- and *HpaII*-digested genomic DNA with a 5S gene probe shows that the loss of methylation in *nprpd1a* and *nprpd1b* mutants, relative to wild-type (WT), is restored in each of three independent *NRPD1a-FLAG* or *NRPD1b-FLAG* transgenic lines.

designated according to which largest subunit is used. Disruption of Pol IV, RDR2, DCL3, or AGO4 genes causes decreased cytosine methylation and siRNA accumulation at endogenous repeats, including 5S ribosomal RNA genes and transposable elements (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005; Xie et al., 2004). However, the order in which these proteins act in the biogenesis of nuclear siRNAs is unclear.

Using RNA fluorescence in situ hybridization (RNA-FISH) together with protein immunolocalization, we present evidence for siRNA processing centers associated with the nucleolus. Within these centers, siRNAs colocalize with a significant portion of the RDR2, DCL3, AGO4, and NRPD1b protein pools. The two subunits of Pol IVa, however, do not localize to the processing centers but colocalize with chromosomal loci that are both sources and targets of siRNAs. A portion of the NRPD1b pool also colocalizes with target loci, as does the SWI2/SNF2 chromatin-remodeling ATPase family member DRD1, a protein required for RNA-directed DNA methylation that acts downstream of siRNA production (Kanno et al., 2004). Based on cytological, biochemical, and genetic evidence, we present a spatial and temporal model for nuclear siRNA biogenesis.

RESULTS

Loss of siRNAs and Cytosine Methylation in Nuclear siRNA Pathway Mutants

In *A. thaliana*, siRNAs homologous to repeated gene families are readily detected on RNA blots, as shown for siRNAs corresponding to the intergenic spacers of 45S or 5S rRNA genes or siRNAs corresponding to *AtSN1* or *Copia* transposable-element families (Figure 1A). Collectively, these endogenous repeats represent genes transcribed by RNA polymerase I (45S rRNA genes), RNA polymerase II (*Copia* elements), and RNA polymerase III (5S genes, *AtSN1* elements). The siRNAs are essentially eliminated upon mutation of the Pol IVa largest subunit, *NRPD1a*, or upon mutation of the second subunit of both Pol IVa and Pol IVb, *NRPD2* (note that the *nripd2a-2 nripd2b-1* double mutant [Onodera et al., 2005] is abbreviated as *nripd2* throughout this paper). siRNAs are also eliminated in *rdr2* mutants. By contrast, siRNAs are reduced in abundance, but not eliminated, in *nripd1b* or *ago4* mutants. A smear of alternatively sized small RNAs is generated in a *dcl3* mutant (Figure 1A) and is probably explained by the action of alternative Dicers (Gascioli et al., 2005). The abundance of siRNAs is also greatly reduced in the *drm1 drm2* mutant, indicating that de novo cytosine methylation plays a role in nuclear siRNA accumulation.

Loss of endogenous siRNAs correlates with loss of cytosine methylation at corresponding DNA sequences. For instance, 5S gene repeats are heavily methylated at CG motifs, making them resistant to digestion by the methylation-sensitive restriction endonuclease HpaII in wild-type *A. thaliana* (Figure 1B, lanes 1 and 8). CG methylation at HpaII sites is decreased to a similar extent in *rdr2*, *ago4*, *nripd1a*, *nripd1b*, and *nripd2* mutants, resulting in more hybridization signal in digested bands nearer the bottom of Southern blots (Figure 1B). Methylation is least affected in a *dcl3* mutant, presumably because other Dicers partially compensate (Gascioli et al., 2005).

CNN methylation is a hallmark of RNA-directed DNA methylation, which is accomplished by the de novo cytosine methyltransferase DRM2 (Cao et al., 2003). At 5S gene loci, sensitivity to digestion by HaeIII reports on CNN methylation. 5S genes are more sensitive to HaeIII digestion in *rdr2*, *nripd1a*, *nripd1b*, and *nripd2* mutants compared to wild-type plants (Figure 1C). Mutation of *DCL3* has a lesser effect on CNN methylation, again suggesting partial compensation by other Dicers. Collectively, the data of Figures 1A–1C indicate that the loss of endogenous repeat siRNAs correlates with the loss of both CG and CNN methylation, implicating RNA-directed DNA methylation (Aufsatz et al., 2002; Cao et al., 2003).

To facilitate cytological and biochemical studies, we developed transgenic lines that express functional, epitope-tagged versions of the proteins involved in the nuclear siRNA pathway. Genomic-clone transgenes expressing NRPD1a, NRPD1b, or DCL3 bearing C-terminal FLAG epitope tags all rescued their corresponding mutations and restored siRNA production, as did a YFP-RDR2 fusion engineered using a full-length *RDR2* cDNA (Figure 1D). The *NRPD1a* and *NRPD1b* transgenes also restored cytosine methylation at 5S gene repeats (Figure 1E). Collectively, these results indicate that the recombinant proteins retain their biological functions.

The Alternative Pol IV Largest Subunits, NRPD1a and NRPD1b, Physically Interact with NRPD2

Genetic evidence suggests that the Pol IV second largest subunit NRPD2 interacts with NRPD1a or NRPD1b within Pol IVa or Pol IVb, respectively (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). To obtain biochemical evidence for such interactions, we exploited transgenic plants expressing FLAG-tagged NRPD1a or NRPD1b and an anti-NRPD2 antibody (Onodera et al., 2005) to ask whether NRPD2 associates with the alternative largest subunits in vivo. Indeed, NRPD2 coimmunoprecipitates with both NRPD1a-FLAG and NRPD1b-FLAG in multiple independent transgenic plants (Figures 1F and 1G). The quantity of

(F) Physical interaction between Pol IVa subunits NRPD1a and NRPD2 detected by coimmunoprecipitation. Proteins from multiple independent NRPD1a-FLAG transgenic lines were immunoprecipitated using anti-FLAG antibody, then subjected to SDS-PAGE and electroblotting. Membranes were sequentially analyzed to detect the FLAG epitope (top) and NRPD2 (bottom).

(G) Physical interaction between NRPD1b and NRPD2. The experiment was performed as for (F) using multiple independent NRPD1b-FLAG transgenic lines.

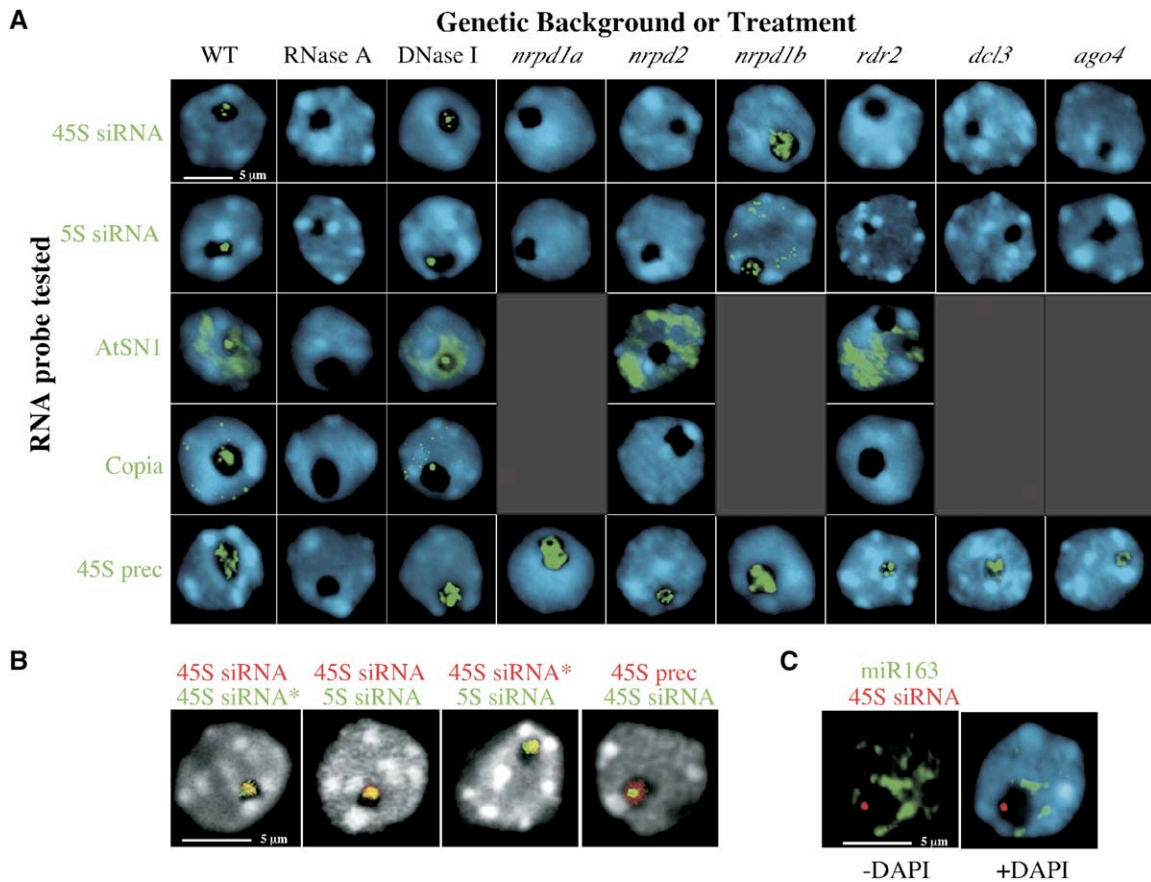


Figure 2. Nuclear Localization of siRNAs

(A) RNA-FISH using the same probe sequences used for the RNA blots of Figure 1A was performed in wild-type, nuclease-treated, or mutant nuclei as indicated. As a control, a probe that detects the 45S rRNA precursor transcripts was also used. Nuclei were counterstained with DAPI (blue). Size bars represent 5 μ m in all panels.

(B) Different siRNAs colocalize within the nucleolus. Simultaneous detection of RNA target pairs was performed using two-color FISH. Three-dimensional projections of five to seven optical sections obtained by multiphoton microscopy are shown. The red or green color of the lettering corresponds to the color of the signal for the indicated probes. Nuclei were counterstained with DAPI (false colored gray in these images). Thirty-five nuclei were observed for each probe combination. In all nuclei examined, at least 50% of the green and red pixels overlapped in the digital images to yield yellow signals.

(C) Two-color FISH using the 45S siRNA probe (red) and miR163 probe (green). Nuclei were counterstained with DAPI (blue). A localization pattern like that shown was observed in all 155 nuclei examined.

coimmunoprecipitated NRPD2 is proportional to the abundance of NRPD1a or NRPD1b in the different lines, as expected of subunits with fixed stoichiometries.

siRNAs Are Concentrated within the Nucleolus

It is not known where endogenous siRNAs are generated or processed within the cell. So, to detect siRNAs or their precursors, we employed RNA fluorescence in situ hybridization (RNA-FISH) with digoxigenin- or biotin-labeled probes (Figure 2A) identical in sequence to those used for siRNA blot hybridization (see Figure 1A). With all siRNA probes, an intense hybridization signal was observed within the nucleolus, which is the region of the nucleus not stained appreciably by the fluorescent DNA binding dye DAPI. This was true of leaf mesophyll cells at interphase, as shown throughout this paper, and in root meri-

stem cells (O.P., unpublished data). In the case of the *AtSN1* probe, a diffuse signal was also observed throughout the nucleoplasm. The nucleolar dots detected with siRNA probes occupy a small portion of the nucleolus when compared to the 45S pre-rRNA precursor transcripts that are generated by RNA polymerase I and processed in the nucleolus (Figure 2A, bottom row).

Hybridization signals detected using different siRNA probes colocalized, as shown using two-color RNA-FISH with probes specific for 45S siRNAs corresponding to opposite DNA strands (45S siRNA and 45S siRNA*) or 5S siRNAs (Figure 2B). These siRNA probe signals are spatially distinct from the signals obtained using a miRNA probe (Figure 2C). Collectively, these data indicate that nuclear siRNA hybridization signals localize within a discrete compartment of the nucleolus, smaller than the

volume occupied by 45S pre-rRNA and distinct from sites where miRNA or their precursors are concentrated.

As shown in [Figure 2A](#), siRNA and pre-rRNA hybridization signals are eliminated if nuclei are treated with ribonuclease A (RNase A) prior to extensive washing and probe hybridization but are not affected by DNase I treatment. These tests suggest that the hybridization signals result from the RNA probes' annealing to RNA targets. Importantly, the nucleolar dot signals are absent in *nRPD2*, *nRPD1a*, *rdr2*, *dcl3*, or *ago4* mutants, and, typically, no signal is observed elsewhere (although low-intensity, dispersed signals occurred infrequently; see [Table S1](#) in the [Supplemental Data](#) available with this article online for quantitative data). The exception is *nRPD1b*, for which dispersal of the nucleolar dot (as shown in [Figure 2A](#)) is more common than complete loss of signal (see [Table S1](#)). In general, these observations are consistent with the RNA blot hybridization data ([Figure 1A](#)). Importantly, 45S pre-rRNAs are unaffected by the siRNA pathway mutations, as expected.

The loss of hybridization signals in the mutants, including *dcl3* and *ago4*, which should act downstream of siRNA precursor formation, suggests that we are detecting siRNAs in the nucleolar dots rather than precursors. Perhaps the latter escape detection because they are dispersed throughout the nucleus and not concentrated in one location. However, the *AtSN1* signals, external to the nucleolus, that persist in the mutants might be precursor RNAs.

Nucleolar siRNA Processing Centers

The detection of nuclear siRNAs prompted us to ask where the proteins of the nuclear siRNA pathway are located. NRPD1a, NRPD1b, RDR2, DCL3, and AGO4 were immunolocalized in transgenic nuclei by virtue of their epitope or YFP tags, whereas native NRPD2 was localized using an anti-peptide antibody ([Figure 3A](#), top row). NRPD1a and NRPD2, the known subunits of Pol IVa, showed similar, punctate localization patterns; significantly, neither protein associates with the nucleolus. By contrast, FLAG-tagged NRPD1b, the largest subunit of Pol IVb, localizes within a nucleolar dot in addition to puncta external to the nucleolus (see also [Li et al., 2006](#) [this issue of *Cell*] and [Table S2](#)). RDR2, DCL3, and AGO4 also display prominent nucleolar dot signals in addition to puncta or diffuse signals outside the nucleolus. RDR2 signals are distinctive in that a ring or crescent at the perimeter of the nucleolus is typically observed in addition to the nucleolar dot, and this is true for both epitope-tagged and native RDR2. Control experiments showed that no immunolocalization signals were detected in transgenic nuclei if primary antibodies were omitted; likewise, no signals were detected in wild-type nuclei using anti-FLAG, anti-Myc, or anti-YFP antibodies (see [Figure S1](#)).

Nucleolar dot signals can be observed at the center or the periphery of the nucleolus, consistent with data of [Li et al. \(2006\)](#) showing that AGO4 colocalizes with markers of nucleolar accessory bodies, or Cajal bodies ([Cioce and](#)

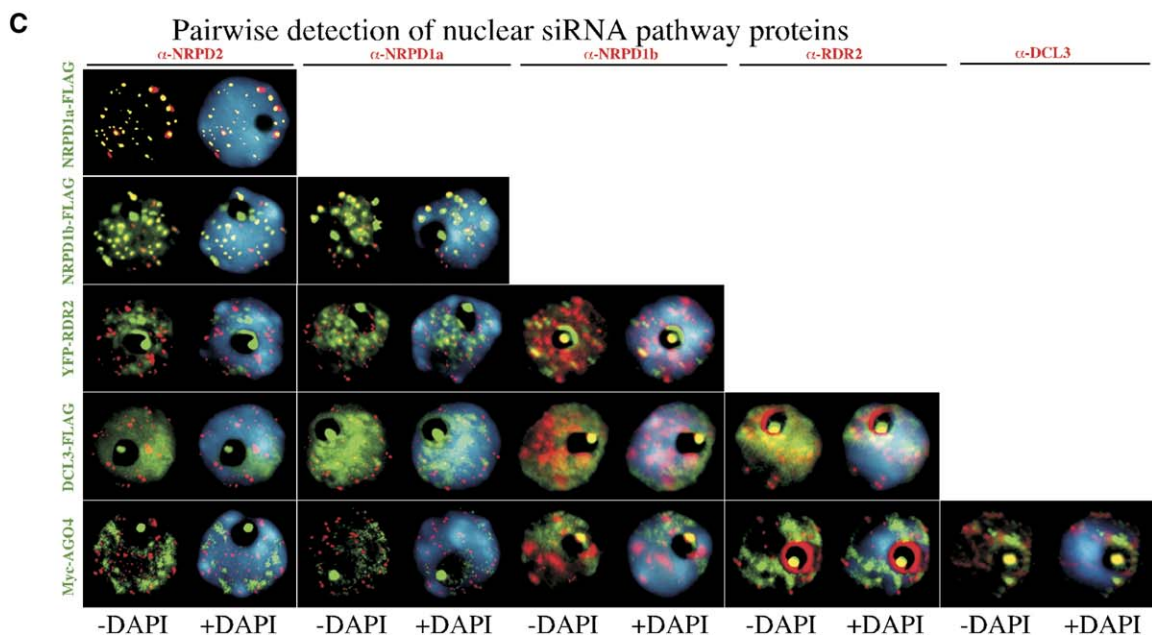
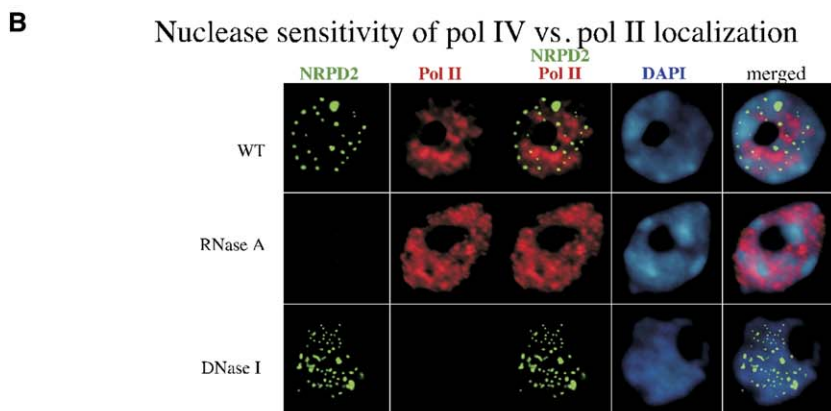
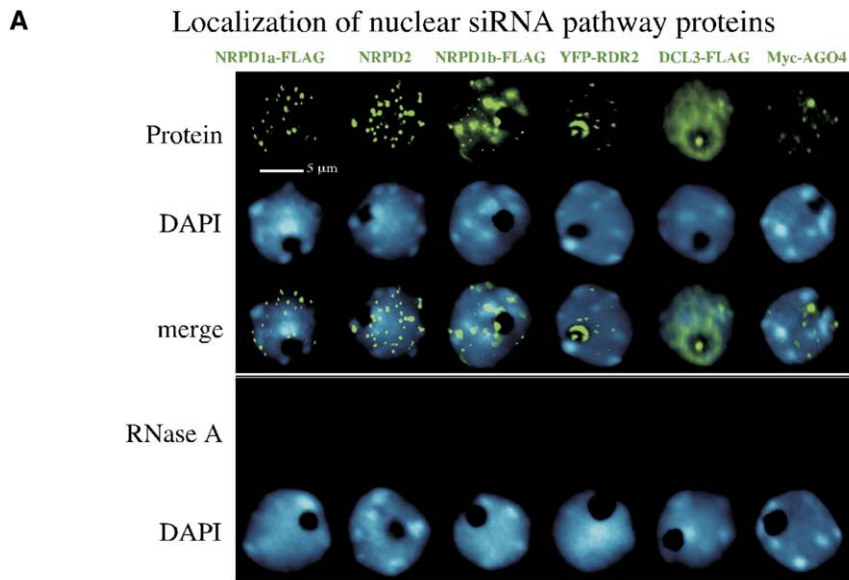
[Lamond, 2005](#)). Cajal bodies are dynamic nuclear organelles that can move in and out of nucleoli ([Boudonck et al., 1999](#)) and are implicated in the assembly of RNA-protein complexes, including snRNPs and snoRNPs ([Cioce and Lamond, 2005](#)). Therefore, what we call nucleolar dots throughout this paper are likely to be Cajal bodies or related entities (see [Li et al., 2006](#)).

Treating nuclei with RNase A prior to antibody incubation caused a complete loss of signal for all of the proteins in the majority of nuclei examined, suggesting that the proteins are not retained in RNA-depleted nuclei ([Figure 3A](#)). However, a minority of the nuclei continued to show wild-type protein localization patterns, albeit at reduced intensity, suggesting that not all nuclei are equally accessible to RNase treatment (see [Table S2](#)). Further analysis showed that, whereas NRPD2, NRPD1a, and NRPD1b signals are lost from RNase A-treated nuclei, the proteins are not lost from DNase I-treated nuclei, although NRPD1b and NRPD2 are partially mislocalized ([Figure 3B](#) and [Figure S2](#), green signals). Conversely, the signals for the second largest subunit of DNA-dependent RNA polymerase II are lost upon DNase, but not RNase, treatment ([Figure 3B](#), red signals). Collectively, these observations suggest that Pol IV interacts with RNA rather than DNA templates, unlike Pol II.

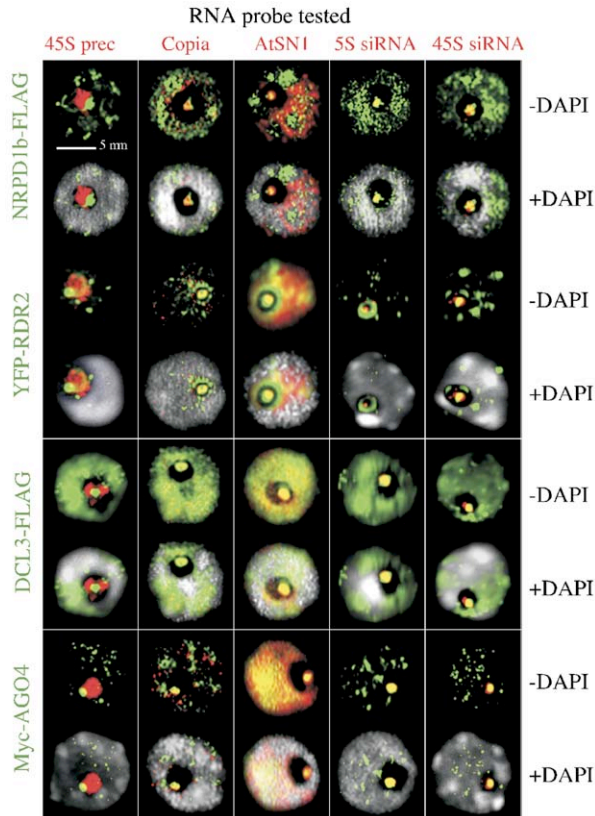
Using anti-epitope antibodies that detect transgene-encoded recombinant proteins, in combination with anti-peptide antibodies recognizing the native proteins, we simultaneously localized pairs of proteins using two-color immunofluorescence ([Figure 3C](#); [Table S3](#)). The native proteins and the recombinant proteins were found to display the same localization patterns, indicating that the anti-peptide antibodies are specific for their targets and that the epitope tags do not disrupt recombinant protein localization. NRPD1a and NRPD2, the subunits of Pol IVa, colocalize precisely, resulting in yellow signals ([Figure 3C](#), top row; note that differences in intensity of the green and red signals influence the apparent extent of overlap). Slightly more than half of the NRPD1b foci external to the nucleolus colocalize with the NRPD1a/NRPD2 foci ([Figure 3C](#), second row from top), suggesting that Pol IVb occurs at approximately half of the Pol IVa foci. However, the remaining NRPD1b foci are spatially distinct from NRPD2 (and NRPD1a). A conclusion from the latter observation is that the Pol IVb largest subunit can exist apart from the second largest subunit, both external to the nucleolus and within the nucleolus, where no NRPD2 is detectable.

External to the nucleolus, NRPD1a, NRPD2, and NRPD1b do not colocalize with RDR2, DCL3, or AGO4. However, the portion of the NRPD1b pool that is nucleolus associated colocalizes with RDR2, DCL3, and AGO4 within the nucleolar dot ([Figure 3C](#)).

We next asked whether the nucleolar dots previously detected by RNA-FISH ([Figure 2](#)) correspond to the same nucleolar dots where NRPD1b, RDR2, DCL3, and AGO4 colocalize ([Figure 3](#)). To address this question, we performed protein immunolocalization followed by



A Dual Protein immunolocalization/RNA-FISH



B AGO4-siRNA immunoprecipitation

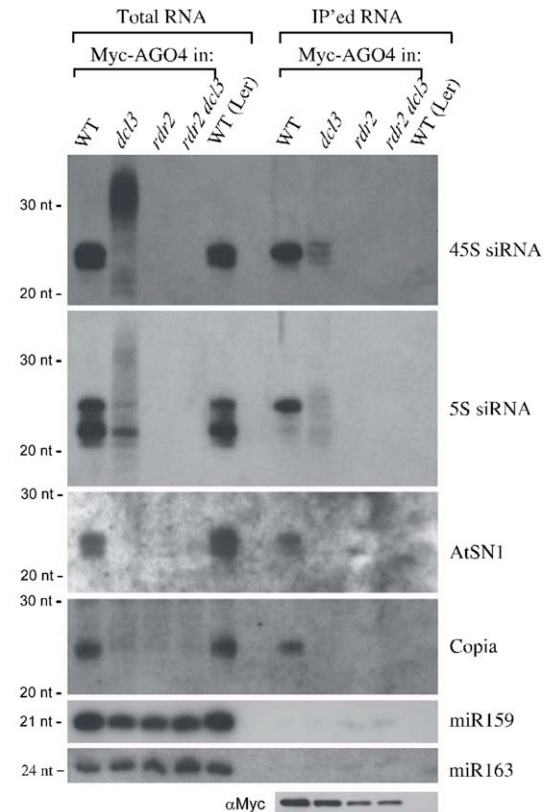


Figure 4. siRNAs Colocalize with NRPD1b, RDR2, DCL3, and AGO4

(A) Nuclei were hybridized with 45S rRNA precursor, *Copia*, *AtSN1*, 5S siRNA, or 45S siRNA probes (red signals). NRPD1b-FLAG, YFP-RDR2, DCL3-FLAG, or Myc-AGO4 was immunolocalized using anti-FLAG, anti-YFP, or anti-Myc antibodies (green signals). Images shown are three-dimensional projections of five to seven optical sections obtained by multiphoton microscopy. Pairs of images are presented for each protein localized, the lowermost image including the DAPI signal (false colored gray) to help reveal the nucleolus.

(B) siRNAs physically associate with AGO4. Total RNA or RNA immunoprecipitated (IP) using anti-Myc antibodies from transgenic plants expressing Myc-AGO4 in wild-type, *dcl3*, *rdr2*, or *dcl3 rdr2* backgrounds was subjected to RNA blot hybridization using 45S siRNA, 5S siRNA, *AtSN1*, *Copia*, and miR159 probes. RNA of nontransgenic wild-type plants (ecotype Ler) served as a control. The presence of AGO4 in immunoprecipitates was confirmed by immunoblotting using anti-Myc antibody.

RNA-FISH (Figure 4A). As is evident by the yellow signals resulting from siRNA probe and protein signal overlap, NRPD1b, RDR2, DCL3, and AGO4 typically colocalize with 45S, 5S, *AtSN1*, and *Copia* siRNAs within the nucleolar dots but do not colocalize precisely with 45S rRNA precursor transcripts (Figure 4A; see also Table S4). We interpret the colocalization of NRPD1b, RDR2, DCL3, AGO4, and siRNAs as evidence of siRNA processing centers in which dsRNAs generated by RDR2 are diced by DCL3 to generate siRNAs that are loaded into RISC effector complexes that contain AGO4 and NRPD1b.

Consistent with the interpretation that siRNAs are stably associated with AGO4, immunoprecipitation of Myc-AGO4 pulls down 45S, 5S, *AtSN1*, and *Copia* siRNAs (Figure 4B). Moreover, in *rdr2* or *rdr2 dcl3* double mutants, siRNAs are no longer found in the Myc-AGO4 immunoprecipitates. In *dcl3* mutants, siRNAs associated with AGO4 are greatly reduced in abundance and variable in size, consistent with the hypothesis that AGO4 is capable of binding siRNAs generated by other Dicers that partially compensate for the loss of DCL3.

Figure 3. Immunolocalization of Nuclear siRNA Pathway Proteins

(A) Epitope-tagged NRPD1a, NRPD1b, DCL3, and AGO4 recombinant proteins that rescue corresponding mutations were immunolocalized (green signals) using anti-FLAG or anti-Myc antibodies. Native NRPD2 was detected using anti-peptide antisera. RDR2-YFP was localized using anti-YFP. Nuclei were counterstained with DAPI.

(B) Immunolocalization of NRPD2 and the Pol II second largest subunit in wild-type untreated, RNase A-, or DNase I-treated nuclei.

(C) Anti-peptide antibodies recognizing native proteins (red signals) were used in combination with antibodies recognizing FLAG-, Myc-, or YFP-tagged recombinant proteins (green signals) in nuclei of transgenic plants. Colocalizing proteins generate yellow signals.

Pol IV and the Putative Chromatin Remodeler DRD1 Colocalize with Endogenous Repeats

To determine where the endogenous DNA repeats are located relative to the nucleolar dots, we used DNA-FISH to localize the 45S rRNA gene loci (i.e., the nucleolus organizer regions; NORs) and 5S rRNA gene clusters. The FISH signals for the highly condensed portions of 45S and 5S rRNA gene loci are not detected within the nucleolus (Figure 5, red signals), indicating that the bulk of the target gene loci, composed mostly of inactive repeats, are distant from the nucleolar dots.

By combining protein immunolocalization (green signals) with DNA-FISH (red signals), we asked whether the Pol IV foci external to the nucleolus correspond to endogenous repeat loci. Indeed, NORs and 5S gene loci were found to colocalize with NRPD1a, NRPD1b, and NRPD2, yielding yellow signals at most, though not all, of the loci (see Table S5 for quantitative data). Some overlap between 5S gene loci and RDR2 or DCL3 signals was also observed, although the diffuse distribution of DCL3 may make the apparent overlap coincidental. We also examined the localization of DRD1, a SWI2/SNF2-related protein that is involved in RNA-directed DNA methylation via a Pol IVb-dependent pathway (Kanno et al., 2005; Kanno et al., 2004). DRD1 is distributed throughout the nucleus, with the exception of the nucleolus, and is concentrated at chromocenters that include NORs and 5S gene loci (Figure 5, bottom row). Collectively, these observations suggest that Pol IVa, Pol IVb, and DRD1 are present at the endogenous repeat loci, presumably acting in the generation of siRNA precursors or in the downstream functioning of siRNA-containing effector complexes.

Mutation-Induced Mislocalization of Nuclear siRNA Pathway Proteins

To deduce the order in which proteins of the nuclear siRNA pathway act, we examined the effect of mutations on each protein's localization, resulting in the matrix of images shown in Figure 6 (see Table S6 for quantitative data). Protein signals were absent upon mutation of the genes that encode the corresponding proteins, as expected, indicating that all of the mutants are protein nulls and that the antibodies are specific for their intended targets. NRPD1a localization is unaffected in *rdr2*, *dcl3*, or *ago4* mutants, as is NRPD2 localization, consistent with Pol IVa acting upstream of RDR2, DCL3, and AGO4. RDR2 localization is dependent on Pol IVa (NRPD1a and NRPD2), but not on NRPD1b, DCL3, or AGO4, indicating that RDR2 acts downstream of Pol IVa, but upstream of Pol IVb, dicing and effector complex assembly.

DCL3 localization is dependent on both Pol IVa and RDR2 but is independent of AGO4 and NRPD1b, suggesting that dicing occurs following double-stranded RNA formation, mediated by RDR2, and upstream of effector complex assembly and Pol IVb function. Consistent with this interpretation, the NRPD1b nucleolar dot is absent in *nRPD1a*, *rdr2*, *dcl3*, and *ago4* mutants but is still present in a *drd1* mutant (see Figure S3), indicating that the nucle-

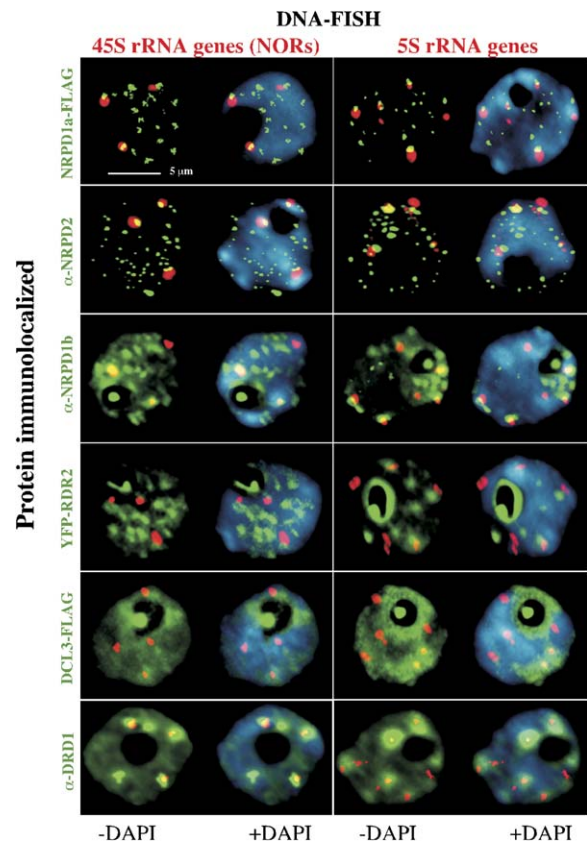


Figure 5. Pol IV Colocalizes with Endogenous Repeat Loci

45S rRNA gene loci (nucleolus organizer regions; NORs) or 5S gene chromosomal loci were visualized using DNA-FISH (red signals), and the indicated proteins were immunolocalized (green signals). Yellow indicates overlapping DNA and protein signals. NRPD1a-FLAG and DCL3-FLAG recombinant proteins were detected in nuclei of transgenic plants using anti-FLAG antibodies; NRPD2, NRPD1b, and DRD1 were detected in nuclei of nontransgenic plants using anti-peptide antibodies recognizing the native proteins; and recombinant YFP-RDR2 was detected using anti-YFP (green signals). Nuclei were counterstained with DAPI (blue). Note that *A. thaliana* has four NORs and six 5S gene loci in the Col-0 ecotype. The NORs tend to coalesce such that only three NORs are observed in most of the images shown.

olar NRPD1b signal is dependent on siRNA processing and effector complex assembly but is formed upstream of steps that involve chromatin remodeling by DRD1. The NRPD1b signals that are outside the nucleolus are unaffected in *rdr2* or *dcl3* mutants but are less punctate and therefore appear more diffuse in the *drd1* mutant, suggesting that DRD1 influences NRPD1b localization at target loci.

DISCUSSION

A Spatial and Temporal Model for the Nuclear siRNA Pathway

RNA-directed DNA methylation requires de novo methyltransferase activity, suggesting that DRM-class cytosine

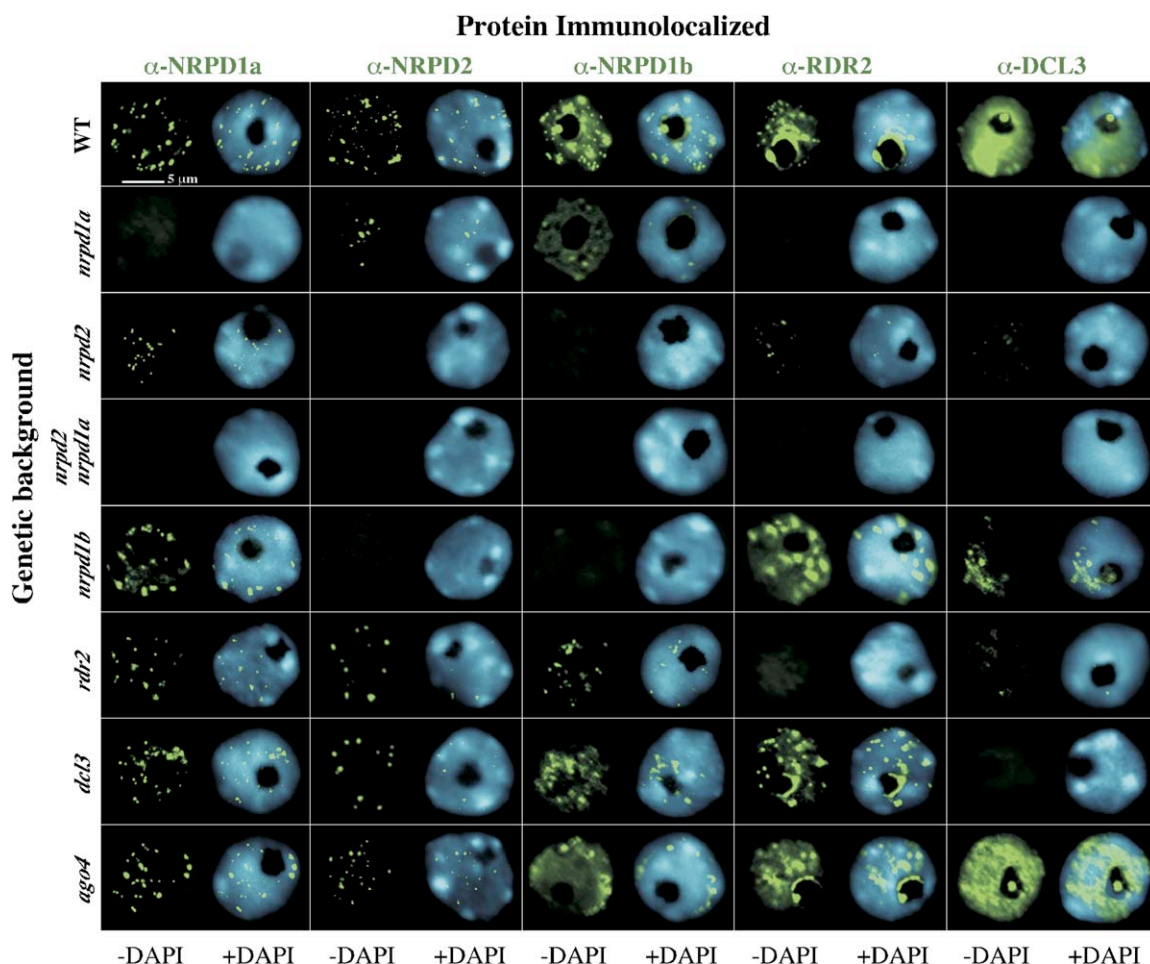


Figure 6. Effects of Mutations on the Localization of Proteins Involved in Nuclear siRNA Biogenesis

The figure shows a matrix of images in which NRPD1a, NRPD2, NRPD1b, RDR2, and DCL3 were immunolocalized using anti-peptide antibodies recognizing the native proteins (green signals) in multiple genetic backgrounds as indicated along the vertical axis. Nuclei were counterstained with DAPI (blue).

methyltransferases (probably *DRM2* only, because *DRM1* is not expressed appreciably) act downstream of siRNA production (Cao et al., 2003). However, endogenous nuclear siRNAs fail to accumulate in *drm* mutants (Xie et al., 2004; Zilberman et al., 2004), suggesting that DRM2 also acts upstream of siRNA production (see also Figure 1A). Our model attempts to address this apparent paradox (Figure 7). Based on a study in *Neurospora* suggesting that methylation impedes RNA polymerase elongation (Rountree and Selker, 1997), we propose that transcripts trailing from polymerases that are stalled or slowed by DRM-mediated methylation (Figure 7, upper left) are sensed as aberrant and, directly or indirectly, become templates for Pol IVa. In this model, Pol IVa is spatially tethered to the DNA by virtue of the RNA template. This aspect of the model accounts for the colocalization of Pol IVa subunits with endogenous repeat loci and their loss in RNase A-treated nuclei. We place Pol IVa first in the pathway because Pol IVa is located directly at the

endogenous repeat loci and because mutation of either Pol IVa subunit (NRPD1a or NRPD2) eliminates siRNA production. By contrast, mutation of NRPD1b, the largest subunit of Pol IVb, which also colocalizes with the endogenous repeat loci, does not eliminate siRNA production but does affect RNA-directed cytosine methylation, suggesting that Pol IVb acts late in the pathway (Kanno et al., 2005; Pontier et al., 2005; Vaucheret, 2005; see also Figures 1A–1C). The fact that siRNA accumulation is reduced in *nrd1b* mutants (see Figure 1A) may be due to the destabilization of the NRPD2 pool upon loss of NRPD1b (see Figure 1G, Figure 6 and Pontier et al., 2005). Loss of NRPD2 would indirectly deplete Pol IVa activity by depriving NRPD1a of its partner catalytic subunit. Alternatively, decreased Pol IVb-dependent cytosine methylation might decrease the incidence of aberrant transcript production at endogenous repeat loci, thereby depleting the pool of Pol IVa templates. These alternative explanations are not mutually exclusive.

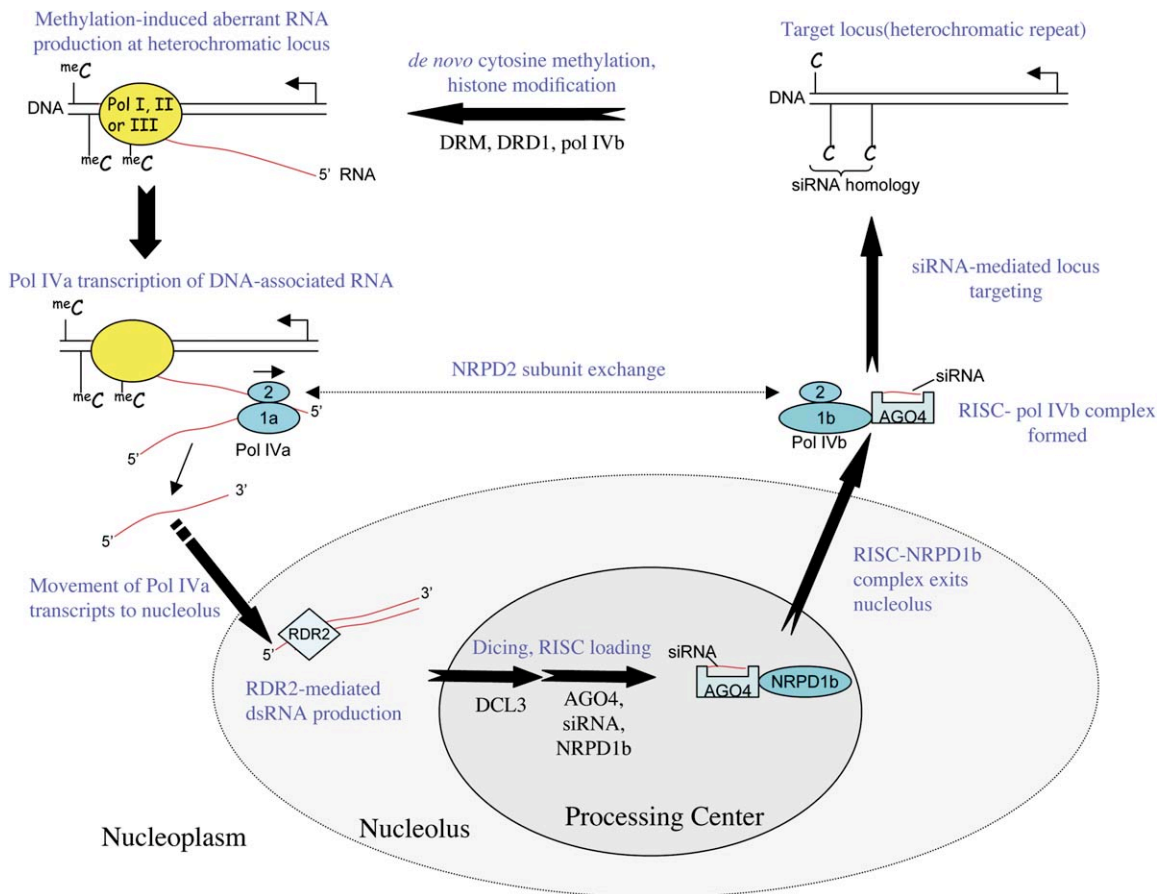


Figure 7. A Spatial and Temporal Model for Nuclear siRNA Biogenesis

Subunits of Pol IVa (abbreviated 1a and 2) colocalize with endogenous repeat loci but are mislocalized upon RNase A treatment, suggesting that Pol IVa transcribes RNA templates whose spatial distribution is influenced by DNA. We propose that cytosine methylation by DRM induces the production of aberrant RNAs, possibly by impeding polymerase elongation, which Pol IVa then uses as templates. Pol IVa transcripts then move, by an unknown mechanism, to the nucleolus, where RDR2, DCL3, and AGO4 are located. In the siRNA processing center, the largest subunit of Pol IVb, NRPD1b, joins the AGO4-containing RISC complex and acquires the NRPD2 subunit to become functional Pol IVb only upon leaving the nucleolus. Formation of Pol IVb is required for the stability of the NRPD2 pool despite the fact that NRPD2 colocalizes more precisely with NRPD1a than with NRPD1b, suggesting that NRPD2 subunits exchange between Pol IVa and b. AGO4, Pol IVb, and DRD1 then play unspecified roles in guiding heterochromatic modifications at the endogenous repeats, including de novo cytosine methylation by DRM. Methylation-dependent production of aberrant RNAs results in a positive feedback loop for maintaining heterochromatin at the DNA repeats.

Like Pol IVa, RDR2 is required for endogenous siRNA production. RDR2 is mislocalized in an *nrdp1a* mutant, whereas the converse is not true (see Figure 6), indicating that RDR2 acts downstream of Pol IVa. RDR2 is not abundant at the endogenous repeats but is concentrated in the nucleolus. Collectively, these observations suggest that Pol IVa generates precursor RNAs at the endogenous repeats and that these transcripts then move to the nucleolus, where their complements are generated by RDR2 transcription. Annealing of these RNAs would produce dsRNAs that are then diced by DCL3 and loaded into an AGO4-containing effector complex, or RISC (RNA-induced silencing complex), within the siRNA processing center. The observation that Pol IVa subunits and RDR2 are not mislocalized in *dcl3* or *ago4* mutants is consistent with Pol IVa and RDR2 acting upstream of DCL3 and

AGO4. Likewise, the absence of siRNAs associated with AGO4 in *rdr2* mutants, the atypical sizes of siRNAs associated with AGO4 in *dcl3* mutants, and the mislocalization of AGO4 in *rdr2* or *dcl3* mutants (see also Li et al., 2006) indicate that AGO4 acts downstream of RDR2 and DCL3.

Two observations suggest that Pol IVb acts downstream of AGO4-RISC assembly. First, the largest subunit of Pol IVb, NRPD1b, colocalizes with the nucleolar dot, but only if siRNAs are being produced and assembled into effector complexes; the nucleolar NRPD1b signal is absent in *nrdp1a*, *rdr2*, *dcl3*, or *ago4* mutants. Second, the NRPD2 subunit is never observed within the nucleolus yet is presumably essential for Pol IVb function based on the genetic screen of Kanno et al. that recovered nine mutant alleles of *NRPD1b* and 12 alleles of *NRPD2a* but no alleles of *NRPD1a* (Kanno et al., 2005). The genetic evidence

strongly predicts that NRPD1b is nonfunctional in the absence of the second largest subunit. We propose that NRPD1b associates with AGO4-RISC, which is supported by our immunolocalization data and the finding that NRPD1b can be coimmunoprecipitated in association with AGO4 (Li et al., 2006). Upon leaving the nucleolus as a subunit of AGO4-RISC, we deduce that NRPD1b can then associate with NRPD2, forming functional Pol IVb. Consistent with this hypothesis, NRPD2 coimmunoprecipitates with AGO4 (J.H. and C.S.P., unpublished data) as well as with NRPD1b (see Figure 1G).

How AGO4-RISC-Pol IVb complexes mediate their effects on chromatin modification at target loci is unclear. One possibility is that AGO4-RISC directs Pol IVb to its target sites. Alternatively, AGO4 might transfer the siRNA to Pol IVb when the NRPD2 subunit joins the NRPD1b subunit, after the AGO4-RISC-NRPD1b complex leaves the nucleolus. The siRNA, or a Pol IVb transcript primed by the siRNA, might then be used to conduct a homology search for target sequences, aided by DRD1 (Kanno et al., 2004), a member of the SWI2/SNF2-related family of chromatin-remodeling ATPases that is within a subfamily most closely related to yeast RAD54. In double-strand DNA break repair, RAD54 is required for helping broken DNA ends conduct a homology search and invade homologous duplex DNA of a sister chromosome, thereby facilitating repair by homologous recombination (Krogh and Symington, 2004). A partnership between Pol IVb and DRD1 could account for their presence at the target loci, the observation that NRPD1b and DRD1 are both essential for cytosine methylation but not siRNA production (Kanno et al., 2004, 2005), and the partial mislocalization of NRPD1b in a *drd1-6* mutant (see Figure S2). Moreover, RNA polymerases and chromatin-remodeling ATPases are nucleotide triphosphate-hydrolyzing molecular motors that can be envisioned working together, with processive movement of the polymerase possibly providing directionality to subsequent chromatin modifications. Resulting de novo DNA methylation by DRM2, which is predicted to contribute to aberrant RNA production, would provide for positive feedback in our model (Figure 7).

As touched upon previously, our observation that NRPD2 signals are severely reduced in *nRPD1b*, more so than in the *nRPD1a* mutant (see Figure 1G and Figure 6), is consistent with previously published immunoblot data (Pontier et al., 2005). Nonetheless, it is surprising given the nearly perfect colocalization of NRPD2 with NRPD1a, as opposed to only ~50% overlap of NRPD2 with NRPD1b (see Figure 3C). Based on these data, one might expect NRPD1a to be most important for NRPD2 stability. To reconcile these findings, we propose that NRPD2 must be able to exchange between Pol IVb and Pol IVa (Figure 7), with NRPD1b interactions somehow more important for the overall stability of the NRPD2 pool.

The idea that incomplete, or otherwise aberrant, transcripts can induce transcriptional silencing at endogenous repeats may have parallels with the silencing of nonproductive human immunoglobulin genes. In this phenome-

non, genes whose transcripts contain premature stop codons following V-D-J recombination are transcriptionally silenced (Buhler et al., 2005), indicating a link between nonsense-mediated decay (NMD) and chromatin modification. In *Arabidopsis*, proteins of the exon-joining complex and NMD pathways were identified within the nucleolar proteome, and some were shown to localize as nucleolar dots (Pendle et al., 2005). Whether these proteins colocalize with the siRNA processing centers is unclear at present.

The nucleolus is best known as the site of 45S pre-rRNA transcription and ribosome assembly. However, small-RNA-directed pre-rRNA cleavage, methylation, and pseudouridylation; biogenesis of signal-recognition particle and telomerase small RNAs; tRNA processing by RNase P; and some pre-mRNA processing also take place within the nucleolus (Bertrand et al., 1998; Filipowicz and Pogacic, 2002; Kiss, 2002; Pederson, 1998). Our findings suggest that processing of endogenous nuclear siRNAs, and possibly RISC storage or sequestration, are additional nucleolar functions to be explored.

EXPERIMENTAL PROCEDURES

Mutant Plant Strains

Arabidopsis rdr2-1 and *dcl3-1* were provided by Jim Carrington, *sgs2-1* (alias *sde1*; *rdr6*) was provided by Herve Vaucheret, and *drd1-6* was provided by Tatsuo Kanno and Marjori Matzke. *drm2-1*, *ago4-1*, and *nRPD1b-11* (SALK_029919) were obtained from the Arabidopsis Biological Resource Center. *nRPD1a* and *nRPD2* mutants were described previously (Onodera et al., 2005).

Generation of Transgenic Lines

Full-length genomic sequences including promoters were amplified by PCR from *A. thaliana* Col-0 DNA using Pfu polymerase (Stratagene) and cloned into pENTR/D-TOPO (Invitrogen). NRPD1a primers were 5'-CACC GG TGTCTC ACATTC CAAAGTCCCC-3' (forward) and 5'-CGGGTTTTCGGAGAAACCACC-3' (reverse). NRPD1b primers were 5'-CACC CGCTACTACA AACGGAAACGGTCA-3' and 5'-TGTCTGCGTCTGGGACGG-3'. Genomic DCL3 was amplified from BAC clone T15B3 using 5'-CACCCCGACCGAAATCCTCATGACCTAA-3' and 5'-CTTTTGATTATGACGATCTTGCGGCGC-3'; the CACC added to forward primers allowed directional cloning into the entry vector. Reverse primers eliminated stop codons to allow epitope-tag fusion. Genes were recombined into pEarleyGate 302 (Earley et al., 2006) to add C-terminal FLAG epitopes. RDR2 coding sequences were amplified by RT-PCR using Pfx Platinum DNA polymerase (Invitrogen) and primers 5'-CACCATGGTGTCTCAGAGACGACGAC-3' and 5'-GGGCAATCAAATGGATACAAGTCC-3'. PCR products captured in pENTR/D-TOPO were recombined into pEarleyGate 104 (Earley et al., 2006), fusing RDR2 sequences C-terminal to YFP expressed from a CaMV 35S promoter. Transformation of constructs into corresponding homozygous mutants was by the floral dip method (Clough and Bent, 1998).

Southern Blotting and Small-RNA Blot Hybridization

Genomic DNA (250 ng) digested with HaeIII or HpaII was subjected to agarose gel electrophoresis, blotted to nylon membranes, and hybridized to a 5S gene probe as described previously (Onodera et al., 2005). Generation of RNA probes labeled with [α -³²P]CTP and small-RNA blot hybridization were also as described previously (Onodera et al., 2005). Specific oligodeoxynucleotides used in T7 polymerase reactions (CTGTCTC hybridized to the T7 promoter adaptor) were as follows: 45S siRNA: 5'-CAATGTCTGTTGGTCCAAAGAGGGAAAAG

GGCCCTGTCTC-3'; 45S prec: 5'-AGTCCGTGGGGAACCCCTTTTTCGGTTCGCCCTGTCTC-3'; 5S siRNA: 5'-AGACCGTGAGGCCAACTTGGCATCCTGTCTC-3'; *Copia*: 5'-TTATTGGAACCCGGTTAGGACCTGTCTC-3', and miR163: 5'-TTGAAGAGGACTTGGAACTTCGATCCTGTCTC-3'.

Antibodies

Rabbit antibodies raised against NRPD2 and Pol II second-largest-subunit peptides were described previously (Onodera et al., 2005). Chicken antibodies recognizing DCL3, NRPD1a, NRPD1b, or RDR2 were generated against peptides conjugated to keyhole limpet hemocyanin. Peptides were as follows: DCL3: SLEPEKMEEGGGNSC; NRPD1a: EELQVPVGLTSLIGC; NRPD1b: MEEESTSEILDGEIC; RDR2: ETTTNRSTVKISNVC; DRD1: NKNVHKRQKQVDDGC. Immunolocalization was performed using 1:200 dilutions of antisera, except that NRPD1b antiserum was diluted 1:500. FLAG-tagged proteins were detected using mouse monoclonal anti-FLAG antibody (Sigma-Aldrich) diluted 1:400. RDR2-YFP was detected using mouse anti-GFP/YFP (BD Biosciences) diluted 1:500.

Immunolocalization

Leaves from 28-day-old plants were harvested and nuclei were extracted as described previously (Onodera et al., 2005). After postfixation in 4% paraformaldehyde/PBS (phosphate-buffered saline), washes in PBS, and blocking at 37°C, slides were exposed overnight to primary antisera in PBS and 0.5% blocking reagent (Roche). After washes in PBS, slides were incubated at 37°C with anti-mouse-FITC diluted 1:100 (Sigma), goat anti-chicken Alexa 488 diluted 1:300 (Molecular Probes), or goat anti-chicken Alexa 543 diluted 1:400 (Molecular Probes). Nuclei were counterstained with 1 µg/ml DAPI (Sigma) in Vectashield (Vector Laboratories).

Immunoprecipitation and Immunoblotting of Epitope-Tagged Proteins

Pol IV immunoprecipitation was performed using protein extracted from 2.0 g of tissue according to Baumberger and Baulcombe (2005), except that homogenates were filtered through two layers of Miracloth and subjected to centrifugation at 16,000 × g for 15 min at 4°C prior to incubation with anti-FLAG M2 affinity gel (Sigma). Proteins eluted in 2× SDS-PAGE loading buffer at 100°C for 2 min were fractionated on 7.5% Tris-glycine SDS-polyacrylamide gels (Cambrex) and electroblotted to PVDF membranes (Millipore). Membranes incubated with peroxidase-linked anti-FLAG M2 antibody diluted 1:2000 (Sigma) were visualized using chemiluminescence detection (Amersham). Membranes were then stripped using 25 mM glycine-HCl (pH 2.0), 1% (w/v) SDS for 30 min with agitation, followed by two 10 min washes in Tris-buffered saline, 0.05% (v/v) Tween 20. NRPD2 immunoblotting was as described in Onodera et al. (2005).

For coimmunoprecipitation of AGO4 and siRNAs, flowers (0.7 g) frozen in liquid nitrogen were homogenized in 2 ml of IP buffer (50 mM Tris-Cl [pH 7.5], 150 mM NaCl, 5 mM MgCl₂, 10% glycerol, 0.1% NP-40) containing fresh DTT (2 mM), PMSF (1 mM), pepstatin (0.7 µg/ml), MG132 (10 µg/ml), and Complete protease inhibitor cocktail (Roche). Following centrifugation, lysates precleared with Protein G-agarose beads (Pierce) for 1 hr at 4°C were incubated with anti-Myc (Upstate) diluted 1:250 for 3 hr at 4°C. Antibody-antigen complexes were captured on Protein G-agarose (60 µl) at 4°C for 2 hr and washed four times with IP buffer. For siRNA detection, beads were treated with Proteinase K and extracted sequentially with TE containing 1.5%, 0.5%, or 0.1% SDS. Pooled supernatants extracted with phenol:chloroform (1:1) followed by chloroform were ethanol precipitated. Total siRNAs and RNA blots were prepared and hybridized as previously described (Mette et al., 2000; Zilberman et al., 2003). DNA probes were used to detect 5S siRNAs, 45S siRNAs, miR157, and miR163; RNA probes were used to detect *AtSN1* and *Copia* siRNAs. Probe sequences were as follows: 5S siRNA: 5'-ATGCCAAGTTTGGCCCTACGGTCT-3'; 45S siRNA: 5'-GTCTGTTGGTGCCAAGAGGGAAG

GGCTAAT-3'; *AtSN1*: 5'-ACCAACGTGTTGTTGGCCAGTGGTAAATCTCTCAGATAGAGG-3'; *Copia*: 5'-TTATTGGAACCCGGTTAGGA-3'; miR159: 5'-TAGAGCTCCCTTCAATCCAAA-3'; miR163: 5'-ATCGAAGTTGGAAGTCCCTCTCAA-3'.

RNA and DNA In Situ Hybridization

RNA probes were labeled by *in vitro* T7 polymerase (Ambion) transcription with digoxigenin-11-UTP or biotin-16-UTP RNA labeling mix (Roche). RNA in situ hybridization was carried out at 42°C overnight using a probe solution containing 1 µg RNA probe, 5 µg yeast tRNA (Roche), 50% dextran sulfate, 100 mM PIPES [pH 8.0], 10 mM EDTA, and 3 M NaCl as described previously (Highett et al., 1993). Slides were washed sequentially in 2× SSC, 50% formamide, 50°C followed by 1× SSC, 50% formamide, 50°C, then 1× SSC 20°C, and finally TBS at 20°C. Where applicable, nuclei were incubated at 37°C for 30 min in a solution of RNase-free DNase I (0.015 U/µl) or in a solution of RNase A (100 µg/ml, Roche). Nuclease reactions were stopped in 10 mM EDTA (pH 7.5) for 2 min followed by three washes in 0.1× SSC.

DNA-FISH using 5S or 45S rRNA gene probes labeled with biotin-dUTP or digoxigenin-dUTP was performed as described (Pontes et al., 2003). Digoxigenin-labeled probes were detected using mouse anti-digoxigenin antibody (1:250, Roche) followed by rabbit anti-mouse antibody conjugated to Alexa 488 (Molecular Probes). Biotin-labeled probes were detected using goat anti-biotin conjugated with avidin (1:200, Vector Laboratories) followed by streptavidin-Alexa 543 (Molecular Probes). DNA was counterstained with DAPI (1 µg/ml) in Vectashield (Vector Laboratories). For dual protein/nucleic acid localization experiments, slides were first subjected to immunofluorescence, then postfixed in 4% formaldehyde/PBS followed by RNA- or DNA-FISH.

Microscopy

Nuclei were routinely examined using a Nikon Eclipse E800i epifluorescence microscope, with images collected using a Photometrics Cool-snap ES Mono digital camera. The images were pseudocolored, merged, and processed using Adobe Photoshop (Adobe Systems). Multiphoton optical-section stacks were collected using a Zeiss LSM 510 Meta microscope. Single optical sections using 40× averaging were acquired by simultaneous scanning to avoid artifactual shift between two optical channels. The 488 nm line of an argon laser was used for detection of FITC FLAG-tagged proteins, and the 543 nm line of a helium-neon laser was used for detection of Alexa 543 siRNA signals. For the detection of DAPI, either a 715 or 750 nm multiphoton tuned titanium-sapphire laser was used. Projections of 3D data stacks were composed using Imaris 4.1 software from Bitplane (<http://www.bitplane.com>).

Supplemental Data

Supplemental Data include three figures and six tables and can be found with this article online at <http://www.cell.com/cgi/content/full/126/1/79/DC1/>.

ACKNOWLEDGMENTS

O.P. performed all microscopy, P.C.N. generated siRNA blots, T.R. performed DNA methylation assays, and J.H. generated epitope-tagged Pol IV lines and colP data. YFP-RDR2 and FLAG-DCL3 lines were generated by O.P. and A.V., respectively; C.S.P. wrote the manuscript. C.F.L. and S.E.J. generated the Myc-AGO4 transgenic line and contributed Figure 4B. O.P. and P.C.N. were supported by fellowships SFRH/BPD/17508/2004 and SFRH/BD/6520/2001, respectively, from the Fundação para a Ciência e Tecnologia (Portugal). Pikaard lab work was supported by NIH grants R01GM60380 and R01GM077590 and the Monsanto Company/Washington University Biology Research Agreement. C.F.L. was supported by Ruth L. Kirschstein National Research Service Award GM07185. S.E.J. is a Howard Hughes Medical

Institute Investigator. Work in the Jacobsen laboratory was supported by NIH grant GM60398. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of NIH, HHMI, or the Monsanto Company. We thank Howard Berg (Donald Danforth Plant Science Center) for multiphoton microscopy training, Tom Juehne and Keming Song (Sigma-Aldrich Company) for Pol IV antibody production, Dr. Wanda Viegas (Instituto Superior de Agronomia, Portugal) for comentoring P.C.N., and Eric Richards (Washington University) for suggestions to improve the manuscript.

Received: March 1, 2006

Revised: March 31, 2006

Accepted: May 17, 2006

Published: July 13, 2006

REFERENCES

- Almeida, R., and Allshire, R.C. (2005). RNA silencing and genome regulation. *Trends Cell Biol.* *15*, 251–258.
- Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J., and Matzke, M. (2002). RNA-directed DNA methylation in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* *99*, 16499–16506.
- Baulcombe, D. (2004). RNA silencing in plants. *Nature* *431*, 356–363.
- Baumberger, N., and Baulcombe, D.C. (2005). *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. USA* *102*, 11928–11933.
- Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* *409*, 363–366.
- Bertrand, E., Houser-Scott, F., Kendall, A., Singer, R.H., and Engelke, D.R. (1998). Nucleolar localization of early tRNA processing. *Genes Dev.* *12*, 2463–2468.
- Boudonck, K., Dolan, L., and Shaw, P.J. (1999). The movement of coiled bodies visualized in living plant cells by the green fluorescent protein. *Mol. Biol. Cell* *10*, 2297–2307.
- Buhler, M., Mohn, F., Stalder, L., and Muhlemann, O. (2005). Transcriptional silencing of nonsense codon-containing immunoglobulin minigenes. *Mol. Cell* *18*, 307–317.
- Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* *13*, 2212–2217.
- Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* *16*, 2733–2742.
- Cioce, M., and Lamond, A.I. (2005). Cajal bodies: a long history of discovery. *Annu. Rev. Cell Dev. Biol.* *21*, 105–131.
- Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* *16*, 735–743.
- Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* *45*, 616–629.
- Filipowicz, W., and Pogacic, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.* *14*, 319–327.
- Gascioli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. (2005). Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr. Biol.* *15*, 1494–1500.
- Grewal, S.I., and Rice, J.C. (2004). Regulation of heterochromatin by histone methylation and small RNAs. *Curr. Opin. Cell Biol.* *16*, 230–238.
- Hannon, G.J. (2002). RNA interference. *Nature* *418*, 244–251.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science* *308*, 118–120.
- Hightett, M.I., Beven, A.F., and Shaw, P.J. (1993). Localization of 5 S genes and transcripts in *Pisum sativum* nuclei. *J. Cell Sci.* *105*, 1151–1158.
- Kanno, T., Mette, M.F., Kreil, D.P., Aufsatz, W., Matzke, M., and Matzke, A.J. (2004). Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr. Biol.* *14*, 801–805.
- Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* *37*, 761–765.
- Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* *109*, 145–148.
- Krogh, B.O., and Symington, L.S. (2004). Recombination proteins in yeast. *Annu. Rev. Genet.* *38*, 233–271.
- Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W.-L., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* *126*, this issue, 93–106.
- Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* *1*, E67.
- Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* *19*, 5194–5201.
- Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* *120*, 613–622.
- Pederson, T. (1998). The plurifunctional nucleolus. *Nucleic Acids Res.* *26*, 3871–3876.
- Pendle, A.F., Clark, G.P., Boon, R., Lewandowska, D., Lam, Y.W., Andersen, J., Mann, M., Lamond, A.I., Brown, J.W., and Shaw, P.J. (2005). Proteomic analysis of the *Arabidopsis* nucleolus suggests novel nucleolar functions. *Mol. Biol. Cell* *16*, 260–269.
- Pontes, O., Lawrence, R.J., Neves, N., Silva, M., Lee, J.H., Chen, Z.J., Viegas, W., and Pikaard, C.S. (2003). Natural variation in nucleolar dominance reveals the relationship between nucleolar organizer chromatin topology and rRNA gene transcription in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* *100*, 11418–11423.
- Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* *19*, 2030–2040.
- Rountree, M.R., and Selker, E.U. (1997). DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*. *Genes Dev.* *11*, 2383–2395.
- Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* *305*, 1434–1437.
- Sontheimer, E.J., and Carthew, R.W. (2004). Molecular biology. Argonaute journeys into the heart of RISC. *Science* *305*, 1409–1410.
- Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. *Genes Dev.* *19*, 517–529.
- Vaucheret, H. (2005). RNA polymerase IV and transcriptional silencing. *Nat. Genet.* *37*, 659–660.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303, 672–676.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297, 1833–1837.

Wassenegger, M. (2005). The role of the RNAi machinery in heterochromatin formation. *Cell* 122, 13–16.

Wassenegger, M., and Krczal, G. (2006). Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci.* 11, 142–151.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* 2, E104.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299, 716–719.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* 14, 1214–1220.

Supplemental Data

The *Arabidopsis* Chromatin-Modifying

Nuclear siRNA Pathway Involves

a Nucleolar RNA Processing Center

Olga Pontes, Carey Fei Li, Pedro Costa Nunes, Jeremy Haag, Thomas Ream, Alexa Vitins, Steven E. Jacobsen, and Craig S. Pikaard

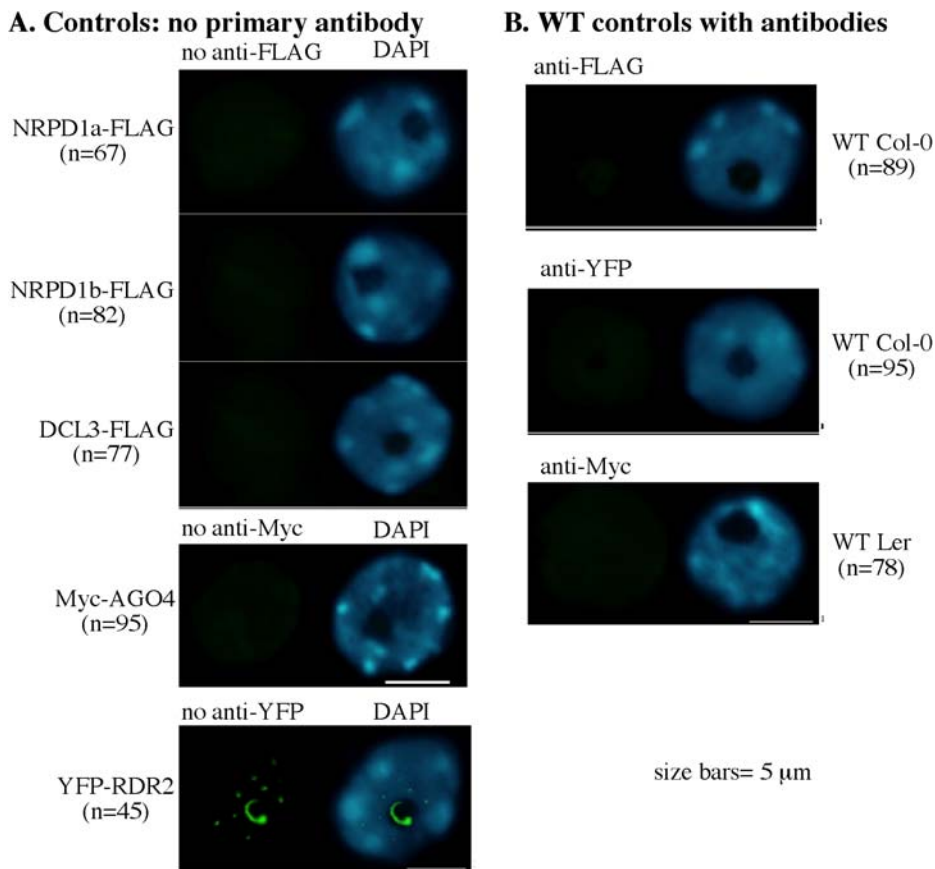


Figure S1. Antibody Specificity Controls

In part A of the figure, nuclei of transgenic lines expressing the indicated epitope tagged proteins were processed for protein immunolocalization as in Figure 3 of the paper except that the primary antibody was omitted prior to incubation with FITC-labeled secondary antibody (green). YFP fluorescence accounts for the YFP-RDR2 signal in the absence of anti-YFP antibody. In part B, non-transgenic, wild-type *A. thaliana* (ecotypes Col-0 or Ler) controls show that no signals are obtained upon immunolocalization using anti-FLAG, anti-YFP or anti-Myc primary antibodies. The images shown are representative of the nuclei observed, with the total number analyzed shown in parentheses. Nuclei were counterstained with DAPI (blue); the size bar corresponds to 5 μ m.

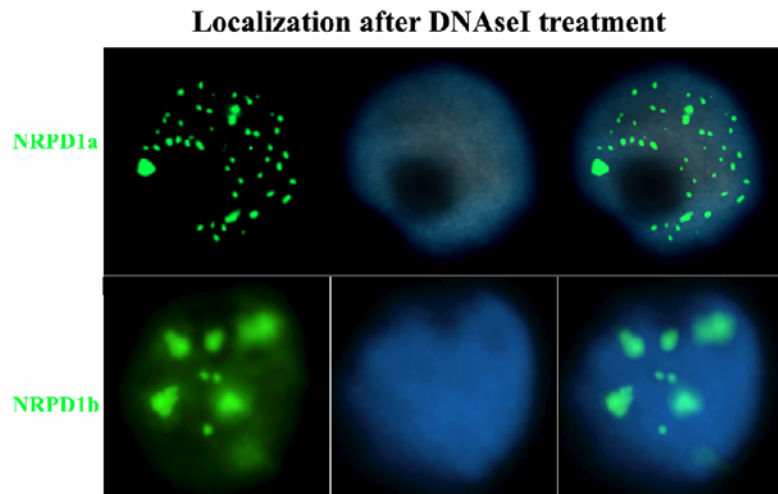


Figure S2. NRPD1a and NRPD1b Immunolocalization Signals Are Not Lost in DNase I-Treated Nuclei

Native NRPD1a and NRPD1b proteins were localized using anti-peptide antibodies in nuclei treated with DNase I as described in Figure 3B of the main paper.

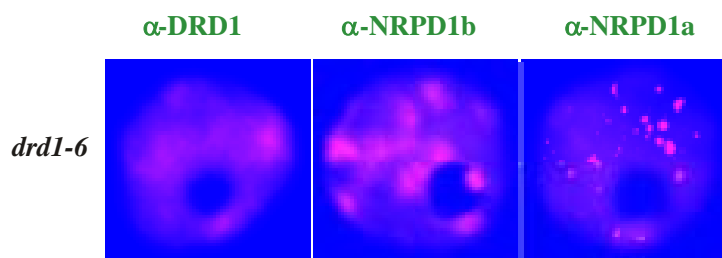


Figure S3. Immunolocalization of DRD1, NRPD1b, and NRPD1a in *drd1-6* Mutant Nuclei

Proteins were detected using anti-peptide antibodies. Note that DRD1 is not detected in the mutant, suggesting that the antibody specifically recognizes DRD1. The *drd1-6* mutation typically does not affect the NRPD1a pattern (85% yield the wild-type pattern for NRPD1a shown below; $n = 90$) but NRPD1b immunolocalization signals are typically more diffuse in *drd1-6* (79%; $n = 79$) than in wild-type, suggesting that DRD1 may act upstream, or at the same step, as NRPD1b.

Table S1. Supporting Data for Figure 2A: siRNA Probe Hybridization Patterns and Frequencies

RNA probe	Localization phenotypes	Frequency (%) of phenotypes observed upon nuclease treatment or in different genetic backgrounds									
		Col	Ler	+RNase A	+DNase I	<i>nrdp1a</i>	<i>nrdp2</i>	<i>nrdp1b</i>	<i>rdr2-1</i>	<i>dcl3-1</i>	<i>ago4-1</i>
45S siR	Nucleolar dot observed:	100	100	0	100	0	0	0	0	0	0
	Dispersed nuclear signal:	0	0	0	0	29	13	56	8	9	29
	No signal:	0	0	100	0	71	87	44	92	91	71
	# nuclei observed	n = 75	n = 71	n = 71	n = 63	n = 65	n = 141	n = 132	n = 62	n = 72	n = 76
5S siR	Nucleolar dot observed:	100	100	0	100	0	0	0	0	0	0
	Dispersed nuclear signal:	0	0	0	0	0	6	75	11	3	17
	No signal:	0	0	100	0	100	94	25	89	97	83
	# nuclei observed	n = 56	n = 48	n = 62	n = 68	n = 81	n = 127	n = 162	n = 85	n = 62	n = 74
AtSN1	Nucleolar dot + nucleoplasm:	74	No data	0	89	No data	0	No data	0	No data	No data
	Nucleoplasm only:	26		0	11		0		0		
	No signal:	0		100	0		100		100		
	# nuclei observed	n = 67		n = 79	n = 85		n = 150		n = 123		
AtCopia4	Nucleolar dot + nuclear spots:	100	No data	0	100	No data	0	No data	0	No data	No data
	No signal:	0		100	0		100		100		
	# nuclei observed	n = 85		n = 53	n = 68		n = 103		n = 91		
45S precursor	Diffuse nucleolar signals:	100	100	100	100	100	100	100	100	100	100
	# nuclei observed	n = 63	n = 57	n = 64	n = 51	n = 86	n = 79	n = 127	n = 72	n = 74	n = 81

The table is organized as in Figure 2A except that the table includes two columns of data for wild-type nuclei (ecotypes Col-0 and Ler) whereas Figure 2A showed only the Col-0 wild-type control.

Table S2. Supporting Data for Figure 3A: Protein Localization and Effects of RNase

	NRPD1a	NRPD2	NRPD1b	RDR2	DCL3	AGO4
protein localization	100% of nuclei display pattern shown <i>n</i> = 82	100% of nuclei display pattern shown <i>n</i> = 245	100% of nuclei show the nucleolar dot. 57% display numerous puncta external to nucleolus, as shown; 43% show <10 puncta <i>n</i> = 77	100% of nuclei display pattern shown <i>n</i> = 87	100% of nuclei display pattern shown <i>n</i> = 125	100% of nuclei display pattern shown <i>n</i> = 96
Effect of RNase A	91% , protein not detectable 9% , WT pattern <i>n</i> = 85	81% , protein not detectable 19% , WT pattern <i>n</i> = 94	65% , protein not detectable 35% , WT pattern <i>n</i> = 93	85% , protein not detectable 15% , WT pattern <i>n</i> = 62	59% , protein not detectable 41% , WT pattern <i>n</i> = 89	72% , protein not detectable 28% , WT pattern <i>n</i> = 61

Table S3. Supporting Data for Figure 3C: Pairwise Detection of Nuclear siRNA Pathway Proteins

Epitope-tagged lines	Antibodies				
	α -NRPD1a	α -NRPD2	α -NRPD1b	α -RDR2	α -DCL3
NRPD1a-FLAG		Majority of the nucleoplasmic signals colocalized <i>n</i> = 93			
NRPD1b-FLAG	Few nucleoplasmic signals colocalized <i>n</i> = 71	Few nucleoplasmic signals colocalized <i>n</i> = 85			
YFP-RDR2	Few nucleoplasmic signals colocalized <i>n</i> = 54	Few nucleoplasmic signals colocalized <i>n</i> = 48	Nucleolar dot + Few nucleoplasmic signals colocalized <i>n</i> = 67		
DCL3-FLAG	Few nucleoplasmic signals colocalized <i>n</i> = 76	Few nucleoplasmic signals colocalized <i>n</i> = 81	Nucleolar dot + Few nucleoplasmic signals colocalized <i>n</i> = 73	Nucleolar dot + Few nucleoplasmic signals colocalized <i>n</i> = 86	
Myc -AGO4	Not colocalized <i>n</i> = 54	Few nucleoplasmic signals colocalized <i>n</i> = 61	Nucleolar dot + Few nucleoplasmic signals colocalized <i>n</i> = 58	Nucleolar dot colocalized <i>n</i> = 45	Nucleolar dot colocalized <i>n</i> = 59

Table S4. Supporting Data for Figure 4: Protein-siRNA Colocalization

RNA probes	Epitope-tagged lines				
		NRPD1b-Flag	YFP-RDR2	DCL3-Flag	cMyc-AGO4
45S siR	Colocalized	81%	82%	79%	91%
	Not colocalized	19% <i>n</i> = 46	18% <i>n</i> = 60	21% <i>n</i> = 75	9% <i>n</i> = 65
siR1003	Colocalized	76%	58%	85%	76%
	Not colocalized	24% <i>n</i> = 57	42% <i>n</i> = 72	15% <i>n</i> = 79	24% <i>n</i> = 57
AtSN1	Colocalized	85%	61%	76%	83%
	Not colocalized	15% <i>n</i> = 74	39% <i>n</i> = 56	34% <i>n</i> = 45	17% <i>n</i> = 56
AtCopia4	Colocalized	82%	54%	78%	72%
	Not colocalized	18% <i>n</i> = 57	46% <i>n</i> = 59	22% <i>n</i> = 49	28% <i>n</i> = 67
45S prec	Colocalized	25%	43%	21%	30%
	Not colocalized	75% <i>n</i> = 81	57% <i>n</i> = 64	79% <i>n</i> = 61	70% <i>n</i> = 75

Colocalization was considered to be when >50% of the RNA probe signal overlapped >50 % of the protein signal.

Table S5. Supporting Data for Figure 5: Localization of Proteins Relative to NORs and 5S Gene Loci

DNA loci		NRPD1a	NRPD2a	NRPD1b	RDR2	DCL3	DRD1
NORs	Colocalized	85%	93%	92%	22%	12%	87%
	Not colocalized	15% <i>n</i> = 71	7% <i>n</i> = 83	8% <i>n</i> = 89	78% <i>n</i> = 55	88% <i>n</i> = 66	13% <i>n</i> = 57
5S gene clusters	Colocalized	68%	72%	81%	13%	27%	72%
	Not colocalized	32% <i>n</i> = 58	28% <i>n</i> = 62	19% <i>n</i> = 76	87% <i>n</i> = 51	73% <i>n</i> = 65	28% <i>n</i> = 61

Colocalization was considered to be when at least two NORs and at least four 5S gene *loci* overlapped half of the protein signals outside the nucleolus.

Table S6. Supporting Data for Figure 6: Protein Localization in Various Nuclear siRNA Pathway Mutants

		NRPD2	NRPD1a	NRPD1b	RDR2	DCL3
WT	Col	100% of nuclei display pattern shown <i>n</i> = 245	100% of nuclei display pattern shown <i>n</i> = 160	71% of nuclei display pattern shown <i>n</i> = 185	77% of nuclei display pattern shown <i>n</i> = 96	100% of nuclei display pattern shown <i>n</i> = 125
	<i>nrd1a</i>	Reduction in labeling intensity <i>n</i> = 181	No signal <i>n</i> = 123	WT pattern <i>n</i> = 87	Very faint to no signal <i>n</i> = 145	Very faint to no signal <i>n</i> = 61
Mutants	<i>nrd2</i>	Not detected <i>n</i> = 155	Reduction in labeling intensity <i>n</i> = 178	Very faint to no signal <i>n</i> = 134	Very faint to no signal <i>n</i> = 141	Very faint to no signal <i>n</i> = 104
	<i>nrd1b</i>	Very faint to no signal <i>n</i> = 138	WT pattern <i>n</i> = 67	No signal <i>n</i> = 149	Nucleolar dot is not detected <i>n</i> = 153	- Very strong reduction in labeling intensity (76%) - Mislocalization of the nucleolar dot to the nucleoplasm (24%) <i>n</i> = 84
	<i>nrd2, nrd1a</i>	Very faint to no signal <i>n</i> = 74	Very faint to no signal <i>n</i> = 81	Very faint to no signal <i>n</i> = 90	Very faint to no signal <i>n</i> = 67	Very faint to no signal <i>n</i> = 57
	<i>rdr2-1</i>	Small reduction in labeling intensity <i>n</i> = 121	WT pattern <i>n</i> = 112	- Nucleolar dot not detected (81%) - Reduction in labeling intensity (19%) <i>n</i> = 157	No signal <i>n</i> = 61	Very faint to no signal <i>n</i> = 87
	<i>dcl3-1</i>	Small reduction in labeling intensity <i>n</i> = 130	WT pattern <i>n</i> = 74	- Nucleolar dot not detected (78%) - Reduction in labeling intensity (22%) <i>n</i> = 72	WT pattern <i>n</i> = 89	No signal <i>n</i> = 91
	<i>ago4-1</i>	Small reduction in labeling intensity <i>n</i> = 109	WT pattern <i>n</i> = 65	- Nucleolar dot not detected (92%) - Reduction in labeling intensity (8%) <i>n</i> = 133	WT pattern <i>n</i> = 122	- WT pattern (67%) - Mislocalization of the nucleolar dot to the nucleoplasm (33%) <i>n</i> = 152

APPENDIX C

SUBUNIT COMPOSITIONS OF THE RNA-SILENCING ENZYMES POL IV AND
POL V REVEAL THEIR ORIGINS AS SPECIALIZED FORMS OF RNA
POLYMERASE II

Published in *Molecular Cell* (2009), 33 (2): 192-203.

My contributions to this work:

In this study, I provided FLAG-tagged transgenic lines for the study of *Arabidopsis thaliana* Pol I, II, III, IV and V complexes. These transgenic lines, along with the NRPD1 and NRPE1 antibodies that I raised, affinity purified, and validated, were critical for Tom Ream's confirmation and extension of the LC-MS/MS results. I performed some of the early *Arabidopsis thaliana* RNAP subunit predictions later used and expanded upon by Tom Ream and provided advice for the phylogenetic analysis. I also provided technical assistance for some of the experiments and offered comments in the editing phase of the paper.

Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II

Thomas S. Ream,¹ Jeremy R. Haag,¹ Andrzej T. Wierzbicki,¹ Carrie D. Nicora,² Angela D. Norbeck,² Jian-Kang Zhu,³ Gretchen Hagen,⁴ Thomas J. Guilfoyle,⁴ Ljiljana Paša-Tolić,² and Craig S. Pikaard^{1,*}

¹Biology Department, Washington University, St. Louis, MO 63130, USA

²Pacific Northwest National Laboratory, Richland, WA 99352, USA

³Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA 92521, USA

⁴Department of Biochemistry, University of Missouri, Columbia, MO 65211, USA

*Correspondence: pikaard@biology2.wustl.edu

DOI 10.1016/j.molcel.2008.12.015

SUMMARY

In addition to RNA polymerases I, II, and III, the essential RNA polymerases present in all eukaryotes, plants have two additional nuclear RNA polymerases, abbreviated as Pol IV and Pol V, that play nonredundant roles in siRNA-directed DNA methylation and gene silencing. We show that *Arabidopsis* Pol IV and Pol V are composed of subunits that are paralogous or identical to the 12 subunits of Pol II. Four subunits of Pol IV are distinct from their Pol II paralogs, six subunits of Pol V are distinct from their Pol II paralogs, and four subunits differ between Pol IV and Pol V. Importantly, the subunit differences occur in key positions relative to the template entry and RNA exit paths. Our findings support the hypothesis that Pol IV and Pol V are Pol II-like enzymes that evolved specialized roles in the production of noncoding transcripts for RNA silencing and genome defense.

INTRODUCTION

In bacteria and Archaea, a single multisubunit RNA polymerase transcribes genomic DNA into RNA. By contrast, eukaryotes have three essential nuclear DNA-dependent RNA polymerases that perform distinct functions. For instance, 45S ribosomal RNA (rRNA) genes are transcribed by RNA polymerase I (Pol I), mRNAs are transcribed by RNA polymerase II (Pol II), and tRNAs and 5S rRNA are transcribed by RNA polymerase III (Pol III) (Grummt, 2003; Schramm and Hernandez, 2002; Woychik and Hampsey, 2002).

Bacterial DNA-dependent RNA polymerase (RNAP) is composed of only four different proteins (β' , β , ω , α ; with two molecules of α in the core enzyme), but archaeal RNAP and eukaryotic Pol I, II, and III are more complex (Cramer et al., 2001; Darst et al., 1998; Hirata et al., 2008). Archaea have a fundamental subunit number of 10, with the caveat that the two largest subunits are generally split into two genes (Werner, 2007). Pol I,

II, and III have 12–17 subunits that include homologs of archaeal polymerase subunits, suggesting their functional diversification from an archaeal progenitor. The crystal structures of bacterial, archaeal, and eukaryotic Pol II are fundamentally similar (Cramer et al., 2001; Darst et al., 1998; Hirata et al., 2008). In each case, the largest and second-largest subunits, corresponding to the β' and β subunits of *E. coli* RNAP, respectively, are the catalytic subunits that interact to form the DNA entry and exit channels, the active site, and the RNA exit channel.

Sequencing of the *Arabidopsis thaliana* genome revealed genes for the expected catalytic subunits of Pol I, II, and III but unexpectedly revealed two atypical largest subunit genes and two atypical second-largest subunit genes (reviewed in Pikaard et al., 2008). Moreover, five subunits of Pol I, II, and III that are typically encoded by single genes in yeast and mammals, namely *RPB5*, *RPB6*, *RPB8*, *RPB10*, and *RPB12* (named according to their discovery as Pol II subunits; aka RNA Polymerase B) (Cramer, 2002; Werner, 2007), are encoded by multi-gene families in *Arabidopsis*, as are the Pol II-specific subunits *RPB3*, *RPB4*, *RPB7*, and *RPB9*. The functional significance of the extensive subunit diversity in plants is unclear.

The genes encoding the atypical largest and second-largest polymerase subunits in *Arabidopsis* are not essential for viability (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005), unlike their Pol I, II, or III counterparts (Onodera et al., 2008). However, the atypical catalytic subunits are nuclear proteins (Onodera et al., 2005; Pontes et al., 2006) required for siRNA-directed DNA methylation and silencing of retrotransposons, endogenous repeats, and transgenes (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). The atypical catalytic subunit genes also play roles in the short-range or long-distance spread of RNA-silencing signals, responses to biotic and abiotic stresses, and the control of flowering time (Borsani et al., 2005; Brosnan et al., 2007; Dunoyer et al., 2007; Katiyar-Agarwal et al., 2007; Pontier et al., 2005; Smith et al., 2007). The atypical largest subunit genes are *NRPD1* and *NRPE1*. *NRPD1* (formerly *NRPD1a*) is the largest subunit of Nuclear RNA polymerase IV (Pol IV; formerly Pol IVa) (Herr et al., 2005; Onodera et al., 2005), whereas *NRPE1* (formerly *NRPD1b*) is the largest subunit of Pol V (formerly Pol IVb) (Kanno et al., 2005; Pontier

et al., 2005). The second-largest subunits of Pol IV and Pol V are encoded by the same gene, designated by the synonymous names *NRPD2a* (*NRPD2* for simplicity) or *NRPE2* (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). Pol IV and Pol V are functionally distinct, with Pol IV required for siRNA production and Pol V generating noncoding transcripts at target loci (Wierzbicki et al., 2008). Our current model is that siRNAs bind to Pol V nascent transcripts to bring the silencing machinery to the vicinity of the chromatin at target loci (Wierzbicki et al., 2008).

Aside from their largest and second-largest subunits, the subunit compositions of Pol IV and Pol V are unknown. Here, we show that Pol IV and Pol V have subunit compositions characteristic of Pol II but make differential use of RPB3, RPB4, RPB5, and RPB7 family variants in addition to having distinct catalytic subunits. Collectively, our results support the hypothesis that Pol IV and Pol V are RNA Pol II derivatives whose molecular niche is the production of noncoding transcripts for RNA-mediated silencing.

RESULTS

Identification of Pol IV, V, and II Subunits Using LC-MS/MS

To affinity purify Pol IV and Pol V from *Arabidopsis thaliana*, we engineered full-length *NRPD1* (*NRPD1a*) and *NRPE1* (*NRPD1b*) genomic clones, including their promoter regions and complete sets of introns and exons, adding a FLAG epitope tag to the protein's C terminus. The transgenes rescue the loss of RNA-directed DNA methylation in their respective null mutants (*nrdp1a-3* or *nrdp1b-11*), indicating that the recombinant proteins are functional (Pontes et al., 2006). *NRPD1*-FLAG and *NRPE1*-FLAG, and their respective associated subunits, were affinity purified on anti-FLAG resin, and tryptic peptides were identified by using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). For both Pol IV and Pol V, their two known catalytic subunits were detected, as expected. However, in each case, ten additional previously unknown subunits were identified, corresponding to the ten noncatalytic subunits of yeast RNA Pol II: RPB3, RPB4, RPB5, RPB6, RPB7, RPB8, RPB9, RPB10, RPB11, and RPB12 (Figure 1; see Table S1 and Figures S1 and S2, available online). The pairs of catalytic subunits specific to RNA Pol I, II, or III were not detected in Pol IV or Pol V samples, ruling out copurification of these polymerases as an explanation for the noncatalytic subunits detected in affinity-purified Pol IV or Pol V. Likewise, coimmunoprecipitation (coIP) data show that Pol IV and Pol V do not associate with each other or with Pol I, II, or III (Figure 2A).

For Pol V, peptide sequence data typically allowed unambiguous identification of subunits that are members of protein families (see Figure S1 for peptide coverage maps and Figures S4–S12 for family alignments). An exception was the RPB8 family, for which the sole peptide identified matched both variants, which are 96% identical. Two RPB3-related variants that are 88% identical are present in *Arabidopsis*, and both proteins are detected in Pol V, resulting in their designation as NRPE3a and NRPE3b (Figure 1, Figure 3A). The single RPB11 subunit encoded by the *Arabidopsis* genome was also detected; hence we

refer to this protein as NRPE11 (Figure 1). Of six homologs of RPB5 in the genome, only one (NRPE5) is detected in Pol V (Figure 1, Figure S5). Two RPB9-like subunits were identified in Pol V (Figures 1 and 2D). These proteins, designated NRPE9a and NRPE9b, are 92% identical. There are four RPB7 homologs in *Arabidopsis*, only one of which is detected in Pol V, NRPE7. One of two RPB4-like subunits (NRPE4), one of two RPB10-like subunits (NRPE10), one of two RPB12-like subunits (NRPE12), and one of two RPB6-like subunits (NRPE6a) were also detected in Pol V (Figure 1).

Analysis of Pol IV's subunit composition revealed similarities and differences compared to Pol V (Figure 1, Figure S2). As with Pol V, peptides for the single RPB11-like subunit were identified. In the context of Pol IV, we refer to this protein as NRPD11; in the context of Pol V, we refer to this same protein as NRPE11. Similar nomenclature rules were adopted for other subunits shared by more than one polymerase (see Figure 1 for synonyms). *NRPD4*, *NRPD6a*, *NRPD8b*, and *NRPD10* subunits were unambiguously identified (Figure 1). Similar to Pol V, both RPB3-like variants were detected in Pol IV, but one is predominant (*NRPD3*; see Figure 1). Interestingly, the RPB5-like subunit of Pol IV, *NRPD5*, is identical to the previously identified *NRPB5* subunit of Pol II but differs from the *NRPE5* subunit of Pol V (Figure 1) (Larkin et al., 1999). The major *NRPD7* subunit detected in Pol IV is 62% identical to the Pol V *NRPE7* subunit, but low-level peptide sequence coverage for the *NRPE7* subunit was detected as well. The Pol IV *NRPD9b* subunit corresponds to *NRPE9b* detected in Pol V (Figures 1 and 2D).

The significant number of Pol II-like subunits in Pol IV and Pol V raised questions concerning the relative similarities of Pol II, Pol IV, and Pol V. Therefore, we affinity purified *Arabidopsis* Pol II by exploiting epitope-tagged *NRPB2* (*NRPB2-FLAG*) expressed from a transgene that rescues the *nrbp2-1* null mutant (Onodera et al., 2008). LC/MS-MS revealed 12 subunits orthologous to their 12 yeast Pol II counterparts, with no contaminating subunits specific to Pol I, III, IV, or V (Figure 1, Figure S3). The same RPB10, RPB11, and RPB12 family subunits found in Pol IV and/or Pol V are present in Pol II (Figure 1). Sequenced peptide coverage for the RPB6, RPB8, and RPB9-like subunits in the Pol II dataset revealed that each of the two genes for these subunits encodes a subunit incorporated into Pol II (Figure S3), suggesting that the genes are redundant. A single RPB3-like subunit, *NRPB3*, is predominant in Pol II, consistent with a previous report (Ulmasov et al., 1996). However, peptides corresponding to the *NRPE3b* subunit were also detected at low frequency. The single RPB5 subunit identified in Pol II corresponds to the expected subunit based on a previous study (Larkin et al., 1999) and is identical to the *NRPD5* subunit of Pol IV but distinct from the *NRPE5* subunit of Pol V. Pol II also makes use of RPB4 and RPB7 variants that are distinct from the corresponding Pol IV and Pol V subunits. These *NRPB4* and *NRPB7* subunits correspond to subunits previously shown to associate with Pol II (Larkin and Guilfoyle, 1998).

Immunological Confirmation of Subunit Associations

To test subunit associations with all five nuclear RNA polymerases, we exploited *Arabidopsis* lines expressing FLAG-tagged

Function	Bacteria	Archaea	Sc Pol II	At Homologs	At Pol II	At Pol IV	At Pol V	Names/Synonyms	
Catalytic	β'	RPOA' RPOA''	RPB1	At4g35800	59			NRPB1	
				At1g63020		58		NRPD1	
				At2g40030			74	NRPE1	
	β	RPOB' RPOB''	RPB2	At4g21710	63			NRPB2	
				At3g23780		18	37	NRPD2/NRPE2	
Assembly	α	RPOD	RPB3	At2g15430	57	28	45	NRPB3/NRPD3/NRPE3a	
				At2g15400	4	4	41	NRPE3b	
	α	RPOL	RPB11	At3g52090	75	56	68	NRPB11/NRPD11/NRPE11	
	α	RPON	RPB10	At1g11475	55	54	55	NRPB10/NRPD10/NRPE10	
				At1g61700				NRPB10-like	
	α	RPOP	RPB12	At5g41010	16	16	16	NRPB12/NRPD12/NRPE12	
				At1g53690				NRPB12-like	
	Auxiliary	ω	RPOK	RPB6	At5g51940	15	15	15	NRPB6a/NRPD6a/NRPE6a
					At2g04630	15	*	*	NRPB6b/NRPE6b
RPOG			RPB8	At1g54250	30	*	*	NRPB8a/NRPE8a	
				At3g59600	30	18	*	NRPB8b/NRPD8b/NRPE8b	
RPOH			RPB5	At3g22320	63	15		NRPB5/NRPD5	
				At3g57080			39	NRPE5	
				At5g57980				NRPB5-like	
				At2g41340				NRPE5-like	
				At3g54490				NRPE5-like	
RPOF			RPB4	At5g09920	61			NRPB4	
				At4g15950		13	8	NRPD4/NRPE4	
RPOE			RPB7	At5g59180	51			NRPB7	
				At4g14660		9	33	NRPE7	
				At3g22900		52		NRPD7	
				At4g14520				NRPB7-like	
TFS/RPOX			RPB9	At3g16980	22		22	NRPB9a/NRPE9a	
				At4g16265	28	22	22	NRPB9b/NRPD9b/NRPE9b	

Figure 1. Relationships of Arabidopsis Pol II, IV, and V Subunits to E. coli, Archaeal, and Yeast RNA Pol II Subunits

Numbers indicate percent protein coverage represented by peptides unique to that protein. "*" indicates that all peptides match both closely related proteins. Unshaded numbers represent alternate subunits detected at trace levels relative to the predominant subunit.

Pol I, II, and III second-largest subunits (NRPA2-FLAG, NRPB2-FLAG, or NRPC2-FLAG) or FLAG-tagged Pol IV and Pol V largest subunits (NRPD1-FLAG, NRPE1-FLAG), each expressed from transgenes that rescue corresponding null mutants (Onodera et al., 2008; Pontes et al., 2006). Plants expressing FLAG-tagged genomic clones of NRPE6a, NRPE8b, NRPE10, or NRPE11 or an NRPE5 cDNA were also engineered. Each recombinant protein could be immunoprecipitated from transgenic plants and detected by immunoblotting using anti-FLAG antibody (Figure 2A). Probing immunoblots with antibodies for NRPE1 and NRPE2 (Onodera et al., 2005) revealed that these Pol V catalytic subunits are present in NRPE1, NRPE6a, NRPE8b, NRPE10, NRPE11, and NRPE5 immunoprecipitates (Figure 2A; see also the anti-NRPE1 specificity control in Figure 2B), consistent with the detection of all of these subunits in Pol V (Figure 1). Controls show that NRPE2 and NRPE1 do not coimmunoprecipitate with Pol I, II, or III; that NRPE1 does not coimmunoprecipitate with Pol IV; and that NRPE2/NRPD2 is present in Pol IV and Pol V, as expected. The anti-NRPE1 antibody consistently reveals multiple NRPE1 isoforms (Figures 2A and 2B); whether

these are degradation, posttranslational modification, or alternative splicing products is unclear.

To test whether NRPE5, NRPE6a, NRPE8b, NRPE10a, and NRPE11 subunits are shared by Pol I, II, and/or III, we used an anti-peptide antibody recognizing an invariant sequence in the Pol I, II, and III second-largest subunits (Onodera et al., 2005); this antibody fails to crossreact with NRPE2/NRPD2 due to a single amino acid substitution. In NRPE6a, NRPE8b, NRPE10, and NRPE11 immunoprecipitated fractions, Pol I, II, or III second-largest subunits are detected, consistent with the LC-MS/MS analysis of Pol II (Figures 1 and 2A). In yeast, RPB6, RPB8, and RPB10 are common to Pol I, II, and III, but RPB11 is Pol II specific. Second-largest subunits of Pol I, II, or III do not coimmunoprecipitate with FLAG-NRPE5, showing that NRPE5 is not a subunit of the essential polymerases (Figure 2A).

The LC-MS/MS data indicate that either of the two RPB8 homologs associate with Pol V. CoIP analysis confirms that NRPE8a or NRPE8b will coimmunoprecipitate with the Pol V catalytic subunits (Figures 2A and 2E). Although LC-MS/MS identified only one RPB6 variant (NRPE6a), its paralog (NRPE6b)

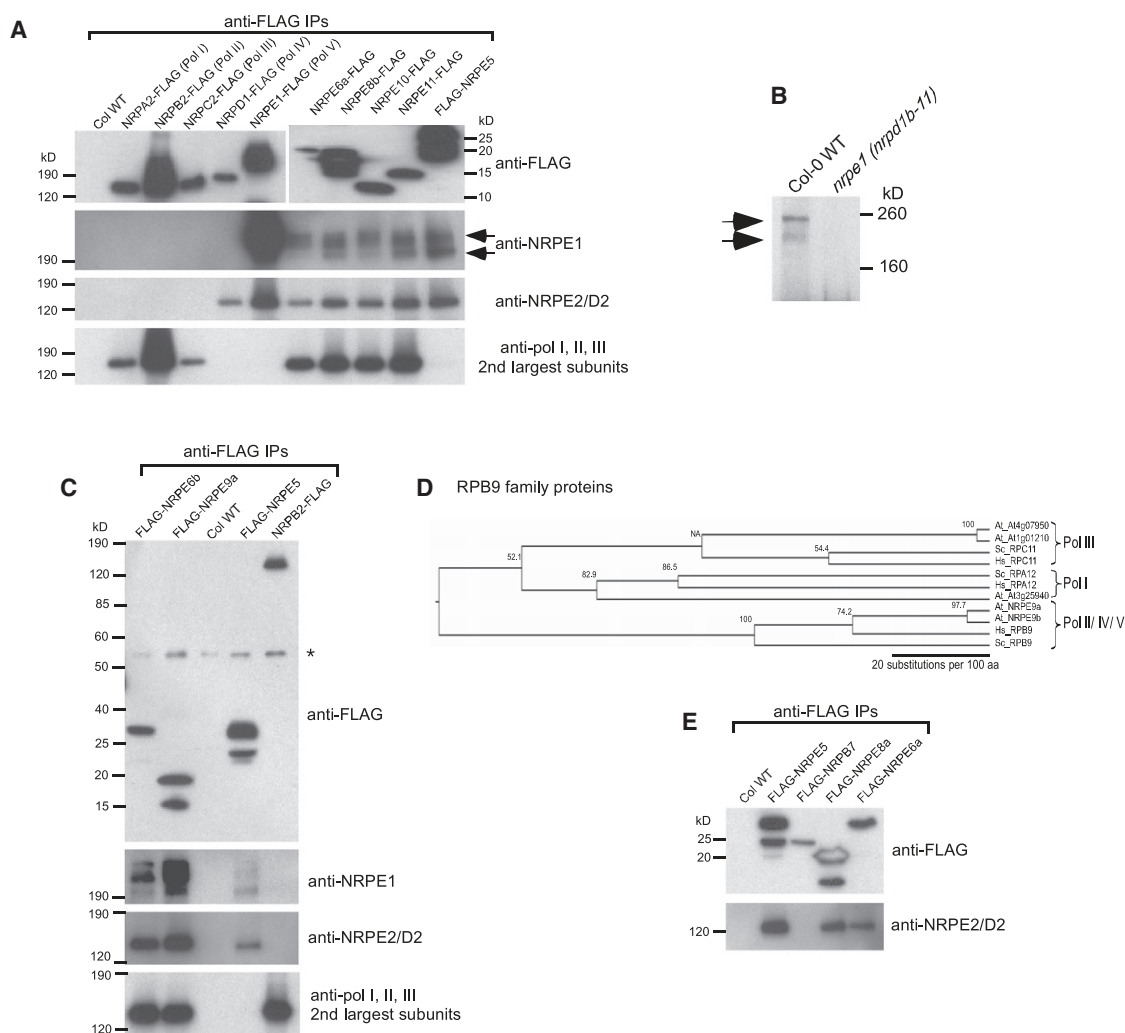


Figure 2. Verification of Pol V Subunit Associations

(A) Pol V includes subunits shared with other polymerases as well as a unique RPB5 family variant. Pol I, II, III, IV, and V were immunoprecipitated by virtue of FLAG-tagged catalytic subunits alongside NRPE6a, NRPE8b, NRPE10, NRPE11, and NRPE5 FLAG-tagged subunits. Duplicate immunoblots were probed with anti-FLAG, anti-NRPE1, anti-NRPE2/NRPD2 (abbreviated anti-NRPE2/D2), or an antibody recognizing the second-largest subunits of Pol I, II, or III. The two panels in the top row are from the same blot but focus on different size ranges.

(B) Control immunoblot showing that the multiple high-molecular-mass bands characteristic of NRPE1 are lost in an *nrpe1* null mutant (allele *nrpd1b-11*), indicating that the antibody is specific for NRPE1.

(C) NRPE6b and NRPE9a are subunits of Pol V as well as Pol I, II, or III. Immunoprecipitation and immunoblot detection was as in (A). NRPE5 and NRPB2 immunoprecipitations serve as controls for Pol V and Pol II, respectively. "*" denotes a nonspecific band detected by the anti-FLAG antibody.

(D) Phylogenetic tree based on a CLUSTALW alignment of *Arabidopsis* RPB9-like proteins with the RPB9 (Pol II), RPC11 (Pol III), and RPA12 (Pol I) subunit equivalents of yeast.

(E) NRPE8a and NRPE6a associate with Pol V. Immunoprecipitation and immunoblot detection was as in (A). NRPE5 and NRPB7 serve as controls for Pol V and Pol II, respectively.

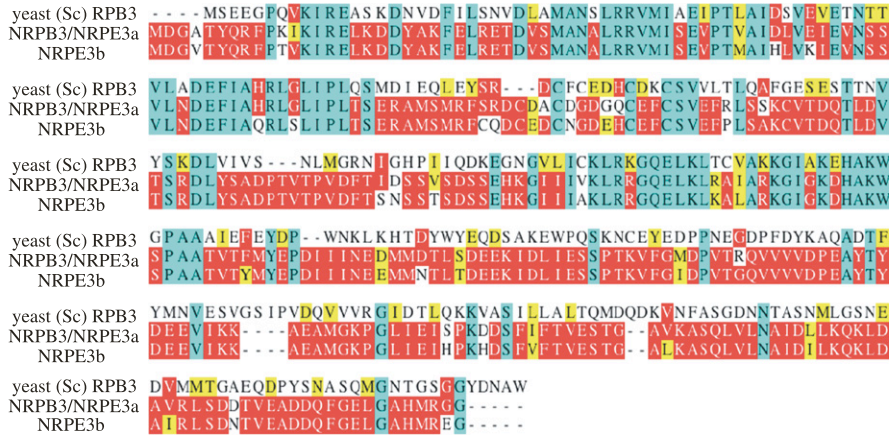
can also associate with Pol V in vivo (Figure 2C). Both Pol II clade RPB9-like subunits (Figure 2D) were detected in Pol V by LC-MS/MS. CoIP analysis confirms that FLAG-NRPE9a associates with the Pol V NRPE1 and NRPE2 catalytic subunits in vivo (Figures 2C and 2D). NRPE6b and NRPE9a also coimmunoprecipitate the second-largest subunits of Pol I, II, or III (Figure 2C).

LC-MS/MS analysis of Pol V identified both potential RPB3 variants (Figure 3A). In confirmation of this result, HA-tagged NRPE3a and NRPE3b both coimmunoprecipitate the Pol V cata-

lytic subunits (Figure 3B). NRPE3a, but not NRPE3b, also coimmunoprecipitates a subunit recognized by the antibody specific for Pol I, II, or III second subunits (Figure 3B); we deduce this to be the Pol II NRPB2 subunit because Pol I and Pol III use third-largest subunits distinct from RPB3. Moreover, the gene encoding NRPE3a was previously shown to encode a NRPB3 (see Figure 1) subunit present in purified Pol II (Ulmasov et al., 1996).

NRPE11, NRPE6a, NRPE8b, NRPE10, and NRPE9a all coimmunoprecipitate with the Pol IV and Pol II largest subunits

A Alignment of the two *A. thaliana* RPB3 family proteins with yeast RPB3



B anti-HA IPs:

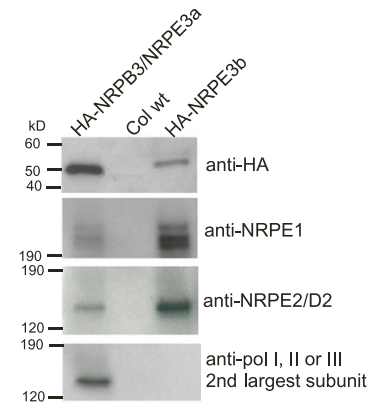


Figure 3. Pol V Utilizes a Distinct RPB3 Variant, NRPE3b, as well as an NRPE3a Variant Corresponding to the Pol II NRPB3 Subunit

(A) Alignment of the two *Arabidopsis* RPB3 family proteins with yeast RPB3.

(B) HA-tagged NRPE3a/NRPE3b and NRPE3b were immunoprecipitated and resulting immunoblots were probed using the indicated antibodies.

(Figures 1 and 4A). Upon immunoprecipitation of NRPE3b, no Pol II is detected in the immunoprecipitated fraction using an antibody recognizing the C-terminal domain (CTD) of the largest subunit. Likewise, Pol IV is detected in only trace amounts using the anti-NRPD1 antibody. We conclude that NRPE3b is used almost exclusively by Pol V (Figures 1 and 4A). In contrast, NRPB3, NRPD3, and NRPE3a are encoded by the same gene. Controls show that the NRPD1 subunit of Pol IV does not coimmunoprecipitate with Pol I, II, III, or V (Figure 4A). Likewise, the NRPB1 subunit of Pol II does not coimmunoprecipitate with Pol I, III, IV, or V (Figure 4A).

Using antibodies specific for NRPB5/NRPD5 or NRPE5 (Larkin et al., 1999), we tested their associations with FLAG-tagged Pol I, II, III, IV, or V (Figures 4B and 4C). Controls show that the NRPD2/NRPE2 subunit common to both Pol IV and Pol V is detected in NRPD1 and NRPE1 IPs, as expected, but not in Pol I, II, or III IPs (Figures 4B and 4C). NRPE5 was detected only in the NRPE1-FLAG immunoprecipitated fraction (Figure 4B), confirming that this subunit is unique to Pol V. By contrast, the NRPB5/NRPD5 subunit is detected in Pol I, II, III, and IV fractions, but not in Pol V (Figure 4C), in agreement with the LC-MS/MS data and previous studies showing that NRPB5/NRPD5 copurifies with Pol I, II, and III (Larkin et al., 1999) (Saez-Vasquez and Pikaard, 1997).

We affinity purified FLAG-tagged NRPE5 expressed in the *nrpe5* mutant background and identified the associated RNA polymerase subunits using LC-MS/MS. The results confirmed association of NRPE5 with all Pol V subunits except NRPE7 (Table S2, Figure S18), which most likely escaped detection in this experiment due to insufficient sample mass.

Collectively, the immunological tests of Figures 2–4 confirm the Pol V association of the NRPE1, NRPE2, NRPE3a, NRPE3b, NRPE5, NRPE6a, NRPE8, NRPE9a, NRPE10, and NRPE11 subunits detected by LC-MS/MS. Likewise, the immunological tests confirm the Pol IV associations of NRPD1, NRPD2, NRPD3, NRPD5, NRPD6a, NRPD8b, NRPD9a, NRPD10, and

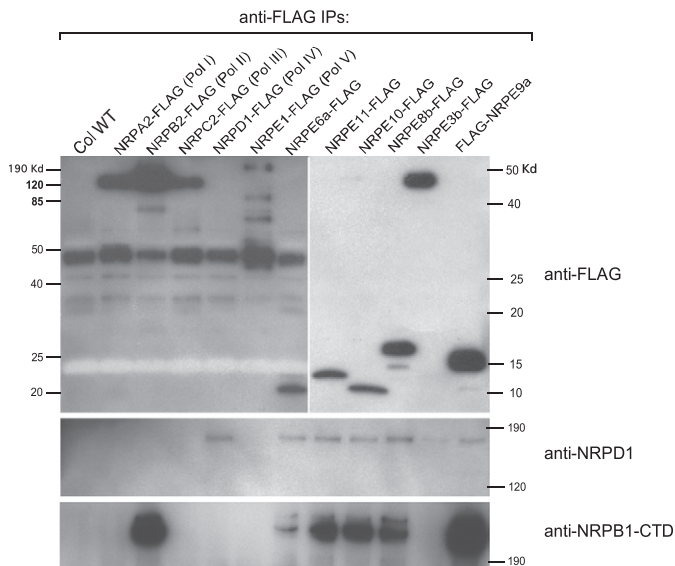
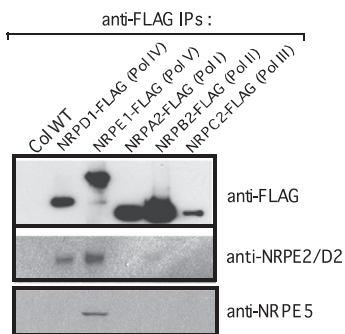
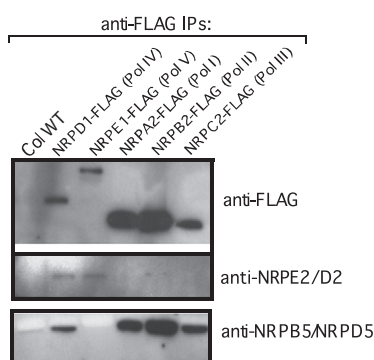
NRPD11. Pol IV and Pol V subunits that are shared with Pol II were also confirmed immunologically.

NRPE5 Is Required for DNA Methylation, siRNA Accumulation, and Gene Silencing at Pol V-Regulated Loci

Of the five full-length homologs of yeast *RPB5* in *Arabidopsis*, RT-PCR analysis shows that only *NRPB5/NRPD5* and *NRPE5* are constitutively expressed; other family members show organ-specific expression patterns (Figure 5A, Figures S5 and S13). Homozygous *nrpe5-1* mutants resulting from a T-DNA insertion (Figure 5B) are viable, as are Pol V *nrpe1* and *nrpe2* mutants. In contrast, homozygous *nrpd5-1/nrpb5-1* T-DNA insertion mutants were not recoverable due to female gametophyte lethality, as shown by reciprocal genetic crosses (Figures S14A and S14B). Female gametophyte lethality is a characteristic of Pol I, II, and III mutants, as demonstrated previously for *nrpa2*, *nrpb2*, *nrpc2*, and *nrpb12* (Onodera et al., 2008). A homozygous *nrpe11* T-DNA insertion mutant was also unrecoverable, consistent with this gene also encoding the Pol II subunit, NRPB11 (Figures S14A and S14B).

Like Pol IV and Pol V catalytic subunit mutants, *nrpe5-1* mutants lack obvious morphological phenotypes but flower later than wild-type plants under short-day conditions (Figure 5C), similar to mutants disrupting the 24 nt siRNA-directed DNA methylation pathway, including *RNA-DEPENDENT RNA POLYMERASE 2 (RDR2)* and *DICER-LIKE 3 (DCL3)* mutants (Chan et al., 2004; Liu et al., 2007; Pontier et al., 2005). Comparison of *nrpe5* and wild-type individuals suggests that the delay in flowering is stochastic, with some individuals showing substantial delays and others flowering at the same time as wild-type plants (Figure S15).

We tested *nrpe5-1* mutants for Pol V-dependent molecular phenotypes, including DNA hypermethylation at 5S rRNA gene clusters and at *AtSN1* and *AtSN2* retroelements. In *nrpd1* (*nrpd1a-3*), *nrpe1* (*nrpd1b-11*), and *nrpd2/nrpe2* mutants, loss

A Test of coIP with NRPD1 (Pol IV) and NRPB1(Pol II)**B Test of NRPE5 coIP with Pools I to V****C Test of NRPB5/NRPD5 coIP with Pools I to V**

of methylation at 5S rDNA repeats results in increased digestion by the methylation-sensitive restriction endonucleases *HpaII* and *HaeIII* compared to wild-type plants (Figure 5D). In the *nrpe5* mutant, methylation at 5S rDNA genes is reduced compared to wild-type, but to a lesser extent than in *nrpe1* or *nrpd2/nrpe2* mutants (Figure 5D). Transformation of the *nrpe5-1* mutant with a 35S:*FLAG-NRPE5* transgene restores methylation to wild-type levels, as shown in three independent transgenic lines (Figure 5D).

To test whether *nrpe5* affects DNA methylation at other Pol V-dependent loci, we examined the SINE retrotransposon families, *AtSN1* and *AtSN2* (Myouga et al., 2001). In wild-type plants, *AtSN1* and *AtSN2* elements are heavily methylated such that their DNA is not cut by *HaeIII* and a PCR product can be obtained (Figures 5E and 5F). In *nrpe1* and *nrpe2/nrpd2* mutants, however, methylation is lost such that *HaeIII* cuts and PCR amplification fails (Figures 5E and 5F). In *nrpe5-1*, decreased *AtSN1* and *AtSN2* methylation occurs, but not as severely as in *nrpe1* or *nrpe2/nrpd2* mutants. Nonetheless, the decreased methylation in *nrpe5-1* plants is rescued by a 35S:*FLAG-NRPE5* transgene (Figures 5E and 5F).

Figure 4. CoIP Tests of Pol V, IV, and II Subunit Associations

(A) Pol I, II, III, IV, and V were immunoprecipitated by virtue of FLAG-tagged catalytic subunits alongside immunoprecipitated NRPE6a, NRPE8b, NRPE9a, NRPE10, NRPE11, and NRPE3b FLAG-tagged subunits. Duplicate immunoblots were probed with anti-FLAG, anti-NRPD1 (Pol IV), or anti-NRPB1-CTD (Pol II). The two panels in the top row show different exposures of the same blot, focused on different size ranges.

(B) Pol I, II, III, IV, and V were immunoprecipitated using the indicated FLAG-tagged subunits and probed with anti-FLAG, anti-NRPE5, or anti-NRPE2/NRPD2.

(C) Immunoprecipitation and immunoblotting using the indicated antibodies were as in (B).

RNA-directed DNA methylation silences *AtSN1* retroelements in wild-type plants such that loss of methylation correlates with increased *AtSN1* transcription (Hamilton et al., 2002; Herr et al., 2005; Kanno et al., 2005). *AtSN1* transcripts are barely detectable in wild-type plants but are abundant in *nrpe5* mutants, as in *nrpe1* or *nrpe2/nrpd2* mutants (Figure 5G). In the *nrpe5-1* genetic background, the 35S:*FLAG-NRPE5* transgene restores *AtSN1* silencing (Figure 5G). Collectively, these results demonstrate that *NRPE5* is important for DNA methylation and silencing of *AtSN1* elements.

In the RNA-directed DNA methylation pathway, Pol IV is required for 24 nt siRNA production (Herr et al., 2005; Onodera et al., 2005) such that siRNAs are eliminated in *nrpd1* and *nrpd2* mutants (Figure 5H). In contrast, siRNAs in *nrpe1* mutants are reduced but not eliminated at 5S rDNA genes and *COPIA* elements (Figure 5H). Consistent with a Pol V mutant phenotype, siRNAs are reduced in *nrpe5* mutants relative to wild-type and are restored by the 35S:*FLAG-NRPE5* transgene (Figure 5H). MicroRNA and *trans*-acting siRNA levels are unaffected in *nrpe5*, *nrpd1*, or *nrpe1* mutants, consistent with the lack of Pol IV or Pol V involvement in these pathways.

Crystallographic studies indicate that yeast RPB5 is composed of an N-terminal jaw domain and a C-terminal assembly domain separated by a short linker (Figures S5, S16, and S17A). These domains appear to be conserved in nearly all plant RPB5 homologs (Figure S16). A feature of *Arabidopsis* NRPE5, and its presumptive orthologs in other plants, is a short N-terminal extension compared to NRPB5 (Figure S16 and S17A). To test the functional significance of this N-terminal extension, we created a 35S:*FLAG-ΔN-NRPE5* construct in which the extension was deleted (Figure S17A). This transgene fails to rescue *nrpe5-1* mutant phenotypes (Figures S17B–S17D). Surprisingly, immunoprecipitation of equal volumes of soluble extracts revealed that the FLAG-ΔN-NRPE5 protein

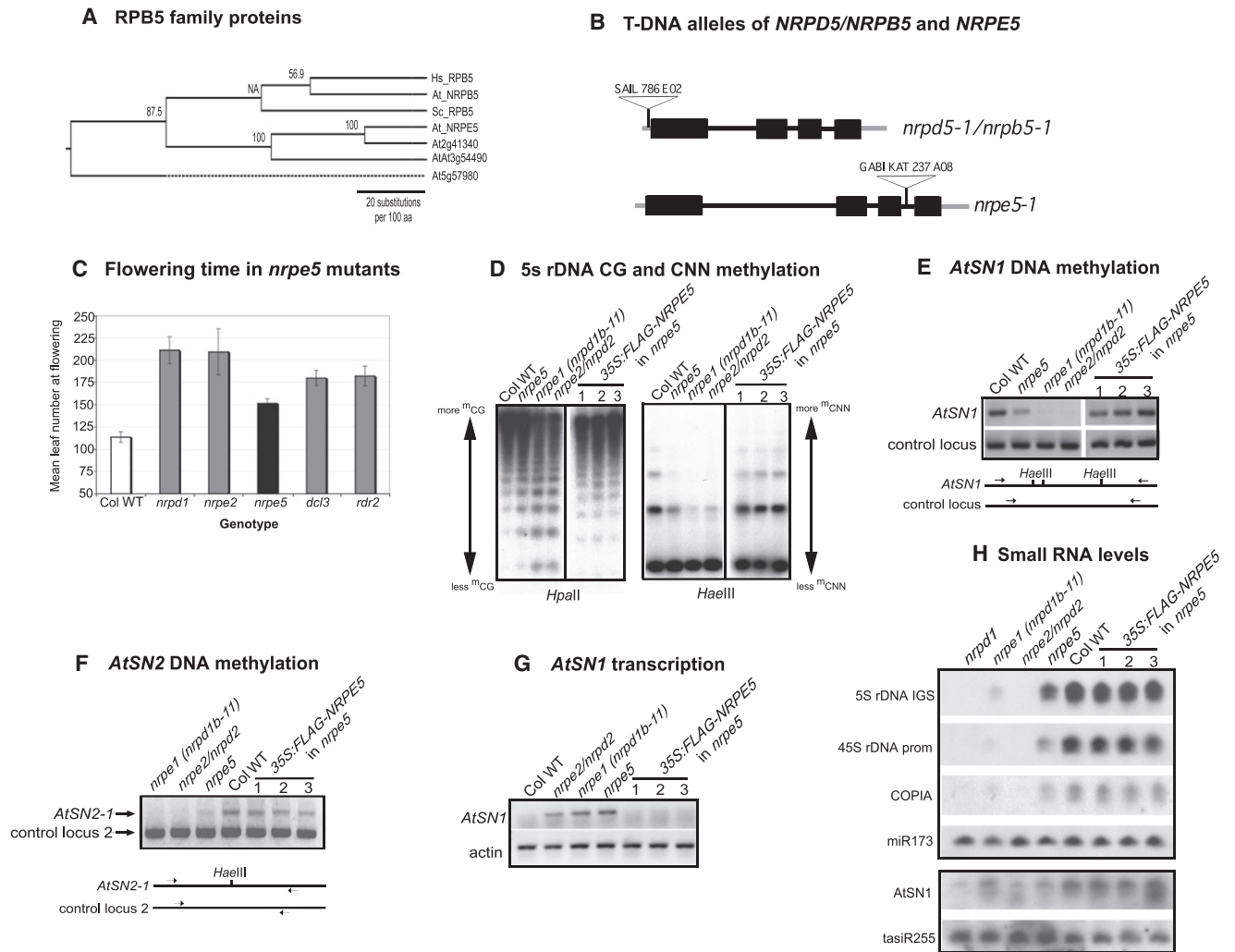


Figure 5. *nrpe5* Mutants Are Defective in RNA-Directed DNA Methylation and Retrotransposon Silencing

(A) Phylogenetic tree based on a CLUSTALW alignment of the five full-length RPB5-like proteins in *Arabidopsis* with the RPB5 subunits of yeast and human.

(B) Locations of T-DNA insertions in the *nrpd5-1/nrpb5-1* and *nrpe5-1* alleles. Black boxes represent exons, black bars represent introns, and gray bars represent 5' and 3'UTRs.

(C) *nrpe5-1* homozygous mutant plants display a delay in flowering under short-day conditions (8 hr light, 16 hr dark). The mean (\pm SEM) number of rosette leaves when the floral bolt reached 10 cm is graphed. All mutants are significantly different from wild-type based on a Student's *t* test ($p < 0.05$).

(D) Methylation-sensitive Southern blot analyses of wild-type, *nrpe1*, *nrpe2/nrpd2*, and *nrpe5* mutants and three different *nrpe5*, 35S:FLAG-NRPE5 transgenic lines. Genomic DNA was digested with either *Hpa*II (left, reports on ^mCG) or *Hae*III (right, reports on ^mCNN) and probed for 5S rDNA repeats. Images for the *Hpa*II or *Hae*III digests are from the same exposures of the same Southern blots; the black vertical lines separate groups of lanes whose order was rearranged for clarity of presentation.

(E and F) PCR-based methylation assay of *AtSN1* and *AtSN2* family retroelements. Genomic DNA was digested with *Hae*III and subjected to PCR using *AtSN1*, *AtSN2-1*, or control primers that amplify sequences lacking *Hae*III sites (*At2g19920* in the case of [E]), and an *AtSN2* family element lacking *Hae*III sites in the case of [F]). Diagrams show the relative positions of the primers flanking the *Hae*III sites.

(G) RT-PCR detection of *AtSN1* and actin transcripts.

(H) Small RNA blot analysis. Blots were probed for siRNAs corresponding to 45S or 5S rDNA genes, Copia or *AtSN1* transposons, and miRNA 173 or *trans*-acting siRNA 255.

is present at very low levels relative to full-length FLAG-NRPE5, despite similar transcript levels (Figure S17E). These data suggest that the N-terminal extension is important for the stability of the NRPE5 protein *in vivo*, possibly because the extended sequence facilitates Pol V-specific subunit interactions.

DISCUSSION

Origins of Pol V

Pol IV and Pol V are plant-specific enzymes that appear to have originated in an algal progenitor of land plants several hundred million years ago (Luo and Hall, 2007). Their specific involvement

Molecular Cell

Subunit Compositions of RNA Polymerases IV and V

in siRNA-mediated transcriptional gene silencing, which also occurs in other metazoans and fission yeast, has begged the question as to which polymerases accomplish the functions of Pol IV and Pol V in other eukaryotes. In fission yeast, Pol II transcripts traverse silenced loci, serving as binding sites for siRNAs and as templates for the sole RNA-dependent RNA polymerase, thereby generating precursors for further siRNA biogenesis (Buhler and Moazed, 2007; Buhler et al., 2006; Grewal and Elgin, 2007; Irvine et al., 2006). Several nonlethal mutations that disrupt siRNA-mediated silencing and/or siRNA accumulation in *S. pombe* have been mapped to the RPB1, RPB2, and RPB7 subunits of Pol II (Djupedal et al., 2005; Kato et al., 2005; Schramke et al., 2005). Our finding that Pol IV and V have Pol II-like subunit compositions fits the hypothesis that Pol IV and Pol V are derivatives of Pol II that evolved specialized roles in RNA silencing but no longer perform Pol II functions essential for viability, in contrast to fission yeast Pol II, which appears to accomplish all of these tasks. Presumably, the subunits of Pol IV/V that are not shared by Pol II, including NRPD1, NRPE1, NRPD2/NRPE2, NRPE3b, NRPD4/NRPE4, NRPE5, NRPD7, and NRPE7, account for Pol IV- or Pol V-specific activities. It is intriguing that most of these subunits occupy key positions with regard to the template channel and RNA exit paths (Figures 6A and 6B).

Previous analyses of Pol IV and Pol V catalytic subunits had pointed to a Pol II connection. In our initial study of Pol IV, we noted that the NRPD2/NRPE2 subunit is more closely related to the second-largest subunit of Pol II than to the corresponding subunits of Pol I or Pol III (Onodera et al., 2005). Moreover, five out of eight intron positions in the beginning of *NRPD1* and *NRPE1* match the intron positions in *NRPB1*, encoding the largest subunit of Pol II (Luo and Hall, 2007). Based on phylogenetic analyses, Luo and Hall proposed that Pol IV came into existence following a duplication of the *NRPB1* gene that generated the *NRPD1* gene. A subsequent duplication of *NRPD1* to generate *NRPE1* is proposed to have led to the evolution of Pol V after the emergence of land plants but prior to the divergence of angiosperms (flowering plants). Our finding that Pol IV utilizes the same RPB5-family subunit as Pol I, II, and III whereas Pol V uses a distinct variant (NRPE5) is consistent with the hypothesis that Pol V is more distantly related to Pol II than is Pol IV.

The fact that Pol IV and Pol V share numerous small subunits with Pol II, including NRPB3, NRPB6, NRPB8, NRPB9, NRPB10, NRPB11, and NRPB12 family subunits, can explain why alleles for these genes have not been identified in genetic screens; loss-of-function mutations in the subunits of essential polymerases cause female gametophyte lethality (Figure S14) (Onodera et al., 2008). Likewise, the use of more than one NRPE3, NRPE6, NRPE8, or NRPE9 variant by Pol IV or Pol V (Figures 6C and 1) can be expected to make identification of mutations in these genes problematic due to functional redundancies (Figure 6C).

Functions for Mystery Subunits

A number of observations in our study fill in gaps concerning the functions of RNA polymerase subunit families in *Arabidopsis*. For instance, Ulmasov et al. reported the existence of two RPB3-like genes in *Arabidopsis*, which they named *AtRPB36a* and

AtRPB36b based on their predicted sizes of ~36 kD (Ulmasov et al., 1996). *AtRPB36a* was found in highly purified Pol II fractions (Ulmasov et al., 1996), but *AtRPB36b* was not, making the function of the latter variant unclear. Our study reveals that *AtRPB36b* is the NRPE3b subunit of Pol V. *AtRPB36a* (now NRPB3) and NRPB11 (formerly *AtRPB13.6*) in Pol II are the homologs and functional equivalents of the two α subunits (α and α') of *E. coli* RNA polymerase. Previous studies demonstrated that NRPB3 and NRPB11 copurify with Pol II in vivo and physically interact in yeast two-hybrid assays (Ulmasov et al., 1996). Interestingly, *AtRPB36b*/NRPE3b also interacted with NRPB11 in yeast two-hybrid assays (Ulmasov et al., 1996), which is likely to be meaningful, occurring in the context of Pol V in a manner equivalent to the interaction of NRPB3 and NRPB11 in Pol II. Interestingly, the *AtRPB36a* variant also associates with Pol V in vivo; therefore, this protein serves as the NRPB3 subunit of Pol II, the NRPD3 subunit of Pol IV, and one of two alternative Pol V NRPE3 subunits (NRPE3a). How these highly similar RPB3-like subunits are differentially assembled into Pol II, IV, or V is a question deserving further study.

Although peptide coverage for the NRPD4/NRPE4 subunit was low in our study, the Jian-Kang Zhu laboratory identified the *nrdp4/nrpe4* gene in a screen for defective RNA-directed DNA methylation and confirmed the Pol IV and Pol V association of the encoded protein (He, X.-J., Hsu, Y.-F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.-S., and Zhu, J.-K., unpublished data). In budding yeast, RPB4 forms a subcomplex with RPB7 that can be dissociated from the ten subunit Pol II core enzyme without abolishing Pol II catalytic activity in vitro (Cramer, 2004), although the subcomplex appears to be more stable in Pol II from plants (Larkin and Guilfoyle, 1998). In vivo, RPB7 is an essential protein in yeast, whereas RPB4 deletion mutants are temperature sensitive (McKune et al., 1993; Woychik and Young, 1989) and are impaired in transcription elongation and mRNA 3' end processing (Runner et al., 2008; Verma-Gaur et al., 2008). It is intriguing that Pol II, IV, and V have unique RPB7-like subunits and that the NRPB4 subunit of Pol II is different from the NRPD4/NRPE4 subunits of Pol IV and Pol V. Given that the RPB4/RPB7 complex is thought to interact with the nascent RNA transcript (see Figure 6), these differences are likely to contribute to the unique functions of Pol II, IV, and V.

Previous studies had shown that one of the two constitutively expressed RPB5 family proteins is a subunit by Pol I, II, and III (Larkin et al., 1999; Saez-Vasquez and Pikaard, 1997). The function of the other variant, formerly designated *AtRPB5b* or *AtRPB23.7*, was unknown. Our study reveals that the latter protein is the NRPE5 subunit of Pol V. By contrast, the NRPD5 subunit of Pol IV is encoded by the same gene that encodes the Pol II NRPB5 subunit and the equivalent subunits of Pol I and III. As we have shown, *nrdp5-1* mutants display defects in DNA methylation, retroelement silencing, siRNA accumulation, and flowering time, similar to *nrdp1* mutants (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). However, *nrdp5-1* mutant phenotypes are typically less severe than *nrdp1* or *nrdp2/nrdp2* mutants. Because the T-DNA insertion is near the 3' end of the gene, *nrdp5-1* may be a partially functional allele. It is also possible that other members of the multigene family are partially redundant with NRPE5, particularly

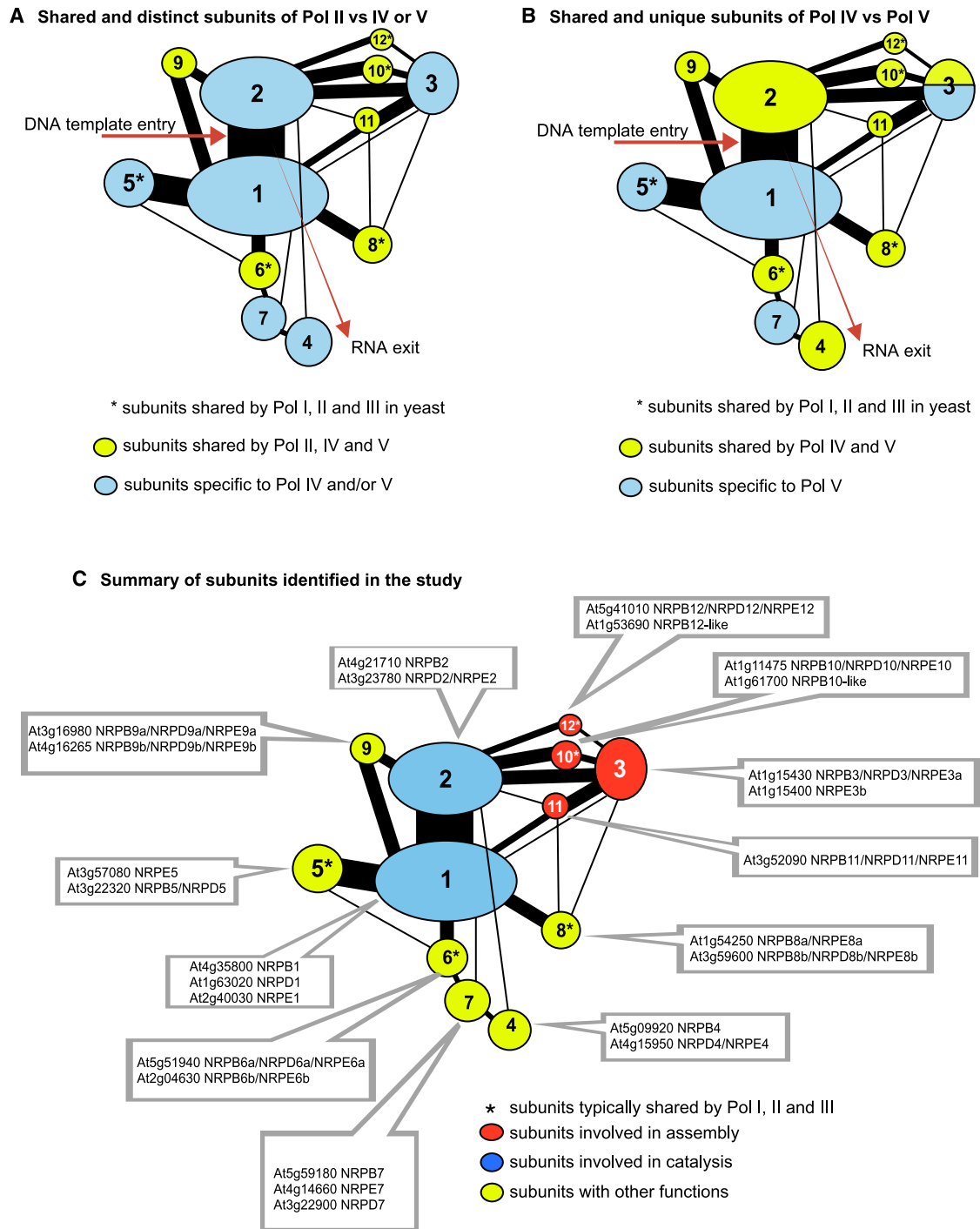


Figure 6. Comparison of RNA Polymerase Subunits in Pol II, IV, and V

(A) Subunits that are unique to Pol IV and/or Pol V compared to Pol II are shown in blue. Subunits common to Pol II, IV, and V are shown in green. The subunit interaction model is based on the yeast Pol II crystal structure (Armache et al., 2005; Cramer et al., 2001; Sampath et al., 2008). The thickness of lines connecting the subunits is proportional to the number of contacts.

(B) Subunits that are unique to Pol V are shown in blue. Subunits common to Pol IV and Pol V are shown in green. The half-blue, half-green shading of the third-largest subunit reflects the fact that Pol V uses the NRPE3b variant that is not used appreciably by Pol IV in addition to the NRPE3a/NRPD3 variant that predominates in Pol IV.

(C) Summary of the *Arabidopsis* genes that encode Pol II, IV, or V subunits.

Molecular Cell

Subunit Compositions of RNA Polymerases IV and V

At2g41340, which shares 70% identity with NRPE5, including the N-terminal extension that is missing in the NRPB5/NRPD5 subunit (Figure 5A and Figure S5). Consistent with this hypothesis, preliminary evidence suggests that a *nrpe5-1 At2g41340* double mutant has a more severe loss of DNA methylation phenotype than does *nrpe5-1* (data not shown). A third possibility is that NRPE5 may not be absolutely required for Pol V transcription. The failure to identify *nrpe5* alleles in genetic screens to date may stem from one or more of these reasons.

The fact that Pol V is unique in using the NRPE5 variant of the RPB5 family is likely to have functional significance. Crystal structures of yeast Pol II reveal that RPB5 interacts with RPB1 and RPB6 to form a mobile “shelf” module that stabilizes the template DNA as it enters the polymerase (Cramer et al., 2001; Gnatt et al., 2001). RPB5 also interacts with hepatitis B transcriptional activator protein X (HBx); the general transcription factor TFIIB; TIP120, a protein which facilitates recruitment of Pol II to the preinitiation complex (Cheong et al., 1995; Lin et al., 1997; Makino et al., 1999); and the yeast chromatin remodeling complex, RSC (Soutourina et al., 2006). Therefore, the differential use of the NRPD5 or NRPE5 subunits in the context of Pol IV or Pol V could mediate different template specificity, locus targeting, or transcriptional activation processes.

EXPERIMENTAL PROCEDURES

Plant Materials

A. thaliana nrpd1 (allele *nrpd1a-3*), *nrpe1* (allele *nrpd1b-11*), and *nrpd2/nrpe2* (*nrpd2a-2 nrpd2b-1*) have been described (Pontes et al., 2006). *nrpe11-1* (*nrpb11-1/nrpd11-1*) is from T-DNA line SALK_100563 (Alonso et al., 2003), *nrpd5-1/nrpb5-1* from T-DNA line SAIL_786_E02 (Sessions et al., 2002), and *nrpe5-1* from GABI-KAT T-DNA line 237A08 (Rosso et al., 2003). Primers for *nrpe11-1*, *nrpd5-1*, and *nrpe5-1* genotyping are listed in Table S3. Callus cultures were induced by germinating sterilized seeds on MS media containing Gamborg's vitamins (Sigma), 5% agargel (Sigma), 0.02 mg/L kinetin (Sigma), and 2 mg/L 2,4-dichlorophenoxyacetic acid (Sigma). Plates were incubated at 23°C. Callus frozen in liquid N₂ was stored at –80°C.

Affinity Purification of Pol IV, V, and II

Frozen callus (115–150 g) expressing FLAG-tagged *NRPE1* or *NRPD1* was ground in extraction buffer (300 mM NaCl, 20 mM Tris [pH 7.5], 5 mM MgCl₂, 5 mM DTT, 1 mM PMSF, and 1:100 plant protease inhibitor cocktail [Sigma]) at 4°C, filtered through two layers of Miracloth (Calbiochem), and centrifuged twice at 10,000 g, 15 min, 4°C. Pol II and NRPE5 were purified with the same protocol from 150 g of leaf tissue expressing FLAG-tagged *NRPB2* or *NRPE5*, respectively. Supernatants were incubated with anti-FLAG-M2 resin for 2–3 hr in a 15 ml tube using 30 μl of resin per 14 ml of extract. Resin was pelleted at 1000 rpm for 2 min and the supernatant incubated with fresh resin for 2–3 hr. Pooled resin was washed five times in 14 ml of extraction buffer containing 0.4% NP-40 (Sigma). Aliquots (125 μl) of resin were then mixed 2 min with 125 μl Ag/Ab Elution Buffer (Pierce) at 4°C. Resin was pelleted, and the eluted complex was pooled. Two ~500 μl batches of pooled complex were concentrated in YM-10 centricon columns (Millipore) at 4°C and desalted using Pierce 500 μl desalting columns. The final elute of ~70 μl containing ~10–50 μg of protein was subjected to LC-MS/MS.

Mass Spectrometry

Samples adjusted to 50% (v/v) 2,2,2-Trifluoroethanol (TFE) (Sigma) were sonicated 1 min at 0°C and then incubated 2 hr at 60°C with shaking at 300 rpm. Proteins were reduced with 2 mM DTT at 37°C for 1 hr, then diluted 5-fold with 50 mM ammonium bicarbonate. CaCl₂ (1 mM) and sequencing-grade modified porcine trypsin (Promega) was added at a 1:50 trypsin-to-protein mass ratio. After 3 hr at 37°C, samples were concentrated to ~30 μl and sub-

jected to reversed-phase liquid chromatography (RPLC) coupled to an electrospray ionization source and LTQ-Orbitrap mass spectrometer (ThermoFisher Scientific). Tandem mass spectra were searched against *A. thaliana* proteins using SEQUEST and filtering criteria, which provided a false discovery rate (FDR) <5%. See the Supplemental Data for details.

Cloning, Vectors, and Transgenic Lines

NRPD1 and *NRPE1* genomic clones (Pontes et al., 2006) were cloned into a Gateway-compatible vector (A.W. and C.S.P., unpublished data) that adds a C-terminal FLAG tag, 3C protease cleavage site, and biotin ligase recognition peptide. *NRPE5*, *NRPE6a*, *NRPE6b*, *NRPE8a*, *NRPE9a*, *NRPB7*, *NRPE3a*, and *NRPE3b* cDNAs were amplified by RT-PCR from poly-T primed cDNA cloned into pENTR-D-TOPO or pENTR-TEV-TOPO. cDNAs were recombined into pEarleyGate 201 (HA tag) or 202 (FLAG tag) (Earley et al., 2006). Genomic *NRPE8b*, *NRPE10*, *NRPE11*, and *NRPE6a* clones were similarly amplified by PCR and cloned into pEarleyGate 302 (FLAG tag). *NRPD1-FLAG*, *NRPE1-FLAG*, *NRPA2-FLAG*, *NRPB2-FLAG*, and *NRPC2-FLAG* transgenes were previously described (Onodera et al., 2008; Pontes et al., 2006).

Methylation Assays

5S rDNA Southern blot methylation assays and *AtSN1* PCR assays were performed using 250 ng–1 μg of DNA as in Onodera et al. (2005).

RT-PCR Analysis of *AtSN1*

For *AtSN1* transcripts, high-molecular-weight RNA was isolated from 300 mg of leaves using a miRVANA (Ambion) kit, and strand-specific RT-PCR was performed as described (Wierzbicki et al., 2008).

Small RNA Northern Blots

Inflorescence small RNA (7.5 μg) was analyzed by northern blot hybridization using *COPIA*, siR1003 (5S rRNA), 45S rRNA, miR173, and tasIR255 probes as described previously (Allen et al., 2005; Onodera et al., 2005; Pontes et al., 2006; Xie et al., 2004). Blots stripped twice with 50% formamide, 0.1 × SSC, and 1% SDS at 65°C for 2 hr were reprobed to generate multiple figure panels.

Antibodies

Anti-NRPE2/NRPD2, anti-NRPB5/NRPD5, and anti-NRPE5 have been described (Larkin et al., 1999; Onodera et al., 2005). Anti-FLAG antibodies were from Sigma. Anti-NRPB1-CTD (8WG16) was purchased from Abcam. NRPE1 antibodies (Covance) recognize peptide N-CDKKNSETESDAAAWG-C. NRPD1 antibodies (Covance) recognize peptide N-CLKNGTLESGGF-SEN-P-C. Anti-NRPA2/NRPB2/NRPC2 antibodies (US Biologicals) recognize N-CGDKKFSRRHGQK-C. Antibodies were affinity purified using immobilized peptides.

Immunoprecipitation and Immunoblotting

Leaves (2–4 g) were ground in extraction buffer (Baumberger and Baulcombe, 2005), filtered through Miracloth, and centrifuged at 10,000 g for 15 min. Supernatants were incubated 3–12 hr at 4°C with 30 μl of anti-FLAG-M2 resin (Sigma). Beads were washed three times in extraction buffer + 0.5% NP-40 (Sigma) and eluted with two bed volumes of 2× SDS sample buffer, and 5–20 μl was subjected to SDS-PAGE and transferred to Immobilon PVDF membranes (Millipore). Blots were incubated with antibodies in TBST + 5% (w/v) nonfat dried milk. Antibody dilutions were as follows: 1:250 (NRPE1), 1:500 (NRPD1), 1:2000 (NRPB1-CTD), 1:750 (NRPB5/NRPD5), 1:750 (NRPE5), 1:250 (NRPD2/NRPE2), 1:500 (anti-Pol I, II, and/or III) and 1:2000–1:10,000 (FLAG-HRP). The secondary antibody was anti-rabbit-HRP, diluted 1:5000–1:20,000; or anti-mouse-HRP, diluted 1:5000 (GE Healthcare, Sigma). Blots were washed four times for 4 min in TBST and visualized by chemiluminescence (GE Healthcare). Blots were stripped for 35 min in 25 mM glycine (pH 2.0), 1% SDS; re-equilibrated in TBST; and probed with additional antibodies.

Alignments

Sequences were aligned using ClustalW and highlighted using BOXSHADE. Construction of phylogenetic trees was performed using MegAlign. Trees are

based on ClustalW alignments of full-length proteins, and bootstrap values are based on 10,000 replicates. Dotted lines represent negative branch lengths.

SUPPLEMENTAL DATA

The Supplemental Data include Supplemental Experimental Procedures, three tables, and 18 figures and can be found with this article online at [http://www.cell.com/molecular-cell/supplemental/S1097-2765\(08\)00858-7](http://www.cell.com/molecular-cell/supplemental/S1097-2765(08)00858-7).

ACKNOWLEDGMENTS

T.S.R. and C.S.P. designed the study and wrote the paper. T.J.G. and G.H. contributed antibodies. J.R.H. generated Pol I, II, and III transgenic lines and NRPD1 and NRPE1 antibodies. A.W. made NRPD1- and NPRED1-FLAG-biotin lines. T.S.R. performed all experiments except LC-MS/MS analyses by C.D.N., A.N., and L.P.-T. at Pacific Northwest National Laboratory (PNNL). J.-K.Z. provided NRPD4/NRPE4 insights. We thank the Washington University greenhouse staff for plant care and Pikaard lab colleagues for discussions. T.S.R. and C.S.P. also thank Biology 4024 students who helped clone cDNAs: Silvano Ciani and Colin Clune (*At2g04630*), Andrew Pazandak and Kariline Bringe (*At1g54250* and *At3g16980*), Caitlin Ramsey and Colin Orr (*At5g59180*), Wan Shi and Soon Goo Lee (*At1g11475*), and Lily Momper and Charu Agrawal (*At5g51940*). Pikaard lab research is supported by National Institutes of Health (NIH) grant GM077590. Any opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the NIH. Portions of this research were supported by the NIH National Center for Research Resources (RR18522), and the W.R. Wiley Environmental Molecular Science Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at PNNL. PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC05-76RL01830.

Received: September 16, 2008

Revised: December 8, 2008

Accepted: December 10, 2008

Published online: December 24, 2008

REFERENCES

- Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207–221.
- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657.
- Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.* **280**, 7131–7134.
- Baumberger, N., and Baulcombe, D.C. (2005). *Arabidopsis* ARGONAUTE1 is an RNA slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. USA* **102**, 11928–11933.
- Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123**, 1279–1291.
- Brosnan, C.A., Mitter, N., Christie, M., Smith, N.A., Waterhouse, P.M., and Carroll, B.J. (2007). Nuclear gene silencing directs reception of long-distance mRNA silencing in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **104**, 14741–14746.
- Buhler, M., and Moazed, D. (2007). Transcription and RNAi in heterochromatic gene silencing. *Nat. Struct. Mol. Biol.* **14**, 1041–1048.
- Buhler, M., Verdell, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873–886.
- Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. *Science* **303**, 1336.
- Cheong, J.H., Yi, M., Lin, Y., and Murakami, S. (1995). Human RPB5, a subunit shared by eukaryotic nuclear RNA polymerases, binds human hepatitis B virus X protein and may play a role in X transactivation. *EMBO J.* **14**, 143–150.
- Cramer, P. (2002). Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.* **12**, 89–97.
- Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. *Curr. Opin. Genet. Dev.* **14**, 218–226.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876.
- Darst, S.A., Polyakov, A., Richter, C., and Zhang, G. (1998). Insights into *Escherichia coli* RNA polymerase structure from a combination of x-ray and electron crystallography. *J. Struct. Biol.* **124**, 115–122.
- Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev.* **19**, 2301–2306.
- Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A., and Voinnet, O. (2007). Intra- and intercellular RNA interference in *Arabidopsis thaliana* requires components of the microRNA and heterochromatic silencing pathways. *Nat. Genet.* **39**, 848–856.
- Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882.
- Grewal, S.I., and Elgin, S.C. (2007). Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399–406.
- Grummt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev.* **17**, 1691–1702.
- Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**, 4671–4679.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118–120.
- Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from *Archaea*. *Nature* **451**, 851–854.
- Irvine, D.V., Zaratiegui, M., Tolia, N.H., Goto, D.B., Chitwood, D.H., Vaughn, M.W., Joshua-Tor, L., and Martienssen, R.A. (2006). Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313**, 1134–1137.
- Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **37**, 761–765.
- Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., and Jin, H. (2007). A novel class of bacteria-induced small RNAs in *Arabidopsis*. *Genes Dev.* **21**, 3123–3134.
- Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309**, 467–469.
- Larkin, R.M., and Guilfoyle, T.J. (1998). Two small subunits in *Arabidopsis* RNA polymerase II are related to yeast RPB4 and RPB7 and interact with one another. *J. Biol. Chem.* **273**, 5631–5637.
- Larkin, R.M., Hagen, G., and Guilfoyle, T.J. (1999). *Arabidopsis thaliana* RNA polymerase II subunits related to yeast and human RPB5. *Gene* **231**, 41–47.
- Lin, Y., Nomura, T., Cheong, J., Dorjsuren, D., Iida, K., and Murakami, S. (1997). Hepatitis B virus X protein is a transcriptional modulator that communicates with transcription factor IIB and the RNA polymerase II subunit 5. *J. Biol. Chem.* **272**, 7132–7139.
- Liu, F., Quesada, V., Crevillen, P., Baurle, I., Swiezewski, S., and Dean, C. (2007). The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase II homolog to downregulate FLC. *Mol. Cell* **28**, 398–407.

- Luo, J., and Hall, B.D. (2007). A multistep process gave rise to RNA polymerase IV of land plants. *J. Mol. Evol.* **64**, 101–112.
- Makino, Y., Yogosawa, S., Kayukawa, K., Coin, F., Egly, J.M., Wang, Z., Roeder, R.G., Yamamoto, K., Muramatsu, M., and Tamura, T. (1999). TATA-binding protein-interacting protein 120, TIP120, stimulates three classes of eukaryotic transcription via a unique mechanism. *Mol. Cell. Biol.* **19**, 7951–7960.
- McKune, K., Richards, K.L., Edwards, A.M., Young, R.A., and Woychik, N.A. (1993). RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability. *Yeast* **9**, 295–299.
- Myouga, F., Tsuchimoto, S., Noma, K., Ohtsubo, H., and Ohtsubo, E. (2001). Identification and structural analysis of SINE elements in the Arabidopsis thaliana genome. *Genes Genet. Syst.* **76**, 169–179.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613–622.
- Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., Ream, T., Ito, Y., and Pikaard, C.S. (2008). Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. *Genetics* **180**, 207–218.
- Pikaard, C.S., Haag, J.R., Ream, T., and Wierzbicki, A.T. (2008). Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci.* **13**, 390–397.
- Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006). The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**, 79–92.
- Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev.* **19**, 2030–2040.
- Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol. Biol.* **53**, 247–259.
- Runner, V.M., Podolny, V., and Buratowski, S. (2008). The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Mol. Cell. Biol.* **28**, 1883–1891.
- Saez-Vasquez, J., and Pikaard, C.S. (1997). Extensive purification of a putative RNA polymerase I holoenzyme from plants that accurately initiates rRNA gene transcription in vitro. *Proc. Natl. Acad. Sci. USA* **94**, 11869–11874.
- Sampath, V., Balakrishnan, B., Verma-Gaur, J., Onesti, S., and Sadhale, P.P. (2008). Unstructured N terminus of the RNA polymerase II subunit Rpb4 contributes to the interaction of Rpb4.Rpb7 subcomplex with the core RNA polymerase II of Saccharomyces cerevisiae. *J. Biol. Chem.* **283**, 3923–3931.
- Schramke, V., Sheedy, D.M., Denli, A.M., Bonila, C., Ekwall, K., Hannon, G.J., and Allshire, R.C. (2005). RNA-interference-directed chromatin modification coupled to RNA polymerase II transcription. *Nature* **435**, 1275–1279.
- Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16**, 2593–2620.
- Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., et al. (2002). A high-throughput Arabidopsis reverse genetics system. *Plant Cell* **14**, 2985–2994.
- Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C. (2007). An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell* **19**, 1507–1521.
- Soutourina, J., Bordas-Le Floch, V., Gendrel, G., Flores, A., Ducrot, C., Dumay-Odelot, H., Soularue, P., Navarro, F., Cairns, B.R., Lefebvre, O., et al. (2006). Rsc4 connects the chromatin remodeler RSC to RNA polymerases. *Mol. Cell. Biol.* **26**, 4920–4933.
- Ulmasov, T., Larkin, R.M., and Guilfoyle, T.J. (1996). Association between 36- and 13.6-kDa alpha-like subunits of Arabidopsis thaliana RNA polymerase II. *J. Biol. Chem.* **271**, 5085–5094.
- Verma-Gaur, J., Rao, S.N., Taya, T., and Sadhale, P. (2008). Genomewide recruitment analysis of Rpb4, a subunit of polymerase II in Saccharomyces cerevisiae, reveals its involvement in transcription elongation. *Eukaryot. Cell* **7**, 1009–1018.
- Werner, F. (2007). Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* **65**, 1395–1404.
- Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**, 635–648.
- Woychik, N.A., and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. *Cell* **108**, 453–463.
- Woychik, N.A., and Young, R.A. (1989). RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. *Mol. Cell. Biol.* **9**, 2854–2859.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104. [10.1371/journal.pbio.0020104](https://doi.org/10.1371/journal.pbio.0020104).

Molecular Cell, Volume 33

Supplemental Data

Subunit Compositions of the RNA-Silencing Enzymes

Pol IV and Pol V Reveal Their Origins

as Specialized Forms of RNA Polymerase II

Thomas S. Ream, Jeremy R. Haag, Andrzej Wierzbicki, Carrie D. Nicora, Angela Norbeck, Jian-Kang Zhu, Gretchen Hagen, Thomas J. Guilfoyle, Ljiljana Paša-Tolić, and Craig S. Pikaard

A. Supplemental Experimental Procedures

Liquid chromatography and mass spectrometry details. Two independent affinity-purified NRPE1 samples and one affinity purified NRPE5 sample were analyzed by LC-MS/MS in order to identify Pol V subunits. Affinity purified NRPD1 and NRPB2 samples were also analyzed to identify Pol IV and Pol II subunits, respectively. In each case, control samples derived from non-transgenic plants were subjected to the affinity purification procedure and analyzed by mass spectrometry.

All samples were prepared for analysis using the following procedure: a Coomassie protein assay (Pierce, Rockford, IL) was performed to determine the initial protein concentration of the sample. 2,2,2-trifluoroethanol (TFE) (Sigma, St. Louis, MO) was then added to the sample for a final concentration of 50% TFE. The sample was sonicated in an ice-water bath for 1 min. and incubated at 60°C for 2 hours with gentle shaking at 300 rpm. The sample was then reduced with 2mM dithiothreitol (DTT) (Sigma, St. Louis, MO) with incubation at 37°C for 1 hr with gentle shaking at 300rpm. Samples were then diluted 5-fold with 50mM ammonium bicarbonate for preparation for digestion. 1mM CaCl₂ and sequencing-grade modified porcine trypsin (Promega, Madison, WI) was added to all protein samples at a 1:50 (w/w) trypsin-to-protein ratio for 3 h at 37°C. The sample was concentrated in a Speed Vac (ThermoSavant, Holbrook, NY) to

a volume of ~30 μ l and was then centrifuged at 14,000 rpm. The supernatant was removed and added to a sample vial for LC-MS/MS analysis.

Peptide samples were analyzed on a custom-built reversed-phase liquid chromatography (RPLC) system coupled via electrospray ionization (ESI) utilizing an ion funnel to a ThermoFisher Scientific LTQ-Orbitrap mass spectrometer (ThermoFisher Scientific, San Jose, CA). Briefly, the capillary RPLC separation was performed under a constant pressure of 10,000 psi, using two ISCO (Lincoln, NE) Model 100 DM high-pressure syringe pumps and a column (60 cm \times 75 μ m i.d.) packed in-house (Pacific Northwest National Laboratory) with Phenomenex (Torrance, CA) Jupiter particles (C18 stationary phase, 5 μ m particles, 300 Å pore size). Mobile phase A consisted of 0.1% formic acid in water, and mobile phase B consisted of 100% acetonitrile. The RPLC system was equilibrated at 10,000 psi with 100% mobile phase A. A mobile phase selection valve was switched 50 min after injection to create a near-exponential gradient as mobile phase B displaced A in a 2.5 mL mixer. A split was used to provide an initial flow rate through the column of ~ 400 nL/min. The column was coupled to the mass spectrometer using an in-house manufactured ESI interface with homemade 20 μ m i.d. chemically etched emitters. The heated capillary temperature and spray voltage were 200° C and 2.2 kV, respectively. Mass spectra were acquired for 80 min over the m/z range 400-2000 at a resolving power of 100K. A maximum of six data-dependent LTQ tandem mass spectra were recorded for the most intense peaks in each survey mass spectrum.

Tandem mass spectra were searched against an *Arabidopsis thaliana* protein file (The Institute for Genomic Research, TIGR 2008 <http://www.tigr.org/plantProjects.shtml>) containing 27,854 protein sequences after the removal of duplicates. Searching was performed using SEQUEST, allowing for a dynamic oxidation of methionine. In addition, peptide cleavage events were limited to fully tryptic sequences. For the spectra acquired in the Orbitrap, the monoisotopic masses were corrected prior to generation of the dta files used for searching using

the program DeconMSN, developed in house. Peptide sequences were considered confident if the scores passed Xcorr and delcn thresholds described by Washburn et al., which gave a False Discovery Rate (FDR) for all identified peptides of less than 5% and averaged 1.5% based on a reversed database search. Proteins with at least 2 filter passing peptides were considered confidently identified.

Generation of transgenic lines. Plants were transformed by *Agrobacterium tumefaciens* strain GV3101 harboring each transgene-bearing plasmid, using the floral dip method (Clough, S.J., and Bent, A.F. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16, 735-743). Transformants were selected by spraying with 0.05% Finale herbicide, containing 5.78% (w/v) glufosinate-ammonium (AgrEvo Environmental Health). Experiments demonstrating rescue of the *nrpe5-1* mutation by the *35S:FLAG-NRPE5* construct were performed for individual T1 transformants. Protein assays in the tagged RNA polymerase subunit lines were performed using 3- to 4-week-old pooled T2 progeny derived from single T1 plants.

Genotyping. One to three leaves were placed in a PCR tube and 125 μ l of extraction buffer was added (200 mM Tris, pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS). Tubes were heated using a thermocycler for 10 min. at 99°C. Tubes were then subjected to centrifugation at 6000 x g for 10 min. The supernatant was transferred to a new PCR tube with 125 μ l of isopropanol and mixed by inversion. After 15 min., the tubes were subjected to centrifugation at 6000 x g for 15 min. The supernatant was removed and the pellet was washed with 125 μ l of 75% ethanol. The tubes were spun for 5 min. at 6000 x g. The supernatant was removed and 75 μ l of TE buffer was added to the pellets. The tubes were incubated in a thermocycler at 55°C for 10 min. 2 μ l of DNA was used in each 20 μ l genotyping reaction with GoTaqGreen polymerase according to the

manufacturer's instructions. Cycling conditions for genotyping *nrbp11-1/nrpd11-1/nrpe11-1*, *nrpd5-1/nrpb5-1* and *nrpe5-1* were: 94°C 2 min. 30 sec., 36 cycles of 94°C 30 sec., 55°C 30 sec. and 72°C 1 min. 15 sec. followed by a final extension of 72°C for 7 min.

Flowering time assay. Mutants tested in the flowering time assay were all in the Columbia ecotype: *nrpd1a-3*, *dcl3-1*, *rdr2-1*. The *dcl3-1* and *rdr2-1* mutants were originally provided by Jim Carrington. Twelve to twenty plants of each genotype were grown under short-day (8 hrs. light, 16 hrs. dark) photoperiod conditions and their positions within the growth chamber were randomized every four to six days to minimize environmental influences. Flowering time was measured as the number of leaves produced in the basal rosette at the time the bolt height reached ten centimeters. P-values were derived from a two-tailed Student's-t-test of significance.

Protein alignments presented as supplemental material. Alignments were performed as described in the main methods. Sequences of RNA polymerase subunits were obtained by BLASTp searches using either *S. cerevisiae* or *A. thaliana* sequences.

RT-PCR. *NRPB5*-family first-strand cDNAs were generated using poly-T primers and PCR-amplified using gene-specific primers. *NRPE5* and *At2g41340* were amplified with the same primers and distinguished using *SpeI* (cleaves *NRPE5*) or *HpaII* (cleaves *At2g41340*).

B. SUPPLEMENTAL DATA

Table S1. Genes whose known or predicted sequences were used in peptide coverage maps and/or protein alignments.

Supplemental Table 1							
Organism	common name/class	Protein	Gene ID	accession no.	source		
Arabidopsis thaliana	thale cress	NRPD1	At1g63020	NP_176490	Genbank		
		NRPE1	At2g40030	NP_181532	Genbank		
		NRPA1	At3g57660	NP_191325	Genbank		
		NRPB1	At4g35800	NP_195305	Genbank		
		NRPC1	At5g60040	NP_200812	Genbank		
		NRPD2a	At3g23780	NP_189020	Genbank		
		NRPD2b	At3g18090	NP_188437	Genbank		
		NRPA2	At1g29940	NP_564341	Genbank		
		NRPB2	At4g21710	NP_193902	Genbank		
		NRPC2	At5g45140	NP_199327	Genbank		
		NRPB3a	At2g15430	NP_179145	Genbank		
		NRPE3b	At2g15400	NP_179142	Genbank		
		NRPB4	At5g09920	ABF58918	Genbank		
		NRPD4/NRPE4	At4g15950	AAT71989	Genbank		
		NRPB5	At3g22320	NP_188871	Genbank		
		NRPE5	At3g57080	NP_191267	Genbank		
		NRPB5-like	At5g57980	NP_200606	Genbank		
		NRPE5-like	At2g41340	NP_181665	Genbank		
		NRPE5-like	At3g54490	NP_191013	Genbank		
		NRPB5-like	At3g16880	NP_188290	Genbank		
		NRPB6a	At5g51940	NP_200007	Genbank		
		NRPB6b	At2g04630	NP_178540	Genbank		
		NRPB7	At5g59180	NP_200726	Genbank		
		NRPE7	At4g14660	NP_193202	Genbank		
		NRPD7	At3g22900	NP_566719	Genbank		
		NRPB7-like	At4g14520	NP_849385	Genbank		
		RPC25-like	At1g06790	NP_200726	Genbank		
		RPA43-like	At1g75670	NP_974148	Genbank		
		NRPB8a	At1g54250	NP_175827	Genbank		
		NRPB8b	At3g59600	NP_191519	Genbank		
		NRPB9a	At3g16980	NP_188323	Genbank		
		NRPB9b	At4g16265	NP_567490	Genbank		
		RPA12-like	At3g29540	ABD38906	Genbank		
		RPC11-like	At4g07950	NP_192535	Genbank		
		RPC11-like	At1g01210	NP_171629	Genbank		
		NRPB10a	At1g11475	NP_849640	Genbank		
		NRPB10-like	At1g61700	NP_176363	Genbank		
		NRPB11	At3g52090	NP_190777	Genbank		
		NRPB12a	At5g41010	NP_198917	Genbank		
		NRPB12-like	At1g53690	NP_175773	Genbank		
		Homo sapiens	human	RPB5		BAA07406	Genbank
		Drosophila melanogaster	fruit fly	RPB5		NP_610630	Genbank
Caenorhabditis elegans	nematode	RPB5		Q9N5K2	Genbank		
Saccharomyces cerevisiae	yeast	RPB3		P16370	Genbank		
		RPB4		NP_012395	Genbank		
		RPB5		CAA85113	Genbank		
		RPB6		CAA37382	Genbank		
		RPB7		AAC60558	Genbank		
		RPB8		CAA99443	Genbank		
		RPB9		CAA96774	Genbank		
		RPB10		CAA99425	Genbank		
		RPB11		NP_014638	Genbank		
		RPB12		AAB68994	Genbank		
Brassica napus	rapeseed	RPB5		AAF81222	Genbank		
Vitis vinifera	grape vine	RPB5a		CAO63075	Genbank		
		RPB5b		CAO42914	Genbank		
		RPB5c		CAO65489	Genbank		
Oryza sativa	rice	RPB5a		NP_001065723	Genbank		
		RPB5b		NP_001066119	Genbank		
		RPB5c		EAY79909	Genbank		
		RPB5d		EAZ13876	Genbank		
		RPB5e		NP_001044564	Genbank		
		RPB5f		CAD41325	Genbank		
		RPB5g		EAZ31161	Genbank		
Zea mays	maize	RPB5a		ACF87172	Genbank		
		RPB5b		ACF81264	Genbank		
		RPB5c		ACF85599	Genbank		
Physcomitrella patens	moss	RPB5a		206246	JGI v1.1		
		RPB5b		231299	JGI v1.1		
		RPB5c		55574	JGI v1.1		
		RPB5d		136486	JGI v1.1		
Medicago trunculata	legume	RPB5a		ABO78350	Genbank		
		RPB5b		ABN07995	Genbank		
		RPB5c		ABD28306	Genbank		
Populus trichocarpa	black cottonwood	RPB5a		584052	JGI v1.0		
		RPB5b		57931	JGI v1.0		
		RPB5c		48513	JGI v1.0		
Ostreococcus lucimarinus	green algae	RPB5		XP_001417617	Genbank		
Chlamydomonas reinhardtii	green algae	RPB5		XP_001697601	Genbank		

Table S2. Subunits of *Arabidopsis* Pol V identified by LC-MS/MS analysis of immunoprecipitated FLAG-NRPE5. Pol V subunit relationships to equivalent subunits of yeast Pol II, archaeal and bacterial RNAP are shown, as in Table 1 of the main text. Numbers denote the % of the protein covered by sequenced peptides that could only have come from the indicated protein; non-unique peptides matching related family members are excluded from the coverage calculation. Asterisks denote the fact that all sequenced peptides could be derived from either of two closely related variants.

function	Bacteria	Archaea	Sc Pol II	At homologs	NRPE5 IP	Names/synonyms		
catalytic	β'	RPOA' RPOA"	RPB1	At4g35800		NRPB1		
				At1g63020		NRPD1		
				At2g40030	22	NRPE1		
	β	RPOB' RPOB"	RPB2	At4g21710		NRPB2		
				At3g23780	24	NRPD2/NRPE2		
assembly	α	RPOD	RPB3	At2g15430	36	NRPB3/NRPD3/NRPE3a		
				At2g15400	4	NRPE3b		
	α	RPOL	RPB11	At3g52090	36	NRPB11/NRPD11/NRPE11		
				RPON	RPB10	At1g11475	28	NRPB10/NRPD10/NRPE10
						At1g61700		NRPB10-like
				RPOP	RPB12	At5g41010	16	NRPB12/NRPD12/NRPE12
						At1g53690		NRPB12-like
	auxillary	ω	RPOK	RPB6	At5g51940	*	NRPB6a/NRPD6a/NRPE6a	
					At2g04630	*	NRPB6b/NRPD6b/NRPE6b	
RPOG			RPB8	At1g54250	*	NRPB8a/NRPD8a/NRPE8a		
				At3g59600	*	NRPB8b/NRPD8b/NRPE8b		
RPOH			RPB5	At3g22320		NRPB5/NRPD5		
				At3g57080	65	NRPE5		
				At5g57980		NRPB5-like		
				At2g41340		NRPE5-like		
				At3g54490		NRPE5-like		
RPOF			RPB4	At5g09920		NRPB4		
				At4g15950	8	NRPD4/NRPE4		
RPOE			RPB7	At5g59180		NRPB7		
				At4g14660		NRPE7		
				At3g22900		NRPD7		
				At4g14520		NRPB7-like		
TFS/RPOX			RPB9	At3g16980	*	NRPB9a/NRPE9a		
				At4g16265	*	NRPB9b/NRPD9b/NRPE9b		

Table S3. List of primer sequences.

Primer	sequence	Used for:
cNRPE5-F	CACC ATG GAA GTG AAA GGG AAA GAG ACA G	cloning <i>NRPE5</i> cDNA
cNRPE5-R	TTA CCA CAC ACA TCG GAA GGC	cloning <i>NRPE5</i> cDNA
NRPA2-F	CACC GCC AAT GCT TTC GAG GAA CGG TTT	cloning <i>NRPA2</i> genomic fragment
NRPA2-R	ATC AGT TAC TCC TTC TCT ATC GCT TAA CTG AAG AGT C	cloning <i>NRPA2</i> genomic fragment
NRPB2-F	CACC TCA CTC TCC GTC TCT CTC TCT CTT	cloning <i>NRPB2</i> genomic fragment
NRPB2-R	CTG TCT GCC TTT AGC CGA TTT CAG G	cloning <i>NRPB2</i> genomic fragment
NRPC2-F	CAC CTG AAT ACA CCC TCC TTA GAG GCC A	cloning <i>NRPC2</i> genomic fragment
NRPC2-R	AGC CTC TGT GAG TTT CAG AGC C	cloning <i>NRPC2</i> genomic fragment
cNRPB3a-F	cacc ATGGACGGTGCCACATACCAAAG	cloning <i>NRPB3a</i> cDNA
cNRPB3a-R	TTA TCC TCC ACG CAT ATG GGC AC	cloning <i>NRPB3a</i> cDNA
cNRPE3b-F	cacc ATGGACGGTGTCACCTACCAAAG	cloning <i>NRPE3b</i> cDNA
cNRPE3b-R	TTA TCC TTC ACG CAT ATG GGC ACC	cloning <i>NRPE3b</i> cDNA
cNRPB6a-F	cacc ATGGCTGACGAAGATTACAACGACG	cloning <i>NRPB6a</i> cDNA
cNRPB6a-R	TTA ATC ACC ACC AAC TTG ACG TTT CC	cloning <i>NRPB6a</i> cDNA
cNRPB6b-F	cacc ATG GCT GAC GAC GAT TAC AAT GAA G	cloning <i>NRPB6b</i> cDNA
cNRPB6b-R	TTA ATC ACC ACC GAC TTG ACG TTT C	cloning <i>NRPB6b</i> cDNA
NRPB6a-F	cacc gcacaaaaaactaataatcacacatc	cloning <i>NRPB6a</i> genomic fragment
NRPB6a-R	ATC ACC ACC AAC TTG ACG TTT C	cloning <i>NRPB6a</i> genomic fragment
cNRPB7-F	cacc ATG TTT TTC CAC ATA GTA TTG GAG CG	cloning <i>NRPB7</i> cDNA
cNRPB7-R	TTA TGC CGC TGC AGG GTC GT	cloning <i>NRPB7</i> cDNA
cNRPB8a-F	cacc ATGGCGAGCAATATCATCTTTGTCG	cloning <i>NRPB8a</i> cDNA
cNRPB8a-R	TTA CAG CTT CCT CAT GAG TAG GAA G	cloning <i>NRPB8a</i> cDNA
cNRPB8b-F	cacc ATGGCGAGCAATATTATCATGTTGCG	cloning <i>NRPB8b</i> cDNA
cNRPB8b-R	TTA AAG CTT CCT CAT GAG TAG AAA GAG	cloning <i>NRPB8b</i> cDNA
cNRPE9a-F	cacc ATGAGTACTATGAAATTTTGCCGCG	cloning <i>NRPE9a</i> cDNA
cNRPE9a-R	TTA TTC TCT CCA GCG ATG ACC AC	cloning <i>NRPE9a</i> cDNA
NRPB10-F	cacc tgttctgtaagcgtagagatcttc	cloning <i>NRPB10</i> genomic fragment
NRPB10-R	ACT GTT GTC TGA TTT CTC CAG AG	cloning <i>NRPB10</i> genomic fragment
NRPB11-F	cacc GTT GTG TCC GAA CAT ACC TCA C	cloning <i>NRPB11</i> genomic fragment
NRPB11-R	AAA CTG ATT CGA AAA CTT GGC C	cloning <i>NRPB11</i> genomic fragment
AISN2-1 F	AGATAGTCACAATGTAAGGCATTCGTG	AISN2/control methylation assay
AISN2-1 R	TTGATCCTTTGTCAATGGAAGATTAC	AISN2/control methylation assay
NRPB5a F	GAG AGG ATC TTG TTA CTC TTA AGG CTA	RT-PCR
NRPB5a R	CGA CCA GCC GTT TCA CTC GGA	RT-PCR
NRPB5c F	CTT GAA AAG AGA AGA GTT TGT TCA GAG G	RT-PCR
NRPB5c R	AAT GAA GTA GCA TCG CTT CGT C	RT-PCR
At3g54490 F	GAG GAG ACA ATG GCC GAA G	RT-PCR
At3g54490 R	CAT TGT TGG AAA TCT GAA TAT GAA GAG CA	RT-PCR
NRPE5 and At2g41340 F	TAC GAA GTC TCC GAC GAA GAT AT	RT-PCR
NRPE5 and At2g41340 R	CTC AAT GCT GAA CTT CTT GAG AAG TG	RT-PCR
At3g16880-F	GTT CTC TTT CTC TCT AGA AAC TTT TG	RT-PCR
At3g16880-R	CAC CAT GAA GAA ATA CAT AGA CCA GTT AAA ATC GGC A	RT-PCR
NRPE5 RNA F span	AAG GTC GAG ATA TTC CAG ATA ACG G	RT-PCR of <i>NRPE5</i> in <i>nrpe5-1</i>
NRPE5 RNA R span	GCG ATT CCG TGA GTT CGC CTC	RT-PCR of <i>NRPE5</i> in <i>nrpe5-1</i>
FLAG F	ATG GAC TAC AAA GAC GAT GAC GAC	RT-PCR of <i>FLAG-NRPE5</i> transgenes
NRPE5 cDNA R	CAG CCC AGT TAT GGT TTC TTG G	RT-PCR of <i>FLAG-NRPE5</i> transgenes
Δ N-NRPE5 F	CACC CTA TCG AGT GAA GAG AGT CAT AGA TAC	cloning Δ N- <i>NRPE5</i> cDNA
NRPE5-R	TTA CCA CAC ACA TCG GAA GGC	cloning Δ N- <i>NRPE5</i> cDNA
SAIL 786E02 LP	AGA GCA CAT GAA TCA GCG ACT	genotyping <i>nrpd5-1</i>
SAIL 786E02 RP	GGA GAG ATC GTC GTA GCA CTG	genotyping <i>nrpd5-1</i>
GABI KAT 237A08 LP	CTT CCC CTG CCC ATT TTT TTG CTA C	genotyping <i>nrpe5-1</i>
GABI KAT 237A08 RP	GTT TAA AGG GTC TGC TTC AAG AAG TG	genotyping <i>nrpe5-1</i>
SALK 100563 LP	GAGAGTATGGGCTGGTGATTG	genotyping <i>nrpe11-1</i>
SALK 100563 RP	AGAGCCTGTTGCTTTGAATTG	genotyping <i>nrpe11-1</i>
NRPE5 RNA 5' F	ATGGAAGTGAAGGGAAAGACACAG	RT-PCR of <i>NRPE5</i> in <i>nrpe5-1</i>
NRPE5 RNA 5' R	GTTCAATGGCTTTCAAGGCTTGATT	RT-PCR of <i>NRPE5</i> in <i>nrpe5-1</i>

Figure S1. Peptide coverage maps of DNA-directed RNA polymerase subunits detected by LC-MS/MS in affinity purified Pol V (NRPE1-FLAG). In the full-length protein sequences that follow, peptides highlighted in yellow or green indicate sequenced tryptic peptides that do not overlap with other sequenced peptides. Cyan highlighting denotes sequences represented by two overlapping peptides. Magenta highlighting indicates regions corresponding to three or more overlapping peptide sequences.

NRPE1 (At2g40030)

MEEESTSEILDGEIVGITFALASHHEICIQSISESAINHPSQLTNAFLGLPLEFGKCESCGAT
 EPDKCEGHFGYIQLPVPIYHPAHVNELKQMLSLCLCLKIKKAKGTSGLADRLLGVC
 CEEASQISIKDRASDGASYLELKLPSRSRLQPGCWNFLERYGYRYGSDYTRPLLAREVKE
 ILRRIPEESRKKLTAKGHIPOEGYILEYLPVPPNCLSVPEASDGFSTMSVDPSEIELKDVLK
 KVIAIKSSRSGETNFESHKAEASEMFRVVDTYLQVRGTAKAARNIDMRYGVSKISDSSSS
 KAWTEKMRTLFIKRGSGFSSRSVITGDAYRHVNEVGIPIEIAQRITFEERVSVHNRGYLQ
 KLVDDKLCLS YTQGSTTYSRLDGSKGHTELKPGQVVHRRVMDGDVVFINRPPTTHKHS
 LQALRVYVHEDNTVKINPLMCSPLSADFDGDCVHLFYPQSLSAKAEVMELFSVEKQLLS
 SHTGQLIQMGSDSLLSLRVMLERVFLDKATAQQLAMYGSLSLPPPALRKSSKSGPAWT
 VFQILQLAFPERLSCKGDRFLVDGSDLLKDFDGV DAMGSINEIVTSIFLEKGPKETLGFFD
 SLQPLLMESLFAEGFSLSLEDLSMSRADMDVIHNLIREISPMVSRLRLSYRDELQLENSIH
 KVKEVAANFMLKSYSIRNLIDIKSNSAITKL VQQTGFLGLQLSDK KKFYTKTLVEDMAIF
 CKRKYGRISSSGDFGIVKGCFFHGLDPYEEMAHSIAAREVIVRSSRGLAEPGTLFKNLMA
 VLRDIVITNDGTVRNTCSNSVIQFKYGVDSERGHQGLFEAGEPVGVLAATAMSNPAYKA
 VLDSSPNSNSSWELMKEVLLCKVNFQNTTNDRRVILYLNECHCGKRFCQENA ACTVRN
 KLNK VSLKDTAVEFLVEYRKQPTISEIFGIDSCLHGHHLNK TLLQDWNISMQDIHQKCE
 DVINSLGQKKKKATDDFKRTSLSVSECCSFRDPCGSKGSDMPCLTFSYNATDPDLERT
 LDVLCNTVYPVLLLEIVIKGDSRICSANIWNSSDMTTWIRNRHASRRGEWVLDVTVEKSA
 VKQSGDAWRVVIDSCLSVLHLIDTKRSIPYSVKQVQELLGLSCAFEQAVQRLSASVRMV
 SKGVLKEHILLANNMTCSGTMLGFNSGGYKALTRSLNIKAPFTEATLIAPRKCFEKAAE
 KCHTDSLSTVVGSCSWGKRVDVGTGSQFELLWNQKETGLDDKEETDVYSFLQMVISTT
 NADAFVSSPGFDVTEEMA EWAE SPERDSALGEPKFEDSADFQNLHDEGKPSGANWEK
 SSSWDNGCSGGSEWGVSKSTGGEANPESNWEKTTNVEKEDAWSSWNTRKDAQESSKS
 DSGGAWGIKTKDADADTPNWETSPAPKDSIVPENNEPTSDVWGHKSVSDKSWDKKN
 WGTESAPAAWGSTDAAVWGSSDKKNSETESDAAAWGSRDKNNSDVGSGAGV LGPWN
 KKSSETESNGATWGSSDKTKSGAAAWNSWDKKNIEDSEPAAWGSGQKKNSETESGP
 AAWGAWDKK KSETEPGPAGWGMGDKK NSETELGPAAMGNWDK KKS DTKSGPAAWG
 STDAAAWGSSDKNNSETESDAAAWGSRNKK TSEIESGAGAWG SWGQPSPTAEDKDTN
 EDDRNPWVSLKETKSREKDDKERSQWGNPAKKFPSSGGWSNGGGADWKGNRNHTPR
 PPRSEDNLAPMFTATRQRLDSFTSEEQELLS DVEPVMRTLKIMHPSAYPDGDPISDDDK
 TFVLEKILNFHPQKETKLGSGVDFITVDKHTIFS DSR CFFV VSTDGAKQDFSYR KSLNNY
 LMKKYPDRAEEFIDKYFTKPRPSGNRDRNNQDATPPGEEQSQPPNQSIGNGGDDFQTQT
 QSQSPSQT RAQSPSQAQAQSPSQTQSQS QSQSQSQSQSQSQSQSQSQSQSQSQSQSQS
 TQTQSPSQTQAQAQSPSSQSPSQTQT

Notes:

1457/1976 amino acids are represented by sequenced peptides =74% coverage.

All peptides are specific to NRPE1 (NRPD1b), meaning that none are identical to any other protein, including NRPD1 (NRPD1a).

NRPE2/NRPD2 (At3g23780)

MPDMDIDVKDLEEFEATTGEINLSELGEGFLQSFCKKAATSFFDKYGLISHQLNSYNYFI
 EHGLQNVFQSFGEMLVEPSFDVVKKKDNDWRYATVKFGEVTVEKPTFFSDDKELEFLP
 WHARLQNMTYSARIKVVNVQVEVFKNTVVKSDKFKTGQDNYVEKKILDVKKQDILIGSI
 PVMVK SILCKTSEK GKENCKKGDCAFDQGGYFVIKGAEKVFIAQEQMCTKRLWISNSP
 WTVSFRSENKRNRNFIVRLSENEKAEDYKRREKVLTVYFLSTEIPVWLLFFALGVSSDKEA
 MDLIAFDGDDASITNSLIASIHVADAVCEAFRCGNNALTYVEQQIKSTKFPPAESVDECL
 HLYLFPGLQSLK KKARFLGYMVKCLLNSYAGKRKCENRDSFRNKRIELAGELLEREIRV
 HLAHARRKMTRAMQK HLSGDGDLKPIEHYLDASVITNGLSRAFASTGAWSHPFKMERV
 SGVVANLGRANPLQTLIDLRRTTRQQVLVTGKVGDARYPHPSHWGRVCFLSTPDGENCG
 LVKNMSLLGLVSTQSLESVVEKLFACGMEELMDDTCTPLFGKHKVLLNGDWVGLCAD
 SESFVAELKSRRRQSELPREMEIKRDKDDNEVRIFTDAGRLLRPLL VVENLQK LKQEKPS
 QYPFDHLLDHGILELIGIEEEEDCNTAWGIKQLLKEPKIYTHCELDSFLLGVSCAVVPPFA
 NHDHGRVLYQSQK HCQQAIGFSSTNPINRCDTLSQQLFYPQKPLFKTLASECLKKEVLF
 NGQNAIVAVNVHLGYNQEDSIVMNKASLERGMFRSEQIRSYKAEVDAKDSEKRRKMD
 ELVQFGKTHSKIGKVDSLEDDGFPIGANMSTGDIVIGRCTESGADHSIKLKHTEGIVQK
 VVLSNDEGKNFAAVSLRQVRSPCLGDKFSSMHGQKGVLGYLEEQQNFPFTIQGIVPDI
 VINPHAFPSRQTPGQLLEAALS KGIACPIQKEGSSAA YTKLTRHATPFSTPGVTEITEQLH
 RAGFSRWGNERNVYNGRSGEMMRSMIFMGPTFYQRLVHMSSEDKVKFRNTGPVHPLTRQ
 PVADRKRFGGIKFGEMERDCLIAHGASANLHERLFTLSDSSQMHCRCCKTYANVIERTP
 SSGRKIRGPYCRVCVSSDHVVRVYVPYGAKLLCQELFSMGITLNFDTKLC

Notes:

434/1172 amino acids represented in sequenced peptides =37% coverage.

155/1172= 13% coverage is accounted for by peptides unique to NRPE2/NRPD2a. The remaining 24% of the peptides match NRPE2/NRPD2a as well as the NRPD2b pseudogene. However, the latter gene is non-functional, and no peptides that would uniquely identify NRPD2b were detected.

NRPE3a/NRPD3/NRPB3 (At2g15430)

MDGATYQRFPKIKIRELKDDYAKFELRETDVSMANALRVMISEVPTVAIDLVEIEVNSS
 VLNDEFIAHRLGLIPLTSERAMSMRFSRDACDGDGQCEFCSEFRLSSKCVTDQTLT
 VTSRDLYSADPTVTPVDFTIDSSVSDSSEHKGIIIIVKLRRGQELKLRAIARKGIGKDHAKW
 SPAATVTFMYEPDIINEDMMDTSLDDEEKIDLISSPTKVFGMDPVTRQVVVVDPPEAYTY
 DEEVIKKAEMGK PGLIEISPKDDSFIFTVESTGAVKASQLVLNAIDLKQKLDVRLSD
 DTVEADDQFGELGAHMRGG

Notes:

184/319 amino acids are represented by sequenced peptides =58% coverage

45% of the coverage corresponds to peptides that match only NRPE3a. The other 13% matches either NRPE3a or NRPE3b.

NRPE3b (At2g15400)

MDGVTYQRFPVTKIRELKDDYAKFELRETDVSMANALRRVMISEVPTMAIHLVKIEVNS
 SVLNDEFIAQRLLSLIPLTSERAMSMRFCQDCEDCNGDEHCEFCSEVFPPLSAKCVTDQTLT
 VTSRDLYSADPTVTPVDFTSNSSSTSDSSEHKGIIIAKLRRGQELKALKALARKGIGKDHAK
 WSPAATVTYMYEPDIINEEMMNTLTDEEKIDLISSPTKVFGIDPVTGQVVVVDPEAYT
 YDEEVIKKAEAMGKPGLEIHPKHDSFVFTVESTGALKASQLVLNAIDILKQKLDAIRLSD
 NTVEADDQFGELGAHMREG

Notes:

170/319 amino acids are represented by sequenced peptides = 53% coverage

131/319=41% coverage corresponds to peptides matching only NRPE3b, whereas the remaining 12% of the coverage matches either NRPE3b or NRPE3a.

NRPD4/NRPE4 (At4g15950)

MSEKGGKGLKSSLKSKDGGKDGSSSTKLKGRKIHFQDQTPPANYKILNVSSDQQPFQSS
 AAKCGKSDKPTKSSKNSLHSFELKDLPENAECEMMDCEAFQILDGIKQGLVGLSEDPSIKI
 PVSYDRALAYVESC VHYTNPQSVRKVLEPLKTYGISDGEMCVIANASSESVDEVLAIFIPS
 LKTKKEVINQPLQDALEELSKLKKSE

17/205=8% coverage

All peptides are unique matches to NRPD4 only.

NRPB4 (At5g09920)

MSGEEEEENAAELKIGDEFLKAKCLMNCEVSLILEHKFEQLQQISEDPMNQVSQVFEKSL
 QYVKRFSRYKNPDAVRQVREILSRHQLTEFELCVLGNLCPETVEEAVAMVPSLKTKGRA
 HDDEAIEKMLNDLSLVKRFE

0/138=0% coverage

No peptides were identified that match this protein sequence.

NRPB5/NRPD5 (At3g22320)

MLTEEELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNMKREDLVTLKA
 KRNDNSDQLYIFFPDEAKVGVKTMKMYTNRMKSENVFRAILVVQQNLTPFARTCISEIS
 SKFHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEKKTLLERYTVKETQLPRIQVTDPIARY
 FGLKRGQVVKIIRPSETAGRYVTYRYVV

0/205 amino acids are represented by sequenced peptides = 0% coverage

No peptides were identified that matched this protein sequence.

NRPE5 (formerly AtRPB5b, AtRPB23.7) (At3g57080)

MEVKGKETASVLSK YVDLSSEESHRYYLARRNGLQMLRDRGYEVSDDEDINLSLHDF
 RTVYGERPDVDRRLRISALHRSDSTKKVKIVFFGTSMVKVNAIRSVVADILSQETITGLILV
 LQNHVTNQALKAIELFSFKVEIFQITDLLVNITKHSLKPQHQVLNDEEKTLLKKFSIEEK
 QLPRISKKDAIVRYYGLEKGVVKNYRGELTESHVAFRCVW

86/222 amino acids are represented by sequenced peptides = 39% coverage

All peptides identified correspond to peptides that match NRPE5 only and no other family member.

NRPB5 family member (At5g57980)

MSDMDDEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCCTMKNVNKEALF
VSANKGPNPADKIYVFYPEGPKVGVVPIKKEVAIKMRDDKVHRGIVVVPMAITAPARM
AVSELNKMILTIEVFEEAELVTNITEHKLVNKYVVLDDQAKKKLLNTYTVQDTQLPRILV
TDPLARYYGLKRGQVVKIRRS DATSLDYTYRFAV

0/210 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE5-like family member (At2g41340)

MEGKGKEIVVGHISISKSSVECHKYLLARRTTMEMLRDRGYDVSDDEDINLSLQQFRALY
GEHPDVDLLRISAKHRFDSSKKISVFCGTGIVKVNAMRVIAADVLSRENITGLILVLQS
HITNQALKAVELFSFKVELFEITDLLVNVSKHVLPRKHQVLNDKEKESLLKKFSIEEKQL
PRLSSKDPIVRYYGLETGQVMKVITYKDELSESHVITYRCVS

0/218 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE5-like family member (At3g54490)

MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNE
AELSLTLSEFRSVFGEKPELERLRICVPLRSDPKKKILVVMGTEPITVKSVRALHIQISNN
VGLHAMILVLQSKMNHFAQKALTFPFTVETFPIDLLVNITKHIQPKIEILNKEEKEQL
LRKHALEDKQLPYLQEKDSFVRYYGLKKKQVVKITYSKEPVGDFVITYRCII

0/233 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPB5 family member (likely pseudogene) (At3g16880)

MKKYIDQLKSANVFRAILVVQDIKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKEHVF
VPEHQALTTEEKQKFLERKRTSFQGFT

0/87 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence. This protein is truncated relative to the other NRPB5-like proteins and likely is a pseudogene.

NRPE6a/NRPB6a/NRPD6a (At5g51940)

MADEDYNDVDDLGYEDEPAEPEIEEGVEEDVEMKENDDVNGEPIEAEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRR^{YL}
PDGSFEEWGVDELIVEDSWK^{RQ}VGGD

48/144 amino acids are represented by sequenced peptides = 33% coverage
22/144 = 15% coverage corresponds to peptides that are NRPE6a-specific, whereas the remaining 18% match either NRPE6a or NRPE6b.

NRPE6b/NRPB6b (At2g04630)

MADDDYNEVDDLGYEDEPAEPEIEEGVEEDADIKENDDVNVDPLETEDKVETEPVQRP
 RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
 PDMSYEEWGVDELIVEDSWKRQVGGD

26/144 amino acids are represented by sequenced peptides = 18% coverage
 0/144=0% of the coverage corresponds to peptides unique to this member of the protein family;
 the sequenced peptide also matches an identical sequence of At5g51940.

NRPE7 (At4g14660)

MFLK VQLPWNVMIPAENMDAKGLMLKRAILVELLEAFASKKATKELGYYVAVTTLDKI
 GEGKIREHTGEVLPVPMFSGMTFKIFKGEIIHGTVVHKVVKHGVFMRCGPIENVYLSYTK
 MPDYK YIPGENPIFMNEKTSRIQVETTVRVVIGIKWMEVEREFQALASLEGDYLGPLSE

58/177 amino acids are represented by sequenced peptides = 33% coverage
 All peptides match At4g14660 and only At4g14660.

NRPB7 (At5g59180)

MFFHIVLERNMQLHPRFFGRNLKENLVSKLMKDVEGTCSGRHGFVVAITGIDTIGKGLIR
 DGTGFVTFPVKYQC VVFRPFKGEILEAVVTLVNKMGGFAEAGPVQIFVSKHLIPDDMEF
 QAGDMPNYTTS DGSVKIQKECEVRLKIIGTRV DATAIFCVGTIKDDDFLGVINDPAAA

0/176 amino acids are represented by sequenced peptides = 0% coverage
 No peptides were identified that match this protein sequence.

NRPD7 (At3g22900)

MFIKVKLPWDVTIPAEDMDTGLMLQRAIVIRLLEAFSKEKATKDLGYLITPTILENIGEGK
 IKEQTGEIQFPVVFNGICFKMFKGEIVHGTVVHKVHKTVFLKSGPYEIIYLSHMKMPGYE
 FIPGENPFFMNQYMSRIQIGARVRFVLDTEWREAEKDFMALASIDGDNLPGF

0/174 amino acids are represented by sequenced peptides = 0% coverage
 No peptides were identified that matched this protein sequence.

NRPB7 family member (At4g14520)

MFSEVEMARDVAICAKHLNGQSPHQILCROLLQDLIHEKACREHGFYLGITALKSIGNNK
 NNNIDNENNHQAKILTFPVSFTCRFLPARGDILQGTVKKVLWNGAFIRSGPLRYAYLSL
 LKMPHYHYVHSPLESEDEKPHFQKDDLSKIAVGVVVRFQVLA VRFKERPHKRRNDYYVL
 ATLEGNGSFGPISLTGSDEPYM

0/200 amino acids are represented by sequenced peptides = 0% coverage
 No peptides were identified that matched this protein sequence.

NRPE8a/NRPB8a/NRPD8a (At1g54250)

MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAVGDKF
 TLALAPTLNLDGTPDTGYFTPGAKK TLADKYEYIMHGKLYKISERDGTKPKAELYVSFG
 GLLMLLKGDPAHISHFELDQRLFLLMRKL

13/146 amino acids are represented by the sequenced peptide = 9% coverage

0/146=0% of the coverage corresponds to peptides unique to this member of the protein family.
This peptide also is an exact match to At3g59600.

NRPE8b/NRPD8b/NRPB8b (At3g59600)

MASNIIMFEDIFVVDKLPDGGKFDKVTREARSHNLEMFHLDVNTEVYPLAVGDKF
TLAMAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTPKAELYVSFG
GLLMLLQGDPAHISHFELDQRLFLLMRKL

13/146 amino acids are represented by the sequenced peptide = 9% coverage
0/146 = 0% of the coverage corresponds to peptides unique to this member of the protein family.
This peptide is also an exact match to At1g54250.

NRPE9a/NRPD9a/NRPB9a (At3g16980)

MSTMKFCRECNILYPKEDKEQKILLYACRNCDHQEVADNSCVYRNEVHHSVSERTQIL
TDVASDPTLPRTKAVRCSKCQHREAVFFQATARGEEMTLFFVCCNPNCGHRWRE

25/114 amino acids are represented by sequenced peptides = 22% coverage
25/114 = 22% coverage corresponds to peptides unique to this member of the protein family.
Two amino acid differences in the identified peptide (underlined) discriminates At3g16980 from
At4g16265.

NRPE9b/NRPD9b/NRPB9b (At4g16265)

MSTMKFCRECNILYPKEDKEQSILLYACRNCDHQEAADNNVCVYRNEVHHSVSEQTQIL
LSDVASDPTLPRTKAVRCAKQCQHGAVFFQATARGEEMTLFFVCCNPNCSTRWRE

25/114 amino acids are represented by the sequenced peptide = 22% coverage
25/114 = 22% coverage corresponds to peptides unique to this member of the protein family.
Two amino acid differences in the identified peptide (underlined) discriminates At3g16980 from
At4g16265.

NRPE10/NRPD10/NRPB10 (At1g11475)

MIIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRRMLMTHVDLIEK
LLNYNTLEKSDNS

50/71 amino acids are represented by sequenced peptides = 70% coverage
39/71= 55% coverage corresponds to peptides that only match this protein, whereas the
remaining 15% match either At1g11475 or At1g61700.

NRPB10 family member (At1g61700)

MIVPVRCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRRMLMTHVDLIE
KLLNYNTMEKSDPN

11/71 amino acids are represented by the sequenced peptide = 15% coverage
0/71= 0% unique. The peptide identified for At1g61700 also matches At1g11475.

NRPE11/NRPD11/NRPB11 (At3g52090)

MNAPERYERFVVEGTTKVSYDRDTKIINAASFTVEREDHTIGNIVRMQLHRDENVLFA
 GYQLPHPLKYKIIIVRIHTTSQSSPMQAYNQAINDLKELDYLNQFEAEVAKFSNQF

79/116 amino acids are represented by sequenced peptides = 68% coverage
 All peptides identified match NRPE11 and only NRPE11.

NRPE12/NRPD12/NRPB12 (At5g41010)

MDPAPEPVTYVCGDCGQENTLKSVDVIQCRCGYRILYKKRTRRVVQYEAR

8/51 amino acids are represented by the sequenced peptide = 16% coverage
 The peptide is unique to this protein.

NRPB12 family member (At1g53690)

MDLQQSETDDKQPEQLVIYVCGDCGQENILKRGDVFQCRDCGFRILYKKRILDKKETRI
 GV

0/62 amino acids are represented by sequenced peptides = 0% coverage
 No peptides were identified that matched this protein sequence.

Figure S2. Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS

analysis of affinity purified Pol IV (NRPD1-FLAG). Highlighting is the same as in Fig. S1.

NRPD1 (At1g63020)

MEDDCEELQVPVGTLSIGFSISNNDRDKMSVLEVEAPNQVTDSRLGLPNPDSVCRITC
 GSKDRKVCEGHFGVINFAYSIINPYFLKEVAALLNKICPGCKYIRK KQFQITEDQPERCRY
 CTLNTGYPLMKFRVTTKEVFRRS GIVVEVNEESLMKLLKRGVLTLPDYWSFLPQDSNI
 DESCLKPTRR IITHAQVYALLL GIDQRLIKK DIPMFNSLGLTSFPVTPNGYR VTEIVHQFN
 GARLIFDERTRIIYKKLVGFEGNTLELSSR VMECMQYSRLFSETVSSSK DSANPYQKSDT
 PKLCGLRFMKDVLLGKRS DHTFRTVVVG DPSLKLNEIGIPESIAK RLQVSEHLNQCNER
 LVTSFVPTLLDNKEMHVRRGDRLVAIQVNDLQTDGKIFRSLMDGDTVLMNRPPSIHQHS
 LIAMTVRILPTTSVVS LN PICCLPFR GDFDGDCLHGYVPQSIQAKVELDELVALDKQLINR
 QNGR NLLSLGQDSLTAAYLVNVEKNCYL NRAQMQLQMYCPFQLPPPAIKASPSSTEP
 QWTGMQLFGMLFPPGFDYTYPLNNVVV SNGELLSFSEGS AWLRDGE GNFIERLLKHDK
 GK VLDIISAQEMLSQWLLMRGLSVSLADLYLSSDLQSRKNL TEEISYGLREAEQVCNK
 QQLMVESWRDFLAVNGEDKEEDSVSDLARFCYERQK SATLSELAVSAFKDAYRDVQA
 LAYRYGDQSN SFLIMSKAGSKGNIGKLVQHSMCIGLQNSAVSLSFGFPRELTCAA WNDP
 NSPLRGAK GK DSTTTESYVPYGV IENSFLTGLNPLESFVHSVTSR DSSFSGNADLPGLTSR
 RLMFFMR DIYAA YDGTVRNSFGNQLVQFTYETDGPVEDITGEALGSL SACALSEAAYSA
 LDQPISLLETSPLLNLK NVLECGSKKGQREQTMSLYLSEYLSK KKHGFYGSLEIKNHLE
 KLSFSEIVSTSMIIFSPSSNTK VPLSPWVCHFISEK VLKRK QLSAESVVS SLNEQYKSRNR
 ELKLDIVDLDIQNTNHCSDDQAMKDDNVCITVTVVEASKHSVLELDAIRLVLIPFLLD S
 PVKGDQGIK KVNILWTD RPKAPKRNGNHLAGELYLKVTMYGDRGKRNCWTALLETCL
 PIMDMIDWGRSHPDNIRQCCSVY GIDAGRSIFVANLES AVSDTGK EILREHLLLVADSL S
 VTGEFVALNAK GWSKQRQVESTPAPFTQACFSSPSQCFLKAAKEGVRDDLQGSIDALA
 WGK VPGFGTGDQFEIISP K VHGFTTPVDVYDLLSSTK TMRRTNSAPKSDKATVQPFGLL

HS AFLK DIK VLDGKGIPMSLLRTIFTWK **NIELLSQSLKRILHSEINELLNERDEGLVKMV**
LQLHPNSVEKIGPGVKGIRVAKSKHGDSCCFEVVRIDGTFEDFSYHK CVLGATKIIAPKK
 MNFYKSKYLKNGTLESGGFSEN

844/1453 amino acids are represented by sequenced peptides =58% coverage

All peptides are specific to NRPD1 (NRPD1a), meaning that none are identical to any other protein, including NRPE1 (NRPD1b).

NRPD2/NRPE2 (At3g23780)

MPDMDIDVKDLEEFATTGEINLSELGEGFLQSFCKKAATSFFDKYGLISHQLNSYNYFI
 EHGLQNVFQSFGEMLVEPSFDVVKKKDNDWRYATVKFGEVTVEKPTFFSDDKELEFLP
 WHAR**LQNMTYSAR**IKVNVQVEVFNKTVVKSDFKFTGQDNYVEKKILDV**KQDILIGSI**
PVMVKSILCKTSEKGENCKKGDCAFQQGGYFVIKGAEKVFIAQEQMCTKRLWISNSP
 WTVSFRSENKRNRFRVRLSENEKAEDYKRREKVLTVYFLSTEIPVWLLFFALGVSSDKEA
 MDLIAFDGDDASITNSLIASIHVADAVCEAFRCGNNALTYVEQQIKSTKFPPAESVDECL
 HLYLFPGLQSLKKARFLGYMVKCLLSYAGKRKCENRDSFRNK**RIELAGELLER**EIRV
 HLAHARRKMTRAMQK**HLSGDGLKPIEHYLDASVITNGLSR**AFSTGAWSHPF**KMERV**
SGVVANLGRANPLQTLIDLRRT**QQVLYTGK**VGDARYPHPSHWGRVCFSTPDGENCG
 LVKNMSLLGLVSTQSLESVVEKLFACGMEELMDDTCTPLFGKHK**VLLNGDWVGLCAD**
SESFVAELKSRRRQSELPREMEIKRDKDNEVRIFTDAGRLLRPLL VVENLQKQKQEKPS
 QYFHDHLLDHGILELIGIEEEEDCNTAWGIKQLLKEPKIYTHCELDLSFLLGVSCAVVPA
 NHDHGRVLYQSQKHCQAIGFSSTNPNI²CDTSLQQLFYPQKPLFKTLASECLKKEVLF
 NGQNAIVAVNVHLGYNQEDSIVMNKASLERGMFRSEQIRSYKAEVDAKDSEK**KKMD**
ELVQFGKTHSKIGK**VDSLEDDGFPIGANMSTGDIVIGR**CTESGADHSIKLKHTEGIVQK
 VVLSNDEGKNFAAVSLRQVRSPCLGDKFSSMHGQK**GVLGYLEEQQNFPFTIQGIVPDI**
VINPHAFPSRQTPGQLLEAALSKGIACPIQKEGSSAA YTKLTRHATPFSTPGVTEITEQLH
 RAGFSRWGNERNVNGRSGEMMRSMIFMGPTFYQRLVHMS**EDKVKFRNTGPVHPLTRQ**
 PVADRKRFGGIKFGEMERDCLIAHGASANLHER**LFTLSDS**SMHICRCK**TYANVIER**TP
 SSGRKIRGPYCRVCVSSDHVVRVYVPYGA**LLCQELFSMGITLNFDTKLC**

211/1172=18% coverage

48/1172=4% coverage is accounted for by peptides unique to NRPE2/NRPD2a. The remaining 14% of the peptides match NRPE2/NRPD2a as well as the NRPD2b pseudogene. However, the latter gene is non-functional, and no peptides that would uniquely identify NRPD2b were detected.

NRPD3/NRPE3a/NRPB3 (At2g15430)

MDGATYQRFPKIKIRELKDDYAKFEL**RETDVSMANALR**RVMISEVPTVAIDLVEIEVNSS
 VLNDEFIAHR**LGLIPLTSER**AMSMRFSRDCDACDGDGQCEFCVSEFRLSSKCVTDQTL
 VTSR**DLYSADPTVTPVDFTIDSSVSDSSEHK**GIIIVKLRRGQELKLRAIARKGIGKDHAKW
 SPAATVTFMYEPDIINEDMMDTSLDDEEKIDLISSPTK**VFGMDPVTR**QVVVVDPEAYTY
 DEEVIKAEAMGK**PGLIEISPK****DDSFIFTVESTGAVK**ASQLVLNAIDLKQKLDVRLSD
DTVEADDQFGELGAHMRGG

101/319=32% coverage

90/319=28% coverage is accounted for by peptides unique to NRPD3. The remaining 4% of the peptides match NRPD3 as well as the NRPD3b variant.

NRPE3b (At2g15400)

MDGVTYQRFPVTKIRELKDDYAKFELR**ETDVSMANALRR**VMISEVPTMAIHLVKIEVNS
 SVLNDEFIAQRLSLIPLTSERAMSMRFCQDCEDCNGDEHCEFCSEVFPPLSAKCVTDQTLT
 VTSRDLYSADPTVTPVDFTSNSSSTSDSSEHKGIIIAKLRRGQELKALKALARKGIGKDHAK
 WSPAATVTYMYEPDIINEEMMNTLTDEEKIDLIESSPTKVFGIDPVTGQVVVDPEAYT
 YDEEVIKKAEAMGKPGLEIHPKHDSFVFTVESTGALK**ASQLVLNAIDILK**QKLDAIRLS
 NTVEADDQFGELGAHMREG

24/319=8% coverage

13/319=4% coverage is accounted for by peptides unique to NRPD3b. The remaining 4% of the peptides match NRPD3 as well as the NRPD3b variant.

NRPD4/NRPE4 (At4g15950)

MSEKGGKGLKSSLKSKDGGKDGSSSTKLKKGKRIHFDQGTTPANYK**ILNVSSDQQPFQSS**
AAKCGKSDKPTKSSK**NSLHSFELK**DLPENAECMMDCFAFQILDGIKQGLVGLSEDPSIKI
 PVSYDRALAYVESC VHYTNPQSVRKVLEPLKTYGISDGEMCVIANASSESVDEVLAIFIPS
 LKTKKEVINQPLQDALEELSKLKKSE

26/205=13% coverage

All peptides are unique matches to NRPD4/NRPE4 only.

NRPB4 (At5g09920)

MSGEEEEENAELKIGDEFLKAKCLMNCEVSLILEHKFEQLQQISEDPMNQVSQVFEKSL
 QYVKRFSRYKNPDAVRQVREILSRHQLTEFELCVLGNLCPETVEEAVAMVPSLKTKGRA
 HDDEAIEKMLNDLSLVKRFE

0/138=0% coverage

No peptides were identified that match this protein sequence.

NRPB5/NRPD5 (formerly AtRPB5a, AtRPB24.3) (At3g22320)

MLTEEELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNMK**REDLVTLKA**
 KRNDNSDQLYIFFPDEAKVGVKTMKMYTNRMKSENVFR**AILVVQQNLTPFARTCISEIS**
 SKFHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEKKTLLERYTVKETQLPRI**QVTDPIARY**
 FGLKRGQVVKIIRPSETAGRYVTYRYVV

31/205=15% coverage

All peptides are unique matches to NRPB5/NRPD5 only.

NRPE5 (formerly AtRPB5b, AtRPB23.7) (At3g57080)

MEVKGKETASVLCLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDEDINLSLHDF
 RTVYGERPDVDRRLRISALHRSDSTKKVKIVFFGTSMVKVNAIRSVVADILSQETITGLILV
 LQNHVTNQALKAIELFSFKVEIFQITDLLVNITKHSKLPQHQLNDEEKTLLKKFSIEEK
 QLPRISKKDAIVRYYGLEKGQVVKVNYRGELTESHVAFRVCVW

0/222=0% coverage

No peptides were identified that match this protein sequence.

NRPB5-like family member (At5g57980)

MSDMDDEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCCKTMNKVNKEALF
 VSANKGPNPADKIYVFYPEGPKVGVVVIKKEVAIKMRDDKVHRGIVVVPMAITAPARM
 AVSELNKMILTIEVFEEAELVTNITEHKLVNKYVVLDDQAKKLLNTYTVQDTQLPRILV
 TDPLARYYGLKRGQVVKIRRSDATSLDYTYRFAV

0/210=0% coverage

No peptides were identified that match this protein sequence.

NRPE5-like family member (At2g41340)

MEGKGKEIVVGHISISKSSVECHKYLLARRTTMEMLRDRGYDVSDDEDINLSLQQFRALY
 GEHPDVDLLRISAKHRFDSSKKISVFCGTGIVKVNAMRVIAADVLSRENITGLILVLQS
 HITNQALKAVELFSFKVELFEITDLLVNVSKHVLPRKHQVLNDKEKESLLKKFSIEEKQL
 PRLSSKDPIVRYYGLETGQVMKVITYKDELSESHVITYRCVS

0/218=0% coverage

No peptides were identified that match this protein sequence.

NRPE5-like family member (At3g54490)

MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNE
 AELSLTLSEFRSVFGEKPELERLRICVPLRSDPKKILVFMGTEPITVKSVRALHIQISNN
 VGLHAMILVLQSKMNHFAQKALTFPFTVETFPIDLLVNITKHIQQPKIEILNKEEKEQL
 LRKHALEDKQLPYLQEKDSFVRYYGLKKKQVVKITYSKEPVGDFVITYRCII

0/233=0% coverage

No peptides were identified that match this protein sequence.

NRPB5 family member (likely pseudogene) (At3g16880)

MKKYIDQLKSANVFRAILVVQDIKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKEHVF
 VPEHQALTTEEKQKFLERKRTSFQGFT

0/87=0% coverage

No peptides were identified that match this protein sequence.

NRPB6a/NRPD6a/NRPE6a (At5g51940)

MADEDYNDVDDLGYEDEPAEPEIEEGVEEDVEMKENDDVNGEPIEAEDKVETEPVQRP
 RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRR^{YL}
 PDGSFEEWGVDELIVEDSWKRQVGGD

48/144=33% coverage

22/144=15% unique

NRPB6b/NRPE6b (At2g04630)

MADDDYNEVDDLGYEDEPAEPEIEEGVEEDADIKENDDVNVDPLETEDKVETEPVQRP
 RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
 PDMSYEEWGVDELIVEDSWKRQVGGD

26/144=18% coverage

0/144=0% unique

This peptide is not a unique match to NRPB6b—it matches either NRPB6a or NRPB6b.

NRPE7 (At4g14660)

MFLKVQLPWNVMIPAENMDAKGLMLKR **AILVELLEAFASK**KATKELGYYYVAVTTLDKI
GEGKIREHTGEVLFVPMFSGMTFKIFKGEIIHGVVHKVLKHGVMRCGPIENVYLSYTK
MPDYKYIPGENPIFMNEKTSRIQVETTVRVVIGIKWMEVEREFQALASLEGDYLGPLSE

13/177=9% coverage

This peptide matches NRPE7 only. This protein might sometimes be used as an alternative NRPD7 subunit.

NRPD7 (At3g22900)

MFIK**VKL**PWDVTIPAEDMDTGLMLQRAIVIR **LLEAF**SK**E**KATK**DLGYLITPTILENIGEGK**
IKEQTGEIQFPVVFNGC**FK**MFK**GEIVHG**V**VHK**VHKTGVFLK**SGPYEIIYLSHMK**MPGYE
FIPGENPFFMNQYMSRIQIGARVRFVVLDTREWAEKDFMALASIDGDNLGPF

90/174=52% coverage

These peptides match NRPD7 only.

NRPB7 family member (At4g14520)

MFSEVEMARDVAICAKHLNGQSPHQILCRLQLDLIHEKACREHGFYLGITALKSIGNNK
NNNIDNENNHQAKILTFPVSFTCRTFLPARGDILQGTVKKVLWNGAFIRSGPLRYAYLSL
LKMPHYHYVHSPLESEDEKPHFQKDDLSKIAVGVVVRFQVLAVRFKERPHKRRNDYYVL
ATLEGNGSFGPISLTGSDEPYM

0/200=0% coverage

No peptides were identified that match this protein sequence.

NRPB7 (At5g59180)

MFFHIVLERNMQLHPRFFGRNLKENLVSKLMKDVEGTCSGRHGFVVAITGIDTIGKGLIR
DGTGFVTFPVKYQC VVFRPFKGEILEAVVTLVNKMGGFAEAGPVQIFVSKHLIPDDMEF
QAGDMPNYTTS DGSVKIQKECEVRLKIIGTRV DATAIFCVGTIKDDDFLGVINDPAAA

0/176=0% coverage

No peptides were identified that match this protein sequence.

NRPD8a/NRPE8a/NRPB8a (At1g54250)

MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAVGDKF
TLALAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGTKPKAELYVSFG
GLLMLLKGDPAHISHFELDQRLFLLMRKL

0/146=0% coverage

No peptides were identified that match this protein sequence.

NRPD8b/NRPE8b/NRPB8b (At3g59600)

MASNIIMFEDIFVVDKLPDYGKKFDKVTREARSHNLEMFMHLDVNTEVYPLAVGDKF
 TLAMAPTLNLDGTPDTPGYFTPGAKKTLADKYEYIMHGKLYKISERDGTKPKAELYVSFG
 GLLMLLQGDPAHISHFELDQRLFLLMRKL

25/146=18% coverage

This peptide is a unique match to At3g59600.

NRPD9a/NRPE9a/NRPB9a (At3g16980)

MSTMKFCRECNILYPKEDKEQKILLYACRNCDHQEVADNSCVYRNEVHHSVSERTQIL
 TDVASDPTLPRTKAVRCSKCQHREAVFFQATARGEEMTLFFVCCNPNCGHRWRE

0/114=0% coverage

No peptides were identified that match this protein sequence.

NRPD9b/NRPE9b/NRPB9b (At4g16265)

MSTMKFCRECNILYPKEDKEQSILLYACRNCDHQEAADNNCVYRNEVHHSVSEQTQI
 LSDVASDPTLPRTKAVRCAKQCQHGEAVFFQATARGEEMTLFFVCCNPNCshrWRE

25/114=22% coverage

This peptide is a unique match to NRPD9.

NRPD10/NRPE10/NRPB10 (At1g11475)

MIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRRMLMTHVDLIEK
 LLNYNTLEKSDNS

39/71=54% coverage

Both peptides are a unique match to NRPD10.

NRPB10 family member (At1g61700)

MIPVRCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRRMLMTHVDLIE
 KLLNYNTMEKSDPN

0/71=0% coverage

No peptides were identified that matched this protein sequence.

NRPD11/NRPB11/NRPE11 (At3g52090)

MNAPERYERFVPEGTTKVSYDRDTKIINAASFTVEREDHTIGNIVR MQLHRDENLFA
 GYQLPHPLKYKIIIVRIHTTSQSSPMQAYNQAINDLKELDYLNQFEAEVAKFSNQF

65/116=56% coverage

All peptides are a unique match to NRPD11.

NRPB12/NRPD12/NRPE12 (At5g41010)MDPAPEPVTYVCGDCGQENTLKSQDVIQCRCGYRILYKKRTRR **VVQYEAR**

7/51=16% coverage

This peptide is unique to At5g41010.

NRPB12 family member (At1g53690)MDLQQSETDDKQPEQLVIYVCGDCGQENILKRGDVFQCRDCGFRILYKKRILDKKETRI
GV

0/62=0% coverage

No peptides were identified that matched this protein sequence.

Figure S3. Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS

analysis of affinity purified Pol II (NRPB2-FLAG). Highlighting is the same as in Fig. S1.

NRPB1 (At4g35800)

MDTR**FPFSPA**EVSKVR**VVQFGILSPDEIR**QMSVIHVEHSETTEK**GKPK****VGGLSDTRLGTI**
DRKVKCETCMANMAECPGHFGYLELAKPMYHVGFMK**TVLSIMR**CVCFNCSKILADEV
 CRSLFRQAMKIKNPKNRLKKILDACKNKTKCDGGDDIDDVQSHSTDEPVKK**SRGGCGA**
 QPKLTIEGMKMIAEYKIQR**KKNDEPDQLPEPAER**KQTLGADR**VLSVLKRISDADCQLL**
GFNPK**FARPDWM**ILEVLPPIPPPPVRPSVMMDATSR**SEDDLTHQLAMI**IRHNENLKRQEK
NGAPAHIIEFTQLLQFHIA**TYFDNELPGQPRATQKSGRPIKSIC**SRLKAKEGRIRGNLMG
KRVDFSAR**TVITPDPTINIDELGVPWSIALNLTYPETVTPYNIERLK**ELVDYGP**PHPPPGKT**
 GAKYIIRDDGQRLDLRYLK**KSSDQHLELGYK**VERHLQDGDVLFNRQPSLHKMSIMGH
RIRIMPYSTFRLNLSVTSPYNADFDGDEMNMHV**PQSFETRAE**VLELMMV**PKCIVSPQAN**
 R**PVMGIVQD**TL**LGC**RKIT**KRDTFIEKDV**FMNTLM**W**WEDFDG**KVP**APAIL**KPRPLWTGK**
QVFNLIIPKQINLLR**YSAWHADTETG**FITPGDTQ**VRIER****GELLAGTLCKK**TLGTSNGSLVH
VIWEEVGPDAARKFLGHTQWL**VNYWLLQNGFTIGIGDTIADSS**TMEK**INETISNAK**TAV
 KDLIRQFQ**GKELDPEPGR**TMRDTFENR**VNQVLNK**ARDDAGSSA**QKSLAETNNL**KAMVT
 AGSKGSFINISQMTACV**GQQNVEGKRIPFG**FDGR**TLPHFTKDDYGPESR**GFVENS**YLRLG**
TPQEFFFHAMGGREGLIDTAVKTSETGYIQ**RRLVK**AMEDIM**VK**YDGTVRNSLGDVIQFL
 YGEDGMDA**VWIESOKLDSLK**MKKSEFDR**TFK**YEIDDENWNPT**YLSDEHLEDLKGIREL**
RDVFDAEYS**KLE**TDRFQLGTEIATNGDST**WPLPVN**IKRHIWNAQ**KTFKIDLRKISDMHPV**
EIVDAVDKLQERLLV**VP**GDDALSVEAQ**KNATLFFNILLRSTLASKR**VLEEYKLSREAFE
 WVIGEIESRFLQSLVAPGEMIGCVAAQSIGEPATQMTLNTFHYAGVSA**KNVTLGVPRLR**
 EIIN**VAKRIK**TPSLSVYL**TPEASKSKEGAK**TVQCALEY**TT**LR**SVTQATEVWYDPDPMSTII**
EEDFEFVRSYYEMPDEDVSPDKISP**WLLRIELNREMMVDK**KL**SMADIAEK**INLEFDDDL
TCIFNDDNAQKLILRIRIMNDEG**PKGELQDESAEDDVFLK**KIESNML**TEMALR**GIPDINK
 VFIKQVRKSR**FDEEGGFK**TSEEWMLDTEGVNLLA**VMCHEDVDPKR**TTSNHLIEIIEVLGI
EAVRRALLDEL**R**VVISFDGSYVNYRHLAILCDTMTYR**GHLMAITRHGINRNDTGPLMRC**
 SFEETVDILLDAAA**YAETDCLR****GVTENIMLGQLAPIGTGDCEL**YLNDEML**KNAIELQLPS**
YMDGLEFGMTPARSPVSGTPYHEGMMSPNYLLSPNMRLSPMSDAQFSPYVGGMAFSPS
 SSPGYSPSSPGYSPTSPGYSPSSPGYSPTSPTYSPPSPGYSPSSPGYSPTSPAYSPTSPSYSPT

SPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPAYSPTSPAYSPTSPAYSPTSPSYSPTSP
 SYSPTSPSYSPTSPSYSPTSPSYSPTSPAYSPTSPGYSPTSPSYSPTSPSYGPTSPSYNPQSAK
 YSPSIAYSPSNARLSPASPYSPSPNYSPSPSYSPTSPSYSPTSPSSPTYSPPSSPYSSGASPDYSP
 SAGYSPTLPGYSPSSTGQYTPHEGDKKDKTGKKDASKDDKGNP

1093/1840=59% coverage

All peptides are a unique match to NRPB1 and do not match the largest subunits of Pol I, III, IV or V.

NRPB2 (At4g21710)

MEYNEYEPQYVEDDDDEEITQEDAWAVISAYFEKGLVRQQLDSFDEFIQNTMQEIV
 DESADIEIRPESQHNPQHQSDFAEITYKISFGQIYLSKPMMTESDGETATLFPK AARLRNL
 TYSAPLYVDVTKRVIKKGHDGEEVTETQDFTK VFIGKVPIMLRSSYCTLFQNSEKDLTEL
 GECPYDQGGYFIINGSEKVLIAQEK MSTNHVYVFKKRQPNKYAYVGEVRSMAENQNR
 PSTMFVRMLARASAKGGSSGQYIRCTLPYIRTEIPIIIVFRALGFVADKDILEHICYDFADT
 QMMELLRPSLEEFVIQNQLVALDYIGKRGATVGVTKERIKYARDILQKEMPLPHVGIG
 EHCETKKAAYYFGYIIHRLLLCALGR RPEDDRDHYGNKRLDLAGPLLGGFRMLFRKLTR
 DVRSYVQKCVDNQKEVNLQFAIKAKTITSLKYSLATGNWQANAAGTRAGVSQVLN
 RLYASTLSHLRLNSPIGREGKLAQPRQLHNSQWGMCPAETPEGQACGLVKNLALM
 VYITVGSAAYPELFLEEWGTENFEEISPSVIPQATKIFVNGMWVGVHRDPDMLVKTLRR
 LRRRVDVNTVEGVVRDIRLRELRIYTDYGRCSRPLFIVDNQKLLIKKRDIALQQRESAE
 EDGWHHLVAKGFIEYIDTEEEETTMMISMTISDLVQARLRPEEAYTENYTHCEIHPSLILGV
 CASIIPFPDHNQSPRNTYQSAMGKQAMGIYVTNYQFRMDTLAYVLYYPQKPLVTTRAM
 EHLHFRQLPAGINAIVAISCYSGYNQEDSVIMNQSSIDR GFFRSLFFRSYRDEEKKMGTLV
 KEDFRPDRGSTMGMRHGSYDKLDDGLAPPGRVSGEDVIIGK TTPISQDEAQQSSR
 YTRRDHSISLRHSETGMVDQVLLTTNADGLRFVKVRVRSVRIPQIGDKFSSRHGQKGT
 GMTYTOEDMPWTIEGVTPDIIVNPH AIPSRMTIGQLIECIMGK VAAHMGKEGDATPFTD
 VTVDNISKALHKCGYQMRGFERMYNGHTGRPLTAMIFLGPTYQRLKHMVDDKIHSR
 GRGPVQILTRQPAEGRSRDGLRFGEMERDCMIAHGAHFLKERLFDQSDAYR VHVCE
 VCGLIAIANLKNSFECRGCKNKTDIVQVYIPYACKLLFQELMSMAIAPRMLTKHLKSA
 KGRQ

750/1188=63% coverage

All peptides are a unique match to NRPB2 and do not match the second-largest subunits of Pol I, III, IV or V.

NRPB3/NRPD3/NRPE3a (At2g15430)

MDGATYQRFPKIKIRELKDDYAKFELRETDVSMANALRRVMISEVPTVAIDLVEIEVNSS
 VLNDEFIAHRLGLIPLTSERAMSMRFSRDCDACDGDGQCEFCVFEFLSSKC VTDQTL
 VTSRDLYSADPTVTPVDFTIDSSVSDSSEHGIIIVKLRRGQELKLRAIARKGIGKDHAKW
 SPAATVTFMYEPDIINEDMMDTSLDDEEKIDLISSPTK VFGMDPVTRQVVVVVDPEAYTY
 DEEVIKKAEMGKPLIEISPKDDSFIFTVESTGAVKASQLVLNAIDLKQKLDVRLSD
 DTVEADDQFELGAHMRGG

230/319=72% coverage

181/319=57% of the peptide coverage is unique to NRPB3a, whereas the other 15% matches either NRPB3a or NRPB3b.

NRPE3b (At2g15400)

MDGVTYQRFPVKIR**ELKDDYAKFELR****ETDVS**MANALRRVMISEVPTMAIHLVKIEVNS
 SVLNDEFIAQRLSLIPLTSERAMSMRFCQDCEDCNGDEHCEFCSEVFPPLSAKC**VTDQTL**
VTSRDLYSADPTVTPVDFTSNSSSDSSEHKGIIIAKLRRGQELKALARKGIGKDHAK
 WSPAATVTYMYEPDIINEEMMNTLTDEEK**IDLI**ESSPTK**VFGIDPVTG****QVVVV**DPEAYT
YDEEVIKAEAMGKPGLEIHPKHDSFVFTVESTGALK**ASQLVLNAIDILK**QKLDAIRLS
 NTVEADDQFGELGAHMREG

72/319=23% coverage

13/319=4% of the peptide coverage is unique to NRPB3b, whereas the other 19% matches either NRPB3a or NRPB3b. This variant may be used infrequently as an alternative NRPB3 subunit.

NRPB4 (At5g09920)

MSGEEENAAELK**IGDEFLKAKCLMNCEVSLILEHKFEQLQQI**SEDPMNQVSQVFEKSL
 QYVKRFSR**YKNPDAVRQVREILSRHQ**LT**EFELCVLGNLCPETVEEAVAMVPSLKT**KGRA
HDDEAIEKMLNDLSL**VKRFE**

84/138=61% coverage

All of the peptides match NRPB4 and only NRPB4.

NRPD4/NRPE4 (At4g15950)

MSEKGGKGLKSSLKSKDGGKDGSSSTKLKKGRIHFDQGTTPANYKILNVSSDQQPFQSS
 AAKCGKSDKPTKSSKNSLHSFELKDLPENAECEMMDCEAFQILDGIKQGLVGLSEDP
 SIKI PVSYDRALAYVESC
 VHYTNPQSVRKVLEPLKTYGISDGEMCVIANASSESVDEVLA
 FIPS LKTKKEVINQPLQDALEELSKLKKSE

0/205=0% coverage

No peptides were found to match this sequence.

NRPB5/NRPD5 (formerly AtRPB5a, AtRPB24.3)(At3g22320)

MLTEEELKRLYRIQK**TLMQMLR****DRGYFIADSELTMTKQQFIRKHGDNMK****REDL**VTLKA
 KRNDNSDQ**LIFFPDEAK**VGVK**TMKMYTNR****MKSENVFR**AILVV**QQNLTPFAR**TCISEIS
 SK**FHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEK**KTLLERYTVKETQLPRI**QVTDPIARY**
 FGLKRGQVVK**IIRPSETAGR**YVTYRYVV

129/205=63% coverage

All peptides match to NRPB5 and only NRPB5.

NRPE5 (formerly AtRPB5b, AtRPB23.7) (At3g57080)

MEVKGKETASVLCLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDEDINLSLHDF
 RTVYGERPDVDRRLRISALHRSDSTKKVKIVFFGTSMVKVNAIRSVVADILSQETITGLILV
 LQNHVTNQAALKAIELFSFKVEIFQITDLLVNITKHSKLPQHQLNDEEKTTLLKKFSIEEK
 QLPRISKKDAIVRYYGLEKGQVVKVNYRGELTESHVAFRCVW

0/222=0% coverage

No peptides were found to match this sequence.

NRPB5-like family member (At5g57980)

MSDMDDEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCCKTMNKVNKEALF
 VSANKGPNPADKIYVFYPEGPKVGVVVIKKEVAIKMRDDKVHRGIVVVPMAITAPARM
 AVSELNKMILTIEVFEEAELVTNITEHKLVNKYVVLDDQAKKKLLNTYTVQDTQLPRILV
 TDPLARYYGLKRGQVVKIRRSDATSLDYTYRFAV

0/210=0% coverage

No peptides were found to match this sequence.

NRPE5-like family member (synonym AtRPB5d) (At2g41340)

MEGKGKEIVVGHISISKSSVECHKYLLARRTTMEMLRDRGYDVSDDEDINLSLQQFRALY
 GEHPDVDLLRISAKHRFDSSKKISVVFCGTGIVKVNAMRVIAADVLSRENITGLILVLQS
 HITNQALKAVELFSFKVELFEITDLLVNVSKHVLPRKHQVLNDKEKESLLKKFSIEEKQL
 PRLSSKDPIVRYYGLETGQVMKVITYKDELSESHVITYRCVS

0/218=0% coverage

No peptides were found to match this sequence.

NRPE5-like family member (At3g54490)

MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNE
 AELSLTLSEFRSVFGEKPELERLRICVPLRSDPKKILVVFIMGTEPITVKSVRALHIQISNN
 VGLHAMILVLQSKMNHFAQKALTTFPFTVETFPIEDLLVNITKHIQQPKIEILNKEEKEQL
 LRKHALEDKQLPYLQEKDSFVRYYGLKKKQVVKITYSKEPVGDFVITYRCII

0/233=0% coverage

No peptides were found to match this sequence.

NRPB5 family member (likely pseudogene) (At3g16880)

MKKYIDQLKSANVFRAILVVQDIKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKEHVF
 VPEHQALTTEEKQKFLERKRTSFQGFT

0/87=0% coverage

No peptides were found to match this sequence.

NRPB6a/NRPD6a/NRPE6a (At5g51940)

MADEDYNDVDDLGYEDEPAEPEIEEGVEEDVEMKENDDVNGEPIEAEDK **VETEPVQRP**
RKTSKFMTKYERARILGTRALQISM **NAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL**
PDGSFEEWGVDELIVEDSWKRQVGGD

58/144=40% coverage

22/144=15% of the coverage is unique to At5g51940, whereas the other 25% matches either At5g51940 or At2g04630.

NRPB6b/NRPD6b/NRPE6b (At2g04630)

MADDDYNEVDDLGYEDEPAEPEIEEGVEEDADIKENDDVNVDPLETEDK **VETEPVQRP**
RKTSKFMTKYERARILGTRALQISM **NAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL**
PDMSYEEWGVDELIVEDSWKRQVGGD

58/144=40% coverage

22/144=15% of the coverage is unique to At2g04630, whereas the other 25% matches either At5g51940 or At2g04630.

NRPE7 (At4g14660)

MFLKVQLPWNVMIPAENMDAKGLMLKRAILVELLEAFASKKATKELGYVAVTTLDKI
GEGKIREHTGEVLFVPMFSGMTFKIFKGEIHHGVVHKVVKHGVFMRCGPIENVYLSYTK
MPDYKYIPGENPIFMNEKTSRIQVETTVRVVIGIKWMEVEREFQALASLEGDYLGPLSE

0/177=0% coverage

No peptides were found to match this sequence.

NRPD7 (At3g22900)

MFIKVKLPWDVTIPAEDMDTGLMLQRAIVIRLLEAFSKEKATKDLGYLITPTILENIGEGK
IKEQTGEIQFPVVFNGICFKMFKGEIVHGVVHKVHKTGVFLKSGPYEIIYLSHMKMPGYE
FIPGENPFFMNQYMSRIQIGARVRFVVDTEWREAEEKDFMALASIDGDNLGP

0/174=0% coverage

No peptides were found to match this sequence.

NRPB7 (At5g59180)

MFFHIVLERNMQLHPRFFGRNLKENLVSKLMKDVEGTCSGRHGFVVAITGIDTIGKGLIR
DGTGFVTFPVKYQCVVFRPFKGEILEAVVTLVNMGFFAEAGPVQIFVSKHLIPDDMEF
QAGDMPNYTSDGSVKIQKECEVRLKIIGTRVDATAIFCVGTIKDDFLGVINDPAAA

89/176=51% coverage

All peptide coverage is unique to NRPB7 only.

NRPB7 family member (At4g14520)

MFSEVEMARDVAICAKHLNGQSPHQILCROLLQDLIHEKACREHGFYLGITALKSIGNNK
NNNIDNENNHQAKILTFVSFTCRTFLPARGDILQGTVKKVLWNGAFIRSGPLRYAYLSL
LKMPHYHYVHSPLESEDEKPHFQKDDLSKIAVGVVVRFQVLAVRFKERPHKRRNDYYVL
ATLEGNGSFGPISLTGSDEPYM

0/200=0% coverage

No peptides were found to match this sequence.

NRPB8a/NRPD8a/NRPE8a (At1g54250)

MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAVGDKF
TLALAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGTTPKAELYVSFG
GLLMLLKGDPAHISHFELDQRLFLLMRKL

96/146=66% coverage

44/146=30% of the coverage is unique to NRPB8a, whereas 33% matches either NRPB8a or NRPB8b.

NRPB8b/NRPD8b/NRPE8b (At3g59600)

MASNIIMFEDIFVVDKLPDGGKFDKVTREARSHNLEMFHMLDVNTEVYPLAVGDKF
 TLAMAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGGKTPKAELYVSFG
 GLLMLLQGDPAHISHFELDQRLFLLMRKL

96/146=66% coverage

40/146=30% of the coverage is unique to NRPB8b, whereas 33% matches either NRPB8a or NRPB8b.

NRPB9a/NRPD9a/NRPE9a (At3g16980)

MSTMKFCRECNNILYPKEDKEQKILLYACRNCDHQEVADNSCVYRNEVHHSVSERTQIL
 TDVASDPTLPRTKAVRCSKCQHREAVFFQATARGEEGMTLFFVCCNPNCGHRWRE

35/114=30% coverage

25/114=22% of the coverage is unique to NRPB9a, whereas the other 8% matches either NRPB9a or NRPB9b.

NRPB9b/NRPD9b/NRPE9b (At4g16265)

MSTMKFCRECNNILYPKEDKEQSILLYACRNCDHQEAADNNCVYRNEVHHSVSEQTQI
 LSDVASDPTLPRTKAVRCAKQHQHGEAVFFQATARGEEGMTLFFVCCNPNCSHRWRE

42/114=37% coverage

32/114=28% of the coverage is unique to NRPB9b, whereas the other 9% matches either NRPB9a or NRPB9b.

NRPB10/NRPD10/NRPE10 (At1g11475)

MIIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRRMLMTHVDLIEK
 LLNYNTLEKSDNS

50/71=70% coverage

39/71=55% of the coverage matches only At1g11475, whereas the remaining 15% matches either At1g11475 or At1g61700.

NRPB10 family member (At1g61700)

MIVPVRFCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRRMLMTHVDLIE
 KLLNYNTMEKSDPN

11/71=15% coverage, matching either At1g11475 or At1g61700.

0/71=0% of the coverage is unique to At1g61700.

NRPB11/NRPD11/NRPE11 (At3g52090)

MNAPERYERFVPEGTTKVSYDRDTKIIINAASFTVEREDHTIGNIVRMQLHRDENVLFA
 GYQLPHPLKYKIIVRIHTTSQSSPMQAYNQAINDLKELDYLNQFEAEVAKFSNQF

87/116=75% coverage

All peptide coverage matches NRPB11 only.

NRPB12/NRPD12/NRPE12 (At5g41010)MDPAPEPVITYVCGDCGQENTLKSGDVIQCRECGYRILYKKRTR**RVVQYEAR**

8/51=16% coverage

This peptide matches only At5g41010.

RPB12 family member (At1g53690)MDLQQSETDDKQPEQLVIYVCGDCGQENILKRGDVFQCRDCGFRILYKKRILDKKETRI
GV

0/62=0% coverage

No peptides were found to match this sequence.

Figures S4-S12

These figures show ClustalW alignments of Arabidopsis and yeast RPB4, 5, 6, 7, 8, 9, 10, 11 and 12 family proteins. Red highlighting denotes invariant residues, yellow denotes conserved residues and cyan denotes similar residues.

Figure S4. RPB4 family alignment

```

Sc_RPB4      MNVSTSTFQTRRRRLKKVEEEENAATLQLGQEFQLKQINHQEEEELIALNLSEARLVIK
At_NRPB4     -----MSGEEEENAAELKIGDEFLKAKCLMNCEVSLILEHKFEQLQQISE
At_NRPD4/NRPE4 MSEKGGKGLKSSLKSKDGGKDGSSSTKLKKGRKIHFDQGTPPANYKILNVSSDQPFQSSA

Sc_RPB4      EALVERRRAFKRSQKKHKKHLKHENANDETTAVEDEDDDLDEDDVNADDDDFMHSETRE
At_NRPB4     DPMNQVSQVFEKS-----
At_NRPD4/NRPE4 AKCGKSDKPTKSSKNSLHSFELKDLPENAECMMDCEAFQILDG-----

Sc_RPB4      KELESIDVLLEQTTGGNNKDLKNTMQYLTNFSRFRDQETVGAVIQLLKSTGLHPFEVAQL
At_NRPB4     -----LQYVKRFSRYKNPDAVRQVREILSRHQLTEFELCVL
At_NRPD4/NRPE4 -IKGQLVGLSEDPSIKIPVSYDRALAYVESCVHYTNPQSVRKVLEPLKTYGISDGEMCVI

Sc_RPB4      GSLACDTADEAKTLIPSLNNK---ISDDELERILKELSNLETLY
At_NRPB4     GNLCPETVEEAVAMVPSLKTKGRAHDDEAIEKMINDLSLVKRFE
At_NRPD4/NRPE4 ANASSESVDEVLAFIPSLKTK-KEVINQPLQDALEELSKLKKSE

```

Figure S5. RPB5 family alignment. For this alignment, the following codes are used:**At3g22320=NRPB5/NRPD5****At5g57980=NRPB5-like****At3g57080=NRPE5****At2g41340=NRPE5-like****At3g54490=NRPE5-like****At3g16880=likely pseudogene**

Start of jaw domain

```

Sc_RPB5      -----MDQENERNISRLWRAFRTVKEMVKDRGYFITQE
Hs_RPB5      -----MDDEEE--TYRLWKIRKTIIMQLCHDRGYLVTQD
At_NRPB5/NRPD5 -----MLTEEELKRLYRIQKTLMQMLRDRGYFIADS
At5g57980    -----MSDMDEITRIFKVRRTVLQMLRDRGYTIEES
At_NRPE5     MEVKGKETASVL-----CLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDE
At2g41340    MEGKGKEIVVGH-----SISK----SSVECHKYLLARRTTMEMLRDRGYDVSDE
At3g54490    MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNEA
At3g16880    -----

```

```

Sc_RPB5      EVELPLEDFKAKYCD--SMGRPQRKMMSFQANPTEESISKFPDMGSLWVEFCDEPSVGVK
Hs_RPB5      ELDQTLLEEFKAQFGDKPSEGRPRRTDLTVLVAHND-----PTDQMFVFFPEEPKVGIK
At_NRPB5/NRPD5 ELTMTKQQFIRKHGDN---MKREDLVTLKAKRNDN-----SDQLYIFFPDEAKVGVK
At5g57980    DLNLKREEFVQRFCKT--MNKVNKEALFVSANKGPN-----PADKIYVFYPEGPKVGVV
At_NRPE5     DINLSLHDFRTVYGER----PDVDRLRISALHRSD-----STKKVKIVFFGTSMVKVN
At2g41340    DINLSLQQFRALYGEH----PDVDLLRISAKHRFD-----SSKKISVVFVCGTGIVKVN
At3g54490    ELSLTLSEFRSVFGEK----PELERLRICVPLRSD-----PKKKILVVFVFMGTEPITVK
At3g16880    -----

```

end of jaw domain |

```

Sc_RPB5      TMK-TFVIHIQEKNFQTGIFVYQNNITPSAMK----LVPSIPPATIEFNEAALVFNITH
Hs_RPB5      TIK-VYCQRMQEEENITRALIVVQGGMTPSAKQS---LVDMAPKYILEQFLEQELLINITE
At_NRPB5/NRPD5d TMK-MYTNRMKSENVFRAILVVQQNLTPFAR---TCISEISSKFHLEVFQEAEMLVNIKE
At5g57980    VIKKEVAIKMRDDKVHRGIVVVPMAITAPARMA---VSELNKMLTIEVFEEAELVTNITE
At_NRPE5     AIRSVVADILSQETITGLILVLQNHVTNQAALKA-----IELFSFKVEIFQITDLLVNITK
At2g41340    AMRVIAADVLSRENITGLILVLQSHITNQAALKA-----VELFSFKVELFEITDLLVNVSK
At3g54490    SVRALHIQISNNVGLHAMILVLQSKMNHFAQKA-----LTTFPFTVETFPIEDLLVNITK
At3g16880    -MK-KYIDQLKSANVFRAILVVQD- IKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKE

```

start of assembly domain

```

Sc_RPB5      HELVPKHIRLSSDEKRELLKRYRLKESQLPRIQRADPVALYLGLKRGEVVKIIRKSETSG
Hs_RPB5      HELVPEHVVMTKEEVSELLARYKLRNQLPRIQAGDPVARYFGIRRGQVVKIIRPSETAG
At_NRPB5/NRPD5 HVLVPEHQVLTTEEKKTLLERYTVKETQLPRIQVTDPIARYFGLKRGQVVKIIRPSETAG
At5g57980    HKLVNKYYVLDQAKKKLLNTYTVQDTQLPRILVTDPLARYYGLKRGQVVKIRRSATSL
At_NRPE5     HSLKPQHQVLDNDEEKTLLKFFSIEEKQLPRISKDAIVRYYGLEKQVVKVNYRGELTE
At2g41340    HVLRPKHQVLDNKEKESLLKFFSIEEKQLPRLSSKDPVRYYGLETGQVMKVITYKDELSE
At3g54490    HIQQPKIEILNKEEKEQLLRKHALEDKQLPYLQEKDSFVRYYGLEKQVVKITYSKEPVG
At3g16880    HVFVPEHQALTTEEKQKFLER---KRTSFQGF-----

```

```

Sc_RPB5      RYASYRICM
Hs_RPB5      RYITYRLVQ
At_NRPB5/NRPD5 RYVTYRYVV
At5g57980    DYYTYRFAV
At_NRPE5     SHVAFRCVW
At2g41340    SHVTYRCVS
At3g54490    DFVTYRCII
At3g16880    -----

```

Figure S6. RPB6 family alignment

```

NRPB6a_At5g51940 MAD--EDYNDVDDLGYEPEAEP-EIEEGVEEDVEMK--ENDDVNGEPIEA-----EDKV
NRPB6b_At2g04630 MAD--DDYNEVDDLGYEPEAEP-EIEEGVEEDADIK--ENDDVNVDPLET-----EDKV
Sc_RPB6          MSDYEEAFNDGNEN-FEDFDVEHFSDEETTYEEKPQFKDGETTDANGKTIVTGGNGPELDFQ

NRPB6a_At5g51940 ETEPVQR-----PRKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLE
NRPB6b_At2g04630 ETEPVQR-----PRKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLE
Sc_RPB6          QHEQIRKRTLKEKAIPKDRATTTPYMTKYERARILGTRALQISMNAPVFDLEGETDPIR

NRPB6a_At5g51940 IAMKELRQRKIPFTIRRYLPDGSFEEWGVDELIVEDSWKRQVGGD
NRPB6b_At2g04630 IAMKELRQRKIPFTIRRYLPDMSYEEWGVDELIVEDSWKRQVGGD
Sc_RPB6          IAMKELAEKKIPLVIRRYLPDGSFEDWSEELIVDL-----

```

Figure S7. RPB7 family alignment

```

Sc_RPA43      MSQVKRANENRETARFIKHKHKQVTNPIDEKNGTNSNCIVRVPIALYVSLAPMYLENPLQG
Atlg75670    -----MEGLKLSAEALMIFIHPSQSRN-VFQ
Sc_RPB7      -----MFFIKDLSLNI TLHPSFFGP---R
NRPB7_At5g59180 -----MFFHIVLERNMQLHPRFFGR---N
Sc_RPC25     -----MFI LSK IADLVRIPPDQFHR----
Atlg06790   -----MFY LSELEHS LRVP PHLNL----
NRPE7_At4g14660 -----MFLKVQLPWNVMI PAENMDAKGLM
NRPD7_At3g22900 -----MFIKVKLPWDVTI PAEDMDT-GLM
At4g14520   -----MFSEVEMARDVAI CAKHLNG--QS

Sc_RPA43      VMKQHLNPLVMKYNNKVGGVVLGYEGLKILDADPLSKEDTSEKLIKITPDTFPFGFTWCHV
Atlg75670    GICRELSLLLFQYNETFDGVLLAYDATVKSQAKILTG-----LHPYFG---VRVNT
Sc_RPB7      MKQYLKTKLLEEVEG-SCTGKFGYI-LCVLDYDNIDIQRG----RILPTDGSAEFNVKY
NRPB7_At5g59180 LKENLVSKLMKDVEG-TCSGRHGFV-VAITGID--TIGKG----LIRDGTGFVTFPVKY
Sc_RPC25     DTISAIHQLNKFKANKIIPNVGLC-ITTYDLLTVEEQ-----LKP GDSSYINVTF
Atlg06790   PLEDAIKSVLQNVFLDKVLADLGLC-VSIYDIKSVEGGF-----VLPGDGAATYKVGL
NRPE7_At4g14660 LKRAILVELLEAFASKKATKELGYY-VAVTTLDKIGEGK-----IREHTGEVLFVPMF
NRPD7_At3g22900 LQRAIVIRLLEAFSKEKATKDLGYL-ITPTILENIGEGK-----IKEQTGEIQFPVVF
At4g14520   PHQPILCRLLQDLIHEKACREHGFY-LGITALKSIGNNKNNNIDNENNHQAKILTFPVSE

Sc_RPA43      NLYVWQPQVQDVLEGYIFIQSASHIGLLIHDAFNASIKKNNIPVDWTFVHNDVEEDADVI
Atlg75670    RLLLFDPKPKSFVEG--KIVKISPEIHHVIVLG----FSAAVITDVIDREEFKYRVR--
Sc_RPB7      RAVVFKPFKGEVVDG--TVVSCSQHGFEVQVG-----PMKV FVTKHLMPQDLTFNAGS-
NRPB7_At5g59180 QCVVFRPFKGEILEA--VVTLVNKMGFFAEAG-----PVQIFVSKHLIPDDMEFQAG--
Sc_RPC25     RAVVFKPFLGEIVTG--WISKCTAEGIKVSLLG---IFDDIFIPQNM LFEGCY YTP E--
Atlg06790   RIVVFRPFVGEVIAA--KFKESDANGRLRLTLG----FFDDIYVPAPLMPKPNRCEPDY
NRPE7_At4g14660 SGMTFKIFKGEI IHG--VVHKV LKHGVFMRCG-----PIENVYLSYTKMPDYKYIPG--
NRPD7_At3g22900 NGICFKMFKGEIVHG--VVHKVHKTG VFLKSG-----PYEIIYLSHMKMPGYEFIPG--
At4g14520   TCRTFLPARGDILQG--TVKKVLWNGAFIRSG-----PLRYAYLSLLKMPHYHYVHSP L

Sc_RPA43      NTDENNGNNNEDNKDSNGGSNSLKGKFSFGNRSLGHWVDSNGEPI DGKLRFTVRNVHTTG
Atlg75670    ---DGE GSFVSRSHKR-----HALKLGTMRLRLQVQSFDEEV
Sc_RPB7      ---NPPSYQSS EDVIT-----IKSR--IRVKIEGCISQV
NRPB7_At5g59180 ---DMPNYTSDG SVK-----IQKECEVRLKIIGTRVDA
Sc_RPC25     ---ESAWI WPMDEETK-----LYFDVNEKIRFRIEREVFVD
Atlg06790   NRKQMIWVWEYGE PKED-----YIVDDACQIKFRVESISYPS
NRPE7_At4g14660 ---ENPIFM-NEKTSR-----IQVETTVRVVVIGIKWME
NRPD7_At3g22900 ---ENPFM-NQYMSR-----IQIGARVRFVVLDT EWRE
At4g14520   SEDEKPHFQ-KDDL SK-----IAGVVVRFQVLA VRKE

Sc_RPA43      RVVSV DGT LISDADEEGNGYNSRSQAESLP IVSNKKIVFDDEVSIENKESHKELD LPEV
Atlg75670    MHIAG-----SLLPENTGCVKWL
Sc_RPB7      SSI-----HAIGSI
NRPB7_At5g59180 TAI-----FCVGTI
Sc_RPC25     VKPKSP-----KERELEERAQLENEIEGKNEETPQNEKPPAYAL L GSC
Atlg06790   VP-----TERAEDAKPFAPMVVTGNM
NRPE7_At4g14660 VER-----EFQALASL
NRPD7_At3g22900 AEK-----DFMALASI
At4g14520   RPHK-----RRNDYYV L ATL

Sc_RPA43      KEDNGSEI VYEENTSESNDGESSDS
Atlg75670    EKKSEEALPTDRDHKRRKLA-----
Sc_RPB7      KED-YLGAI-----
NRPB7_At5g59180 KDD-FLGVINDPAAA-----
Sc_RPC25     QTD-GMGLVSWWE-----
Atlg06790   DDD-GLGPVSWWDSYEQVDQEE----
NRPE7_At4g14660 EGD-YLGPLSEE-----
NRPD7_At3g22900 DGD-NLGPF-----
At4g14520   EGN SFGPISLTGSDEPYM-----

```

Figure S8. RPB8 family alignment

```

NRPB8a_At1g54250 MASNIILMFEDIFVVDQLDEPDGKKFKDKVTRVQATSHNLEMF-MHLDVNTFVYPLAVGDKFT
NRPB8b_At3g59600 MASNIIMFEDIFVVDKLDDEPDGKKFKDKVTRVEARSHNLEMF-MHLDVNTFVYPLAVGDKFT
Sc_RPB8          MSN--TLFDDIFQVSEVDPG--RYNKVCRIEAASTTQDQCKLTLDINVELFPVAAQDSL

NRPB8a_At1g54250 LALAPTLNLDGTPDPTG-----YFTP--GAKKTLADKYEYIMHGKLYKISERDGTTPKAEL
NRPB8b_At3g59600 LAMAPTLNLDGTPDPTG-----YFTP--GAKKTLADKYEYIMHGKLYKISERDGTTPKAEL
Sc_RPB8          VTIASSLNLLEDTPANDSSATRSWRPQAGDRSLADDYDVMYGTAYKFEVVS--KDLIAV

NRPB8a_At1g54250 YVSFGGLMLLKGDPAHISHFELDQRLFLLMRKL
NRPB8b_At3g59600 YVSFGGLMLLQGDPAHISHFELDQRLFLLMRKL
Sc_RPB8          YVSFGGLMLRLEGNYRNLLNNLKQEN-AYLLIRR-

```

Figure S9. RPB9 family alignment

```

Sc_RPA12          --MSVVGSLIFCLDCGDLEENPNAVLG--SNVECSQCKATYPKQSFSNLKVVVTTTADDAF
At3g25940        MEKSRESEFLFCNLCGTMLVLKST----KYAECPHCKTTRNAKDIIDKEIAYTVSAEDI
Sc_RPB9-         -----MTTFRFCRDCNNMLYPREDKENRLLFECRTCSYVEEAGS-PLVYRHELITNIGE
NRPE9a_At3g16980 -----MSTMKFCRECNNILYPKEDKEQKILLYACRNCDHQEVADN-SCVYRNEVHHSVSE
NRPE9b_At4g16265 -----MSTMKFCRECNNILYPKEDKEQSILLYACRNCDHQEAADN-NCVYRNEVHHSVSE
Sc_RPC11         -----MLSFCPSCNNMLLITSGDS-GVYTLACRSCPVEFPIEG-IEIYDRKKLPRKEV
At4g07950        -----MEFCPTCGNLLRYEGG---GSSRFFCSTCPYVANIERRVEIKKKQLLVKCSI
At1g01210        -----MEFCPTCGNLLRYEGG---GNSRFFCSTCPYVAYIQRQVEIKKKQLLVKCSI

Sc_RPA12          PSSLRAKKSVVKTSLKKNELKDGATIKEKCPQCGNEMNYHTLQLRSADEGATVFFYTTTS
At3g25940        RRELGISLFGKTKQAEAEALPKI---KKACEKQHPPELVYTTROT RSADEGQTTYTTCPN
Sc_RPB9-         TAGVVQDIGSDPTLPRSDRE-----CPKCHSRENVEFFQSQRKDTSMVLFVFCLS
NRPE9a_At3g16980 RTQILTDVASDPTLPRTKAVR-----CSKQCHREAVFFQATARGE--GMTLFFVCCN
NRPE9b_At4g16265 QTQILSDVASDPTLPRTKAVR-----CAKQCHGEAVFFQATARGE--GMTLFFVCCN
Sc_RPC11         DDVLG-GGWDNVDTKTQCPN-----YDTCGGESAYFFQLQIRSADEPMTTFYKCVN
At4g07950        EPVVTKDDIPTAAETEAPCP-----RCGHDKAYFKSMQIRSADEPESRFYRCLK
At1g01210        EAVVTKDDIPTAAETEAPCP-----RCGHDKAYFKSMQIRSADEPESRFYRCLK

Sc_RPA12          --CGYKFRINN-----
At3g25940        --CAHRFTEG-----
Sc_RPB9-         --CSHIFTSQKNKRTQFS
NRPE9a_At3g16980 PNCG-----
NRPE9b_At4g16265 PNCshrwrEhrwrE----
Sc_RPC11         --CGHRWKEN-----
At4g07950        --CEFTWREE-----
At1g01210        --CEFTWREE-----

```

Figure S10. RPB10 family alignment

```

NRPB10/NRPD10/NRPE10 MIIIVRCFTCGKVIENKWDQYLDLLQLD-YTEGDALDALQLVRYCCRRMLMTHVDLIEKL
At1g61700          MIVPVRCFTCGKVICNKWDTYLELLQAD-YAEGDALDALGLVRYCCRRMLMTHVDLIEKL
Sc_RPB10          MIVPVRCFSCGKVVGDKWESYLNLLQEDDELDECTALSRLGLKRYCCRRMILTHVDLIEKF

NRPB10/NRPD10/NRPE10 LNYNTLEKSDNS
At1g61700          LNYNTMEKSDPN
Sc_RPB10          LRYNPLEKRD--

```

Figure S11. RPB11 alignment

```

NRPB11_At3g52090 MNAPERYERFVVPEGTKVSYDRDTKIIINAASFIVEREDHTIGNIVRMQLHRDENVLFAG
Sc_RPB11          MNAPDRFELFLLGEGESLKKIDPDTKAPNAVVITFEKEDHTLGNLIRAEILLNRDKVLFAA

NRPB11_At3g52090 YQLPHPLKYKIIIVRIHTTSQSSPMQAYNQAINDLKELDYLKNQFEAEVAKFS----NQF
Sc_RPB11          YKVEHPFFARFKLRIQTTEGYDPKDALKNACNSIINKLGALKTNFETEWNLQTLAADDAA

```

Figure S12. RPB12 family alignment

```

NRPB12/NRPD12/NRPE12 -----MDP-----APEP-VTVVCGDGGQENTLKSGDVIQCRECGYRILYKK
At1g53690                -----MDLQQSEITDDKQPEQLVIYVCGDGGQENILKRGDVFQCRDCGFRILYKK
Sc_RPB12                 MSREGFQIPTNLDAAAAGTSQARTAT-LKYICAECSKLSLSRTDAVRCKDCGHRILLKA

NRPB12/NRPD12/NRPE12  RTRRVVQYEAR-
At1g53690              RILDKKETRIGV
Sc_RPB12               RTKRLVQFEAR-

```

Figure S13. Expression patterns of the RPB5 family. RT-PCR detection of mRNAs

corresponding to the six *Arabidopsis* genes homologous to yeast RPB5. Actin served as a control to show that similar amounts of RNA were isolated from the tissues tested.

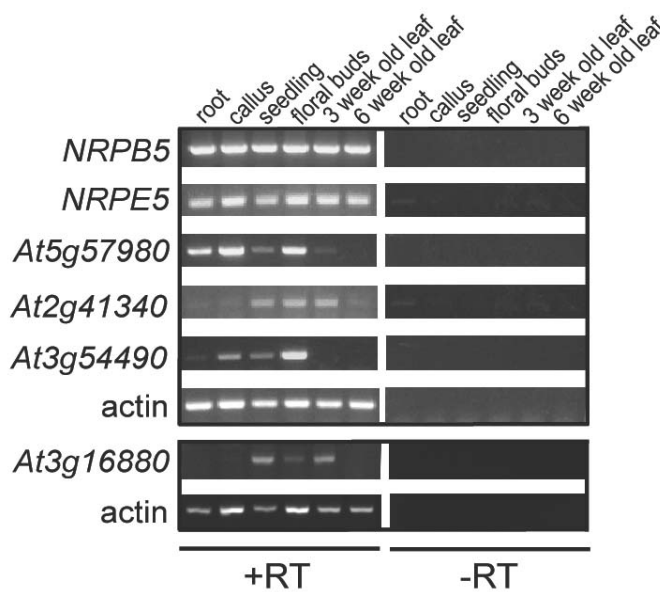
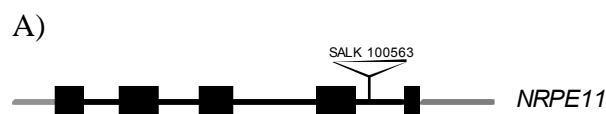


Figure S14. Analysis of *nrpd5-1* and *nrpe11-1* T-DNA insertion mutants. A. Gene structure of *NRPE11* and location of the T-DNA insertion. B. Genotyping results for offspring from a selfed *nrpd5/+* and *nrpe11/+* heterozygotes (top) and genotyping results of F1 offspring of reciprocal crosses between *nrpd5/+* heterozygotes and wild-type plants *nrpe11/+* heterozygotes and wild-type plants (bottom). *nrpd5-1* homozygotes are not recovered due to female gametophyte lethality, as shown by reciprocal crosses, whereas *nrpe11-1/nrpd11-1/nrpb11-1* homozygous mutants (abbreviated as *nrpe11-1* below) appear to be embryo lethal since the T-DNA is passed through both the male and female gametophyte. C. RT-PCR of transcript levels in Col wt vs. *nrpe5-1* mutants using primers that span the T-DNA insertion or are upstream of the T-DNA insertion. Actin served as a control to show that similar amounts of RNA were loaded in each genotype.



B)
Genetic analysis of RNA polymerase subunits.

Progeny of:	+/+	+/-	-/-	total
<i>nrpd5-1/+</i>	80	21	0	101
<i>nrpe11-1/+</i>	33	63	0	96

F1 progeny of female x male:	+/+	+/-	total
<i>nrpd5-1/+</i> x <i>+/+</i>	78	0	78
<i>+/+</i> x <i>nrpd5-1/+</i>	23	10	33
<i>nrpe11-1/+</i> x <i>+/+</i>	13	4	17
<i>+/+</i> x <i>nrpe11-1/+</i>	25	25	50

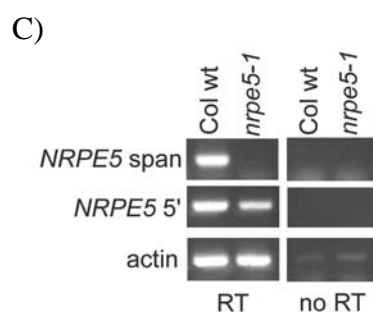


Figure S15.

Flowering time of individual plants from wild-type (ecotype Col-0) and *nrpe5-1* populations.

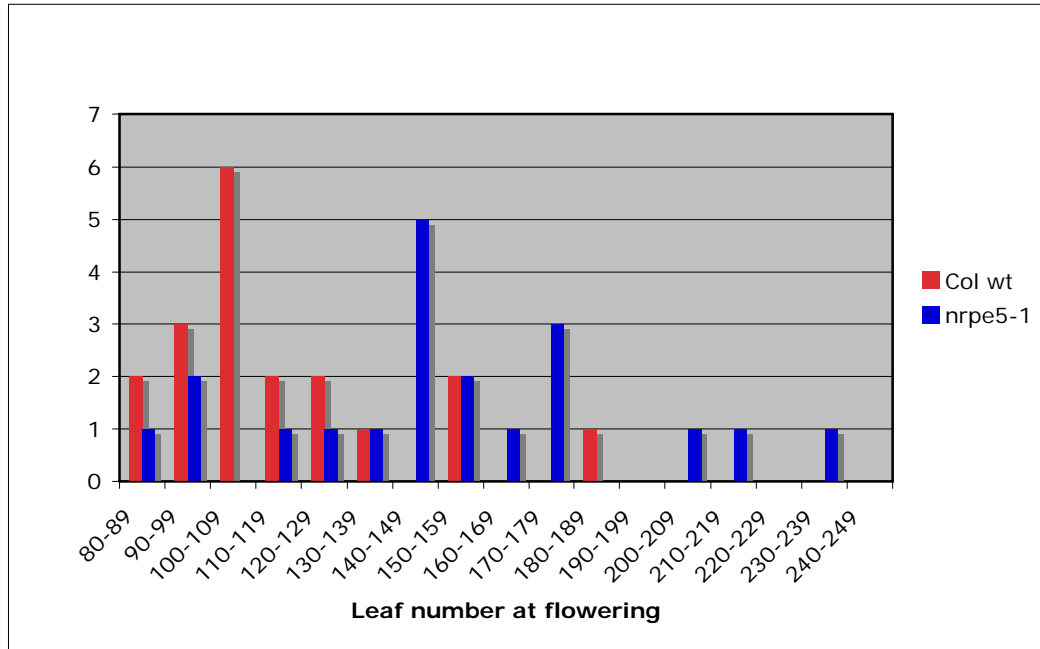


Figure S16. Alignment of RPB5 family variants in diverse plants with non-plant RPB5s. Red: absolutely conserved residues; yellow: consensus residues; cyan: similar residues. Locations of the jaw and assembly domains are indicated by arrows. Hs= *Homo sapiens*; Dm = *Drosophila melanogaster*; Ce = *Caenorhabditis elegans*; Sc = *Saccharomyces cerevisiae*

```

→Start of Jaw domain
C_reinhardtii_XP_001697601 -----MDN-
O_lucimarinus_XP_001417617 -----MSND
Hs_RPB5_BAA07406 -----MDDE
Dm_RPB5_NP_610630 -----MDDE
Ce_RPB5_Q9N5K2 -----MADDE
Populus_trichocarpa_584052 -----MTLTE
Vitis_vinifera_CAO65489 -----MSASE
NRPB5/NRPD5_At3g22320 -----MLTE
Medicago_truncatula_ABO78350 -----MVFSE
Oryza_sativa_EAZ13876 -----MSAGLVTEE
Oryza_sativa_NP_001044564 -----MSAGLVTEE
Zea_mays_ACF85599 -----MSAGLVTDE
Oryza_sativa_CAD41325 -----MAS-E
Oryza_sativa_EAZ31161 -----MAS-E
Zea_mays_ACF81264 -----MASPD
Physcomitrella_patens_206246 -----MSGQSLD
Physcomitrella_patens_55574 -----MSGQSLD
Physcomitrella_patens_231299 -----MAEHVLD
Physcomitrella_patens_136486 -----MAEHVLD
Sc_RPB5_CAA85113 -----MDQENE
NRPB5-like_At5g57980 -----MSDMD
NRPE5-like_At2g41340 -----MEGKG-----KEIVVGHHSISK---SS-
Brassica_napus_AAF81222 -----MEGKG-----KELAVGSGLSKSLDESR-
NRPE5_At3g57080_NP_191267 -----MEVKG-----KETASVLCLSKYVDLSS-
Populus_trichocarpa_57931 -----MES-----LGRCLSSFVDEGS-
Vitis_vinifera_CAO63075 -----MDGGGWFDDLNDFEVKRLSSFFVDEGR-
Medicago_truncatula_ABN07995 -MATENGGGQNGTTETAITTMEIENGDIITQPQLQEQPQCLFTKKNDS-
Populus_trichocarpa_48513 -----MAATTETFNNGASFHGVLDLDRCLTDFVDEGS-
Vitis_vinifera_CAO42914 -----MESQAGSH-GNGS-----CITADMEQGS-
NRPE5-like_At3g54490 -----MEETMAEEGCENVESTFDDGTNCISKTEDTGG-
Medicago_truncatula_ABD28306 -----MAMIENGNET-----RSECLVVICNEESN
Oryza_sativa_NP_001065723 -----MAAEMEVDV--DVHEVPECIASMIDRG-S
Oryza_sativa_NP_001066119 -----MAAEMEVDV--DVHEVPECIASMIDRG-S
Oryza_sativa_EAY79909 -----MAAEMEADDV--DVHEVPECIASMIDRG-S
Zea_mays_ACF87172 MESAESTAAAAARASNGAARAVVEDDEDD--DVPEVAACISTMLDRGGS
consensus E

```

(cont'd below)

C_reinhardtii_XP_001697601 --LTRLWRVRRCTLQMLNDRGYLVSQEEIGTTKDQFRDRFGENP-----R
O_lucimarinus_XP_001417617 -KRTRLFRVRKTIHKMLAARGYLVSAKELERLDSFTEDFGEEP-----K
Hs_RPB5_BAA07406 EETYLRLWLRIRKTIIMQLCHDRGYLVTDQDELDTLEEFKAQFGDKPSEGRPR
Dm_RPB5_NP_610630 AETYKLWLRIRKTIIMQLSHDRGYLVTDQDELDTLEEFKEMFGDKPSEKRPA
Ce_RPB5_Q9N5K2 LETYRLWRIRKTIIMQMVHDRGYLVQAQDELDTLEEFKQVQYGDPRSEKKPA
Populus_trichocarpa_584052 EEIKRLLRIRKTIIMQMLKDRGYFVGDFFETKMTREQFESKYGNM-----K
Vitis_vinifera_CA065489 EEISRLFRIRKTIIMQMLKDRGYFVGDFFENMTKHQFVSKFGENM-----K
NRPB5/NRPD5_At3g22320 EELKRLYRIQKTIIMQMLRDRGYFTADSELTMTKQFIRKHGDNM-----K
Medicago_truncatula_ABO78350 EEITRLYRIRKTIIMQMLKDRNYLVGDFELNMSKHDFKDKYGENM-----K
Oryza_sativa_EAZ13876 VMVGRLLVIRRTVMQMLRDRGYLVVEHELAMGRDRFLRKYGESF-----H
Oryza_sativa_NP_001044564 VMVGRLLVIRRTVMQMLRDRGYLVVEHELAMGRDRFLRKYGESF-----H
Zea_mays_ACF85599 ATVGRLLYRIRRTVMQMLRDRGYLVVDHELATSRDRFLRKYGESF-----H
Oryza_sativa_CAD41325 EETSRLFRIRRTVMQMLRDRGYLVTELDIDLPRGDFVARFGDPV-----D
Oryza_sativa_EAG31161 EETSRLFRIRRTVMQMLRDRGYLVTELDIDLPRGDFVARFGDPV-----D
Zea_mays_ACF81264 DEISRLFRIRRTVYEMLRDRGYVVRDEQIKLERHKFIERYGPNV-----R
Physcomitrella_patens_206246 EQSARLYRIRKTIIMEMLRDRDYVVADELTLSEKQFREKYGDEP-----K
Physcomitrella_patens_55574 EQCARLYRIRKTIIMEMLRDRDYVVADELTLSEKQFREKYGDEP-----K
Physcomitrella_patens_231299 RQSTHLYQVRKVKLEMMRDLDYVADNELTLTNEQFCEKYREDP-----K
Physcomitrella_patens_136486 RQSTHLYQVRKVKLEMMRDLDYVADNELTLTNEQFCEKYREDP-----K
Sc_RPB5_CAA85113 RNISRLWRARFTVKEMVDRGYFITQEEVELPLEDFKAKYCDMSG-----D
NRPB5-like_At5g57980 DEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCKTMN----KVN
NRPB5-like_At3g54490 VECHKYYLARRTMEMLRDRGYDVSDEDINLSLQDFRALYGEHP-----D
Brassica_napus_AAF81222 VDSHSYLLARRTMEMLRDRGYDISNEDINLSLQDFRALYGEHP-----D
NRPB5_At3g57080_NP_191267 EESHRYLLARRNGLQMLRDRGYEVSDDEDINLSLHDFRTVYGERP-----D
Populus_trichocarpa_57931 TESHRYLLSRRTVLEMLKDRGYSVPSSEIDISLQDFRGVYQNP-----D
Vitis_vinifera_CA063075 IESHRYLLARRTLEMLRDRGYSIPALDIDISLQDFRSFYSQKP-----D
Medicago_truncatula_ABN07995 IESHRYLLSRRTVLEMLKDRGYSIPSDEIQLSLDDFRQIHGQSP-----D
Populus_trichocarpa_48513 AESYRYIISRTVLEMLKDRGYDVLDSSELNRSLEFRSDFVFGNSP-----D
Vitis_vinifera_CA042914 IESYRYLLSRRTLFQMLSDRGYNVPHSELTRSLSDFRASFGHNP-----D
NRPB5-like_At3g54490 IESKRFLYARRTAFEMLRDRGYEVNEAELSLTLESEFRSDFVGEKP-----E
Medicago_truncatula_ABD28306 IETIRYFECRKTLMMDLHDRGYNVSESDLTLSLSEFRSDFGFEFP-----K
Oryza_sativa_NP_001065723 VESHRLFLARRTAMEMLRDRGYSVPEAEIARTLPEFRAWWAEKP-----G
Oryza_sativa_NP_001066119 VESHRLFLARRTAMEMLRDRGYSVPEAEIARTLPEFRAWWAEKP-----G
Oryza_sativa_EAY79909 VESHRLFLARRTAMEMLRDRGYSVPEAEIARTLPEFRAWWAEKP-----G
Zea_mays_ACF87172 VESHRLFLARRTAMEMLRDRGYAVPEEELARTLPEFRAWWEYRP-----E
consensus ES RLYRIRRTVMEMLRDRGY V E EL LTL DFR KYGE P

C_reinhardtii_XP_001697601 KDDLITLVPRQDDPTEQIFVFFP-----EFQKVGKTIK-LLAERM
O_lucimarinus_XP_001417617 RESLITLAPKRDDPSENIFVFFP-----DEEKVGKTIK-DLAKRM
Hs_RPB5_BAA07406 RTDLITLVVAHNDPDTQDMFVFFP-----EEPKIGIKTIK-VYQCQR
Dm_RPB5_NP_610630 RSDLITLVVAHNDPDTQDMFVFFP-----EEPKIGIKTIK-TYCTRM
Ce_RPB5_Q9N5K2 RSDLITLVVAHNDPADQDMFVFFP-----EDAKIGIKTIK-AICQQM
Populus_trichocarpa_584052 REDLVINKTKRNDSSDQIYVFFP-----EEAKVGKTIK-TYTNRM
Vitis_vinifera_CA065489 REDLVINKAKRTDSSDQIYVFFP-----EQKVGKTIK-TYTNRM
NRPB5/NRPD5_At3g22320 REDLVINKAKRNDSSDQIYVFFP-----DEAKVGKTIK-MYTNRM
Medicago_truncatula_ABO78350 REDLVINKTKKDKPSSDQIYVFFP-----EEAKVGKTIK-TYTNRM
Oryza_sativa_EAZ13876 REDLLINKYKKNPSSDQIYVFFP-----NDDKVMKHIK-KYVEMM
Oryza_sativa_NP_001044564 REDLLINKYKKNPSSDQIYVFFP-----NDDKVMKHIK-KYVEMM
Zea_mays_ACF85599 REDLLINKYKKNPSSDQIYVFFP-----NDDKVMKHIK-KYVEMM
Oryza_sativa_CAD41325 RDHLVFSRHKKDNGADQIYVFFP-----KDAKPGVKTIR-SYVERM
Oryza_sativa_EAZ31161 RDHLVFSRHKKDNGADQIYVFFP-----KDAKPGVKTIR-SYVERM
Zea_mays_ACF81264 RDELTFNATKLNQPSDQIYVFFP-----NEAKPGVKTIR-NYVEKM
Physcomitrella_patens_206246 REDLVIQKPRRSNNAEHIFVFFP-----EEAKVGKTIK-TYVDRM
Physcomitrella_patens_55574 REDLVIQKPKRSNNAEHIFVFFP-----EEAKVGKTIK-TYVDRM
Physcomitrella_patens_231299 QEDLMILPKKSNNAEHGPKTGG-----KGRVGLKTIK-TCKKRM
Physcomitrella_patens_136486 QEDLMILPKKSNNAEHVMVFHEF-----FSPFPTLVGLKTIK-TCKKRM
Sc_RPB5_CAA85113 RPQRKMMSFQANPTEESISKFPDMGSLWVEFCDEPSVGVKTIK-TFVIHI
NRPB5-like_At5g57980 KEALFVSANKGPNPADKIYVFFP-----EGPKVGPVIKKEVAIKM
NRPB5-like_At3g41340 VDRLRISAKHRFDSKIKSVVFC-----GTGIVKVNAMRVIAADVL
Brassica_napus_AAF81222 VDRLRISAQHCSDSSKIIAVVFC-----GSGIVKVS AIRDIAADVL
NRPB5_At3g57080_NP_191267 VDRLRISALHRSDSTKKVKIVFF-----GTSMVKNVAIRSVVADIL
Populus_trichocarpa_57931 IELLKFSATHKSDPSKRMLVIFC-----GLGVVKVGMIRLITVQIT
Vitis_vinifera_CA063075 PDRLRISAAALRSDPSKIKLVIFC-----GPDVVKNVAIRSIATQIV
Medicago_truncatula_ABN07995 VDRLRLTATHATNP SKRILVVFS-----GPGIVK VNGVRDIAGQIV
Populus_trichocarpa_48513 LDSLRFSVSLRSIPHKKTLMVFL-----GTDEIKTANIRTVYGGQIL
Vitis_vinifera_CA042914 PSRLRILCLPLISSPKIKLVVFC-----GTDEIRKAVIRVIF-QQI
NRPB5-like_At3g54490 LERLRICVPLRSDPKKIKLVVFM-----GTPEITVKSVALHIQIS
Medicago_truncatula_ABD28306 PHTLVGSVSLRSNPSIKVQVFFP-----GTDDIRKSNLIVIQSQIV
Oryza_sativa_NP_001065723 IERLAFTTTLVSDPSKVKVLFVFC-----PPEPVKIATIREIYLQTK
Oryza_sativa_NP_001066119 IERLAFTTTLVSDPSKVKVLFVFC-----PPEPVKIATIREIYLQTK
Oryza_sativa_EAY79909 IERLAFTTTLVSDPSKVKVLFVFC-----PPEPVKIATIREIYLQTK
Zea_mays_ACF87172 LERLAFSTTLTSDPSKVKVFFC-----PPGPVKIAAIRLIYTEVK
consensus RE L I RSDPSD IYVFFP E KVGKTIK Y M

```

end Jaw domain→
C_reinhardtii_XP_001697601 KDEKVNRAIMVTPSKFTPFSAKSALEDNR-PKYHIEHFLESELLVNITEHV
O_lucimarinus_XP_001417617 KDENVFRAIIVVQASLTPFAKQSLLECQTQKFYIEQFQETELLVNIIDHV
Hs_RPB5_BAA07406 QEENITRALIVVQGMTPSAKQSLVDMA-PKYIIEQFLEQELLNITEHE
Dm_RPB5_NP_610630 QEENIHRAIVVQGMTPSAKQSLVDMA-PKYIIEQFLESELLNITEHE
Ce_RPB5_Q9N5K2 QEQNISRAIIVVQGMTPSAKQSIGDMA-PKYMLEHFLEAELLVNITEHE
Populus_trichocarpa_584052 KSENVFRAILLVVQNLTPFARTCINEIS-TKFHLEVFQEAELLVNIKEHV
Vitis_vinifera_CA065489 KSENVFRAILLVVQNLTPFARTCINEIS-TKFHLEVFQEAELLVNIKEHV
NRPB5/NRPD5_At3g22320 KSENVFRAILLVVQNLTPFARTCISEIS-SKFHLEVFQEAELLVNIKEHV
Medicago_truncatula_ABO78350 NSENVYRAILVVCQTSLTPFAKTCVSEIA-SKFHLEVFQEAELLVNIKEHV
Oryza_sativa_EAZ13876 KAENVSRAVLVQLQNLTPFARSFLOELE-PKIHLEIFQEAELLVNIKEHV
Oryza_sativa_NP_001044564 KAENVSRAVLVQLQNLTPFARSFLOELE-PKIHLEIFQEAELLVNIKEHV
Zea_mays_ACF85599 THENVSRAVLVQLQNLTPFAKSFLEIELE-PKIHLEIFQEAELLVNIKEHV
Oryza_sativa_CAD41325 KQESVFNGLIVVQALSAFARSAVQEV-SKFHLEVFQEAELLVNIKDHT
Oryza_sativa_EAZ31161 KQESVFNGLIVVQALSAFARSAVQEV-SKFHLEVFQEAELLVNIKDHT
Zea_mays_ACF81264 KNENVFAGILVVQALSAFARSAVQEV-SKYHLEVFQEAELLVNIKDHT
Physcomitrella_patens_206246 KTEENVHRAILVVQNLTPFARQCVSEMA-SKYHLEVFQEAELLVNIKEHV
Physcomitrella_patens_55574 KTEENVHRAILVVQNLTPFARQCVSEMS-SKYHVEVFQEAELLVNIKDHT
Physcomitrella_patens_231299 KRENVPRAVFVVQOHITPLSKQYISRKA-QKYHLEVFLEPEFLVNITECY
Physcomitrella_patens_136486 KRENVPRAVFVVQOHITPLSKQYISRKA-QKYHLEVFLEPEFLVNITECY
Sc_RPB5_CAA85113 QEKNFQGTGFVYQNNITPSAMKLVPSIP--PATIETFNAAALVNVITHEE
NRPB5-like_At5g57980 RDDKVRHGIIVVPMAITAPARMAVSELN-KMLTIEVFQEAELLVNITEHK
NRPE5-like_At2g41340 SRENITGLILVQLSHITNQALKAV-ELF--SFKVELFEITDLLVNVSKHV
Brassica_napus_AAF81222 GRENLITGLILVQLSDITNQALKAV-ELF--SFKVELFQITELLVNITKHV
NRPE5_At3g57080_NP_191267 SQETITGLILVQLNHVTNQALKAI-ELF--SFKVEIFQITDLLVNITKHS
Populus_trichocarpa_57931 DRDSLITGLILVQLNNITNQAMKAL-DLF--KFKIEIFQITDLLVNITKHI
Vitis_vinifera_CA063075 NKDSLKILVQLNHITSQALKAV-DLF--SFQVEKQITDLLVNITKHV
Medicago_truncatula_ABN07995 NRESLITGLILVQNLITNSQALKAV-NLL--SFKVEIFQITDLLVNITKHV
Populus_trichocarpa_48513 NKESLHGLILVQLSKMNHFAKKEK-ELF--PFKVEVFQITDLLVNITKHV
Vitis_vinifera_CA042914 NREGLHRLILVQLSKMNSHARKVV-DEY--PIKVEFFQITELLVNITKHV
NRPE5-like_At3g54490 NNVGLHAMILVQLSKMNHFAQKAL-TTF--PFTVETFPIDLLVNITKHI
Medicago_truncatula_ABD28306 DKERLSRLILVLMQSKMNTSYARKEL-ENC--PFKVEIIFQITDLLVNITKHV
Oryza_sativa_NP_001065723 E-ENLSRLVLILQSKILSRAREAIKEIF--KFKVDIFQATDLLVNITKHV
Oryza_sativa_NP_001066119 E-ENLSRLVLILQSKILSRAREAIKEIF--KFKVDIFQATDLLVNITKHV
Oryza_sativa_EAY79909 E-ENLSRLVLILQSKILSRAREAIKEIF--KFKVDIFQATDLLVNITKHV
Zea_mays_ACF87172 D-ENLSRLILVQLGKIMSTTRESIKEIF--RFKVDITFQITELLVNITKHV
consensus ENV RAILVVQQ IT AR V EL KF LEVFQE ELLVNITEHV
Start of Assembly domain
C_reinhardtii_XP_001697601 LVPEHRILSPDEKRTLLDRYKIKETQ-----LPRIQASDAVA
O_lucimarinus_XP_001417617 LVPEHILLSDQKRTLLDRYKVKDTQ-----LPRIQMHDPPIA
Hs_RPB5_BAA07406 LVPEHVVMTKEEVSELLARYKLENQ-----LPRIQAGDPVA
Dm_RPB5_NP_610630 LVPEHVVMTVEEKQELLRYKLENM-----LMRIQAGDPVA
Ce_RPB5_Q9N5K2 LVPEHVMTAEKAEELLARYKLDKQ-----LPRIQQCDPVA
Populus_trichocarpa_584052 LVPEHQVLNNEEKTLLERYTVKETQ-----LPRIQITDPIA
Vitis_vinifera_CA065489 LVPEHQVLTSEEKTLLERYTVKETQ-----LPRIQVSDPIA
NRPB5/NRPD5_At3g22320 LVPEHQVLTSEEKTLLERYTVKETQ-----LPRIQVTDPIA
Medicago_truncatula_ABO78350 LVPEHQVLTSEEKTLLERYTVKETQ-----LPRIQVTDPIA
Oryza_sativa_EAZ13876 LVPEHQVLTSEEKTLLERYTVKETQVYIHDHMLGEIIFLRRSHVNDPMA
Oryza_sativa_NP_001044564 LVPEHQVLTSEEKTLLERYTVKETQ-----LPRIQITDPIA
Zea_mays_ACF85599 LVPEHQVLTSEEKTLLERYTVKETQ-----LPRIQITDPIA
Oryza_sativa_CAD41325 LVPEHELLTPEQKTLLERYTVKETQ-----LPRIQITDPIA
Oryza_sativa_EAZ31161 LVPEHELLTPEQKTLLERYTVKETQILSLTQLV-KCVNLPRIQITDPIA
Zea_mays_ACF81264 LVPEHVLLTPEDKTLLERYTVKETQ-----LPRIQITDPIA
Physcomitrella_patens_206246 LVPLHEVLTPEDEKTLLERYTVKETQ-----QLPRMQENDPVA
Physcomitrella_patens_55574 LVPQHEVLNAAEKITLLQRYTVKETQ-----QLPRMQENDPVA
Physcomitrella_patens_231299 LVPLHEILTPEEKNTLLERYTEGNPVM-----VLLPVMQHNDPVA
Physcomitrella_patens_136486 LVPLHEILTPEEKNTLLERYTEGNP-----LPVMQHNDPVA
Sc_RPB5_CAA85113 LVPKHIRLSDEKRELLKRYRLKESQ-----LPRIQRADPVA
NRPB5-like_At5g57980 LVNKKYVLDQAKKLLNNTYTVQDTQ-----LPRILVTDPIA
NRPE5-like_At2g41340 LRPKHQVLDNDEKESLLKFKFSIEEKQ-----LPRLSKDPVIV
Brassica_napus_AAF81222 LRPKHVHLNDEKESLFFKFSIQEQQ-----LPKLLKDDPTA
NRPE5_At3g57080_NP_191267 LKPQHQVLDNDEKESLFFKFSIEEKQ-----LPRISKDDAIV
Populus_trichocarpa_57931 LKPKHQVLSAQKRLKLYKFSIEEKQ-----LPRLLKDDAIV
Vitis_vinifera_CA063075 LKPKHQVLSAQKRLKLYKFSIEEKQ-----LPRMLQDDAIV
Medicago_truncatula_ABN07995 LKPKHQVLDNDEKESLFFKFSIEEKQ-----LPRMLQDDAIV
Populus_trichocarpa_48513 LQPQMDILTAEQKQVVMNKYKLEDKQ-----LPRMLESDAIV
Vitis_vinifera_CA042914 LVPKHEVLSAQKRLKLYKFSIEEKQ-----FPIMQKDDAIV
NRPE5-like_At3g54490 QPKIEILNKEEKEQLLRKHALEDKQ-----LPYLQEKDSFV
Medicago_truncatula_ABD28306 LQPKYEVLTANEKQKLLNKYKVEEKQ-----LPHMLRTDPIA
Oryza_sativa_NP_001065723 LKPKHEVLSADQKAKLLKEYNVEDSQ-----LPRMLETDAVA
Oryza_sativa_NP_001066119 LKPKHEVLSADQKAKLLKEYNVEDSQ-----LPRMLETDAVA
Oryza_sativa_EAY79909 LKPKHEVLSADQKAKLLKEYNVEDSQ-----LPRMLETDAVA
Zea_mays_ACF87172 LKPKHEVLTAEQKAKLLKEYNVEDSQ-----LPRMLENDAVA
consensus LVP H VLT EEK TLL RYTVKETQ LPRIQ DPIA

```

```

C_reinhardtii_XP_001697601 RYLQ--LQRGQVVRIVRP-SETAGRYVVTYRFCPPLWR
O_lucimarinus_XP_001417617 RYYG--MRRGQVVRIRP-SETAGRYVVTYRLCV----
Hs_RPB5_BAA07406 RYFG--IRRGQVVKIIRP-SETAGRYITTYRLVQ----
Dm_RPB5_NP_610630 RYFG--LKRQVVKIIRS-SETAGRYISYRLVC----
Ce_RPB5_Q9N5K2 RYFG--LRRGQVVKIIRP-SETAGRYITTYRLVV----
Populus_trichocarpa_584052 RYYG--LKRQVVKIIRP-SETAGRYVVTYRYVI----
Vitis_vinifera_CAO65489 RYFG--LKRQVVKIIRP-SETAGRYITTYRYVV----
NRPB5/NRPD5_At3g22320 RYFG--LKRQVVKIIRP-SETAGRYVVTYRYVV----
Medicago_truncatula_ABO78350 RYYG--LKRQVVKIIRP-SETAGRYVVTYRFVV----
Oryza_sativa_EAZ13876 VIVENLNYLSHIQLAIAPNMSTYGYKYMAGLVP---
Oryza_sativa_NP_001044564 RYYG--LRRGQVVKIIRP-SETAGRYVVTYRYVV----
Zea_mays_ACF85599 RYYG--LRRGQVVKIIRP-SETAGRYVVTYRYVV----
Oryza_sativa_CAD41325 RYYG--MKRGQVVKIIRA-SETAGRYVVTYRYVV----
Oryza_sativa_EAZ31161 RYYG--MKRGQVVKIIRA-SETAGRYVVTYRYVV----
Zea_mays_ACF81264 RYYG--MKRGQVVKITRA-SETAGRYITTYRYVV----
Physcomitrella_patens_206246 RYYG--LKRQVVKIIRP-SETAGRYVVTYRFVV----
Physcomitrella_patens_55574 RYYG--LKRQVVKIIRP-SETAGRYVVTYRFVV----
Physcomitrella_patens_231299 RYYG--INPGQVVKIIQS-SETAGRYVVTYRLFV----
Physcomitrella_patens_136486 RYYG--INPGQVVKIIQS-SETAGRYVVTYRLFV----
Sc_RPB5_CAA85113 LYLQ--LKRGEVVKIIRK-SETSGRYASYRICM---
NRPB5-like_At5g57980 RYYG--LKRQVVKIIRS-DATSLDYTYRFAV----
NRPE5-like_At2g41340 RYYG--LETGQVMKVTYKDELSES-HVTYRCVS----
Brassica_napus_AAF81222 KYYG--LEKQVVEVVTYKGESESDHVSYRCAW----
NRPE5_At3g57080_NP_191267 RYYG--LEKQVVKVNYRGELTES-HVAFRCVW----
Populus_trichocarpa_57931 RYYG--LERGQVVKVTYDGDITGS-HVTYRCVW----
Vitis_vinifera_CAO63075 RYYG--LEKQVVKVIYNGEITGS-HVTYRCVW----
Medicago_truncatula_ABN07995 RYYG--LQRGQVVKVITYTGEITQM-HVTYRCVW----
Populus_trichocarpa_48513 QYYG--LQRGQVVKIITYSGEIVDH-LVTYRCVT----
Vitis_vinifera_CAO42914 RYYG--LEKQVVKIITYKGGMTDS-LVTYRCVS----
NRPE5-like_At3g54490 RYYG--LKKKQVVKIITYSKEPVGDFVTYRCIT----
Medicago_truncatula_ABD28306 SYYG--LEKQVVKIISHSGEMFNS-LVMYRCVV----
Oryza_sativa_NP_001065723 RYYG--FDKGTVVKVIYDGELTGK-RVAYRCVF----
Oryza_sativa_NP_001066119 RYYG--FDKGTVVKVTYDGELTGK-RVAYRCVF----
Oryza_sativa_EAY79909 RYYG--FDKGTVVKVIYDGELTGK-RVAYRCVF----
Zea_mays_ACF87172 RYYG--LGKGTVVKVIYDSELTGN-HVTYRCIT----
consensus RYYG LKRQVVKIIR SETAGRYVVTYR VV

```

Figure S17. The N-terminal extension of NRPE5 is required for the protein's stability and function.

A. Diagram highlighting the jaw and assembly domains and the short N-terminal extension present in NRPE5 but absent in NRPB5/NRPD5. Underlined amino acids were deleted in the *35S:FLAG-ΔN-NRPE5* transgene.

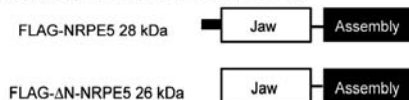
B. *AtSN1* retrotransposon expression in Pol V mutants, wild-type, and *35S:FLAG-ΔN-NRPE5 nrpe5* lines assayed by strand-specific RT-PCR.

C. *AtSN1* methylation in *35S:FLAG-NRPE5 nrpe5*, *35S:FLAG-ΔN-NRPE5 nrpe5* lines and Pol V mutants compared to wild-type.

D. Methylation-sensitive Southern blot analysis of 5S rRNA genes in Pol V mutants, wild-type, and *35S:FLAG-ΔN-NRPE5 nrpe5* lines.

E. RT-PCR and immunoblot analysis of mRNA and protein levels in T2 generation plants of *35S:FLAG-NRPE5 nrpe5* and *35S:FLAG-ΔN-NRPE5 nrpe5* lines. The upper panels show RT-PCR reactions, including actin and no reverse transcriptase (no RT) controls. In the bottom panel, equal amounts of tissue homogenate were subjected to anti-FLAG IP and immunoblot detection of the tagged proteins.

A. Deletion of NRPE5 N terminus

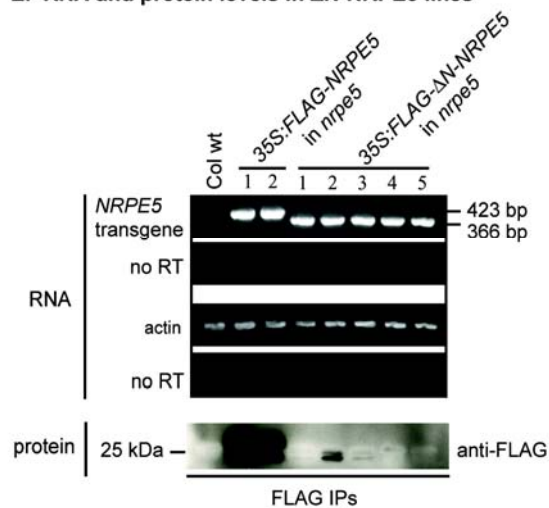


MEVKGKETASVLCLSKYYDL SSEESHRYLA RRINGLQMLRDRGYEV SDE
 DINLSLHDFRTVYGERPDV DRLRISALHRSDSTKKVKIVFFGTSMVKVNAI
 RSVVADILSQETITGLILVLQNHVTNQALKAIELFSPKVEIFQITDLLVNI
 TKHSLKPQHQLNDEEKTLLKKFSIEBKQLPRISKDAIVRYGLEGKQ
 VVKVNYRGETESHVAFRCVW

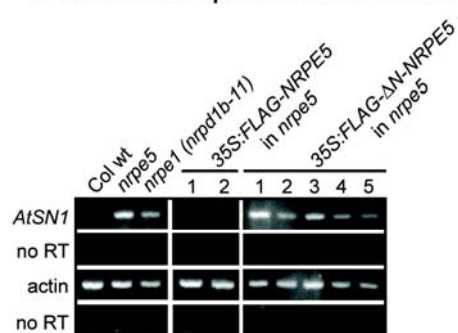
C. *AtSN1* methylation in ΔN-NRPE5 lines



E. RNA and protein levels in ΔN-NRPE5 lines



B. *AtSN1* transcription in ΔN-NRPE5 lines



D. 5s rDNA methylation in ΔN-NRPE5 lines

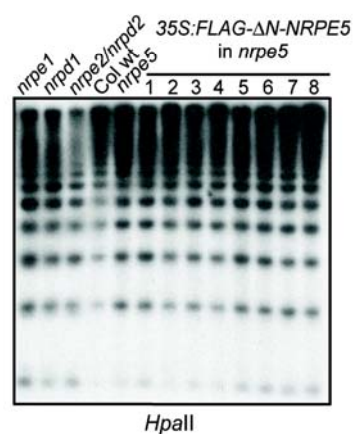


Figure S18. Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS in affinity purified FLAG-NRPE5 samples. In the full-length protein sequences that follow, peptides highlighted in yellow or green indicate sequenced tryptic peptides that do not overlap with other sequenced peptides. Cyan highlighting denotes sequences represented by two overlapping peptides. Magenta highlighting indicates regions corresponding to three or more overlapping peptide sequences.

NRPE1 (At2g40030)

MEEESTSEILDGEIVGITFALASHHEICIQSISESAINHPSQLTNAFLGLPLEFGKCESCGAT
 EPDKCEGHFGYIQLPVPIYHPAHVNELKQMLSLCLKCLKIKKAKGTSGGLADRLLGVC
 CEEASQISIKDRASDGASYLELKLPSRSRLQPGCWNFLERYGYRYGSDYTRPLLAREVKE
 ILRRIPEESRKKLTAKGHIPQEGYILEYLPVPPNCLSVPEASDGFSTMSVDPRIELKDVLK
 KVIAIKSSRSGETNFESHKAEASEMFRVVDTYLQVRGTAKAARNIDMRYGVSKISDSSSS
 KAWTEKMRTLFIKRGSGFSSRSVITGDAYRHVNEVGPIEIAQRITFEERVSVHNRGYLQ
 KLVDDKLCLSYTQGSTTYSLRDGSKGHTELKPGQVVHRRVMDGDVVFINRPPTTHKHS
 LQALRVYVHEDNTVKINPLMCSPLSADFDGDCVHLFYQSLSAKAEVMELFSVEKQLLS
 SHTGQLILQMGSDSLLSLRVMLERVFLDKATAQQLAMYGSLSLPPPALRKSSKSGPAWT
 VFQILQLAFPERLSCKGDRFLVDGSDLLKDFDGDAMGSIINEIVTSIFLEKGPKETLGFDD
 SLQPLLMESLFAEGFSLLEDLSMSRADMDVIHNLIIREISPMVSRRLRSYRDELQLENSIH
 KVKEVAANFMLKSYSIRNLIDIKSNSAITKLVQQTGFLGLQLSDKKKFYTKTLVEDMAIF
 CKRKYGRISSSGDFGIVKGCFFHGLDPYEEMAHSIAAREVIVRSSRGLAEPGTLFKNLMA
 VLRDIVITNDGTVRNTCSNSVIQFKYGVDSERGHQGLFEAGEPVGVLAAATAMSNPAYKA
 VLDSSPNSNSSWELMKEVLLCKVNFQNTTNDRRVILYLNECHCGKRFCQENA ACTVRN
 KLNK VSLKDTAVEFLVEYRKQPTISEIFGIDSLHGHHLNKTLLQDWNISMQDIHQKCE
 DVINSLGQKKKKKATDDFKRTSLSVSECCSFRDPCGSKGSDMPCLTFSYNATDPLERT
 LDVLCNTVYPVLLIIVIKGDSRICSANIIWNSSDMTTWIRNRHASRRGEWVLDVTVEKSA
 VKQSGDAWRVVIDSCLSVLHLIDTKRSIPYSVKVQVQELLGLSCAFEQAVQRLSASVRMV
 SKGVLKEHIILLANNMTCSGTMLGFNSGGYKALTRSLNIKAPFTEATLIAPRKCFEKA
 KCHTDSLSTVVGSCSWGKRVDVGTGSQFELLWNQKETGLDDKEETDVYSFLQMVISTT
 NADAFVSSPGFDVTEEEMA EWAESPERDSALGEPKFEDSADFQNLHDEGKPSGANWEK
 SSSWDNGCSGGSEWGVSKSTGGEANPESNWEKTTNVEKEDAWSSWNTRKDAQESSKS
 DSGGAWGIKTKDADADTPNWETSPAPKDSIVPENNEPTSDVWGHKSVSDKSWDKKN
 WGTESAPAAWGSTDAAVWGSSDKNSETESDAAAWGSRDKNNSDVGSGAGVLGPWN
 KKSSETESNGATWGSSDKTKSGAAAWNSWDKKNIETDSEPAAWGSQGKNSETESGP
 AAWGAWDKKKSETEPGPAGWGMGDKNSETELGPAAMGNWDKKSSTKSGPAAWG
 STDAAAWGSSDKNSETESDAAAWGSRNKTSEIESGAGAWGSWGQPSPTAEDKDTN
 EDDRNPWVSLKETKSREKDDKERSQWGNPAKKFPSSGGWSNGGADWKGNRNHTPR
 PPRSEDNLPMTATRQR LDSFTSEEQELSDVEPVMR TLRKIMHPSAYPDGDPISDDDK
 TFVLEKILNFHPQKETKLGSGVDFITVDKHTIFSDSRCFFVSTDGAKQDFS YRKS LNNY
 LMKKYPDRAEEFIDKYFTKPRPSGNRDRNNQDATPPGEEQSQPPNQSIGNGGDDFQTQT

QSQSPSQTRAQSPSQAQAQSPSQTQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQS
TQTQSPSQTQAQAQSPSSQSPSQTQT

Notes:

427/1976 amino acids are represented by sequenced peptides =22% coverage.

All peptides are specific to NRPE1 (NRPD1b), meaning that none are identical to any other protein, including NRPD1 (NRPD1a).

NRPE2/NRPD2 (At3g23780)

MPDMDIDVKDLEEFATTGEINLSELGEGFLQSFCKKAATSFDDKYGLISHQLNSYNYFI
EHGLQNVFQSFGEMLVEPSFDVVKKKDNDWRYATVKFGEVTVEKPTFFSDDKELEFLP
WHARLQNMTYSARIKVVNVQVEVFKNTVVKSDKFKTGQDNYVEKKILDVKKQDILIGSI
PVMVKSILCKTSEK GKENCKKGDCAFDOGGYFVIKGAEKVFIAQEQMCTKRLWISNSP
WTVSFRSENKRNRFRIVRLSENEKAEDYKRREKVLTVYFLSTEIPVWLLFFALGVSSDKEA
MDLIAFDGDDASITNSLIASIHVADAVCEAFRCGNNALTYVEQQIKSTKFPPAESVDECL
HLYLFPGLQSLKKKARFLGYMVKCLLSYAGKRKCENRDSFRNKRIELAGELLERREIRV
HLAHARRKMTRAMQKHLSGDGDLPKPIEHYLDASVITNGLSRAFSTGAWSHPPFRKMERV
SGVVANLGRANPLQTLIDLRRTROQVLYTGKVGDARYPHPSHWGRVCFLSTPDGENCG
LVKNMSLLGLVSTQSLESVVEKLFACGMEELMDDTCTPLFGKHKVLLNGDWVGLCAD
SESFVAELKSRRRQSELPREMEIKRDKDDNEVRIFTDAGRLLRPLL VVENLQKQKQEKPS
QYPFDHLLDHGILELIGIEEEEDCNTAWGIKQLLKEPKIYTHCELDLSFLLGVSCAVVPPFA
NHDHGRVLYQSQKHCQQAIGFSSTNPNI RCDTLSQQLFYPQKPLFKTLASECLKKEVLF
NGQNAIVAVNVHLGYNQEDSIVMNKASLERGMFRSEQIRSYKAEVDAKDSEKRRKMD
ELVQFGKTHSKIGKVDSDLEDDGFPIGANMSTGDIVIGRCTESGADHSIKLKHTEGIVQK
VVLSSNDEGKNFAAVSLRQVRSPCLGDKFSSMHGQKGVLYGYLEEQQNFPFTIQGIVPDI
VINPHAFPSRQTPGQLEEAALSKGIACPIQKEGSSAA YTKLTRHATPFSTPGVTEITEQLH
RAGFSRWGNERNVNGRSGEMMRSMIFMGPTFYQRLVHMSSEDKVKFRNTGPVHPLTRQ
PVADRKRFGGKIFGEMERDCLIAHGASANLHERLFTLSDSQQMHICRCKCTYANVIERTP
SSGRKIRGPYCRVCVSSDHVVRVYVPYGA KLLCQELFSMGITLNFDTKLC

Notes:

281/1172 amino acids represented in sequenced peptides =24% coverage.

72/1172= 6% coverage is accounted for by peptides unique to NRPE2/NRPD2a. The remaining 18% of the peptides match NRPE2/NRPD2a as well as the NRPD2b pseudogene. However, the latter gene is non-functional, and no peptides that would uniquely identify NRPD2b were detected.

NRPE3a/NRPD3/NRPB3 (At2g15430)

MDGATYQRFPKIKIRELKDDYAKFELRETDVSMANALRVMISEVPTVAIDLVEIEVNSS
VLNDEFIAHRLGLIPLTSERAMSMRFSRDCDACDGDGQCEFCVSEFRLSSKCVTDQTLT
VTSRDLYSADPTVTPVDFTIDSSVSDSSEHKGIIIVKLRGQELKLRAIARKGIGKDHAKW
SPAATVTFMYEPDIINEDMMDTLDSEKIDLIESSPTKVFGMDPVTRQVVVVDPPEAYTY
DEEVIKKAEMGKPLIEISPKDDSFIFTVESTGAVKASQLVLNAIDLKQKLDVRLSD
DTVEADDQFGE LGAHMRGG

Notes:

155/319 amino acids are represented by sequenced peptides =48% coverage

115/319=36% unique coverage. 36% of the coverage corresponds to peptides that match only NRPE3a. The other 12% matches either NRPE3a or NRPE3b.

NRPE3b (At2g15400)

MDGVTYQRFPTVKIRELKDDYAKFELRETDVSMANALRRVMISEVPTMAIHLVKIEVNS
SVLNDEFIAQRLSLIPLT SERAMSMRFCQDCEDCNGDEHCEFCSEVFPPLSAKCVTDQTLT
VTSRDLYSADPTVTPVDFTS NSSTSDSSEHKGIII AKLRRGQELK LKALARKGIGKD HAK
WSPAATV TYMYEPDIINEEMNTLTDEEKIDLIESPTK VFGIDPVTG QVVVDPEAYT
YDEEVIKKAEAMGK PGLIEIHPKHDSFVFTVESTGALKASQLVLNAIDILKQKLD AIRLSD
NTVEADDQFGELGAHMREG

Notes:

53/319 amino acids are represented by sequenced peptides = 16% coverage

13/319=4% coverage corresponds to peptides matching only NRPE3b, whereas the remaining 12% of the coverage matches either NRPE3b or NRPE3a.

NRPE4/NRPD4 (At4g15950)

MSEKGGKGLKSSLKSKDGGKDGSS TKLKKGRKIHFDQGT PPANYKILNVSSDQQPFQSS
AAKCGKSDKPTKSSKNSLHSEFELKDLPENAE CMMDC EAFQILDG IKGQLVGLSEDP SIKI
PVS YDRALAYVESC VHYTNPQSVR KVL EPLKTYGISD GEMCVIANASSES VDEV LAFIPS
LTKKKEVINQPLQDALEELSKLKKSE

17/205 amino acids are represented by sequenced peptides=8% coverage. All peptides sequenced match only At4g15950 and no other RPB4-like protein.

NRPB4 (At5g09920)

MSGEEEEENAAELKIGDEFLKAKCLMNCEVSLILEHKFEQLQQISEDPMNQVSQVFEKSL
QYVKRFSRYKNPDAVRQVREILSRHQLTEFELCVLGNLCPETVEEAVAMVPSLKTKGRA
HDDEAIEKMLNDLSLVKRFE

0/138 amino acids are represented by sequenced peptides=0% coverage. No peptides were identified that matched this protein sequence.

NRPB5/NRPD5 (formerly AtRPB5a, AtRPB24.3) (At3g22320)

MLTEEELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNMKREDLVT LKA
KRNDNSDQLYIFFPDEAKVGVKTMKMYTNRMKSENVFRAILVVQQNLTPFARTCISEIS
SKFHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEKKTLLERYTVKETQLPRIQVTDPIARY
FGLKRGQVVKIIRPSETAGRYVTYRYVV

0/205 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE5 (formerly AtRPB5b, AtRPB23.7)(At3g57080)

MEVKGKETASVLCLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDDEDINLSLHDF
RTVYGERPDVDRLRISALHRSDSTKKVKIVFFGTSMVKVNAIRSVVADILSQETITGLILV

LQNHVTNQALKAIELFSFKVEIFQITDLLVNITKHSCLKPQHQLNDEEKTTLKFSIIEK
QLPRISKKDAIVRYYGLEKGVVKNYRGELTESHVAFRCVW

145/222 amino acids are represented by sequenced peptides = 65% coverage

All peptides identified correspond to peptides that match NRPE5 only and no other family member.

NRPB5-like family member (synonym AtRPB5c) (At5g57980)

MSDMDDEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCCKTMNKVNKEALF
VSANKGPNPADKIYVFYPEGPKVGVVVIKKEVAIKMRDDKVHRGIVVVPMAITAPARM
AVSELNKMILTIEVFEEAELVTNITEHKLVNKYVYVLDQAKKLLNTYTVQDTQLPRILV
TDPLARYYGLKRGQVVKIRRSDATSLDYTYRFAV

0/210 amino acids are represented by sequenced peptides = 0% coverage

No peptides were identified that matched this protein sequence.

NRPE5-like family member (synonym AtRPB5d) (At2g41340)

MEGKGKEIVVGHISISKSSVECHKYLLARRTMEMLRDRGYDVSDDEDINLSLQQFRALY
GEHPDVDLLRISAKHRFDSSKISVFCGTGIVKVNAMRVIAADVLSRENITGLILVLQS
HITNQALKAVELFSFKVELFEITDLLVNVSKHVLVLPKHQVLNDKEKESLLKKFSIIEKQL
PRLSSKDPIVRYYGLETGQVMKVTYKDELSESHVITYRCVS

0/218 amino acids are represented by sequenced peptides = 0% coverage

No peptides were identified that matched this protein sequence.

NRPE5-like family member (At3g54490)

MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNE
AELSLTLSEFRSVFGEKPELERLRICVPLRSDPKKILVVFMGTEPITVKSVRALHIQISNN
VGLHAMILVLQSKMNHFAQKALTFPFTVETFPIDLLVNITKHIQQPKIEILNKEEKEQL
LRKHALEDKQLPYLQEKDSFVRYYGLKKKQVVKITYSKEPVGDFVITYRCII

0/233 amino acids are represented by sequenced peptides = 0% coverage

No peptides were identified that matched this protein sequence.

NRPB5 family member (likely pseudogene) (At3g16880)

MKKYIDQLKSANVFRAILVVQDIKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKEHVF
VPEHQALTTEEKQKFLERKRTSFQGFT

0/87 amino acids are represented by sequenced peptides = 0% coverage

No peptides were identified that matched this protein sequence. This protein is truncated relative to the other NRPB5-like proteins and likely is a pseudogene.

NRPE6a/NRPD6a/NRPB6a (At5g51940)

MADEDYNDVDDLGYEDEPAEIEEGVEEDVEMKENDDVNGEPIEAEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
PDGSFEEWGVDELIVEDSWKRQVGGD

26/144 amino acids are represented by sequenced peptides = 18% coverage

0/144 = 0% coverage corresponds to peptides that are NRPE6a-specific, the sequenced peptide also matches At2g04630.

NRPE6b/NRPD6b/NRPB6b (At2g04630)

MADDDYNEVDDLGYEDEPAEPEIEEGVEEDADIKENDDVNVDPLETEDKVETEPVQRP
 RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
 PDMSYEEWGVDELIVEDSWKRQVGGD

26/144 amino acids are represented by sequenced peptides = 18% coverage

0/144=0% of the coverage corresponds to peptides unique to this member of the protein family; the sequenced peptide also matches an identical sequence of At5g51940.

NRPE7 (At4g14660)

MFLKVQLPWNVMIPAENMDAKGLMLKRAILVELLEAFASKKATKELGYVAVTTLDKI
 GEGKIREHTGEVLFVPMFSGMTFKIFKGEIHHGVVHKVLKHGVFMRCGPIENVYLSYTK
 MPDYKYIPGENPIFMNEKTSRIQVETTVRVVIGIKWMEVEREFQALASLEGDYLGPLSE

0/177 amino acids are represented by sequenced peptides = 0% coverage

No peptides were identified that match this protein sequence.

NRPB7 (At5g59180)

MFFHIVLERNMQLHPRFFGRNLKENLVSKLMKDVEGTCSGRHGFVVAITGIDTIGKGLIR
 DGTGFVTFPVKYQC VVFRPFKGEILEAVVTLVNKMGFFAEAGPVQIFVSKHLIPDDMEF
 QAGDMPNYTTS DGSVKIQKECEVRLKIIGTRV DATAIFCVGTIKDDFLGVINDPAAA

0/176 amino acids are represented by sequenced peptides = 0% coverage

No peptides were identified that match this protein sequence.

NRPD7 (At3g22900)

MFIKVKLPWDVTIPAEDMDTGLMLQRAIVIRLLEAFSKEKATKDLGYLITPTILENIGEGK
 IKEQTGEIQFPVVFNGICFKMFKGEIVHGVVHKVHKTGVFLKSGPYEIIYLSHMKMPGYE
 FIPGENPFFMNQYMSRIQIGARVRFVVLDTREWAEKDFMALASIDGDNLGPF

0/174 amino acids are represented by sequenced peptides = 0% coverage

No peptides were identified that matched this protein sequence.

NRPB7 family member (At4g14520)

MFSEVEMARDVAICAKHLNGQSPHQILCROLLQDLIHEKACREHGFYLGITALKSIGNNK
 NNNIDNENNHQAKILTFPVSFTCRFTLPARGDILQGTVKKVLWNGAFIRSGPLRYAYLSL
 LKMPHYHYVHSPLESEDEKPHFQKDDLSKIAVGVVVRFQVLA VRFKERPHKRRNDYYVL
 ATLEGNGSFGPISLTGSDEPYM

0/200 amino acids are represented by sequenced peptides = 0% coverage

No peptides were identified that matched this protein sequence.

NRPE8a/NRPD8a/NRPB8a (At1g54250)

MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAVGDKF
 TLALAPTLNLDGTPDTGYFTPGAKK **TLADKYEYIMHGK**LYKISERDGTKPKAELYVSFG
 GLLMLLKGDPAHISHFELDQRLFLLMRKL

13/146 amino acids are represented by sequenced peptides = 9% coverage
 0/146=0% of the coverage corresponds to peptides unique to this member of the protein family.
 This peptide also is an exact match to At3g59600.

NRPE8b/NRPB8b/NRPD8b (At3g59600)

MASNIIMFEDIFVVDKLDPDGKKFDKVTRVEARSHNLEMFMHLDVNTEVYPLAVGDKF
 TLAMAPTLNLDGTPDTGYFTPGAKK **TLADKYEYIMHGK**LYKISERDGTKPKAELYVSFG
 GLLMLLQGDPAHISHFELDQRLFLLMRKL

13/146 amino acids are represented by sequenced peptides = 9% coverage
 0/146 = 0% of the coverage corresponds to peptides unique to this member of the protein family.
 This peptide is also an exact match to At1g54250.

NRPE9a/NRPD9a/NRPB9a (At3g16980)

MSTMKFCRECNILYPKEDKEQKILLYACRNCDHQEVADNSCVYRNEVHHSVSERTQIL
TDVASDPTLPRTKAVRCSKCQHRE**EA**VFFQATAR**G**EEGMTLFFVCCNPNCGHRWRE

10/114 amino acids are represented by sequenced peptides = 9% coverage
 0/114 = 0% coverage corresponds to peptides unique to this member of the protein family. Two
 amino acid differences in the identified peptide (underlined) discriminates At3g16980 from
 At4g16265.

NRPE9b/NRPD9b/NRPB9b (At4g16265)

MSTMKFCRECNILYPKEDKEQSILLYACRNCDHQEAADNNCVYRNEVHHSVSEQTQI
LSDVASDPTLPRTKAVRCAKQCQHG**EA**VFFQATAR**G**EEGMTLFFVCCNPNCshrWRE

10/114 amino acids are represented by sequenced peptides = 9% coverage
 0/114 = 0% coverage corresponds to peptides unique to this member of the protein family. Two
 amino acid differences in the identified peptide (underlined) discriminates At3g16980 from
 At4g16265.

NRPE10/NRPB10/NRPD10 (At1g11475)

MIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRR**MLMTHVDLIEK**
LLNYNTLEKSDNS

20/71 amino acids are represented by sequenced peptides = 28% coverage
 20/71= 28% coverage corresponds to peptides that only match this protein and not At1g61700.

NRPB10 family member (At1g61700)

MIPVRCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRR**MLMTHVDLIE**
KLLNYNTMEKSDPN

11/71 amino acids are represented by sequenced peptides = 15% coverage
 0/71= 0% unique. The peptide identified for At1g61700 also matches At1g11475.

NRPE11/NRPB11/NRPD11 (At3g52090)

MNAPERYERFVVPEGTKKVS YDRDTKIINAASFTVEREDHTIGNIVR MQLHRDENVLFAGYQLPHPLKYKIIVRIHTTSQSSPMQAYNQAINDLKELDYLKNQFEAEVAKFSNQF

42/116 amino acids are represented by sequenced peptides = 36% coverage
 All peptides identified match NRPE11 and only NRPE11.

NRPE12/NRPB12/NRPD12 (At5g41010)

MDPAPEPVITYVCGDCGQENTLKS GDVIQCRECGYRILYKKRTRRVVQYEAR

8/51 amino acids are represented by the sequenced peptide = 16% coverage
 The peptide is a unique match to this protein.

NRPB12 family member (At1g53690)

MDLQQSETDDKQPEQLVIYVCGDCGQENILKRGDVFQCRDCGFRILYKKRILDKKETRI
 GV

0/62 amino acids are represented by sequenced peptides = 0% coverage
 No peptides were identified that matched this protein sequence.

APPENDIX D

NONCODING TRANSCRIPTION BY RNA POLYMERASE POL IVb/POL V
MEDIATES TRANSCRIPTIONAL SILENCING OF OVERLAPPING AND
ADJACENT GENES

Published in *Cell* (2008), 135 (4): 635-648.

My contributions to this work:

I cloned, generated and validated the Pol II, Pol V, and Pol V active site mutant transgenic lines used in the analysis. I also provided technical assistance for the Western blot data and comments during the editing phase of the article.

Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes

Andrzej T. Wierzbicki,¹ Jeremy R. Haag,¹ and Craig S. Pikaard^{1,*}

¹Biology Department, Washington University, 1 Brookings Drive, St. Louis, MO 63130, USA

*Correspondence: pikaard@biology.wustl.edu

DOI 10.1016/j.cell.2008.09.035

SUMMARY

Nuclear transcription is not restricted to genes but occurs throughout the intergenic and noncoding space of eukaryotic genomes. The functional significance of this widespread noncoding transcription is mostly unknown. We show that *Arabidopsis* RNA polymerase IVb/Pol V, a multisubunit nuclear enzyme required for siRNA-mediated gene silencing of transposons and other repeats, transcribes intergenic and noncoding sequences, thereby facilitating heterochromatin formation and silencing of overlapping and adjacent genes. Pol IVb/Pol V transcription requires the chromatin-remodeling protein DRD1 but is independent of siRNA biogenesis. However, Pol IVb/Pol V transcription and siRNA production are both required to silence transposons, suggesting that Pol IVb/Pol V generates RNAs or chromatin structures that serve as scaffolds for siRNA-mediated heterochromatin-forming complexes. Pol IVb/Pol V function provides a solution to a paradox of epigenetic control: the need for transcription in order to transcriptionally silence the same region.

INTRODUCTION

Nuclear transcription in eukaryotes is not restricted to messenger RNAs (mRNAs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), or genes required for their processing. In humans, such conventional genes account for less than 2% of the genome, yet ~90% of the genome is transcribed (Kapranov et al., 2007; Prasanth and Spector, 2007; Willingham et al., 2006). Much of the noncoding RNA (ncRNA) pool corresponds to intergenic sequences or antisense transcripts of unknown function. However, the potential for noncoding RNAs (ncRNAs) to epigenetically regulate adjacent genes is increasingly clear (Prasanth and Spector, 2007). Long ncRNAs that regulate adjacent genes include the *Xist* and *Tsix* RNAs involved in X chromosome inactivation in mammals (Masui and Heard, 2006; Yang and Kuroda, 2007), the H19 and Air ncRNAs involved in imprinting at mouse and human *Igf2* and *Igf2r* loci, respectively (Pauler et al., 2007), and the *roX* ncRNAs involved in X chromosome dosage compensation in flies (Bai

et al., 2007). The persistence of *Xist* and *roX* transcripts at affected loci indicates a role in the assembly of repressive or activating chromatin states, respectively (Bai et al., 2007; Herzig et al., 1997). Likewise, at the *Drosophila Ultrabithorax (Ubx)* locus, intergenic ncRNAs serve as scaffolds for the recruitment of Ash1, a histone methyltransferase that modifies the adjacent chromatin to switch on *Ubx* transcription (Sanchez-Elsner et al., 2006).

In diverse eukaryotes, establishment of DNA methylation and/or repressive heterochromatic histone modifications are ncRNA-directed processes (Buhler et al., 2007; Grewal and Elgin, 2007; Zaratiegui et al., 2007). In plants and fission yeast, small interfering RNAs (siRNAs) of 20–25 nt that are generated from long double-stranded RNA (dsRNA) precursors by dicer endonuclease(s) bind to argonaute (AGO) proteins and guide chromatin modifications to homologous DNA sequences (Baulcombe, 2006; Brodersen and Voinnet, 2006; Peters and Meister, 2007). Noncoding transcripts in fission yeast serve at least two functions: acting as precursors of siRNAs and as scaffolds to which siRNAs bind in order to recruit the chromatin-modifying machinery (Buhler et al., 2006, 2007; Irvine et al., 2006). AGO-mediated slicing of scaffold transcripts coupled with RNA-dependent RNA polymerase-mediated dsRNA production generates additional siRNAs, thereby perpetuating heterochromatin formation (Irvine et al., 2006; Locke and Martienssen, 2006). RNA-mediated heterochromatin formation requires that an affected region be transcribed (Buhler et al., 2006; Djupedal et al., 2005; Irvine et al., 2006; Kato et al., 2005), presenting an intriguing paradox as to how transcription and transcriptional silencing can occur at the same locus (Grewal and Elgin, 2007).

The paradox of transcription-dependent gene silencing in plants might be explained by the existence of two structurally and functionally distinct plant-specific RNA polymerases: RNA polymerases IVa/Pol IV and Pol IVb/Pol V (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). Pol IVa/Pol IV and Pol IVb/Pol V are not essential for viability in *Arabidopsis* but participate in multiple small RNA-mediated gene silencing pathways (Pikaard et al., 2008). Pol IVa/Pol IV and Pol IVb/Pol V have distinct largest subunits that have been named either NRPD1a and NRPD1b (Herr et al., 2005; Onodera et al., 2005) or RPD1 and RPE1 (Luo and Hall, 2007). The latter terminology has been adopted, in modified form, to allow the naming of Pol IVa/Pol IV subunits using the Nuclear RNA polymerase D (NRPD) gene symbol and Pol IVb/Pol V subunits using the Nuclear RNA polymerase E (NRPE) prefix. The transition to the

Pol IV and Pol V nomenclature in place of Pol IVa and Pol IVb has been made necessary by the need for a systematic nomenclature defining their numerous subunits (T. Ream and C.S.P., unpublished data) and reflects the fact that the two activities are functionally nonredundant as well as structurally distinct. Therefore, we refer to Pol IVa and Pol IVb as Pol IV and Pol V for the remainder of this paper. The revised nomenclature denotes the largest subunits of Pol IV and Pol V as NRPD1 and NRPE1. Pol IV and Pol V both utilize a second-largest subunit that is encoded by a single gene bearing the synonymous names *NRPD2* or *NRPE2*. In the siRNA-directed DNA methylation pathway, Pol IV is required for siRNA production, whereas Pol V acts primarily downstream of siRNA production (Kanno et al., 2005; Mosher et al., 2008; Pontes et al., 2006; Pontier et al., 2005; Zhang et al., 2007). Pol IV or Pol V transcripts have not been identified in vivo or in vitro, but the catalytic subunits of Pol IV and Pol V have amino acids that are invariant at the active sites of multisubunit RNA polymerases and are essential for Pol IV and Pol V biological functions (J.R.H. and C.S.P., unpublished data).

By pursuing the hypothesis that Pol IV and/or Pol V might synthesize ncRNAs required for transcriptional gene silencing, we identified intergenic regions where Pol V-dependent transcripts are detectable by RT-PCR. Pol V (Pol IVb) physically associates with loci that give rise to these transcripts and also physically associates with the RNA transcripts themselves. Moreover, production of the Pol V-dependent transcripts is lost upon mutation of the conserved active site of NRPE1/NRPD1b, suggesting that the RNAs are Pol V transcripts. The putative chromatin remodeler DRD1 is required for Pol V to physically associate with intergenic loci and generate transcripts that suppress adjacent transposons via the establishment of repressive heterochromatin. Importantly, Pol V transcription alone is not sufficient for transposon silencing; instead, the combination of Pol V transcription and siRNA production is required. Collectively, our data indicate that Pol V (Pol IVb) transcription occurs independently of siRNA biogenesis and support a model whereby Pol V transcripts serve as scaffolds for the binding of siRNAs that guide heterochromatin formation. Pol V's role in gene silencing provides a solution in plants to the paradox of how transcription can be required for transcriptional gene silencing.

RESULTS

Identification of Pol V-Dependent Transcripts in Intergenic Noncoding Regions

A heterochromatic knob, or chromomere, on the northern arm of *A. thaliana* chromosome 4 is a well-characterized interval rich in transposons and other heterochromatic repeats (Fransz et al., 2000; Lippman et al., 2004). Within this domain are intergenic noncoding (IGN) regions at which RNA transcripts have not been detected using tiling DNA microarrays (Lippman et al., 2004). Nonetheless, siRNAs and DNA hypermethylation often map to these regions (Kasschau et al., 2007; Lippman et al., 2005; Lister et al., 2008), suggesting that low-abundance transcripts might serve as siRNA precursors. Therefore, we used RT-PCR to search for IGN RNAs present in wild-type plants but missing in Pol IV or Pol V mutants. Of 14 IGN regions examined, six had RNAs that were lost or reduced in Pol V mutants (Figures 1

and S1 available online). For instance, at intergenic noncoding regions 5 and 6 (*IGN5* and *IGN6*) (Figures 1A and 1B), transcripts detected in wild-type (ecotype Col-0) and *nRPD1* mutants are depleted in *nRPE1* (*nRPD1b-11*) or *nRPD2* mutants (Figure 1E, top three rows), indicating that Pol V, but not Pol IV, is required for their production. However, *AtSN1* family retrotransposons are derepressed (activated) in both the Pol IV and Pol V mutants (Figure 1E, fourth row from the top). Actin 2 mRNA abundance is unaffected by the mutations (Figure 1E).

IGN5 and *IGN6* are located in regions rich in transposon-derived elements, siRNA production, and DNA hypermethylation (Lister et al., 2008), all characteristic of heterochromatic domains. Pol V-dependent transcripts are also detected at *IGN7* and *IGN17* (Figure 1F), which are located in pericentromeric heterochromatic regions (Figure S1). However, *IGN10* and *IGN15* are present in gene-rich environments with relatively few transposon-related repeats (Figures 1C and 1D) yet also give rise to Pol V-dependent transcripts (Figure 1F). Collectively, these data suggest that Pol V contributes to IGN transcription in both heterochromatic and euchromatic environments.

Characterization of Pol V-Dependent Transcripts

To determine whether Pol V-dependent RNAs initiate at specific sites, we performed 5' RACE at *IGN5* and *IGN6* (Figures 2A–2C). Resulting PCR-amplified RACE products yielded distinct bands upon agarose gel electrophoresis (Figure 2C), but excising the bands and cloning and sequencing of the cDNAs revealed heterogeneity at the 5' ends. At *IGN5*, top-strand clones initiated at two sites seven nucleotides apart (Figures 2A and S2). An *IGN5* bottom-strand-specific primer yielded five different 5' ends spanning a 33 nt interval (Figures 2A and S2). At *IGN6*, clones derived from the gel-purified upper and lower bands collectively revealed four distinct 5' ends spanning a 94 nt interval (Figures 2B and S2). Bottom-strand-specific transcripts were not detected at *IGN6*.

It is noteworthy that the 5' terminal nucleotides of all RACE products were adenosine or guanosine (Figure S2), given that transcripts of eukaryotic Pol I, II, III, and bacterial RNA polymerase typically begin with purines (Smale and Kadonaga, 2003; Sollner-Webb and Reeder, 1979; Zecherle et al., 1996). To test whether RACE 5' ends represent transcription start sites or cleavage sites, we exploited the fact that initiating nucleotides have 5' triphosphate groups (Pol I, Pol III) or 7-methylguanosine caps (Pol II). By contrast, cleaved RNAs have 5' monophosphate or hydroxyl groups. Terminator exonuclease (Epicentre Biotechnologies) is a 5' → 3' exonuclease that degrades RNAs having 5' monophosphates, but not RNAs that have 5' triphosphate groups, 5' hydroxyl groups, or 7-methylguanosine caps. Total RNA treated with Terminator endonuclease was subjected to RT-PCR using *IGN5*-specific primers (Figure 2D; interval A is depicted in Figure 2A). In agreement with Figure 1, *IGN5* transcripts were detected in wild-type (Col-0) plants but were absent in the Pol V mutant (*nRPE1/nRPD1b-11*). Terminator exonuclease treatment prior to RT-PCR caused an ~70% reduction in the Pol V-dependent *IGN5* transcript signal, suggesting that the majority of the transcripts amplified by PCR are 5' monophosphorylated; however, the remaining transcripts are resistant to the exonuclease (Figure 2D). Treatment of the RNA with Tobacco Acid

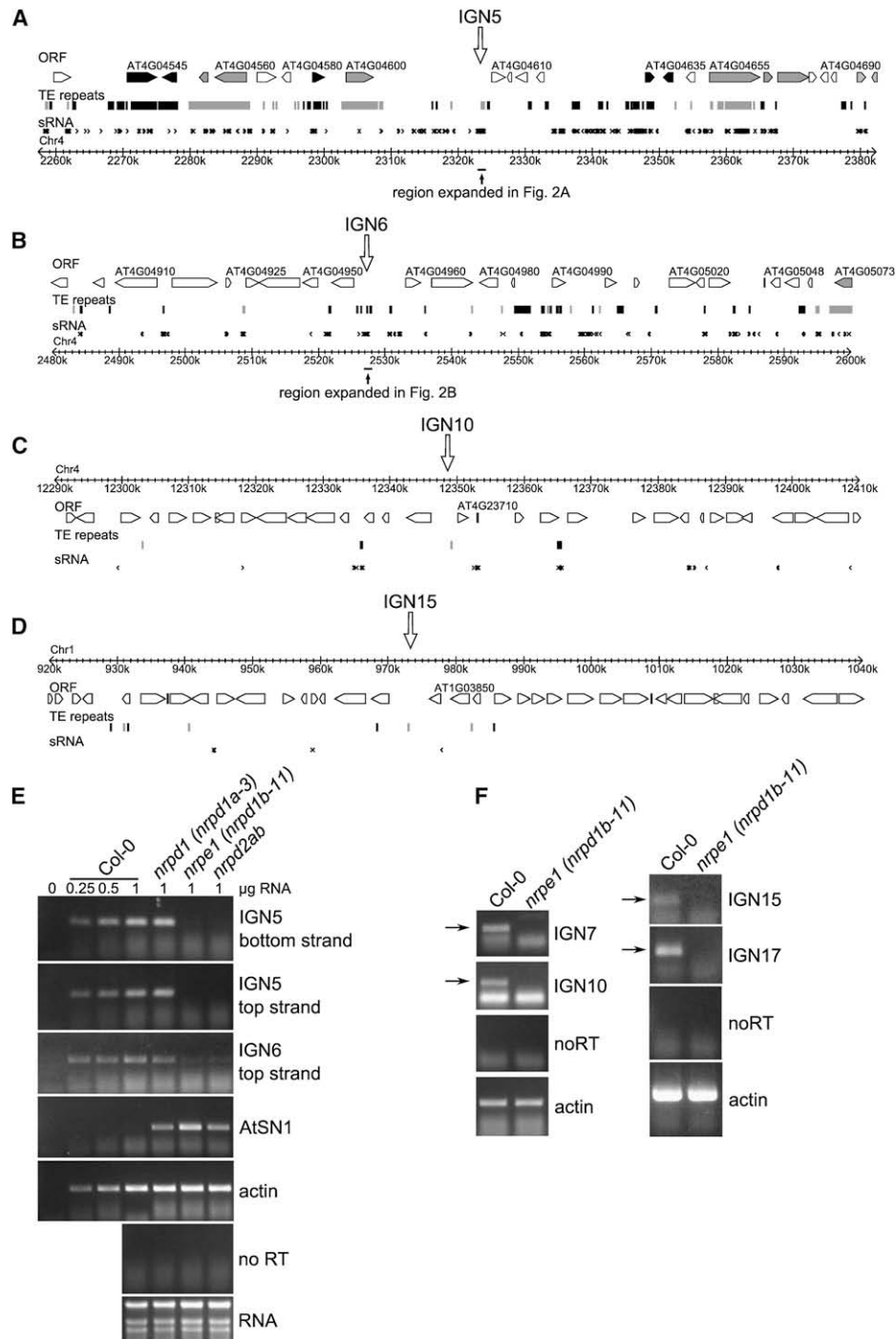


Figure 1. Detection of Intergenic Pol V-Dependent Transcripts

(A–D) Chromosomal contexts of intergenic regions *IGN5*, *IGN6*, *IGN10*, and *IGN15*. Open reading frames (ORF), transposable element (TE)-derived repeats, and small RNAs (sRNA) in the MPSS database (<http://mpss.udel.edu/at/>) are shown. Single-copy genes are marked in white; retrotransposons, in gray; and DNA transposons, in black. Diagrams derive from <http://chromatin.cshl.edu/cgi-bin/gbrowse/arabidopsis5/>.

(E) Strand-specific RT-PCR analysis of *IGN5*, *IGN6*, and *AtSN1* transcripts in wild-type (ecotype Col-0), *nprp1a-3*, *nprp1* (*nprp1b-11*), and *nprp2a-2 nprp2b-1* mutants. Actin RT-PCR products and ethidium bromide-stained rRNAs resolved by agarose gel electrophoresis serve as loading controls. Dilutions of Col-0 RNA show that PCR results are semiquantitative. To control for background DNA contamination, a reaction using *IGN5* top-strand primers, but no reverse transcriptase (no RT), was performed. No RNA (0 µg) controls are provided for all primer pairs.

(F) RT-PCR analysis of Pol V-dependent transcripts at intergenic regions *IGN7*, *IGN10*, *IGN15*, and *IGN17* in wild-type (Col-0) and *nprp1* mutants.

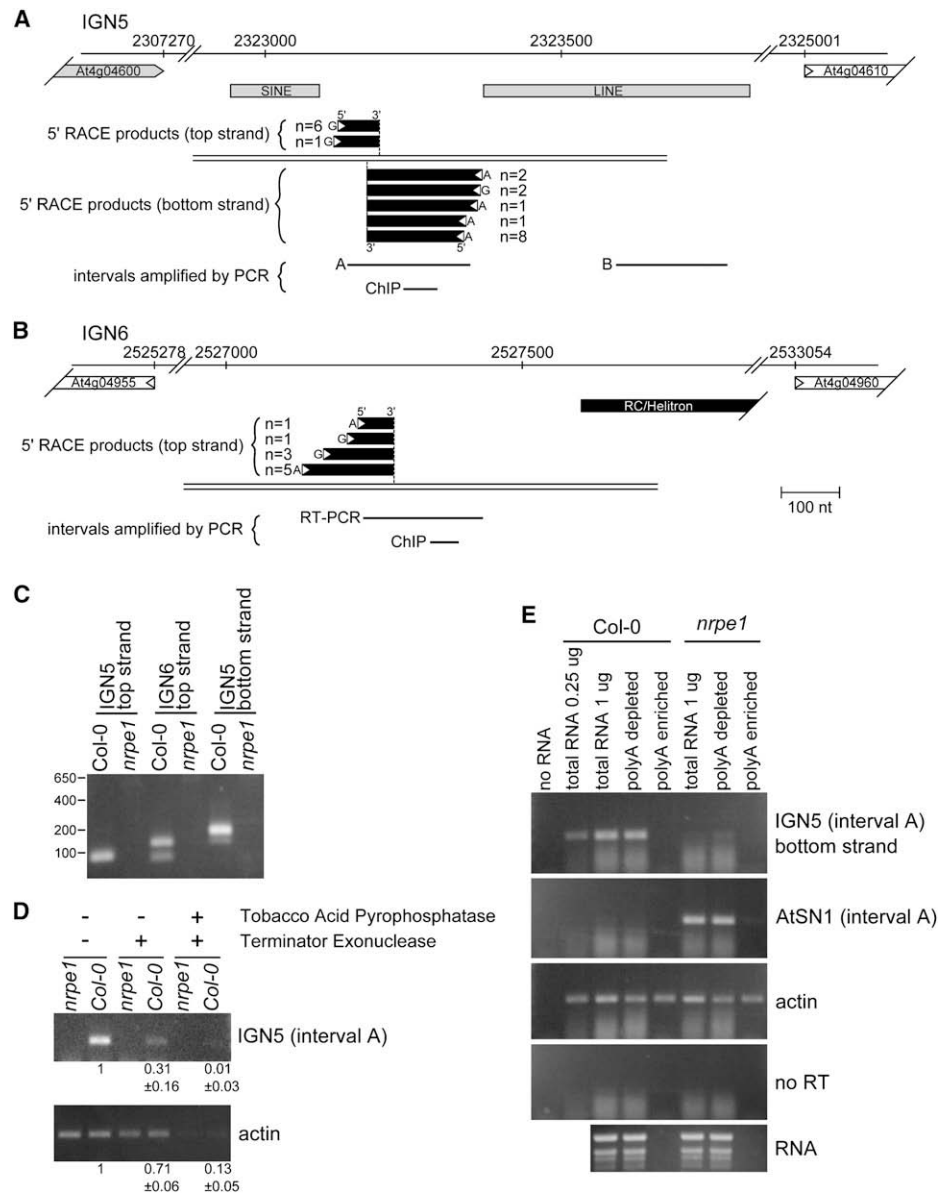


Figure 2. Characterization of Pol V-Dependent Transcripts

(A and B) Local contexts of *IGN5* (A) and *IGN6* (B), showing neighboring genes or transposons, 5' RACE products, and intervals amplified by PCR. Color coding of annotated genes and TE elements is the same as in Figure 1. For RACE products, the 5' terminal nucleotide and number of clones (n) sharing that 5' end are shown.

(C) Ethidium bromide-stained agarose gel of 5' RACE products.

(D) 5' end analysis for Pol V-dependent *IGN5* transcripts. RT-PCR was performed on total RNA or RNA treated with Terminator exonuclease, Tobacco Acid Pyrophosphatase, or both enzymes. Numbers below the panels are relative densitometric band intensities relative to the untreated control. The mean and standard deviation resulting from three independent experiments is shown.

(E) Pol V-dependent transcripts are not polyadenylated. Poly A-enriched and poly A-depleted RNA fractions were subjected to RT-PCR using *IGN5*, *AtSN1*, and actin primer pairs followed by agarose gel electrophoresis and ethidium bromide staining. Controls include no RT (*IGN5* bottom-strand primers) and no RNA (all primer pairs) reactions.

Pyrophosphatase, which removes 7-methylguanosine caps or triphosphates and leaves a 5' monophosphate, rendered the *IGN5* transcripts and actin control fully susceptible to Terminator exonuclease digestion. Therefore, *IGN5* transcripts that require Tobacco Acid Pyrophosphatase in order to be made Terminator

susceptible are deduced to be triphosphorylated or capped (Figure 2D), indicative of transcription start sites. It is noteworthy that 5' RACE requires a 5' monophosphate for adaptor ligation. RACE products were only obtained upon treating RNA with Tobacco Acid Pyrophosphatase, but not upon treating RNA with

T4 polynucleotide kinase and ATP (data not shown), which would have converted 5' hydroxyls to phosphates and allowed their cloning. Collectively, our observations suggest that the 5' ends detected by RACE are transcription start sites. However, much of the RNA detected by RT-PCR consists of processed RNAs.

To test whether Pol V-dependent transcripts are polyadenylated, total RNA was fractionated using oligo d(T) magnetic beads. *IGN5* transcripts were detected in total RNA and poly A-depleted fractions of wild-type Col-0 but were not detected in poly A-enriched RNA (Figure 2E), unlike *Actin 2* mRNA. *AtSN1* transcripts produced in *nrpe1* (*nrpd1b-11*) mutants were present in total and poly A-depleted, but not poly A-enriched, RNA, consistent with Pol III transcription of *AtSN1* (see below).

Collectively, the assays of Figure 2 suggest that Pol V-dependent transcripts can be at least ~200 nt in size, can initiate from multiple sites, have triphosphates or 7meG caps at their 5' ends, and lack poly A tails.

Evidence that Pol V Synthesizes IGN Transcripts

The largest subunits of Pol IV and Pol V include sequences that are invariant among DNA-dependent RNA polymerases, including a DFDGD at the active site (metal A site) that coordinates a magnesium ion essential for nucleoside polymerization (Cramer, 2004). We tested the importance of the presumptive NRPE1 metal A site by analyzing *nrpe1* (*nrpd1b-11*) mutants transformed with a wild-type *NRPE1* transgene or a transgene in which the invariant aspartates were changed to alanines (active site mutant [ASM]) (Figure 3A). Both transgenes utilized the native *NRPE1* promoter, included their full complement of introns and exons, and were similarly expressed, as shown by immunoblot detection of the FLAG epitope tags added to their C termini (Figure 3B, bottom row). Moreover, the wild-type and ASM mutant proteins both coimmunoprecipitate NRPD2/NRPE2, the second-largest subunit of both Pol IV and Pol V, suggesting that the ASM mutation does not disrupt Pol V subunit assembly (J.R.H. and C.S.P., unpublished data). The wild-type *NRPE1* transgene restored Pol V-dependent *IGN5* and *IGN6* transcripts in the *nrpe1* (*nrpd1b-11*) mutant background, but the *NRPE1*-ASM transgene did not (Figure 3B), indicating that synthesis of Pol V-dependent transcripts requires the conserved active site.

To determine whether NRPE1 physically interacts with loci giving rise to Pol V-dependent transcripts, we performed chromatin immunoprecipitation (ChIP) of FLAG-tagged NRPE1 as well as FLAG-tagged NRPB2, the second-largest subunit of RNA polymerase II (Figure 3C). Subsequent quantitative real-time PCR showed that NRPE1 physically associates with *IGN5*, whereas NRPB2 does not. A retrotransposon-derived solo long terminal repeat (LTR) shown to be silenced in a Pol V-dependent manner (Huettel et al., 2006) is also occupied by NRPE1. The *solo LTR* most likely programs Pol II transcription, and Pol II is detected at this locus above background (defined as ChIP signals obtained with Col-0 plants that lack a FLAG-tagged transgene) but at lower levels than at the *actin 2* gene locus *At3g18780*. Collectively, the ChIP data indicate that Pol V is present at loci that give rise to Pol V-dependent RNAs.

We next asked whether Pol V-dependent RNAs could be immunoprecipitated (IPed) in association with NRPE1. Formaldehyde-crosslinked chromatin preparations of nontransgenic

Col-0 or *nrpe1* (*nrpd1b-11*) lines expressing FLAG-tagged NRPE1 were IPed using anti-FLAG antibody. Following DNase I treatment, samples were tested by RT-PCR (Figure 3D). *IGN5*, *IGN6*, *AtSN1*, and *solo LTR* RNAs were all enriched in IP fractions of NRPE1-FLAG plants compared to nontransgenic Col-0 controls that were also subjected to anti-FLAG IP (Figure 3D). Background levels of abundant actin mRNA were equivalent in Col-0 and NRPE1-FLAG IP fractions, indicating that the enrichment of the IGN and transposon RNAs in NRPE1-FLAG IP fractions compared to Col-0 reflects specific interaction of these RNAs with Pol V. Because Pol V-dependent transcripts require the presumptive NRPE1 active site, NRPE1 physically associates with loci giving rise to these transcripts, and NRPE1 physically associates with the transcripts themselves, we deduce that Pol V synthesizes the transcripts.

Pol V Transcription Is Necessary in Order to Silence Overlapping and Adjacent Genes

Transcriptional silencing of *AtSN1* retroelements requires both Pol IV and Pol V (see Figure 1E). *AtSN1* family elements are short interspersed nuclear elements (SINEs) that possess A box and B box elements (see diagram in Figure 4A) typical of the internal promoters of Pol III-transcribed genes (Myouga et al., 2001). In wild-type (Col-0) plants, *AtSN1* elements are silenced, but, in *nrpe1* (*nrpd1b-11*) mutants, they are derepressed (Figures 1E and 4C, interval A). *AtSN1* silencing is restored in *nrpe1* mutants by the full-length *NRPE1* transgene, but not by the active site mutant *NRPE1*-ASM transgene (Figure 4C, top row), indicating that Pol V transcription is required for *AtSN1* silencing. In the intergenic region and overlapping the expected Pol III transcription start site (see Figures 4A and S3), IGN transcripts corresponding to both DNA strands can be detected by RT-PCR. These transcripts, within intervals B and C, are readily detected in wild-type plants but are absent, or much reduced, in *nrpe1* mutants (Figure 4C, rows 2–5). The interval B and C transcripts are restored in *nrpe1* mutants by the wild-type *NRPE1* transgene, but not by the *NRPE1*-ASM transgene. Collectively, the data indicate that *AtSN1* transcripts are only generated if Pol V transcripts are absent.

Like *AtSN1*, a long interspersed nuclear element (LINE), *At5g27845*, which overlaps the *solo LTR* (see Figure 4B), is silenced in a Pol V (Pol IVb)-dependent manner (Huettel et al., 2006). Transcription of this LINE is low in wild-type plants but increases substantially in the *nrpe1* (*nrpd1b-11*) mutant (Figure 4C, RT-PCR interval A). Silencing is restored by the wild-type *NRPE1* transgene, but not by the *NRPE1*-ASM transgene (Figure 4D).

In wild-type plants, transcripts are detected from both strands upstream of the LINE and *solo LTR* (interval B), including intergenic sequences and overlapping an adjacent transcription unit, *At5g27850* (see Figure 4B). These RNAs in wild-type plants might be Pol V transcripts. However, unlike the intergenic region adjacent to *AtSN1*, where transcripts disappear in *nrpe1* (*nrpd1b-11*) mutants, suggesting that Pol V is the sole polymerase transcribing the region, transcript abundance in the region adjacent to the *solo LTR* increases dramatically in *nrpe1* or *NRPE1*-ASM transgenic plants (Figure 4D). This increased transcription is attributable to RNA polymerase II, as shown by ChIP (Figure 4F). Whereas Pol II occupancy of the locus is low in wild-type plants,

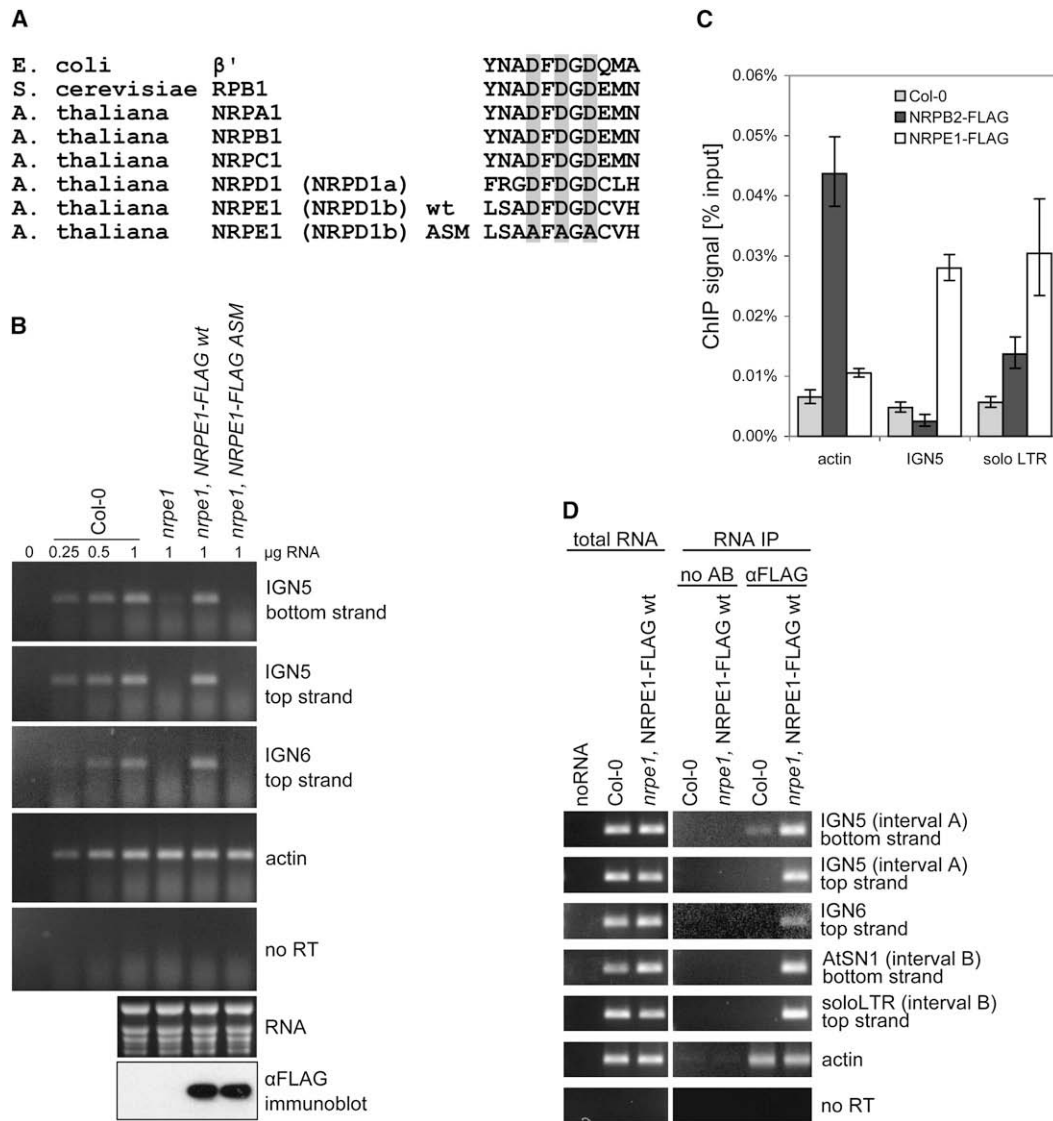


Figure 3. Evidence that Pol V Synthesizes IGN Transcripts

(A) Multiple alignments of DNA-dependent RNA polymerase largest subunits surrounding the metal A active site. Invariant aspartates are marked in gray. (β') Largest subunit of *E. coli* polymerase; (RPB1) Largest subunit of yeast Pol II; (NRPA1) Largest subunit of *Arabidopsis* Pol I; (NRPB1) Largest subunit of *Arabidopsis* Pol II; (NRPC1) Largest subunit of *Arabidopsis* Pol III; (NRPD1) Largest subunit of *Arabidopsis* Pol IV (also known as NRPD1a); (NRPE1 WT) Largest subunit of *Arabidopsis* Pol V (also known as NRPD1b); (NRPE1-ASM) Active site mutant of NRPE1.

(B) Strand-specific RT-PCR analysis of *IGN5* and *IGN6* transcripts in Col-0 wild-type, *nrpe1* (*nrpd1b-11*), and *nrpe1* mutants transformed with a wild-type (WT) FLAG-tagged *NRPE1* transgene or the *NRPE1*-ASM transgene. Actin RT-PCR reactions and ethidium bromide-stained rRNAs serve as loading controls. Dilutions of Col-0 wild-type RNA demonstrate that PCR results are semiquantitative. No RT (*IGN5* top-strand primers) and no RNA (all primer pairs) controls are included. Equal expression of transgenic wild-type and active site mutant NRPE1 was verified by immunoprecipitation followed by α FLAG immunoblot detection (bottom row).

(C) ChIP of FLAG-tagged Pol II or Pol V at the *actin 2* gene, *IGN5*, or a solo retroelement LTR silenced by Pol V. Wild-type Col-0 plants or plants expressing FLAG-tagged NRPE1 or FLAG-tagged NRPE1 were subjected to ChIP using anti-FLAG antibody followed by real-time PCR. Histograms show mean values \pm SD obtained for three independent PCR amplifications.

(D) RNA immunoprecipitation. Wild-type (nontransgenic) Col-0 and *nrpe1* (*nrpd1b-11*) mutants expressing the NRPE1-FLAG transgene were subjected to RNA-IP using anti-FLAG antibody. Following DNase treatment, *IGN5*, *IGN6*, *AtSN1*, *solo LTR*, or *actin 2* RNAs were detected by RT-PCR. *AtSN1* and *solo LTR* PCR-amplified intervals are shown in Figure 4; *IGN5* and *IGN6* PCR-amplified intervals are shown in Figure 2. Total RNA controls, assayed prior to immunoprecipitation, show that the RNAs are present in equivalent amounts in wild-type Col-0 and NRPE1-FLAG transgenic plants. No RT controls used *IGN5* top-strand primers. No signals were obtained following RNA IP in the absence of anti-FLAG antibody (no AB columns). Background signal for actin RNA shows that equal RNA amounts were tested.

it increases dramatically in the *nrpe1* mutant. Transformation of *nrpe1* with the wild-type *NRPE1* transgene reduces Pol II occupancy of the locus, whereas the *NRPE1-ASM* mutant is ineffective (Figure 4F). Taken together, the data indicate that derepression of Pol II transcription in the *solo LTR* region occurs in the absence of Pol V transcription.

A LINE element located to the right of *IGN5* is expressed at low levels in wild-type plants but is derepressed in the *nrpe1* mutant (Figure 4E). Silencing is restored by the wild-type *NRPE1* transgene, but not by the *NRPE1-ASM* mutant transgene. Collectively, the data of Figure 4 indicate that intergenic Pol V transcription plays a direct role in suppressing transcription from overlapping or adjacent LINE and SINE transposons.

Pol V Transcription Is Necessary for Heterochromatin Formation at Affected Loci

We next examined histone modifications and cytosine methylation at Pol V affected loci (Figure 5). ChIP using an antibody specific for histone H3 lysine 27 monomethylation (H3K27me1), a heterochromatic mark previously shown to be dependent on Pol V (Pol IVb) (Huettel et al., 2006), resulted in significant enrichment of *IGN5*, the *solo LTR* region, and *AtSN1* relative to the actin gene control (Figure 5A). Decreased H3K27me1 at the *IGN5*, *solo LTR*, and *AtSN1* loci in *nrpe1 (nrpd1b-11)* was restored by the *NRPE1* transgene, but not the *NRPE1-ASM* transgene (Figure 5A). ChIP controls in which antibody was omitted yielded negligible background signals (Figure S4). ChIP using an antibody specific for dimethylated histone H3 lysine 9 (H3K9me2), also a heterochromatic mark, showed association of *IGN5* and the *solo LTR* region that was reduced in *nrpe1* and rescued by the wild-type *NRPE1* transgene, but not the *NRPE1-ASM* transgene (Figure 5B). Interestingly, Pol V mutations did not significantly affect H3K9me2 at *AtSN1* despite their pronounced effect on H3K27me1 at the locus.

Diacetylation of histone H3 on lysines 9 and 14 (abbreviated H3Ac2) is a characteristic of active, euchromatic genes, such as actin (Figure 5C). At the *solo LTR*, H3Ac2 levels increased significantly in the *nrpe1 (nrpd1b-11)* mutant and were restored by the wild-type *NRPE1* transgene, but not the *NRPE1-ASM* transgene (Figure 5C). These results parallel increased Pol II occupancy of the locus in the absence of functional NRPE1 (see Figure 4F). H3Ac2 levels at *IGN5* and *AtSN1* were not influenced by NRPE1. Differences in histone hyperacetylation at the loci may reflect the different RNA polymerases transcribing them; *IGN5* is transcribed by Pol V, and *AtSN1* is presumably transcribed by Pol III, whereas Pol II transcribes the *solo LTR*.

We assayed *IGN5*, *IGN6*, and *solo LTR* DNA methylation status based on *McrBC* endonuclease sensitivity (Figure 5D). *McrBC* specifically cleaves methylated DNA, preventing its subsequent amplification by PCR. In wild-type Col-0, methylcytosine levels are high at *IGN5*, *IGN6*, and the *solo LTR*, such that *McrBC* digestion reduces their PCR amplification by ~80% (Figure 5D). At *IGN5* and the *solo LTR*, DNA methylation is significantly reduced in the *nrpe1 (nrpd1b-11)* mutant and in a null mutant for RNA-dependent RNA polymerase 2 (RDR2), a protein required for 24 nt siRNA biogenesis (Xie et al., 2004). In the *nrpe1* mutant background, *IGN5* and *solo LTR* methylation are restored by the wild-type *NRPE1* transgene, but not by the *NRPE1-ASM*

transgene. The data indicate that Pol V transcription, like RDR2, is needed for siRNA-directed DNA methylation at these loci.

Unlike *IGN5* and the *solo LTR*, DNA methylation at *IGN6* does not require Pol V or RDR2 but does require DDM1 (decrease in DNA methylation 1), a SWI/SNF family chromatin remodeler that acts primarily in the maintenance, rather than RNA-mediated establishment of cytosine methylation (Jeddeloh et al., 1999). DDM1 also affects maintenance methylation at *IGN5* but has no appreciable effect at the *solo LTR*, which may rely exclusively on RNA-directed DNA methylation.

Loss of DNA methylation at the *AtSN1*, *IGN5*, and *solo LTR* loci in the *nrpe1 (nrpd1b-11)* mutant was also demonstrated using methylation-sensitive restriction endonucleases (Figures 5E and 5F). Methylation of *HaeIII* or *AluI* recognition sites blocks the enzymes from cutting the DNA, allowing PCR amplification of the region. However, unmethylated sites are cleaved such that PCR amplification fails. DNA methylation was lost at *HaeIII* or *AluI* sites of the *AtSN1*, *IGN5*, and *solo LTR* loci in the *nrpe1* mutant and was restored by the wild-type *NRPE1* transgene, but not by the *NRPE1-ASM* transgene (Figures 5E and 5F). At *IGN6*, no effect of *nrpe1* was observed on methylation of the sole *AluI* site tested (Figure 5F). Collectively, the data indicate that Pol V mediates the establishment of heterochromatic histone modifications and DNA methylation changes that correlate with the silencing of Pol II- or Pol III-transcribed genes that overlap the Pol V-transcribed regions.

Pol V-Dependent Transcription Does Not Require Small RNA Biogenesis

Because Pol V is required for siRNA-dependent DNA methylation, we asked whether mutations in genes required for siRNA biogenesis, RNA-directed gene silencing, or DNA methylation affect Pol V transcription (Figure 6A). At *IGN5* and *IGN6*, Pol V transcripts lost in *nrpe1 (nrpd1b-11)* and *nrpd2* mutants were unaffected by mutation of the four dicers that process double-stranded RNA precursors into siRNAs, including a quadruple mutant that combines a hypomorphic *dcl1* allele with null alleles of *dcl2*, *dcl3*, and *dcl4*. Pol V-dependent transcripts were also unaffected in mutants defective for RNA-dependent RNA polymerases (*rdr2*, *rdr1*, and *rdr6*) implicated in generating siRNA precursors or in mutants affecting cytosine methylation (*drm2*, *met1*, and *ddm1*). However, many of these mutants interfere with *AtSN1* silencing, including the dicer quadruple mutant, *rdr2*, *nrpd1a*, *drm2*, and *drd1* (Figure 6A, row 4). Collectively, the results reveal that Pol V transcription occurs independently of small RNA biogenesis, de novo cytosine methylation (*drm2*), or maintenance cytosine methylation (*met1*, *ddm1*). However, Pol V and siRNA biogenesis are both required for *AtSN1* silencing.

DRD1 Facilitates the Association of Pol V with Chromatin

As shown in Figure 6A, Pol V transcripts are lost in *drd1-6* mutants. DRD1 is a member of the SWI2/SNF2 family of ATP-dependent chromatin remodelers and was identified in a genetic screen that also identified *nrpe1 (nrpd1b)* and *nrpd2* alleles, suggesting that DRD1 and Pol V act in collaboration (Huettel et al., 2007). ChIP of FLAG-tagged NRPE1 in wild-type or *drd1* mutant backgrounds was conducted to determine whether

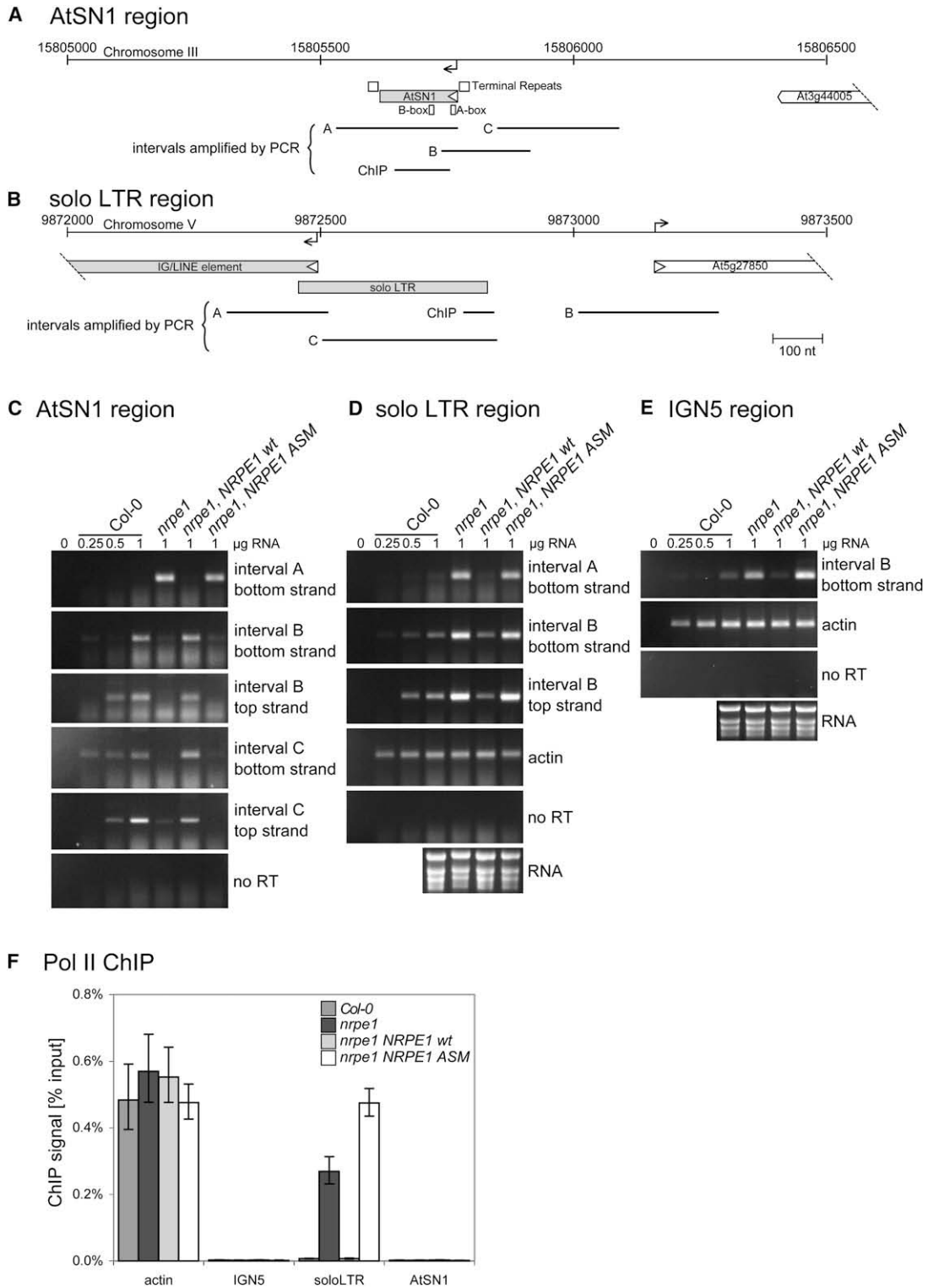


Figure 4. RNA Polymerase Activity of Pol V Is Necessary for Silencing Adjacent Transposons and Repetitive Elements

(A and B) *AtSN1* (A) and *solo LTR* (B) regions, including neighboring genes, repetitive elements, and regions amplified by PCR. The diagram for the *solo LTR* region is based on analysis of transcription units by Huettel et al. (2006).

(C) Strand-specific RT-PCR analysis of transcription from the *AtSN1* region in Col-0 wild-type, *nrpe1* (*nrpd1b-11*), and the *nrpe1* mutant expressing a wild-type *NRPE1* transgene or the *NRPE1-ASM* transgene. Intervals amplified by RT-PCR are depicted in (A). No RT (interval A bottom-strand primers) and no RNA controls (all primer pairs) are included.

DRD1 regulates Pol V association with chromatin (Figure 6B). NRPE1-FLAG protein levels were similar in both genetic backgrounds (Figure 6C). In *nrpe1* plants that are wild-type at the *DRD1* locus, the NRPE1-FLAG protein physically associates with *IGN5*, *IGN6*, and the *solo LTR* locus (Figure 6B). However, in the *drd1* mutant background, NRPE1 association with these loci is reduced to background levels resembling the actin gene control (Figure 6B). We conclude that DRD1 mediates Pol V recruitment to chromatin.

DISCUSSION

Polymerase Activity of Pol V

RNA polymerase activity has not yet been demonstrated for Pol IV or Pol V in vitro. However, our study provides in vivo evidence for Pol V polymerase activity by demonstrating the existence of Pol V-dependent transcripts, by showing that these RNAs require the conserved polymerase active site, by showing that Pol V physically associates with DNA loci corresponding to Pol V-dependent transcripts, and by showing that Pol V physically associates with the transcripts themselves. The most parsimonious explanation for the results is that Pol V transcribes DNA into RNA, which fits with the crosslinking of Pol V to both DNA and RNA and with the requirement for the putative chromatin remodeler DRD1 in order for Pol V to associate with transcribed loci. DRD1 and Pol V do not appear to physically interact, based on colP experiments (T. Ream, A.T.W., and C.S.P., unpublished data), suggesting that DRD1 functions upstream of Pol V, presumably by remodeling chromatin to facilitate Pol V recruitment to the DNA. If Pol V were to utilize RNA templates, a prediction is that Pol V-dependent transcript abundance would increase in accord with the abundance of RNAs serving as templates. However, mutations that derepress transposons, including *rdr2*, *drm2*, *met1*, or *ddm1*, have no effect on Pol V transcript abundance. Likewise, Pol V transcripts do not decrease in mutants for the major RNA-directed RNA polymerases, *rdr2* or *rdr6*, which could potentially generate RNA templates for Pol V.

Detection of multiple Pol V transcript 5' ends using RACE suggests that Pol V may initiate transcription in a promoter-independent fashion. How sites of Pol V initiation are chosen is unclear. One hypothesis is that specific DNA methylation patterns or histone modifications recruit Pol V. However, Pol V transcripts are detectable in both heterochromatic, transposon-rich regions as well as gene-rich, presumably euchromatic environments. Moreover, mutants affecting siRNA production or DNA methylation have no effect on Pol V transcript abundance. An alternative possibility, which we favor, is that Pol V initiates transcripts throughout the genome, both in silenced and non-silenced regions, and these transcripts are necessary, but not sufficient, for gene silencing. Instead, we envision that Pol V transcription renders a locus competent for silencing, but silencing

only occurs if siRNAs complementary to the locus are also produced (see below).

The Role of Pol V Transcription in Transcriptional Gene Silencing

ncRNAs originating in intergenic regions are prevalent in eukaryotes, including *Arabidopsis*, but their functional significance is mostly unknown. Our results indicate that Pol V-transcribed ncRNAs play direct roles in silencing overlapping or adjacent genes. At the *AtSN1* locus, Pol V transcripts and retrotransposon transcripts presumably generated by Pol III are mutually exclusive, suggesting that Pol V transcription prevents Pol III transcription. Likewise, at the *solo LTR* locus, Pol II association is low in wild-type plants but increases 35-fold in *nrpe1* mutants. Similar increases in transcription of the LINE element adjacent to *IGN5* occur in *nrpe1* mutants. Collectively, the data indicate that Pol V transcription facilitates the silencing of overlapping genes as a result of repressive chromatin modifications, including H3K9 methylation, H3K27 methylation, and cytosine hypermethylation.

Pol V transcription is necessary, but not sufficient, to silence *AtSN1* and *solo LTR* elements. Other necessary proteins include Pol IV, RDR2, one or more DCL proteins, AGO4, DRD1, and DRM2 (see Figure 6), which are components of the 24 nt siRNA-directed DNA methylation pathway. Because mutants that disrupt siRNA biogenesis (e.g., *nrpd1*, *rdr2*, *dicer*) have no effect on the production of Pol V-dependent transcripts, our results suggest that Pol V transcription and siRNA production occur independently but collaborate in gene silencing. This hypothesis fits with the observation that Pol V is not required for siRNA production at the majority of the ~4000 loci giving rise to 24 nt siRNAs (Mosher et al., 2008), including the *AtSN1* (Kanno et al., 2005; Pontes et al., 2006) and *solo LTR* (Huettel et al., 2006) loci we have examined. At other endogenous repeat loci giving rise to siRNAs, all of which require Pol IV, Pol V is apparently required (Mosher et al., 2008). However, this does not necessarily imply that Pol V transcripts serve as siRNA precursors. Instead, Pol V-dependent heterochromatin formation may stimulate Pol IV-dependent production of siRNAs in a positive feedback loop that enforces gene silencing (Li et al., 2006; Pontes et al., 2006).

In our alternative models (Figure 7), we envision that chromatin remodeling by DRD1 is required for Pol V transcription initiation. In parallel, siRNAs produced by the combined actions of Pol IV, RDR2, and DCL3 are incorporated into AGO4. Our favored model is that Pol V transcripts base pair with siRNAs that are associated with AGO4 (Figure 7A), similar to the way that Pol II transcripts reading through silenced fission yeast pericentromeric regions are proposed to interact with the siRNA-AGO moiety of the RNA-induced transcriptional silencing (RITS) complex (Buhler et al., 2006; Irvine et al., 2006). The interaction of the siRNA with the nascent transcript might then direct the silencing machinery, including the de novo cytosine methyltransferase DRM2 and/or

(D) Strand-specific RT-PCR analysis of transcription at the *solo LTR* region. No RT (interval B bottom-strand primers) controls are included.

(E) Strand-specific RT-PCR analysis of transcription from a LINE element flanking *IGN5*. Figure 2A shows the location of interval B amplified by PCR. No RT (interval B bottom-strand primers) controls are included.

(F) Pol II occupancy of *actin 2*, *IGN5*, *solo LTR*, and *AtSN1* loci detected using ChIP. Col-0 wild-type, *nrpe1* (*nrpd1b-11*), and *nrpe1* mutant plants transformed with the wild-type *NRPE1* transgene or the *NRPE1-ASM* transgene were subjected to ChIP using α NRPB2 antibody and detected by real-time PCR. Histograms show the means \pm SD obtained from three independent amplifications.

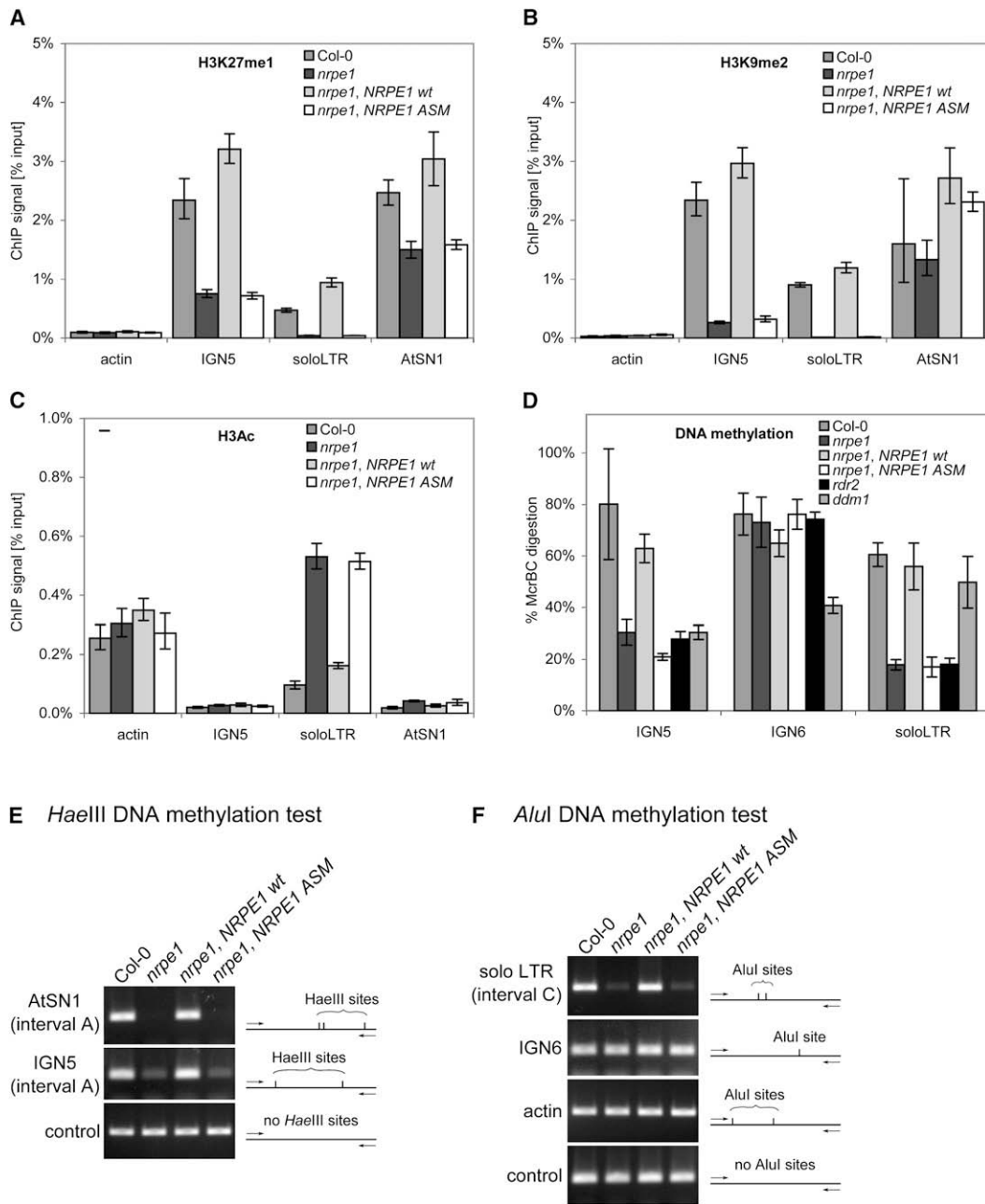


Figure 5. Pol V-Dependent Transcription Is Necessary for Heterochromatin Formation

(A–C) ChIP using α H3K27me1 (A), α H3K9me2 (B), or α H3Ac (C) antibodies and chromatin of Col-0 wild-type, *nrpe1* (*nrpd1b-11*), or *nrpe1* mutants transformed with the wild-type *NRPE1* transgene or *NRPE1*-ASM transgene. Histograms show the means \pm SD from three independent amplifications.

(D) DNA methylation analysis at the indicated loci performed by digestion of genomic DNA with *McrBC* followed by quantitative real-time PCR. Comparison to undigested DNA allowed the fraction susceptible to *McrBC* to be calculated.

(E and F) DNA methylation analysis at the *AtSN1*, *IGN5*, *IGN6*, and *solo LTR* loci performed by digesting purified DNA with the methylation-sensitive restriction endonucleases *HaeIII* (E) or *AluI* (F) followed by PCR. Sequences lacking *HaeIII* (actin; [E]) or *AluI* (*IGN5* interval A; [F]) sites served as controls to show that equivalent amounts of DNA were tested in all reactions.

histone-modifying activities, to the adjacent DNA. Alternatively, Pol V transcripts may directly bind to AGO4 and stabilize siRNA-DNA interactions (Figure 7B). It is also possible that Pol V transcripts or the act of transcription itself influence structural features of heterochromatin that are required by AGO4 for efficient

interactions with target loci (Figure 7C). In each of these scenarios, AGO4 recruitment is expected to be cotranscriptional and may involve direct interactions between AGO4 and the C-terminal domain of NRPE1/NRPD1b (El-Shami et al., 2007; Li et al., 2006). A prediction of all of the models is that transcriptional

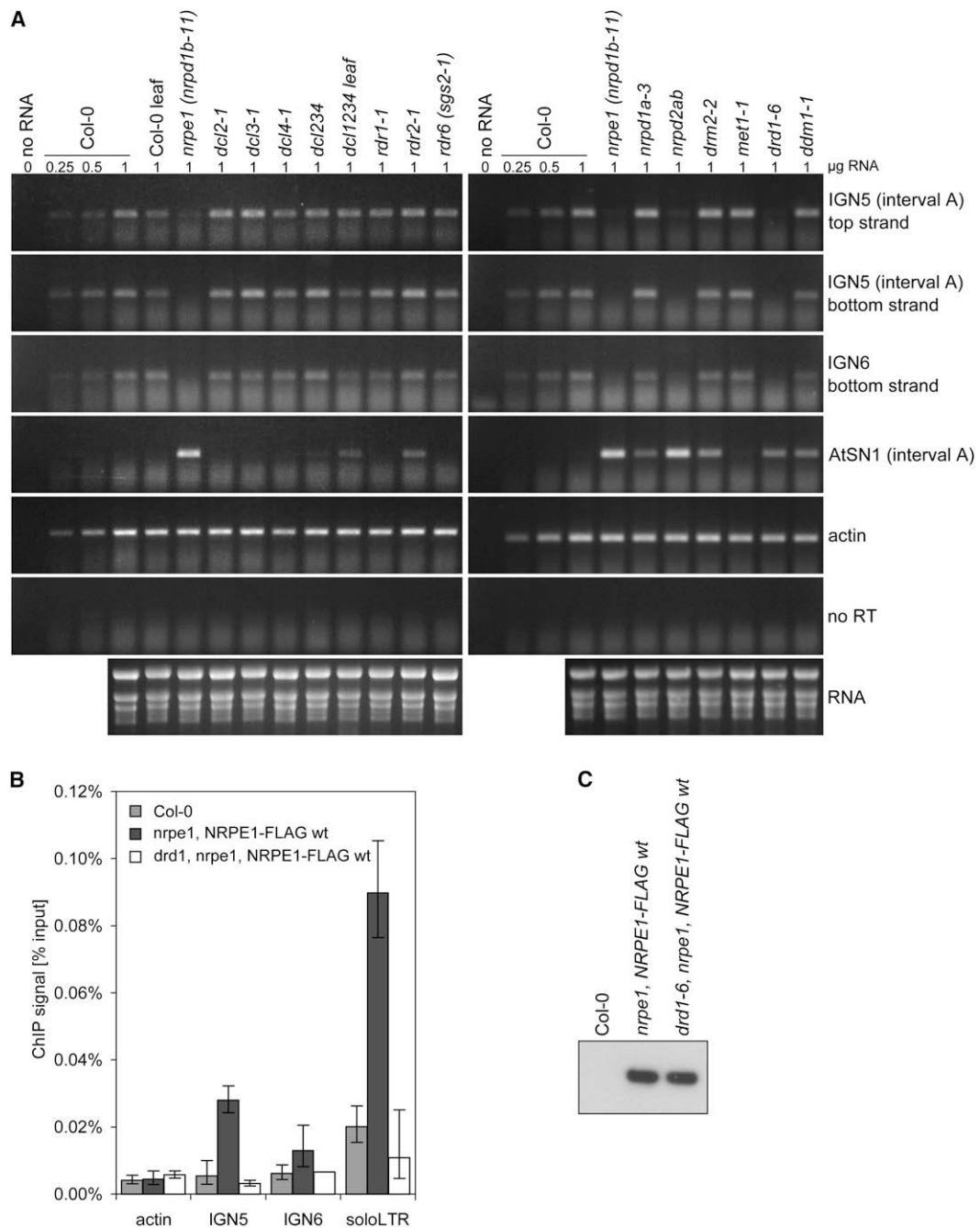


Figure 6. Pol V-Dependent Transcription Requires the Chromatin Remodeler DRD1, but Not siRNA Production or DNA Methylation

(A) Strand-specific RT-PCR analysis of *IGN5* and *IGN6* transcription in mutants disrupting dicer (*dcl1*, *dcl2*, *dcl3*, *dcl4*), RNA-dependent RNA polymerase (*rdr1*, *rdr2*, *rdr6*), Pol IV (*nrpd1*, *nrpd2*), Pol V (*nrpe1/nrpd1b-11*, *nrpd2*) DNA methylation (*met1*, *ddm1*, *drm2*) or chromatin remodeling (*ddm1*, *drd1*) activities. Detection of *AtSN1* retroelement transcripts indicates a loss of *AtSN1* silencing. Col-0 RNA dilutions show that results are semiquantitative. No RT controls used *IGN5* top-strand primers.

(B) DRD1 is required for Pol V to interact with chromatin. ChIP with α FLAG antibody was performed using chromatin isolated from Col-0 wild-type, *nrpe1* (*nrpd1b-11*) plants expressing the *NRPE1-FLAG* transgene or *drd1 nrpe1* double mutants expressing the *NRPE1-FLAG* transgene. *Actin 2*, *IGN5*, *IGN6*, and *solo LTR* loci were detected using quantitative real-time PCR. Histograms show the means \pm SD obtained from three independent amplification reactions.

(C) Immunoblot with α FLAG antibody showing that equivalent amounts of NRPE1-FLAG recombinant protein are immunoprecipitated in the *nrpe1* (*nrpd1b-11*) and *drd1 nrpe1* genetic backgrounds.

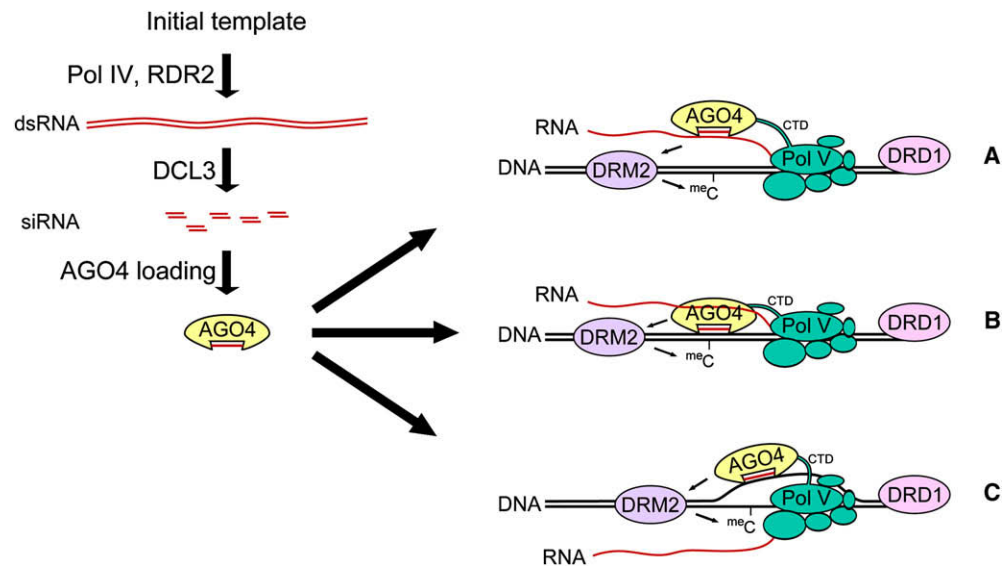


Figure 7. Possible Modes of Action for Pol V in RNA-Directed Transcriptional Silencing

Pol V transcription and siRNA production occur independently but collaborate in silencing transposons such as *AtSN1*. 24 nt siRNAs are produced by Pol IV, RDR2, and DCL3 and loaded into AGO4. Chromatin remodeling by DRD1 is required for Pol V to associate with chromatin, and physical interactions may occur between the Pol V C-terminal domain (CTD) and AGO4. In (A), which we favor, siRNAs bound to AGO4 interact with nascent Pol V transcripts, thereby recruiting chromatin-modifying activities, including histone-modifying enzymes and the de novo cytosine methyltransferase DRM2, to the adjacent DNA. In (B), AGO4 interacts with the nascent transcripts, but the siRNA base pairs with DNA. In (C), the siRNA associated with AGO4 interacts with DNA in a manner dependent upon Pol V-mediated chromatin perturbation.

silencing does not occur everywhere that Pol V transcription occurs but only at sites where Pol V transcription and siRNA production overlap. Testing this hypothesis on a whole-genome basis is a goal for future studies.

EXPERIMENTAL PROCEDURES

Plant Strains

A. thaliana nprd1a-3 (nprd1), *nprd1b-11 (nrpe1)*, and *nprd2a-2 nprd2b-1* mutants were described previously (Onodera et al., 2005; Pontes et al., 2006), as were *nrdp1b-11 NRPD1b-FLAG (NRPE1-FLAG)* (Pontes et al., 2006) and *NRPB2-FLAG* (Onodera et al., 2008) transgenic lines. *NRPE1* mutagenesis and production of transgenic lines expressing Pol IV and Pol V active site mutants will be described elsewhere (J.R.H. and C.S.P., unpublished data). *rdr1-1*, *rdr2-1*, *dcl2-1*, and *dcl3-1* were provided by J. Carrington; *sgs2-1 (rdr6)* and *dcl4-1* were provided by H. Vaucheret; *drd1-6* was provided by M. Matzke; *met1-1* and *ddm2-1* were provided by E. Richards; *dcl234 (dcl2-5 dcl3-1 dcl4-2)* and *dcl1234 (dcl1-9 dcl2-5 dcl3-1 dcl4-2)* were provided by T. Blevins; *drm2-2 (SAIL_70_E12)* was provided by E. Richards.

RNA Analysis

RNA was isolated from 2-week-old plants using an RNeasy Kit (QIAGEN). The 5' RACE was performed using a GeneRacer Kit (Invitrogen) with two nested amplification steps; see Table S1 for primers. The 5' RACE products were gel purified and cloned into TOPO-TA (Invitrogen). Tobacco Acid Pyrophosphatase (Invitrogen) or Terminator exonuclease (Epicentre) treatments followed manufacturers' instructions. Polyadenylated RNA was purified using a Fast-Track MAG Kit (Invitrogen). For RT-PCR, 1 μ g of RNA digested with DNase I (Invitrogen) was reverse transcribed 30 min at 55°C using 60 units SuperScript III Reverse Transcriptase (Invitrogen), 1.5 units Platinum Taq (Invitrogen), and a gene-specific primer. After heat inactivation of reverse transcriptase, the second primer was added and PCR was performed. Alternatively, the One-Step RT-PCR Kit (QIAGEN) was used. Table S1 shows primer pairs.

ChIP and RNA-IP

ChIP was performed by adapting existing protocols (Lawrence et al., 2004; Nelson et al., 2006), as was RNA-IP (Gilbert and Svejstrup, 2006; Martianov et al., 2007). Details are provided in the Supplemental Data. All ChIP and RNA IP experiments were reproduced at least twice.

Real-Time Quantitative PCR

DNA was amplified using an Applied Biosystems model 7500 thermocycler with 0.5 units of Platinum Taq (Invitrogen), SYBR Green I (Invitrogen), and Internal Reference Dye (Sigma). Primer pairs are shown in Table S1. Results were analyzed using the comparative C_T method (Livak and Schmittgen, 2001) relative to input or undigested samples.

Antibodies

Anti-FLAG M2 mouse monoclonal and rabbit polyclonal antibodies were purchased from Sigma-Aldrich. Anti-Pol II (anti-NRPB2) was described previously (Onodera et al., 2005). Anti-H3K27me1 antibody no. 8835 (Peters et al., 2003) was provided by Thomas Jenuwein. Antibody against diacetyl-H3 (K9 and K14) was obtained from Upstate Biologicals (cat. no. 06599, lot no. 31994). Anti-H3K9me2 was obtained from Abcam (cat. no. ab7312, lot no. 133588).

SUPPLEMENTAL DATA

The Supplemental Data include Supplemental Experimental Procedures, four figures, and one table and can be found with this article online at [http://www.cell.com/supplemental/S0092-8674\(08\)01192-6](http://www.cell.com/supplemental/S0092-8674(08)01192-6).

ACKNOWLEDGMENTS

A.T.W. and C.S.P. conceived the work, A.T.W. performed all described experiments and contributed all figures, J.R.H. generated the NRPE1 active site mutant, and A.T.W. and C.S.P. wrote the paper. We thank Keith Earley and Eric Richards for helpful discussions. This work was supported by NIH grant GM077590.

Received: May 19, 2008
 Revised: July 29, 2008
 Accepted: September 15, 2008
 Published: November 13, 2008

REFERENCES

- Bai, X., Larschan, E., Kwon, S.Y., Badenhorst, P., and Kuroda, M.I. (2007). Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in *Drosophila melanogaster*. *Genetics* **176**, 1491–1499.
- Baulcombe, D.C. (2006). Short silencing RNA: the dark matter of genetics? *Cold Spring Harb. Symp. Quant. Biol.* **71**, 13–20.
- Brodersen, P., and Voinnet, O. (2006). The diversity of RNA silencing pathways in plants. *Trends Genet.* **22**, 268–280.
- Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873–886.
- Buhler, M., Haas, W., Gygi, S.P., and Moazed, D. (2007). RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* **129**, 707–721.
- Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. *Curr. Opin. Genet. Dev.* **14**, 218–226.
- Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev.* **19**, 2301–2306.
- El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T. (2007). Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* **21**, 2539–2544.
- Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Druenen, C., Dean, C., Zabel, P., Bisseling, T., and Jones, G.H. (2000). Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* **100**, 367–376.
- Gilbert, C., and Svejstrup, J.Q. (2006). RNA immunoprecipitation for determining RNA-protein associations in vivo. In *Current Protocols in Molecular Biology*, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. (New York: John Wiley & Sons).
- Grewal, S.I., and Elgin, S.C. (2007). Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399–406.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118–120.
- Herzing, L.B., Romer, J.T., Horn, J.M., and Ashworth, A. (1997). Xist has properties of the X-chromosome inactivation centre. *Nature* **386**, 272–275.
- Huetzel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A.J., and Matzke, M. (2006). Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J.* **25**, 2828–2836.
- Huetzel, B., Kanno, T., Daxinger, L., Bucher, E., van der Winden, J., Matzke, A.J., and Matzke, M. (2007). RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. *Biochim. Biophys. Acta* **1769**, 358–374.
- Irvine, D.V., Zaratiegui, M., Tolia, N.H., Goto, D.B., Chitwood, D.H., Vaughn, M.W., Joshua-Tor, L., and Martienssen, R.A. (2006). Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313**, 1134–1137.
- Jeddeloh, J.A., Stokes, T.L., and Richards, E.J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat. Genet.* **22**, 94–97.
- Kanno, T., Huetzel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **37**, 761–765.
- Kapranov, P., Willingham, A.T., and Gingeras, T.R. (2007). Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* **8**, 413–423.
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol.* **5**, e57.
- Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309**, 467–469.
- Lawrence, R.J., Earley, K., Pontes, O., Silva, M., Chen, Z.J., Neves, N., Viegas, W., and Pikaard, C.S. (2004). A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Mol. Cell* **13**, 599–609.
- Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126**, 93–106.
- Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476.
- Lippman, Z., Gendrel, A.V., Colot, V., and Martienssen, R. (2005). Profiling DNA methylation patterns using genomic tiling microarrays. *Nat. Methods* **2**, 219–224.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408.
- Locke, S.M., and Martienssen, R.A. (2006). Slicing and spreading of heterochromatic silencing by RNA interference. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 497–503.
- Luo, J., and Hall, B.D. (2007). A multistep process gave rise to RNA polymerase IV of land plants. *J. Mol. Evol.* **64**, 101–112.
- Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670.
- Masui, O., and Heard, E. (2006). RNA and protein actors in X-chromosome inactivation. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 419–428.
- Mosher, R.A., Schwach, F., Studholme, D., and Baulcombe, D.C. (2008). PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl. Acad. Sci. USA* **105**, 3145–3150.
- Myoung, F., Tsuchimoto, S., Noma, K., Ohtsubo, H., and Ohtsubo, E. (2001). Identification and structural analysis of SINE elements in the *Arabidopsis thaliana* genome. *Genes Genet. Syst.* **76**, 169–179.
- Nelson, J.D., Denisenko, O., and Bomsztyk, K. (2006). Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.* **1**, 179–185.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613–622.
- Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., Ream, T., Ito, Y., and Pikaard, C.S. (2008). Sex-biased lethality or transmission of defective transcription machinery in *Arabidopsis*. *Genetics* **180**, 207–218.
- Pauler, F.M., Koerner, M.V., and Barlow, D.P. (2007). Silencing by imprinted noncoding RNAs: is transcription the answer? *Trends Genet.* **23**, 284–292.
- Peters, L., and Meister, G. (2007). Argonaute proteins: mediators of RNA silencing. *Mol. Cell* **26**, 611–623.
- Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577–1589.
- Pikaard, C.S., Haag, J.R., Ream, T., and Wierzbicki, A.T. (2008). Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci.* **13**, 390–397.

- Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006). The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**, 79–92.
- Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev.* **19**, 2030–2040.
- Prasanth, K.V., and Spector, D.L. (2007). Eukaryotic regulatory RNAs: an answer to the ‘genome complexity’ conundrum. *Genes Dev.* **21**, 11–42.
- Sanchez-Elsner, T., Gou, D., Kremmer, E., and Sauer, F. (2006). Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. *Science* **311**, 1118–1123.
- Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**, 449–479.
- Sollner-Webb, B., and Reeder, R.H. (1979). The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in *X. laevis*. *Cell* **18**, 485–499.
- Willingham, A.T., Dike, S., Cheng, J., Manak, J.R., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Duttagupta, R., Ganesh, M., et al. (2006). Transcriptional landscape of the human and fly genomes: nonlinear and multifunctional modular model of transcriptomes. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 101–110.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104.
- Yang, P.K., and Kuroda, M.I. (2007). Noncoding RNAs and intranuclear positioning in monoallelic gene expression. *Cell* **128**, 777–786.
- Zaratiegui, M., Irvine, D.V., and Martienssen, R.A. (2007). Noncoding RNAs and gene silencing. *Cell* **128**, 763–776.
- Zecherle, G.N., Whelen, S., and Hall, B.D. (1996). Purines are required at the 5' ends of newly initiated RNAs for optimal RNA polymerase III gene expression. *Mol. Cell. Biol.* **16**, 5801–5810.
- Zhang, X., Henderson, I.R., Lu, C., Green, P.J., and Jacobsen, S.E. (2007). Role of RNA polymerase IV in plant small RNA metabolism. *Proc. Natl. Acad. Sci. USA* **104**, 4536–4541.

Cell, *Volume 135*

Supplemental Data

**Noncoding Transcription by RNA Polymerase
Pol IVb/Pol V Mediates Transcriptional Silencing
of Overlapping and Adjacent Genes**

Andrzej T. Wierzbicki, Jeremy R. Haag, and Craig S. Pikaard

Supplemental Material

Chromatin Immunoprecipitation Details

Three grams of above-ground tissue of 2-week old plants was crosslinked with 0.5% formaldehyde for 10 min by vacuum infiltration, followed by addition of glycine to 80 mM. Plants were rinsed with water, frozen in liquid nitrogen, ground into powder using a mortar and pestle, suspended in 25 ml of Honda Buffer (20 mM HEPES-KOH pH 7.4, 0.44 M sucrose, 1.25% ficoll, 2.5% Dextran T40, 10 mM MgCl₂, 0.5% Triton X-100, 5 mM DTT, 1 mM PMSF, 1% plant protease inhibitors (Sigma)), filtered through two layers of Miracloth and centrifuged at 2000 x g for 15 min. Nuclear pellets were washed three times with 1ml of Honda buffer, resuspended in Nuclei Lysis Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, 1 mM PMSF, 1% Plant Protease Inhibitors) and sonicated as described (Lawrence et al. 2004). After centrifugation at 16,000 x g for 10 min., the supernatant was diluted 10-fold with 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl. 25 µl of protein A agarose/salmon sperm DNA (Upstate Biologicals) and the appropriate antibody was added. Samples were then incubated overnight at 4°C on a rotating mixer. Agarose-antibody complexes were washed five times, 5 min each, with binding/washing buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 1 mM PMSF) and washed twice for 5 min each with 10 mM Tris-HCl pH 8.0, 1 mM EDTA. 100µl of 10% (w/v) Chelex (Bio Rad) resin, in water, was then added to the beads and crosslinking was reversed at 99 °C for 10 min. Samples were digested with 20 µg of proteinase K (Invitrogen) for 1h at 43 °C followed by heat-inactivation at 95 °C for 10 min.

RNA Immunoprecipitation Details

RNA IP was based on ChIP with the following modifications. RNase OUT RNase inhibitor (Invitrogen) was included in all buffers. IP was performed for 3h followed by four washes with Binding/Washing buffer. Immune complexes were eluted with 100 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS for 10 min at room temperature followed by a second elution at 65 °C. Crosslinking was reversed at 65 °C for 1h in the presence of 20 µg Proteinase K (Invitrogen). RNA was purified by extraction with acidic phenol:chloroform and ethanol precipitation.

Supplemental Table and Figures

Table S1. Oligonucleotides Used in This Study

Target	Name	Sequence (5' – 3')	Application
Actin 2 At3g18780	ACTmaiFW	TCATACTAGTCTCGAGAGATGACTCAGATCATGTTTGAG	RT-PCR
	ACTmaiRV	TCATTCTAGAGGCGCGCCACAATTTCCCGTTCTGCGGTAG (Herr et al, 2005)	
	A118	GAGAGATTCAGATGCCCAGAAGTC	real time PCR
	A119	TGGATTCCAGCAGCTTCCA	
	A65	CGAGCAGGAGATGGAAACCTCAA	Chop-PCR
A66	AAGAATGGAACCCAGATCCAGACA		
AtSN1	A122	CCAGAAATTCATCTTCTTTGGAAAAG	real time PCR
	A123	GCCCAGTGGTAAATCTCTCAGATAGA	
AtSN1 (A)	ATS15	ACCAACGTGCTGTTGGCCAGTGGTAAATC	RT-PCR Chop-PCR
	AtSN1-F4	AAAATAAGTGGTGGTTGTACAAGC (Herr et al, 2005)	
AtSN1 (B)	A205	TGAGAGATTTACCACTGGGCCAACAA	RT-PCR
	A206	TGAGGAGCTCAACACATAAATGGCAATA	
AtSN1 (C)	A207	CCTTTCCAAGACACCATCTCAACAAC	RT-PCR
	A208	TCCTCAACAAAAATAATTCCGAACGAC	
IGN5 (A)	A28	TCCCGAGAAGAGTAGAACAAATGCTAAAA	RT-PCR Chop-PCR
	A29	CTGAGGTATTCCATAGCCCCTGATCC	
IGN5 (B)	A293	CGCAGCGGAATTGACATCCTATC	RT-PCR
	A294	TCGGAAAGAGACTCTCCGCTAGAAA	
IGN5	A193	AAGCCCAAACCATACTAATAATCTAAT	real time PCR
	A194	CCGAATAACAGCAAGTCCTTTTAATA	
IGN5 bottom strand	A69	TCATGCGGCCCAATAACCAACAAAAC	5' RACE
	A70	TGAAGAAAGCCCAAACCATACT	
IGN5 top strand	A60	TGTTGGTTATTGGGCCGCATGATACA	5' RACE
	A67	AGCATTGTCTACTCTCTCGGGAAC	
IGN6	A30	GGGACATCTATTGGGTTTAGGCTGGATG	RT-PCR Chop-PCR
	A31	TTTGTAATTCTCAGTTCGGGTATCTGCTTG	
	A162	TTTCGCCGTCACATAACATGTAATG	real time PCR
	A163	GAAGTAGCTTTTCGGTCCAGTTC	
IGN6 top strand	A62	TCGGTTGCTATGTTTGC GGATCATGC	5' RACE
	A71	CCAGCCTAAACCAATAGATGTCC	
IGN7	A44	CATCCACAACCTCTATTGCTTTGTTTACC	RT-PCR
	A45	TTTTCTTTGAGTTGGTCATTGTTGTTT	
IGN10	A50	TCTAACGCTTTGGTTGTGTATAGTGTGC	RT-PCR
	A51	ACCGGTATCTTAGTTCCTCCACGTGTC	
IGN15	A110	CCATAGCATAGAACTTGCGGATATATGAA	RT-PCR
	A111	CGGAAAAGGTAAGGTGGTTGGAAAA	
IGN17	A114	AACCCTAGCCTTTCATTAAAACCCTCTC	RT-PCR
	A115	CATAGATAGGAACTCAATCTCTTCGCATTT	
solo LTR A	A221	ATCAATTATTATGTCATGTTAAAACCGATTG	RT-PCR
	A222	TGTTTCGAGTTTTATTCTCTCTAGTCTTCATT	
solo LTR B	A217	CATATAACCGAAGCCGAAGGATGTGAAA	RT-PCR
	A218	CAGAAACCTAAGGAACATTACACGCTAAACC	
solo LTR C	A211	ATAAACTCGAAACAAGAGTTTTCTTATTGCTTTC	Chop-PCR
	A212	TAATGGTATTATTTTGATCAGTGTATAAACCGGA	
solo LTR	A142	GGATAGAGATGAATGATGGATAATGACA	real time PCR
	A143	TTATTTTGATCAGTGTATAAACCGGATA	

Figure S1. The chromosomal contexts of *IGN7* and *IGN15* loci at which Pol V-dependent transcripts have been identified (Fig 1F). Shown are open reading frames (ORF), repetitive elements (TE repeats) and small RNAs from the MPSS database (sRNA). Single copy genes are marked in white, retrotransposons in grey and transposons in black. Data were obtained from <http://chromatin.cshl.edu/cgi-bin/gbrowse/arabidopsis5/>.

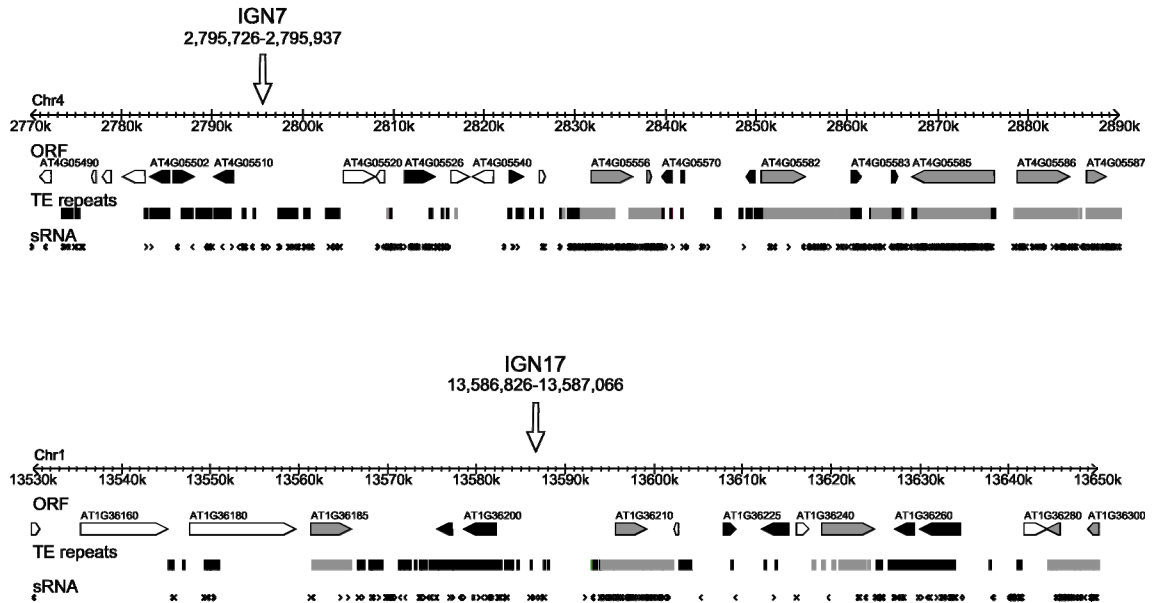


Figure S2. 5' ends of Pol V-dependent transcripts identified by 5' RACE. The terminal nucleotides of cloned 5' RACE products are marked with short arrows and n indicates the number of independent clones obtained for each 5' end. 5' RACE primers as well as nested primers used for amplification are marked with long arrows. Annotations above the DNA sequence refer to top strand-specific RACE clones and those below the DNA sequence refer to bottom strand clones.

IGN5

2323500- TCAAAAATGTGTGGTGGTCCTTCAGGAAGAAGTCCAATCTGAAACATTTTGGGCCGTTTCTAGAAGAACT

2323570- GATTGGATCAGCTGATAGAATTTTATTAACAAAAAAGACTATTTGⁿ⁼¹TATATTGⁿ⁼⁶ATTTTCATTTTTAGTTC

2323640- CCGAGAAGAGTAGAACAAATGCTAAAATGTAATCATGCGGCCCAATAACCAACAAAAGTAACTAGTTGAAAGATG
 nested primer ← 5' RACE primer
 5' RACE primer →

2323710- GATCAATGGTTTTTACATGAAGAAAGCCCAAACCATACACTAATAATCTAATATTTCTATTTAAAAGGACTT
 nested primer →

2323780- GCTGTTATTCGGCCCAATAGCCAACAAAACCTAATTGAAAGATGGATCAGGGGCTATGGAATACCTCAGAA
 n=8 n=1

2323850- TAAAAAGTGTATTTCATTGCAGAGGACCCCTTAAGCGGACATGGTTGGGTCCTTGTTCGACAAGACTTAG
 n=1 n=2 n=2

2323920- TCCTCCATCTTGGCCTCAAGAGTGCTCGACGAAGTTTTATCACCGCTTCATGCGGAATTTTTATTCTTGCT

IGN6

2527000- CTGATGATCATCGAACTTCTTGGATTCCAGCTTCTTTACTATCTAGGTATCATCAATCTCATTACATTC

2527070- TCTCCAATCAATTCATCCCAATTTTTAGTTCTTTTTTCGCTAATTTCTGAGGAATCTCAGGAATTTTAGA
 n=5

2527140- ACCAATAAAGATCTTATACCTTTGAATATCGTAGGAATAGAGGTACCAAAGAAGACAGTTCTTCGAGGTA
 n=3 n=1

2527210- GGTCCCATATCAATTCCTTGGGACATCTATTTGGGTTAGGCTGGATGCATGATCCGCAAACATAGCAAC
 nested primer ← 5' RACE primer

2527280- CGAGAACTCTTGAATTGAAAATCTGATAGGCGCTTCACCAAATCCATCCAAATCTCGTGCTTTTTTCGCC

2527350- GTCACTAACATGTAATGAGCGAACTGGACCGAAAAGCTACTTCTCGACCGTGCTCAAGCAGATACCCGA

2527420- ACTGAGAATTACAAAGCTGTGACCATTCTTTCTCTTGCTTATTCAACGCATACTTCACAGTGTCTTGATA

Figure S3. The chromosomal contexts of the *AtSN1* and *solo LTR* loci tested in our study. Shown are open reading frames (ORF), repetitive elements (TE repeats) and small RNAs in the MPSS database (sRNA). Single copy genes are marked in white, retrotransposons in grey and transposons in black. Data were obtained from <http://chromatin.cshl.edu/cgi-bin/gbrowse/arabidopsis5/>.

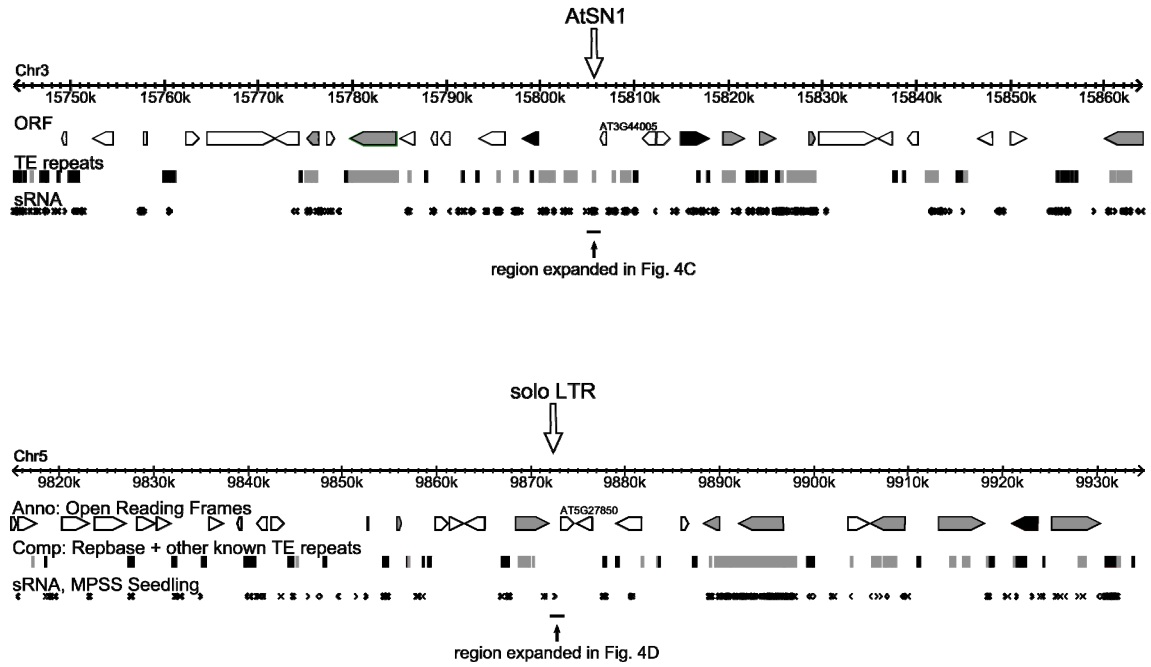
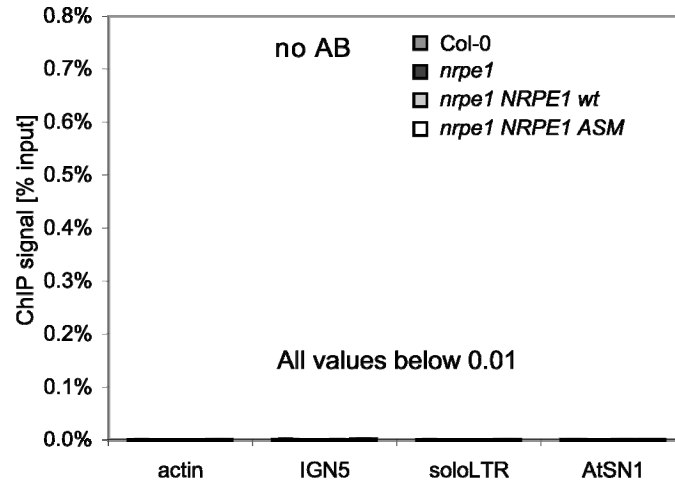


Figure S4. Quantitative PCR of control reactions in which no antibody was included in the chromatin immunoprecipitation (ChIP) experiments shown in Fig. 4F and Figs. 5A-C. Mean values for reactions performed in triplicate are essentially baseline in all cases.



APPENDIX E

RNA POLYMERASE V TRANSCRIPTION GUIDES ARGONAUTE 4 TO
CHROMATIN

Published in *Nature Genetics* (2009), 41 (5): 630-634.

My contributions to this work:

In this study I provided Western blot data demonstrating that the AGO4 protein is unaffected in *pol V* mutants, but is absent when components of the siRNA biogenesis pathway are mutated, namely *pol IV* and *rdr2* (Figure 4D). While *dcl3* mutants still retain low but detectable AGO4 protein levels, AGO4 is absent in the *dcl2,3,4* triple mutant (this experiment was initially performed by me but the experimental result depicted in Figure 4C was generated by Andrzej Wierzbicki). This data not only builds upon the results initially published by the Jacobsen lab (Li et al, 2006), but also establishes that AGO4 protein production and/or stability requires siRNA production. The finding that AGO4 protein levels are unaffected in *nrpe1* mutants was critical to the interpretation of chromatin immunoprecipitation experiments reported. I also made comments on the manuscript and provided technical assistance.

RNA polymerase V transcription guides ARGONAUTE4 to chromatin

Andrzej T Wierzbicki, Thomas S Ream, Jeremy R Haag & Craig S Pikaard

Retrotransposons and repetitive DNA elements in eukaryotes are silenced by small RNA-directed heterochromatin formation. In *Arabidopsis*, this process involves 24-nt siRNAs that bind to ARGONAUTE4 (AGO4) and facilitate the targeting of complementary loci^{1,2} via unknown mechanisms. Nuclear RNA polymerase V (Pol V) is an RNA silencing enzyme recently shown to generate noncoding transcripts at loci silenced by 24-nt siRNAs³. We show that AGO4 physically interacts with these Pol V transcripts and is thereby recruited to the corresponding chromatin. We further show that DEFECTIVE IN MERISTEM SILENCING3 (DMS3), a structural maintenance of chromosomes (SMC) hinge-domain protein⁴, functions in the assembly of Pol V transcription initiation or elongation complexes. Collectively, our data suggest that AGO4 is guided to target loci through base-pairing of associated siRNAs with nascent Pol V transcripts.

Arabidopsis Pol V, AGO4 (ref. 5), DMS3 (ref. 4) and the putative chromatin remodeller DRD1 (ref. 6) function in the silencing of siRNA-homologous loci at one or more steps downstream of siRNA biogenesis^{3,7–10}. Recently, we showed that DRD1 facilitates Pol V transcription of noncoding RNAs at target loci, revealing a functional relationship between these two activities³. However, the functional relationships, if any, between AGO4, DMS3 and Pol V transcription are unclear.

Mutations disrupting *NRPE1* (encoding the largest Pol V subunit), *AGO4* or *DMS3* cause similar losses of RNA-directed DNA methylation at *AtSN1* retrotransposons, *IGN5* (*INTERGENIC REGION 5*) and a retroelement *solo LTR* locus (Fig. 1a,b). Likewise, histone H3 lysine 27 monomethylation (H3K27me1), a characteristic of silenced heterochromatin, is reduced at these loci in *nrpe1*, *ago4* and *dms3* mutants compared to wild-type plants (ecotype Col-0) (Fig. 1c). These results indicate that Pol V, AGO4 and DMS3 collaborate in the establishment of repressive chromatin modifications. At the *solo LTR* locus transcribed by RNA polymerase II (Pol II), chromatin immunoprecipitation (ChIP) shows that levels of diacetylated histone H3 (H3Ac2; acetylated on lysines 9 and 14), a mark of active chromatin, increase in the mutants (Fig. 1d), coincident with increased Pol II occupancy of the locus (Fig. 1e; compare to no-antibody controls in Fig. 1f). At *IGN5* and *AtSN1*, which lack associated Pol II (Fig. 1e), no increase in

histone H3 acetylation is observed in the mutants (Fig. 1d). *AtSN1* elements are thought to be transcribed by Pol III; therefore, differences in H3 acetylation at the *solo LTR* and *AtSN1* loci may reflect the different polymerases involved.

AGO4 and Pol V colocalize in a nucleolus-associated Cajal body^{7,8} that is distant from the target loci subjected to siRNA-mediated silencing. These observations have suggested that AGO4–siRNA complexes might guide Pol V to the target loci^{7,8}. To test this hypothesis, we asked whether production of Pol V transcripts is AGO4 dependent. At intergenic regions *IGN5* and *IGN6* (ref. 3), Pol V transcripts are lost or substantially reduced in the Pol V mutant (*nrpe1*) but not in the *ago4* mutant (Fig. 2a); in fact, *IGN5* transcript levels increase by ~50% in *ago4* (Fig. 2b). This increase in transcript levels is dependent on Pol V, as shown by analysis of the *nrpe1 ago4* double mutant (Fig. 2a). In the *rdr2* (*rna-dependent rna polymerase 2*) mutant, which abolishes 24-nt siRNA biogenesis^{11,12}, or in an *rdr2 ago4* double mutant, Pol V transcript levels are unaffected compared to wild-type (Col-0) plants. We conclude that AGO4–siRNA complexes are dispensable for Pol V transcription at target loci, arguing against the hypothesis that AGO4–siRNA complexes guide Pol V to target loci. The functional significance of AGO4 and Pol V colocalization in Cajal bodies is unclear but could reflect independent protein processing/assembly or storage functions that are unrelated to RNA-induced silencing complex (RISC) assembly.

To test an alternative hypothesis, that AGO4–siRNA complexes are recruited to chromatin in a Pol V-dependent manner, we assayed AGO4 associations with target loci using ChIP (Fig. 3). In wild-type (Col-0) plants, *solo LTR*, *IGN5*, *AtSN1* and *IGN6* loci are all enriched upon AGO4-ChIP, whereas only background levels are observed in *ago4* or *nrpe1* mutants or in control ChIP reactions lacking antibody to AGO4 (anti-AGO4, Fig. 3a). These findings indicate that AGO4 interacts with target locus chromatin and does so in a Pol V-dependent manner. AGO4–chromatin interactions are not diminished by mutation of *DRM2* (Fig. 3a), which encodes the *de novo* DNA methyltransferase that carries out siRNA and AGO4-dependent cytosine methylation^{13,14}. Collectively, these data indicate that Pol V, but not preexisting DNA methylation, is required to recruit AGO4 to chromatin.

To test whether Pol V enzymatic activity is required for AGO4 binding to chromatin, we examined AGO4–chromatin associations in *nrpe1* mutants that had been transformed with either a full-length,



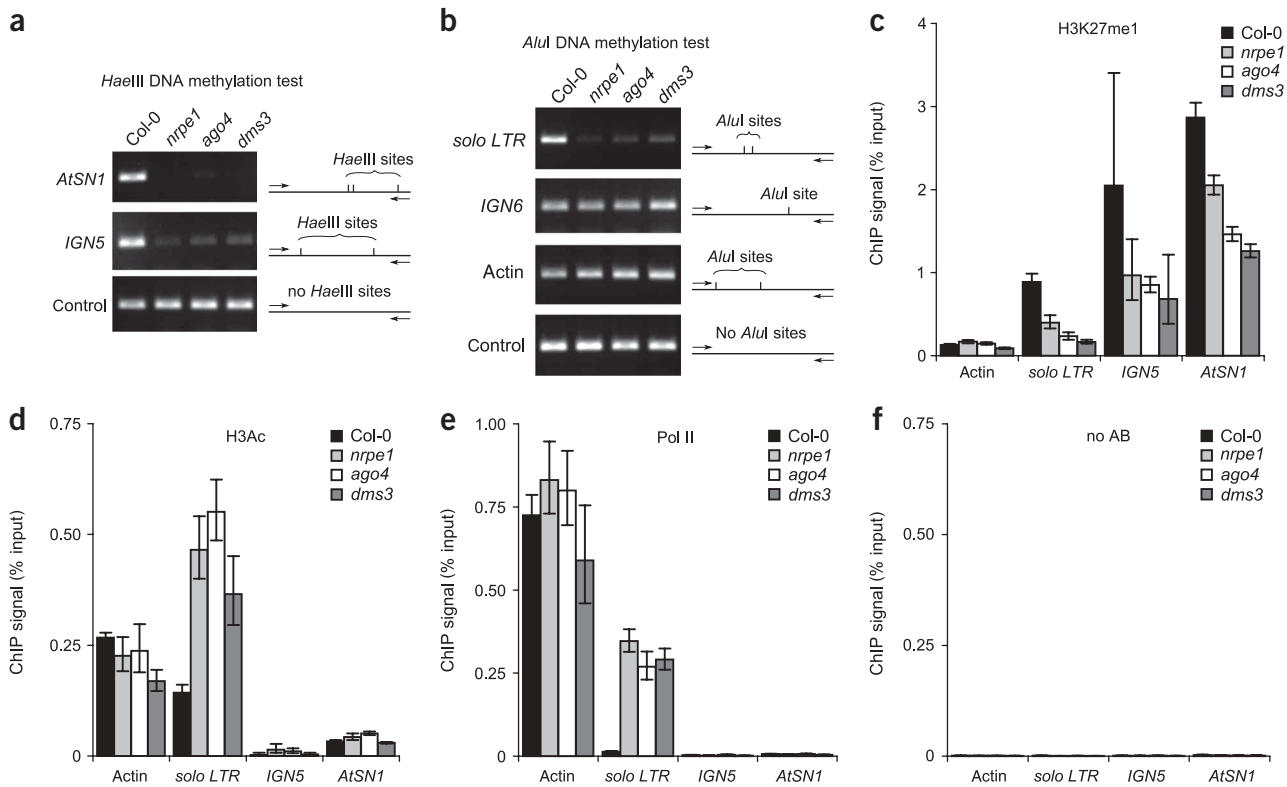
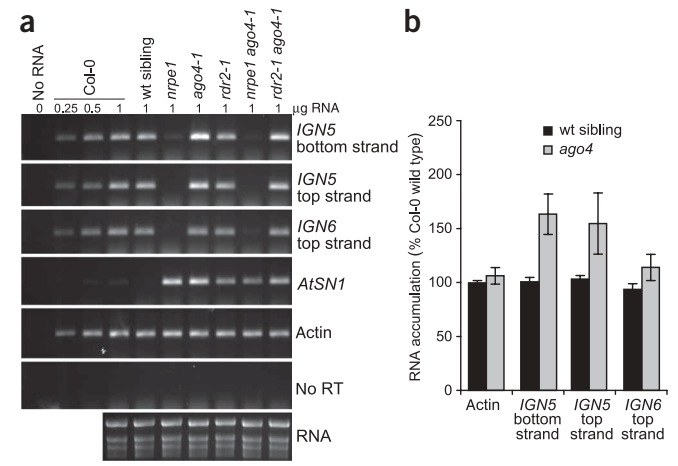


Figure 1 Pol V, AGO4 and DMS3 work nonredundantly in heterochromatin formation. (a,b) DNA methylation analysis at the *AtSN1*, *IGN5* and *solo LTR* loci in *nrpe1*, *ago4* and *dms3* mutants. Genomic DNA was digested with *HaeIII* (a) or *AluI* (b) methylation-sensitive restriction endonucleases followed by PCR. Sequences lacking *HaeIII* sites (actin 2; a) or *AluI* sites (*IGN5*, b) served as controls to show that equivalent amounts of DNA were tested in all reactions. (c,d) ChIP analysis of H3K27me1 (c) and H3Ac2 (d) levels in *nrpe1*, *ago4* and *dms3* mutants. Histograms show means \pm s.d. obtained from three independent amplifications. (e) ChIP analysis of Pol II binding to chromatin in *nrpe1*, *ago4* and *dms3* mutants. Histograms show means \pm s.d. obtained from three independent amplifications. (f) Control ChIP reactions carried out in the absence of antibody reveal background signal levels.

wild-type *NRPE1* transgene or an equivalent transgene bearing point mutations within the metal A motif of the active site (*NPRE1 ASM* transgene). The active site point mutations do not affect *NRPE1* stability or its association with the second-largest subunit but eliminate Pol V transcripts and Pol V biological activity^{3,15}. Whereas the wild-type *NRPE1* genomic transgene (*NRPE1 wt*) restored AGO4 interaction with the *solo LTR*, *IGN5*, *AtSN1* and *IGN6* loci in the *nrpe1* mutant background (Fig. 3b), the active-site mutant (*NRPE1 ASM*) failed to do so. Immunoblotting ruled out the trivial explanation that AGO4 protein levels might be differentially affected by the *nrpe1* mutation or the *NRPE1* transgenes (Fig. 3c) and also demonstrated that the antibody specifically recognizes AGO4, which is absent in the *ago4* mutant. Collectively, the data indicate that Pol V transcriptional activity is required to recruit AGO4 to chromatin.

Base-pairing between AGO4-associated siRNAs and nascent Pol V transcripts could be a mechanism by which Pol V transcription recruits AGO4 to target loci. To test this hypothesis, we used RNA immunoprecipitation to ask whether AGO4 associates with Pol V transcripts *in vivo*. In wild-type (Col-0) plants, anti-AGO4 immunoprecipitates *IGN5* and *IGN6* Pol V transcripts³ (Fig. 4a). Important controls show that Pol V transcripts are not immunoprecipitated in the *ago4* or *nrpe1* mutant backgrounds. Anti-AGO4

Figure 2 AGO4 is not required for Pol V transcription. (a) Strand-specific RT-PCR of Pol V transcription at *IGN5*, *IGN6* and *AtSN1* in *ago4* and *rdr2* mutants as well as *nrpe1 ago4* and *rdr2 ago4* double mutants. Wild-type sibling is a wild-type sibling of the *ago4* mutant identified in a segregating family. Actin RT-PCR products and ethidium bromide-stained rRNAs resolved by agarose gel electrophoresis serve as loading controls. To control for background DNA contamination, we carried out a reaction using *IGN5* top strand primers but no reverse transcriptase (no RT). No-RNA (0 μ g) controls are provided for all primer pairs. (b) Densitometric analysis of RT-PCR data for the *ago4* mutant presented in a. The histogram provides mean band intensities relative to wild type Col-0, \pm s.d. obtained from three independent experiments.



© 2009 Nature America, Inc. All rights reserved.



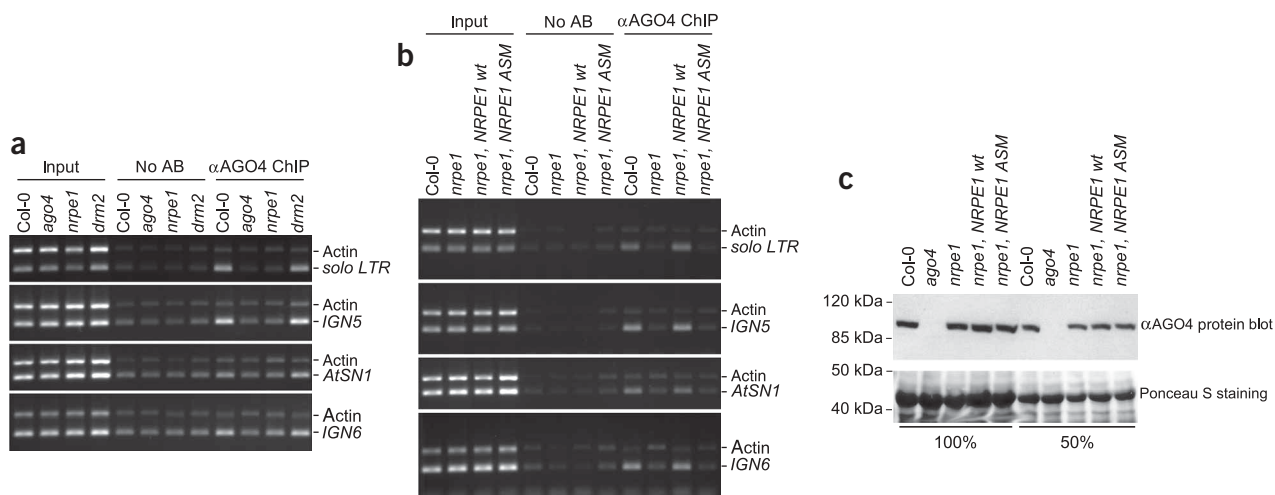


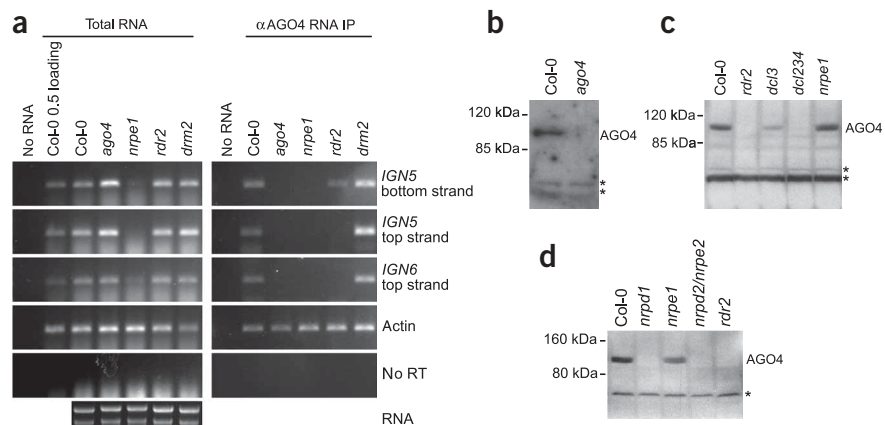
Figure 3 Pol V transcription is necessary for AGO4–chromatin interactions. **(a)** ChIP data showing AGO4 binding to chromatin at *solo LTR*, *IGN5*, *AtSN1* and *IGN6* loci in *ago4*, *nrpe1* and *drm2* mutants. DNA purified from input chromatin samples, chromatin subjected to the immunoprecipitation procedure in the absence of antibody (no Ab) and chromatin immunoprecipitated using anti-AGO4 (α AGO4) was amplified by PCR using locus-specific primers. Primers amplifying the *Actin2* locus served as an internal control. **(b)** ChIP data showing AGO4 binding to chromatin at *solo LTR*, *IGN5*, *AtSN1* and *IGN6* loci in *nrpe1* mutant, *nrpe1* mutant transformed with a wild-type *NRPE1* transgene (*NRPE1 wt*), and *nrpe1* mutant transformed with an *NRPE1* active site mutant transgene (*NRPE1 ASM*). **(c)** Immunoblot detection of AGO4 in protein extracts of wild type (Col-0), *ago4*, *nrpe1*, or *nrpe1* transformed with either a wild-type *NRPE1* transgene (*NRPE1 wt*) or an *NRPE1* active site mutant transgene (*NRPE1 ASM*). Ponceau S staining revealed equal loading of lanes; 100% and 50% sample loadings indicate that the assay is semiquantitative.

immunoprecipitation of *IGN5* or *IGN6* RNAs was also reduced or eliminated in *rdr2* mutant plants, indicating that AGO4–Pol V transcript interactions are dependent on siRNAs. However, in the absence of siRNA biogenesis, as in the *rdr2*, *nrpd1*, *nrpd2/nrpe2* or *dcl2,3,4* mutants, AGO4 protein levels drop below the limits of immunoblot detection^{7,8} (Fig. 4b–d). By contrast, AGO4 protein levels are unaffected in *nrpe1* (Fig. 4b–d) or *drm2* mutants (ref. 7), which act downstream of siRNA biogenesis. The instability of AGO4 in the absence of siRNAs complicates the interpretation of these results. Although we favor the hypothesis that siRNA–Pol V transcript base-pairing is responsible for AGO4 association with Pol V transcripts, we cannot rule out the possibility that AGO4 binds Pol V transcripts directly, with siRNAs merely being required for AGO4 stability.

DMS3 was recently identified as a gene required for RNA-directed DNA methylation that acts at an unspecified step downstream of

siRNA biogenesis⁴. The encoded protein shares sequence similarity with the hinge-domain regions of SMC proteins, such as the core proteins of cohesin and condensin complexes¹⁶, suggesting a chromatin-related function. We found that at *IGN5*, *IGN6* and *AtSN1* loci, Pol V transcripts are substantially reduced or absent in *dms3* mutant plants, as in *nrpe1* (Fig. 5a) or *drd1* mutants³. Likewise, transcriptional suppression of *AtSN1* and *solo LTR* elements is similarly disrupted in *dms3* and *nrpe1* mutants (Fig. 5b). ChIP using an antibody to NRPE1 revealed that, in the *dms3* mutant, Pol V–chromatin associations are reduced to background levels, resembling the actin and *nrpe1* mutant controls (Fig. 5c). Collectively, these data (Fig. 5) indicate that DMS3 is required for Pol V transcription, as shown previously for the chromatin remodeller DRD1 (ref. 3). The loss of detectable Pol V–chromatin association in *dms3* or *drd1* mutants suggests that these chromatin proteins participate in the assembly of Pol V transcription complexes.

Figure 4 AGO4 physically interacts with Pol V transcripts. **(a)** RNA immunoprecipitation using anti-AGO4 (α AGO4). Immunoprecipitated RNA isolated from the indicated mutants was digested with DNaseI and amplified by RT-PCR. Total RNA controls show that the Pol V transcripts are present in equivalent amounts in all mutants tested except *nrpe1*. Ethidium bromide–stained rRNAs (bottom left) show that equal amounts of RNA were tested. The no reverse transcriptase (no RT) control was done with *IGN5* bottom-strand primers. No-RNA controls were carried out for all primer pairs tested. RT-PCR amplification of actin RNA serves as a loading control. **(b)** Immunoblot detection of AGO4 in protein extracts of wild-type (Col-0) plants or *ago4* mutant. Asterisks denote nonspecific bands. **(c)** Immunoblot detection of AGO4 in protein extracts of wild-type (Col-0), *rdr2*, *dcl3*, *dcl234* or *nrpe1* mutants. Asterisks denote nonspecific bands. **(d)** Immunoblot detection of AGO4 in protein extracts of wild-type (Col-0), *nrpd1* (Pol IV), *nrpe1* (Pol V), *nrpd2/nrpe2* (shared subunit of Pol IV and Pol V) or *rdr2* mutants. Asterisks denote nonspecific bands.



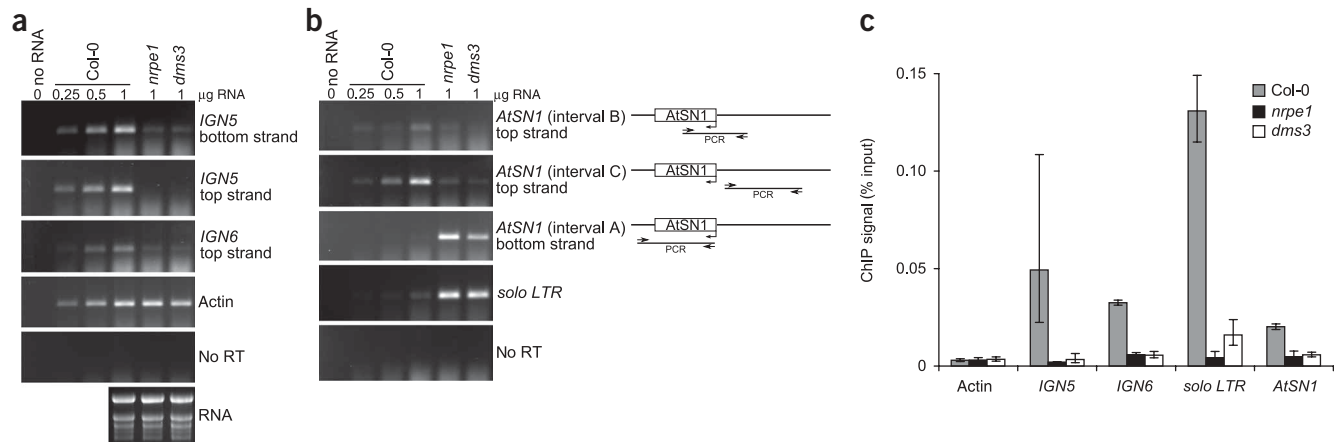


Figure 5 The SMC hinge-domain protein DMS3 is required for Pol V transcription and detectable Pol V-chromatin interactions. **(a,b)** Strand-specific RT-PCR detection of Pol V transcripts at *IGN5* and *IGN6* **(a)** and *AtSN1* **(b)** in wild-type (Col-0) and *nrpe1* and *dms3* mutants. Derepression of Pol II transcripts at the solo LTR and putative Pol III transcripts at *AtSN1* in the *nrpe1* and *dms3* mutants is shown in the right panel. Actin RT-PCR products and ethidium bromide-stained rRNAs resolved by agarose gel electrophoresis serve as loading controls. To control for background DNA contamination, we carried out a reaction using *IGN5* bottom strand **(a)** or *AtSN1* (interval B) primers **(b)** but no reverse transcriptase (no RT). No-RNA (0 μ g) controls are provided for all primer pairs. **(c)** ChIP with anti-NRPE1 in Col-0 wild-type, *nrpe1* and *dms3* mutants followed by real-time PCR. Histograms show means \pm s.d. obtained from three independent amplifications.

Our results suggest that siRNAs and Pol V transcripts are produced by independent pathways that intersect to bring about heterochromatin formation and gene silencing (**Fig. 6**). In one pathway, Pol IV, RDR2 and DCL3 collaborate to produce 24-nt siRNAs that associate with AGO4 (ref. 1). Independent of this pathway, DRD1 and DMS3 facilitate noncoding Pol V transcription at target loci. AGO4's interaction with Pol V transcripts, and the fact that AGO4 association with chromatin requires the Pol V active site, suggests that siRNA-AGO4 complexes are guided to target loci by interacting with Pol V transcripts. It has also been reported that AGO4 can interact with the C-terminal domain (CTD) of NRPE1 *in vitro*^{7,17} and *in vivo*⁷, suggesting that Pol V might recruit AGO4 directly, in an RNA-independent manner. However, we have been unable to detect

AGO4-Pol V associations *in vivo* using immunoprecipitation and subsequent immunoblotting nor by mass spectrometric analysis of affinity-purified Pol V (data not shown), suggesting that any interactions between AGO4 and Pol V may be weak or transient. We suggest that AGO4 recruitment to chromatin is primarily an RNA-mediated process but may also involve protein-protein interactions.

In fission yeast, artificial tethering of the RNA-induced transcriptional silencing (RITS) complex to *ura4* pre-mRNAs is sufficient to induce heterochromatin formation at the normally euchromatic *ura4⁺* locus¹⁸. These and other results are consistent with the hypothesis that fission yeast silencing complexes are guided to chromatin via associations with nascent Pol II transcripts¹⁹. Our findings suggest that plants and yeast are fundamentally similar in their use of RNA guidance mechanisms for recruiting Argonaute-containing transcriptional silencing complexes to target loci. It is intriguing that plants should have evolved a unique RNA polymerase, Pol V, whose specialized role seems to be the generation of noncoding RNAs that can serve as scaffolds for Argonaute recruitment.

METHODS

Plant strains. *Arabidopsis thaliana nrpe1* (*nprp1b-11*) was described previously⁸. The *dms3-4* mutant (SALK_125019C) of locus At3g49250 was obtained from the *Arabidopsis* Biological Resource Center. The *dcl2, dcl3, dcl4* triple mutant (*dcl2,3,4*) was provided by T. Blevins (Washington University, St. Louis). The *ago4-1* mutant (Ler ecotype background) was provided by S. Jacobsen (University of California, Los Angeles) and was introgressed into the Col-0 background by three rounds of backcrossing.

Antibodies. Anti-Pol II (anti-NRPB2) was described previously²⁰. Anti-H3K27me1 #8835 (ref. 21) was provided by T. Jenuwein (Max Planck Institute of Immunobiology). Antibody against diacetyl-H3 (K9 and K14) was obtained from Millipore (cat. #06599, lot #JBC1349702). Rabbit anti-NRPE1 has been described⁹. Rabbit anti-AGO4 was raised against a C-terminal portion of the protein (amino acids 573–924) expressed in bacteria.

RNA and DNA analysis. RNA isolation, RT-PCR and real-time quantitative PCR were carried out as described³ except that real-time quantitative PCR analysis of the *IGN5* locus was done using the following oligonucleotide primers: A195, 5'-ACATGAAGAAAGCCCAACCA-3'; A196,

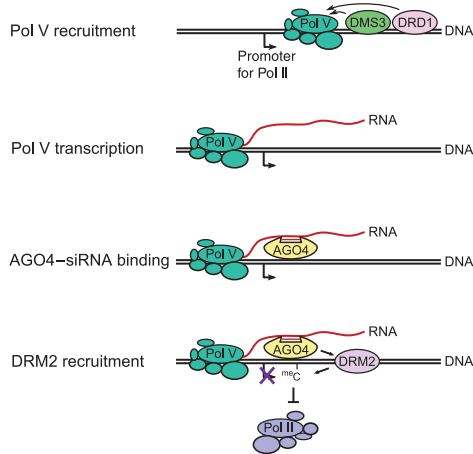


Figure 6 A model for Pol V and siRNA-dependent heterochromatin formation. DMS3 and DRD1 mediate the assembly of Pol V initiation and/or elongation complexes and the production of Pol V transcripts. AGO4-siRNA complexes recognize target loci via base-pairing of siRNAs with nascent Pol V transcripts. AGO4 subsequently recruits chromatin modifying activities including the de novo DNA methyltransferase DRM2 and histone modifying enzymes via unknown mechanisms.

5'-GGCCGAATAACAGCAAGTCCT-3'. Densitometric analysis of DNA resolved by agarose gel electrophoresis was performed using ImageJ.

ChIP and RNA IP. ChIP and RNA IP were carried out as described³ except that for ChIP with anti-AGO4, RNase A was added during immunoprecipitation, washes with TE buffer were omitted, immune complexes were eluted with 100 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS for 10 min at room temperature and a second elution at 65 °C was performed. Crosslinking was reversed at 65 °C for 1 h in the presence of 40 µg Proteinase K (Invitrogen). DNA was purified by extraction with phenol:chloroform and ethanol precipitation. DNA recovery was assayed by PCR using 1.5 u Platinum Taq (Invitrogen).

ACKNOWLEDGMENTS

Our work is supported by US National Institutes of Health grant GM077590. The content of the paper is the sole responsibility of the authors and does not necessarily reflect the views of the NIH. We thank T. Blevins (Washington University, St. Louis), S. Jacobsen (University of California, Los Angeles) and T. Jenuwein (Max Planck Institute of Immunobiology) for providing reagents.

AUTHOR CONTRIBUTIONS

T.S.R. generated anti-AGO4; J.R.H. and T.S.R. assayed NRPE1-AGO4 interactions; J.R.H. produced **Figure 4d**; A.T.W. performed all remaining experiments. A.T.W. and C.S.P. wrote the manuscript.

Published online at <http://www.nature.com/naturegenetics/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Chapman, E.J. & Carrington, J.C. Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* **8**, 884–896 (2007).
- Girard, A. & Hannon, G.J. Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol.* **18**, 136–148 (2008).
- Wierzbiński, A.T., Haag, J.R. & Pikaard, C.S. Noncoding transcription by RNA Polymerase Pol IVb/ Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**, 635–648 (2008).
- Kanno, T. *et al.* A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat. Genet.* **40**, 670–675 (2008).
- Vaucheret, H. Plant ARGONAUTES. *Trends Plant Sci.* **13**, 350–358 (2008).
- Kanno, T. *et al.* A SNF2-like protein facilitates dynamic control of DNA methylation. *EMBO Rep.* **6**, 649–655 (2005).
- Li, C.F. *et al.* An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126**, 93–106 (2006).
- Pontes, O. *et al.* The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**, 79–92 (2006).
- Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* **19**, 2030–2040 (2005).
- Kanno, T. *et al.* Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **37**, 761–765 (2005).
- Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, e104 (2004).
- Kasschau, K.D. *et al.* Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol.* **5**, e57 (2007).
- Cao, X. *et al.* Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* **13**, 2212–2217 (2003).
- Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
- Haag, J.R., Pontes, O. & Pikaard, C.S. Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS ONE* **4**, e4110 (2009).
- Peric-Hupkes, D. & van Steensel, B. Linking cohesin to gene regulation. *Cell* **132**, 925–928 (2008).
- El-Shami, M. *et al.* Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* **21**, 2539–2544 (2007).
- Bühler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873–886 (2006).
- Bühler, M. & Moazed, D. Transcription and RNAi in heterochromatic gene silencing. *Nat. Struct. Mol. Biol.* **14**, 1041–1048 (2007).
- Onodera, Y. *et al.* Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613–622 (2005).
- Peters, A.H.F.M. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577–1589 (2003).

APPENDIX F

SEX-BIASED LETHALITY OR TRANSMISSION OF DEFECTIVE
TRANSCRIPTION MACHINERY IN ARABIDOPSIS

Published in *Genetics* (2008), 180 (1): 207-218.

My contributions to this work:

The DNA-dependent RNA Polymerase I, II and III FLAG-tagged lines (NRPA2, NRPB2 and NRPC2, respectively) were cloned, dipped and validated in the wild type background by me. The NRPC2-FLAG construct was also dipped into the heterozygous *nrpc2* mutant background. Under my supervision, Diane Pikaard screened the progeny of these individuals and identified NRPC2-FLAG transformants in the homozygous *nrpc2* mutant background. Yasuyuki Onodera crossed the NRPA2-FLAG and NRPB2-FLAG transformants into the *nrpa2* and *nrpb2* heterozygous mutant backgrounds, respectively, and identified transformants in the homozygous mutant backgrounds [Note, the *nrpa2* and *nrpb2* SAIL mutant lines already contained the BASTA selectable marker precluding selection for successful Agrobacterium-mediated vector transformants]. I also provided comments during the editing of this work.

Sex-Biased Lethality or Transmission of Defective Transcription Machinery in Arabidopsis

Yasuyuki Onodera,^{*1} Kosuke Nakagawa,^{*1} Jeremy R. Haag,[†] Diane Pikaard,[†]
Tetsuo Mikami,^{*} Thomas Ream,[†] Yusuke Ito^{*} and Craig S. Pikaard^{†,2}

^{*}Division of Applied Bioscience, Research Faculty of Agriculture, Hokkaido University, Kita 9, Nishi 9, Kita-ku, Sapporo 060-8589, Japan and [†]Biology Department, Washington University, St. Louis, Missouri 63130

Manuscript received April 23, 2008

Accepted for publication June 24, 2008

ABSTRACT

Unlike animals, whose gametes are direct products of meiosis, plant meiotic products undergo additional rounds of mitosis, developing into multicellular haploid gametophytes that produce egg or sperm cells. The complex development of gametophytes requires extensive expression of the genome, with DNA-dependent RNA polymerases I, II, and III being the key enzymes for nuclear gene expression. We show that loss-of-function mutations in genes encoding key subunits of RNA polymerases I, II, or III are not transmitted maternally due to the failure of female megaspores to complete the three rounds of mitosis required for the development of mature gametophytes. However, male microspores bearing defective polymerase alleles develop into mature gametophytes (pollen) that germinate, grow pollen tubes, fertilize wild-type female gametophytes, and transmit the mutant genes to the next generation at moderate frequency. These results indicate that female gametophytes are autonomous with regard to gene expression, relying on transcription machinery encoded by their haploid nuclei. By contrast, male gametophytes make extensive use of transcription machinery that is synthesized by the diploid parent plant (sporophyte) and persists in mature pollen. As a result, the expected stringent selection against nonfunctional essential genes in the haploid state occurs in the female lineage but is relaxed in the male lineage.

IN flowering plants, three rounds of postmeiotic mitosis and development give rise to an eight-nucleate female gametophyte, one cell of which is the egg cell (SCHNEITZ *et al.* 1995; GROSSNIKLAUS and SCHNEITZ 1998; DREWS and YADEGARI 2002). Pollen, the male gametophyte, consists of three haploid cells, two of which are sperm cells. The three pollen cells are clonally related and are all descended from a single haploid meiotic product of a pollen mother cell (McCORMICK 1993, 2004). The male gametophyte can survive independent of the sporophyte (the parent plant) and upon landing on a receptive flower, the pollen germinates and develops a pollen tube that elongates through the transmitting tract of the pistil, the female floral organ, to reach the ovary. Within the ovary, the pollen tube grows toward chemical signals emanating from the two synergid cells of the female gametophyte (HIGASHIYAMA 2002; HIGASHIYAMA *et al.* 2001, 2003; JOHNSON and PREUSS 2002). Upon reaching a synergid cell, adjacent to the egg, the pollen tube ruptures, releasing the sperm. One sperm cell fuses with the egg to give rise to the diploid embryo. The

second sperm cell fuses with the female gametophyte's central cell, giving rise to the endosperm. Proper development of both embryo and endosperm as a result of double fertilization is required for seed maturation (RUSSELL 1993; GROSSNIKLAUS and SCHNEITZ 1998; YADEGARI *et al.* 2000).

Large-scale analyses of cDNA libraries generated from mRNAs purified from maize and wheat female gametophytes have shown that thousands of genes are expressed in female gametophytes (SPRUNCK *et al.* 2005; YANG *et al.* 2006). Comparative microarray-based transcript profiling analyses using ovules of Arabidopsis wild-type plants and mutants lacking embryo sacs have similarly identified large numbers of female gametophyte-specific genes (YU *et al.* 2005; JOHNSTON *et al.* 2007; JONES-RHOADES *et al.* 2007; STEFFEN *et al.* 2007). Collectively, expression-profiling studies combined with analyses of female gametophytic mutants (PAGNUSSAT *et al.* 2005) provide evidence for extensive transcriptional regulatory networks that are critical for the proper development of female gametophytes.

In Arabidopsis, ~62% of all genes in the genome are expressed during at least one stage of male gametophyte development, with ~10% of these transcripts being pollen specific (HONYS and TWELL 2003, 2004). Moreover, labeled UTP is incorporated into RNA in pollen and the transcription inhibitor, actinomycin D inhibits

¹These authors contributed equally to this work.

²Corresponding author: Department of Biology, Washington University, Campus Box 1137, 1 Brookings Dr., St. Louis, MO 63130.
E-mail: pikaard@biology2.wustl.edu

pollen tube growth (MASCARENHAS 1989, 1993; HONYS and TWELL 2004). These observations indicate that male gametophytes are actively engaged in the transcription of their haploid genomes.

The enzymes central to nuclear gene expression are DNA-dependent RNA polymerases I, II, and III (Pol I, Pol II, and Pol III), each of which is composed of between 12 and 17 subunits. Pol I is responsible for transcribing the 45S preribosomal RNAs (rRNAs) that are then processed into the 18S, 5.8S, and 25–28S (the latter size depends on the species) rRNAs that form the catalytic core of ribosomes. Pol II transcribes messenger RNAs (mRNAs) as well as RNAs that do not encode proteins, such as micro RNAs and small nuclear RNAs that guide mRNA and rRNA processing events. Pol III is primarily responsible for transcribing transfer RNAs (tRNAs) and repetitive 5S rRNA genes (KASSAVETIS *et al.* 1994; PAULE and WHITE 2000).

For purposes of gene and subunit nomenclature, Arabidopsis Pol I is denoted as nuclear RNA polymerase A (NRPA), Pol II is denoted as NRPB, and Pol III is denoted as NRPC. Their second-largest subunits, denoted as NRPA2, NRPB2, and NRPC2, respectively, are homologs of the β -subunits of eubacterial RNA polymerase. Together with the largest subunits, the β -like second-largest subunits help form the active sites of the enzymes and are essential for RNA synthesis. In *Arabidopsis thaliana*, the Pol I, Pol II, and Pol III second-largest subunits are encoded by single-copy genes located on chromosomes 1, 4, and 5, respectively (LARKIN and GUILFOYLE 1993; ONODERA *et al.* 2005); see also phylogenetic analyses by Craig S. Pikaard and Jonathan Eisen discussed in ARABIDOPSIS GENOME INITIATIVE (2000).

Contrary to our expectation that loss-of-function mutations in NRPA2, NRPB2, or NRPC2 genes would be unrecoverable due to lethality in both the haploid male and female gametophytes, transgenic lines hemizygous for T-DNA disruptions of each gene can be identified and maintained. Detailed analysis of these lines revealed that the mutant RNA polymerase alleles are not transmitted through the female lineage due to the failure of mutant female gametophytes to complete their development. By contrast, the mutant alleles are transmitted to subsequent generations through the male gametophyte at moderate efficiency compared to wild type. Our data indicate that pollen can develop to maturity, grow pollen tubes, and carry out fertilization in the absence of functional RNA polymerase genes, apparently by utilizing transcription machinery synthesized premeiotically in pollen mother cells. By contrast, female gametophyte development is autonomous and requires transcription machinery generated *de novo* in the haploid state.

MATERIALS AND METHODS

Plant strains and growth conditions: *Arabidopsis thaliana* wild-type and T-DNA insertion mutants (ecotype Columbia in

both cases) were grown at 22° with a 16-hr photoperiod. Gene locus identifiers for NRPA2, NRPB2, and NRPC2 are At1g29940, At4g21710, and At5g45140, respectively. The T-DNA insertion alleles we named *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, and *nrpb2-2* are carried within Torrey Mesa Research Institute (San Diego) transgenic lines: GARLIC_726_H01, GARLIC_918_C10, GARLIC_859_B04, and GARLIC_110_G08, respectively [GARLIC is the former name of the Syngenta Biotechnology's SAIL collection of T-DNA lines, available from the Arabidopsis Biological Resource Center (ABRC) at Ohio State University]. The parental line for GARLIC_110_G08 was homozygous for the *qrt1-2* allele of the *QUARTET* gene (ecotype Columbia) (PREUSS *et al.* 1994); other GARLIC lines are wild type at the *QRT* locus. The T-DNA allele *nrpc2-1* is present in Salk line 007865 (ALONSO *et al.* 2003) obtained from the ABRC. Seeds of plants bearing the *nrpc2-2* (GABI_131_B09) allele were obtained from GABI-Kat (Rosso *et al.* 2003). The transgenic Arabidopsis line (SAIL_100_H07) carrying a *LAT52::GUS* reporter gene(s) inserted in an intergenic region was obtained from ABRC.

Genotyping: To identify T-DNA disrupted alleles in segregating families, PCR was carried out using primers complementary to the T-DNA left border (5'-GCATCTGAATTTCA TAACCAATCTC-3', 5'-CGTCCGCAATGTGTTATTAAG-3', or 5'-CCCATTGGACGTGAATGTAGACAC-3') and primers specific for NRPA2 (5'-AGAGAGGTAGAGAACTCAGC-3' or 5'-ATAAACAGTTAGGCAAGCGAA-3'), NRPB2 (5'-CGATTTGAG CTTCTACCGTTT-3' or 5'-CCTAGAACATACCATGCCGAAA-3') or NRPC2 (5'-CTCGCACAATGAAGGATGTTT-3' or 5'-TAATTC TTGCCGCAAATTGAC-3'). Wild-type alleles of NRPA2, NRPB2, and NRPC2 were identified using the gene-specific primers above in combination with 5'-GATGAGTTGGATAACACGA AC-3' or 5'-AGCACCCCTTTAAGCTACAAAG-3' for NRPA2; 5'-CCATCAGACTCTGTATCATA-3' or 5'-ACGAAGGGTAA GCATGCAGTT-3' for NRPB2; and 5'-AGCTACTCCAGGGGA GATTAT-3' or 5'-GGCAAGTACTATAGCCCCCTG-3' for NRPC2.

The unique genomic DNA/T-DNA junction sequences at both ends of the single T-DNA loci in *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles were amplified by PCR and verified by sequencing.

Production of transgenic plants: Genomic sequences for NRPA2 (positions -1433 to +7346 relative to the translation start site), NRPB2 (positions -338 to +6514), or NRPC2 (positions -1947 to +10295) were amplified by PCR. Amplified gene sequences included promoter regions and all introns and exons. Resulting PCR products were captured in pENTR/D-TOPO and recombined into the Gateway recombination (Invitrogen)-compatible expression vector pEarley-Gate 302 (EARLEY *et al.* 2006). Resulting NRPA2, NRPB2, or NRPC2 full-length transgenes were introduced into hemizygous plants bearing a corresponding mutant allele (+/*nrpa2-1*, +/*nrpb2-1*, or +/*nrpc2-1*). Progeny of transgenic plants that were homozygous for the *nrpa2-1*, *nrpb2-1*, or *nrpc2-1* mutations and were rescued by the full-length transgenes were identified by PCR genotyping.

Confocal laser scanning microscopy: Examination of specimens was carried out using a Zeiss LSM confocal microscope system equipped with a Helium/Neon laser. Images were processed using Adobe Photoshop 7.0 software. Floral stages were defined according to BOWMAN (1994). Developmental stages of female gametophytes were defined according to CHRISTENSEN *et al.* (1997).

Cytological and histochemical analysis of pollen: *In vitro* pollen germination was carried out as described by HASHIDA *et al.* (2007). Pollen were stained with 1 μ g/ml DAPI in 20 mM Tris-HCl pH 7.65, 0.5 mM EDTA, 1.2 mM spermidine, 7 mM 2-mercaptoethanol, 0.4 mM phenylmethylsulfonyl fluoride, 0.1 mg/ml FDA in 0.5 M sucrose, or Alexander solution (ftp://

ftp.arabidopsis.org/home/tair/Protocols/EMBOmanual/ch1.pdf). Pollen and self-pollinated pistils were incubated at 37°C for 12 hr in GUS staining solution (50 mM sodium phosphate pH 7.2, 0.2% Triton X-100, 2 mM potassium ferrocyanide, 2 mM potassium ferricyanide, and 1 mg/ml X-Gluc).

RESULTS

Sex-biased defects in the transmission of mutant alleles encoding RNA polymerase I, II, and III second-largest subunits: We used a PCR-based strategy to verify the existence of T-DNA-disrupted alleles for the catalytic second-largest subunits of RNA polymerase I (alleles *nrpa2-1* and *nrpa2-2*), RNA polymerase II (alleles *nrbp2-1* and *nrbp2-2*), or RNA polymerase III (alleles *nrpc2-1* and *nrpc2-2*) (Figure 1A). We then genotyped the progeny resulting from self-fertilization of plants bearing these alleles. In all cases, individuals that carried a mutant RNA polymerase allele also carried a corresponding wild-type allele (Figure 1B and data not shown), indicating that these plants were hemizygous for the mutations. No plants homozygous for the Pol I (*nrpa2-1*, *nrpa2-2*), Pol II (*nrbp2-1*, *nrbp2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) mutant alleles were recovered, indicating that the alleles are all severe loss-of-function mutations in essential genes, consistent with the essential roles of Pol I, Pol II, and Pol III in nuclear gene expression.

Hemizygotes should outnumber homozygous wild-type siblings 67%:33% (2:1) among the progeny of a hemizygous parent bearing one copy of a defective essential gene, assuming that the homozygous mutant is inviable. However, as shown in Table 1, PCR-based genotyping revealed that only 8–38% of the progeny were hemizygous for Pol I (*nrpa2-1* or *nrpa2-2*), Pol II (*nrbp2-1* or *nrbp2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) mutant alleles (Table 1). Instead, the majority of the progeny possessed only wild-type alleles, indicating a defect in the transmission of the mutant RNA polymerase alleles.

To test for sex-biased defects in the transmission of the mutant alleles through the male or female gametophytes, Pol I hemizygotes (+/*nrpa2-1* or +/*nrpa2-2*), Pol II hemizygotes (+/*nrbp2-1* or +/*nrbp2-2*, *qrt1-2*; the latter is a Pol II mutant hemizygote in a homozygous quartet mutant background), or Pol III hemizygotes (+/*nrpc2-1* or +/*nrpc2-2*) were reciprocally crossed with wild-type (+/+) plants by hand-pollinating emasculated flowers. Resulting progeny were then genotyped by PCR. None of the mutant polymerase alleles were found to be transmitted to the progeny via the maternal parent (Figure 1, C–E; Table 2); instead all progeny of hemizygous (+/–) female plants crossed with wild-type (+/+) males were homozygous wild type (+/+). By contrast, the *nrpa2-1*, *nrpa2-2*, *nrbp2-1*, *nrbp2-2*, *nrpc2-1*, and *nrpc2-2* alleles were all pollen transmissible, such that 13–38% of the progeny inherited a mutant allele from the hemizygous paternal parent when crossed with a wild-type female (Table 2). Note, however, that equal

numbers of hemizygous (+/–) and homozygous (+/+) progeny are expected from a (+/+) × (+/–) cross if the wild-type and mutant alleles are transmitted with equal efficiency; the male-transmitted Pol I, II, and III mutant alleles were not inherited at such high levels.

The reciprocal crossing data summarized in Tables 1 and 2 indicate a lack of transmission of the mutant polymerase second-largest subunit alleles through female gametophytes and a partial defect in their transmission through the male gametophyte. Similar allele transmission behavior was observed for the RNA polymerase subunit mutant *nrbp12a* (supplemental Table S1). The homolog of *NRPB12a* in yeast is a single-copy gene whose encoded protein is incorporated into all three nuclear polymerases (Pol I, II, and III). As was the case for the second-largest subunit mutants, homozygous *nrbp12a* mutants were not recoverable. Moreover, *nrbp12a* mutant alleles were transmitted via pollen but not through the female gametophytes. Collectively, our results indicate that male-specific transmissibility of defective RNA polymerase alleles is a general characteristic of RNA polymerase subunit genes and not a peculiarity of second-largest subunit genes.

Defective RNA polymerase alleles cause female gametophyte developmental arrest: Lack of maternal transmission of the Pol I (*nrpa2-1* or *nrpa2-2*), Pol II (*nrbp2-1* or *nrbp2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) alleles prompted an examination of siliques (seed pods) of self-pollinated hemizygous +/*nrpa2-1*, +/*nrpa2-2*, +/*nrbp2-1*, +/*nrbp2-2*, +/*nrpc2-1*, +/*nrpc2-2*, or +/*nrbp2-2*, *qrt1-2* plants. Siliques of these plants contain small unfertilized ovules interspersed with an equal number of normal seeds; as an example, a silique from a +/*nrpa2-1* plant is shown in Figure 2A. Whereas wild-type plants produce 51–58 seeds per silique, siliques of Pol I (*nrpa2-1* or *nrpa2-2*), Pol II (*nrbp2-1* or *nrbp2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) mutant hemizygotes contain only 25–27 mature seeds (Figure 2B).

Defects in seed set caused by the polymerase mutations were rescued by transforming Pol I, Pol II, or Pol III hemizygotes with full-length *NRPA2*, *NRPB2*, or *NRPC2* genomic clone transgenes expressed from their endogenous promoters (Figure 2B). Southern blot and segregation analyses showed that the transgenes in each case were integrated in multiple copies at a single locus (data not shown) such that the plants tested in Figure 2B were hemizygous for the polymerase mutant alleles as well as being hemizygous for the rescuing transgene loci. As a result, seed set is rescued by the transgenes to a level intermediate between the mutant and wild-type phenotypes. This is due to the independent segregation of the transgenes and polymerase alleles such that only half of the gametophytes bearing a mutant polymerase allele inherit a rescuing transgene. Collectively, our data indicate that functional RNA polymerases are essential for one or more critical aspects of female gametophyte development, fertilization, or seed development.

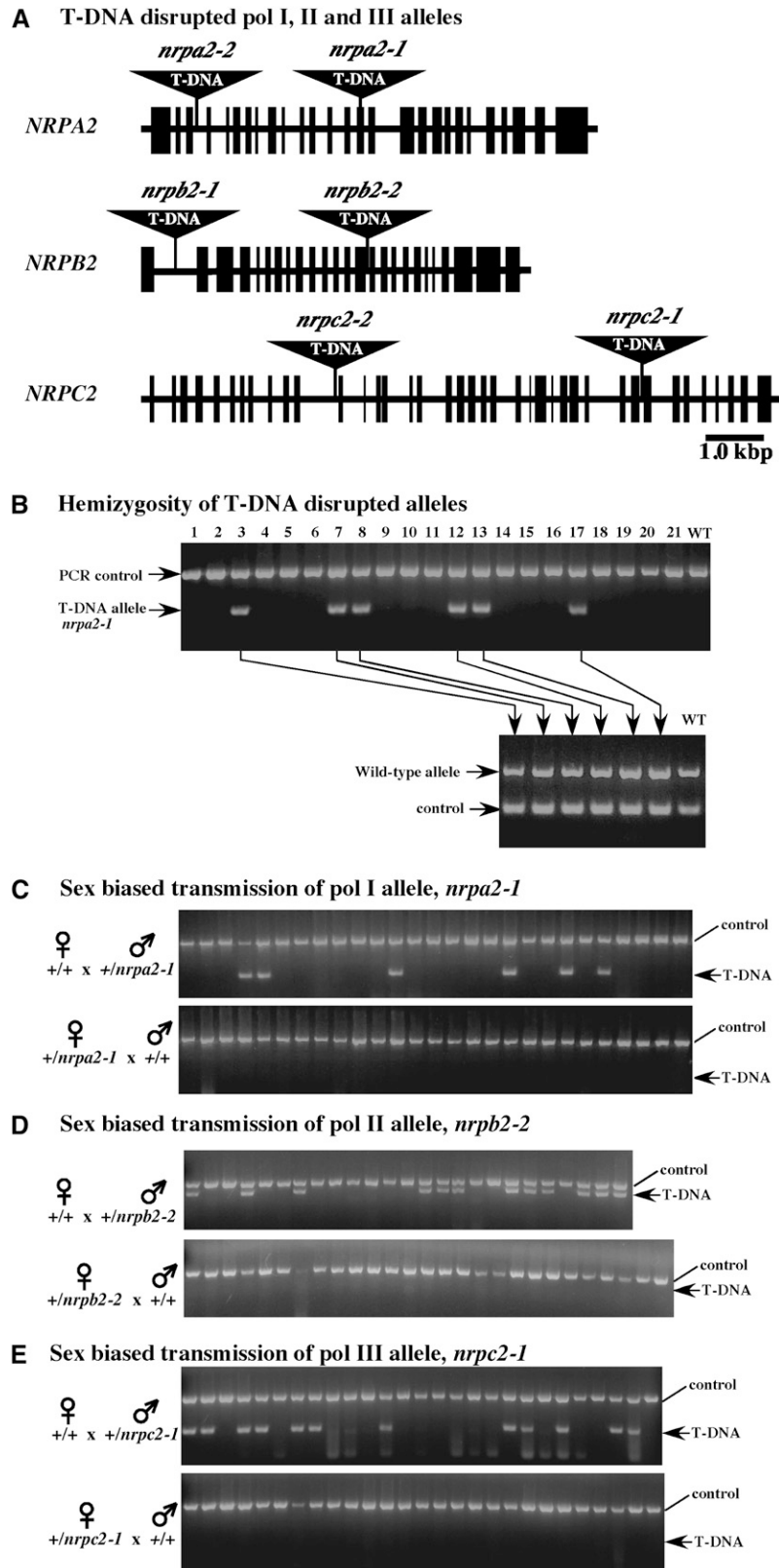


FIGURE 1.—Sex-biased transmission of disrupted alleles for second-largest subunits of RNA polymerases I, II, and III (*NRPA2*, *NRPB2*, and *NRPC2*, respectively). (A) Structures of the *NRPA2*, *NRPB2*, and *NRPC2* genes showing the positions of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* T-DNA insertions. Solid boxes represent exons. (B) PCR-based genotyping of progeny of a self-fertilized $+/\textit{nrpa2-1}$ hemizygote. Disrupted alleles were detected using a T-DNA-specific primer in conjunction with a gene-specific primer. Wild-type alleles were detected using primers that flank the T-DNA insertion site. (C–E) PCR-based detection of T-DNA disrupted alleles in progeny generated from reciprocal crosses between wild-type ($+/+$) and $+/\textit{nrpa2-1}$, $+/\textit{nrpb2-2}$, and $+/\textit{nrpc2-1}$ hemizygotes.

To further investigate the defects in ovule development and female transmission of mutant alleles (Figure 1, C–E; Table 2), ovaries of flowers at floral stage 13 (BOWMAN 1994), a stage just prior to flower opening,

were examined by confocal laser scanning microscopy (CLSM). Female gametophytes develop relatively synchronously (CHRISTENSEN *et al.* 1997) such that gametophytes that have undergone all three rounds of mitosis

TABLE 1
Genotypes of progeny of Pol I, II, and III hemizygotes

Parental genotype	% homozygous wt (+/+)	% hemizygous (+/-)	% homozygous mutant
+/ <i>nrpa2-1</i>	76 (62/82)	24 (20/82)	0 (0/82)
+/ <i>nrpa2-2</i>	63 (32/51)	37 (19/51)	0 (0/51)
+/ <i>nrpb2-1</i>	86 (18/21)	14 (3/21)	0 (0/21)
<i>qrt1-2</i> , +/ <i>nrpb2-2</i>	80 (67/84)	20 (17/84)	0 (0/84)
+/ <i>nrpc2-1</i>	62 (39/63)	38 (24/63)	0 (0/63)
+/ <i>nrpc2-2</i>	92 (45/49)	8 (4/49)	0 (0/49)

Mutant alleles *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* are underrepresented among the progeny of self-fertilized hemizygotes. Numbers in parentheses represent the number of individuals displaying a given genotype and the total number of individuals examined. wt, wild type.

(female gametophyte stages FG5–FG7; see Figure 3A) are observed at floral stage 13 in wild-type pistils (Figure 3B and Table 3). By contrast, in floral stage 13 pistils of hemizygous plants segregating mutant alleles for Pol I (+/*nrpa2-1* or +/*nrpa2-2*), Pol II (+/*nrpb2-1* or +/*nrpb2-2*, *qrt1-2*), or Pol III (+/*nrpc2-1* or +/*nrpc2-2*), ~50% of the female gametophytes arrest after only one or two rounds of mitosis (2–4 nuclei), at developmental stages FG2–FG4 (Table 3, Figure 3, C–G, and supplemental Figure S1). The other ~50% of the gametophytes in these ovaries display normal development, as in wild-type plants, consistent with the 1:1 segregation of wild-type and mutant alleles within the siliques of plants hemizygous for the mutations.

Detailed examination of ovules within +/*nrpa2-1* plants indicated that female gametophytes lacking

functional Pol I arrest most frequently at the two-nucleus stage (FG2 and FG3; Figure 3, C and D and Table 3) and were not observed to progress beyond the four-nucleus stage. Similar results were observed for hemizygous plants bearing the *nrpa2-2* Pol I mutant allele (Table 3 and supplemental Figure S1, A and B).

As shown in Table 3, Figure 3, E–G, and supplemental Figure S1, most of the *nrpb2-1*, *nrpb2-2*, and *nrpc2-1* female gametophytes arrested after the second mitotic division (FG4), at the four-nucleus stage, whereas the majority of *nrpc2-2* female gametophytes displayed developmental arrest at the two-nucleus stage (FG2 and FG3). The difference in the severity of the *nrpc2-1* and *nrpc2-2* alleles is presumably due to the relative locations of the T-DNA insertions, with the T-DNA in the stronger *nrpc2-2* allele occurring in an earlier intron (see Figure 1).

Collectively, the microscopic analyses suggest that female gametophytes carrying defective alleles for RNA polymerases I, II, or III arrest early in development, at or prior to the four-nucleus stage, FG4.

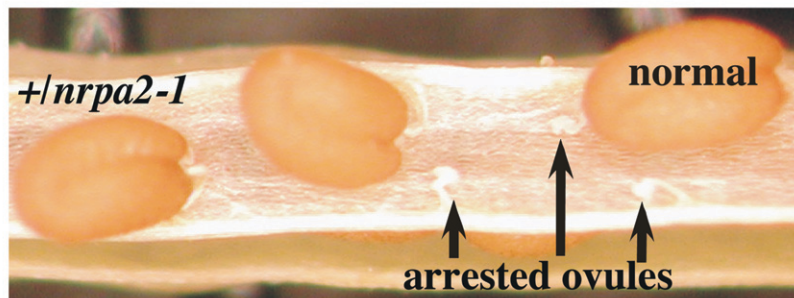
Certation explains reduced male transmissibility of defective polymerase alleles: As shown in Table 2 and Figure 1, C–E, *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles are all transmitted via the male gametophyte. However, homozygous wild-type individuals outnumber hemizygous individuals among the progeny of self-fertilized hemizygotes or among the progeny of wild-type females outcrossed with a hemizygous male (Tables 1 and 2). These data indicate that male gametophytes bearing wild-type RNA polymerase alleles are either more viable or more successful at fertilization than are male gametophytes bearing mutant polymerase alleles.

TABLE 2
Male-specific transmission of Pol I, II, and III mutant alleles

Parental genotype		Genotypes of progeny	
Female parent	Male parent	% homozygous wt (+/+)	% hemizygous (+/-)
+/ <i>nrpa2-1</i>	+/+	100 (55/55)	0 (0/55)
+/ <i>nrpa2-2</i>	+/+	100 (46/46)	0 (0/46)
+/ <i>nrpb2-1</i>	+/+	100 (52/52)	0 (0/52)
<i>qrt1-2</i> , +/ <i>nrpb2-2</i>	+/+	100 (42/42)	0 (0/42)
+/ <i>nrpc2-1</i>	+/+	100 (56/56)	0 (0/56)
+/ <i>nrpc2-2</i>	+/+	100 (47/47)	0 (0/47)
+/+	+/ <i>nrpa2-1</i>	75 (42/56)	25 (14/56)
+/+	+/ <i>nrpa2-2</i>	62 (24/39)	38 (15/39)
+/+	+/ <i>nrpb2-1</i>	79 (38/48)	21 (10/48)
+/+	<i>qrt1-2</i> , +/ <i>nrpb2-2</i>	70 (19/27)	30 (8/27)
+/+	+/ <i>nrpc2-1</i>	67 (36/54)	33 (18/54)
+/+	+/ <i>nrpc2-2</i>	87 (45/52)	13 (7/52)

Paternally biased transmission of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles. Wild-type (+/+) plants were reciprocally crossed with +/*nrpa2-1*, +/*nrpa2-2*, +/*nrpb2-1* (in *qrt1-2* mutant background); +/*nrpb2-2*, +/*nrpc2-1*, and +/*nrpc2-2* and resulting progeny were genotyped. Numbers in parentheses are the number of progeny displaying the specified genotype out of the total number of progeny examined.

A



B

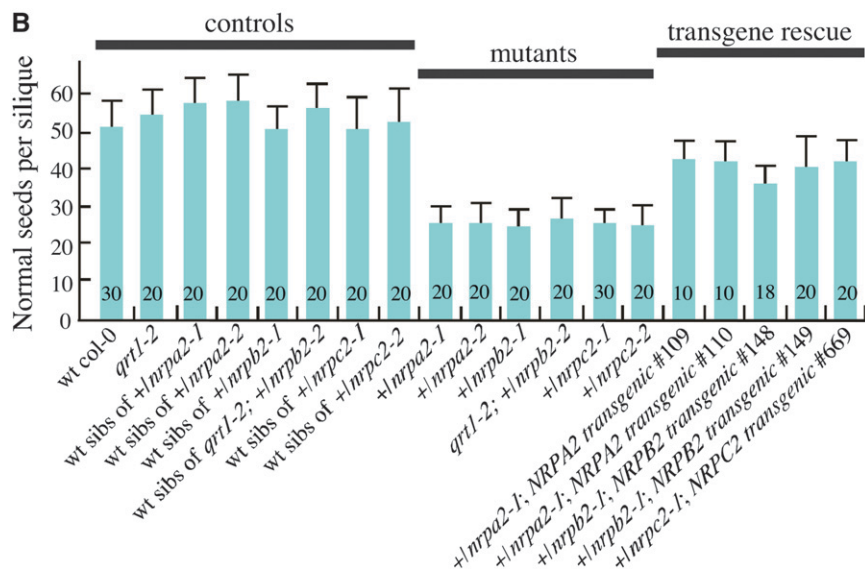


FIGURE 2.—Failed seed development in siliques of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* hemizygotes. (A) A silique of a hemizygous *+/nrpa2-1* plant. Normal seeds and undeveloped (arrested) ovules occur in a silique of a hemizygous plant. (B) Average amounts of normal seeds per silique from wild-type and hemizygous plants. Numbers of siliques examined are indicated.

To investigate the influence of defective RNA polymerase alleles on pollen development and viability using tetrad analysis, we generated lines that carry a Pol I (*nrpa2-1*), Pol II, (*nrpb2-1*), or Pol III (*nrpc2-1*) mutant allele in the *quartet* (*qrt*) mutant background. The *quartet* mutation causes the four pollen that develop from the four meiotic products (microspores) to remain associated with one another, rather than dissociating into individual pollen grains. Thus, pollen tetrads of plants hemizygous for the polymerase mutants include two pollen-bearing mutant polymerase alleles and two bearing wild-type polymerase alleles.

Pollen tetrads were examined by DAPI (4',6-diamidino-2-phenylindole), FDA (fluorescein diacetate), or Alexander staining (Figure 4, A–H). DAPI staining of chromatin in pollen of *quartet* (*qrt1-2*) mutant plants; Pol I hemizygote *quartet* (*+/nrpa2-1; qrt1-2*), Pol II hemizygote *quartet* (*+/nrpb2-1; qrt1-2* as well as *+/nrpb2-2; qrt1-2*), or Pol III hemizygote *quartet* (*+/nrpc2-1; qrt1-2*) plants revealed the normal pattern of one diffuse vegetative cell nucleus and two compact sperm cell nuclei in each of the four attached pollen (Figure 4, B and F, and data not shown). FDA and Alexander staining detected no differences in viability among the individual pollen in tetrads of wild-type or mutant plants (Figure 4, C, D, G, and H, and data not shown).

Two of the pollen in each tetrad of a polymerase mutant hemizygote carry defective RNA polymerase alleles and lack wild-type alleles. In the case of the *nrpb2-2* hemizygotes, the mutant alleles are tagged by a *LAT52::GUS* reporter gene that is present within the T-DNA inserted into the Pol II *NRPB2* gene (Figure 4, I and J). The *LAT52* promoter is specifically expressed in mature pollen and pollen tubes, thereby allowing the pollen bearing the mutant *nrpb2-2* alleles to be visualized by GUS staining. Equal numbers of GUS-positive (blue) and GUS-negative pollen are present in *nrpb2-2/+* pollen quartets, indicating that wild-type and mutant pollen develop in equal abundance and that the *nrpb2-2* mutant allele segregates normally (Figure 4, I and J).

It is noteworthy that mRNA-encoded proteins, such as the GUS enzyme, are synthesized by RNA polymerase II and require the distinctive 5'7-methylguanosine caps and poly A tails of Pol II transcripts to be translated. Pol I and Pol III transcripts lack these features and are not translated. Despite the disruption of the gene encoding the essential Pol II second-largest subunit (NRPB2), the GUS enzyme is clearly expressed from the *LAT52* promoter in *nrpb2* mutant pollen (Figure 4, I and J). Expression of the GUS gene cannot be attributed to stored GUS mRNA transcribed premeiotically; if so, it would be

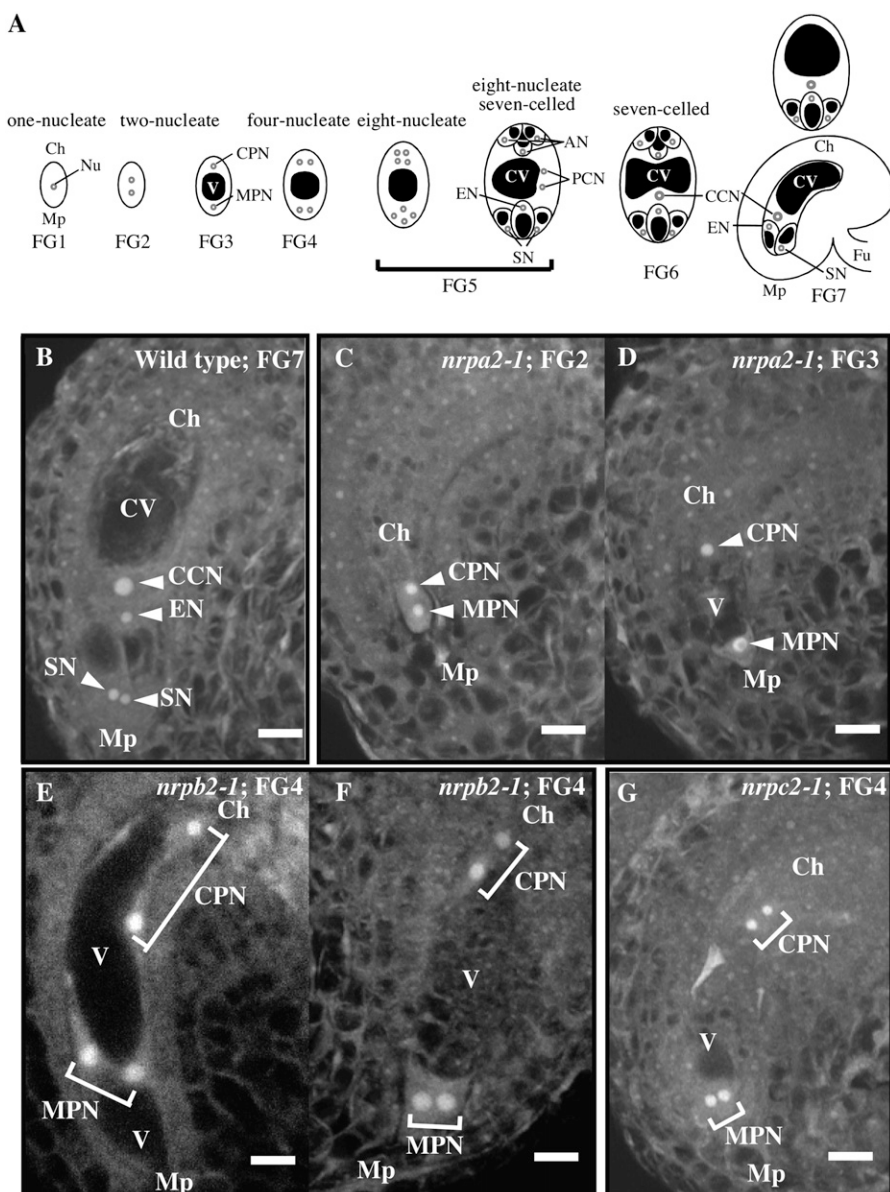


FIGURE 3.—Developmental arrest of mutant female gametophytes in flowers just prior to anthesis was visualized by confocal fluorescence microscopy. (A) Stages of female gametophyte development (FG1–FG7), according to CHRISTENSEN *et al.* (1997). Mp, micropylar pole; Ch, chalazal pole; Nu, nucleus; V, vacuole; CPN, chalazal pole nucleus; MPN, micropylar nucleus; AN, antipodal cell nucleus; CV, central cell vacuole; EN, egg cell nucleus; PCN, polar cell nucleus; CCN, central cell nucleus; SN, synergid cell nucleus; Fu, funiculus. (B) A wild-type female gametophyte, at floral stage 13, that is fully developed (FG7). The nuclei and vacuoles for the 2N central cell, the egg cell, and two synergid cells are apparent. (C and D) *nrpa2-1* female gametophytes arrested at the two-nucleate stage (FG2 and FG3). (E and F) *nrpb2-1* female gametophytes arrested at the four-nucleate stage. (G) A *nrpc2-1* female gametophyte arrested at the four-nucleate stage. Scale bars, 10 μm.

present in all four pollen of the tetrad. Moreover, the *LAT52* promoter has previously been shown to be expressed only postmeiotically, making it a useful male-gametophyte-specific marker (EADY *et al.* 1994; TWELL *et al.* 1990). We conclude that Pol II transcription takes place in *nrpb2-2* mutant pollen despite the lack of a functional *NRPB2* allele.

Examination of pollen germination and pollen tube growth *in vitro* revealed no differences among pollen tubes that grew from pollen quartets consisting of two pollen-bearing defective RNA polymerase alleles and two pollen-bearing wild-type alleles, at least up to a pollen tube length of 100–150 μm (Figure 4, K and L, and data not shown). Self-pollinated pistils of *qrt1-2*; +/- *nrpb2-2* plants stained for GUS also reveal pollen tube growth from pollen bearing the disrupted allele *in vivo* (Figure 4, M and N). Most of the GUS-stained tubes

from *nrpb2-2* pollen are observed at the stigma and upper portions of the ovary (Figure 4, M and N; Figure 5C). However, in rare cases, tubes from *nrpb2-2* pollen are observed in the distal portion of the ovary (Figure 4N, images at top right and bottom). Collectively, these observations suggest that in pollen that do not encode endogenous functional RNA polymerase II, Pol II-dependent GUS activity is sustained during pollen development and early pollen tube growth.

To test the hypothesis that pollen bearing Pol I (*nrpa2-1* or *nrpa2-2*), Pol II (*nrpb2-1* or *nrpb2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) mutant alleles are at a competitive disadvantage compared to wild-type pollen, we determined the distribution of seeds bearing mutant alleles within the siliques of self-pollinated hemizygous plants. Due to the previously demonstrated lethality of the 50% of female gametophytes that inherit a

TABLE 3
Female gametophyte development in polymerase mutants

Plant genotype	Pistil identification no.	No. of female gametophytes at specified developmental stages							Total
		FG1	FG2	FG3	FG4	FG5	FG6	FG7	
wt col-0	1					7	2	8	17
	2					1	2	11	14
	3					3	1	12	16
	4						1	14	15
<i>qrt1-2</i>	1						1	12	13
	2					5	3	9	17
	3							13	13
+/ <i>nrpa2-1</i>	1		4	4	1		2	4	15
	2		2	4	2		1	3	12
	3		4	4	2	7	2	4	23
	4		1	6			1	9	17
+/ <i>nrpa2-2</i>	1		3	6	1	2	1	7	20
	2		2	7	1	4	1	1	16
	3		1	4	1	7	3	1	17
	4		1	6	4	2	1	7	21
+/ <i>nrpb2-1</i>	1			2	8		1	4	15
	2			1	4		2	2	9
	3			3	3	1	1	7	15
	4				9		2	8	19
	5			1	10		2	9	22
	6			4	10			8	22
<i>qrt1-2</i> , +/ <i>nrpb2-2</i>	1			2	6	2	1	4	15
	2			5	8	3	2	3	21
	3			1	3			8	12
+/ <i>nrpc2-1</i>	1			4	12	2	1	7	26
	2			2	7	2	1	5	17
	3			2	3	3	1	7	16
	4			4	4	1	2	6	17
+/ <i>nrpc2-2</i>	1		3	10	1	3	1	6	24
	2		2	10			1	8	21
	3		3	6	1		3	6	19
	4		1	8			1	9	19
	5		2	7	1		1	8	19

Developmentally arrested *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* female gametophytes. Pistils from flowers just prior to anthesis (flower opening) were fixed, and female gametophytes within these pistils were classified according to their developmental stage (FG1–FG7). wt, wild type.

mutant polymerase allele (depicted as ovules with an “X” through them in Figure 5A), only the 50% of female gametophytes that bear wild-type alleles are available to be fertilized. Therefore, any mutant alleles detected in the seeds are inherited via the male gametophytes (refer to Figure 1, C–E, and Table 2). Seeds were collected from the top one-third of the silique, which is nearest to the stigma where the pollen germinates to initiate the growth of pollen tubes, or from the middle or bottom one-third of the silique. Following germination of the seeds, resulting plants were genotyped (Figure 5B). This test revealed that mutant alleles were found most frequently among seeds that developed within the top one-third of the siliques; 35–50% of these seeds develop as hemizygotes (note that a frequency of 50% is ex-

pected if there is no difference in the fitness of wild-type and mutant pollen). The frequency of hemizygous seeds within the middle portions of the siliques were significantly reduced (11–21%) in comparison with the top one-third, except for the *nrpa2-2* allele that was detected in 16 of the 23 sibs examined. In the bottom one-third of the siliques, where fertilization of the ovules would require the growth of the longest pollen tubes, hemizygotes represented only a small proportion of the seeds (0–11%).

The extent of mutant pollen tube growth fits with the distribution of hemizygous seeds following fertilization. A nonmutant transgenic line in which a T-DNA bearing the *LAT52::GUS* reporter gene inserted into an intergenic region was used as a control for comparison to

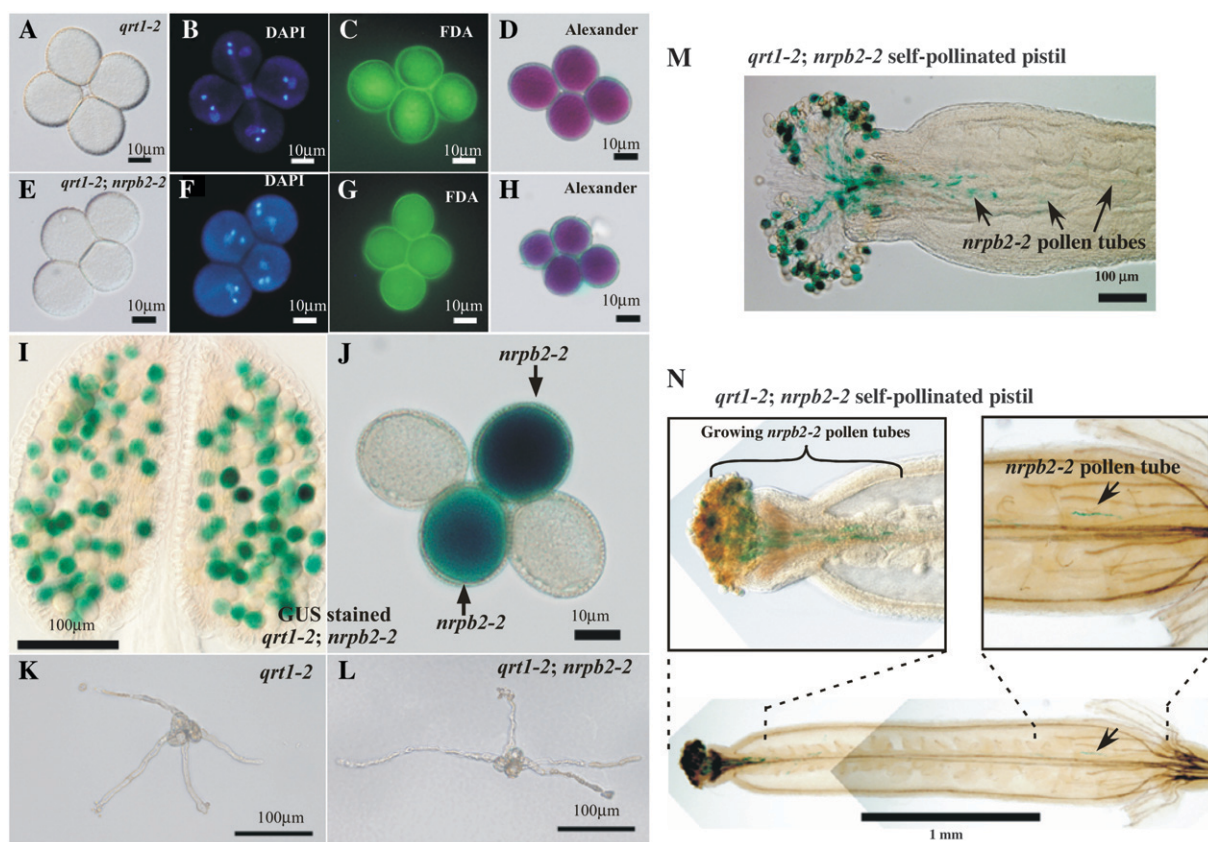


FIGURE 4.—Development and early tube elongation of pollen are unaffected by defects in RNA polymerases. (A–H) Cytological examination of mature pollen from *qrt1-2* (a–d) and *qrt1-2; nrpb2-2* (e–h). (a and e) Bright-field microscopy; (b and f) DAPI staining test; (c and g) FDA staining test; (d and h) Alexander staining test. (I and J) *LAT52::GUS* expression in pollen defective for the Pol II subunit (*nrpb2-2* pollen). (K and L) Germinating *qrt1-2* (k) and *qrt1-2; nrpb2-2* (l) pollen. Pollen was incubated for 18 hr at 22 ° in a germination medium and its images were captured. Note that four tubes of quartet pollen from wild-type (k, *qrt1-2*) and mutant (l, *qrt1-2; nrpb2-2*) plants grew equally in this assay, to a length of ~100–150 mm. (M and N) Self-pollinated pistils from *qrt1-2; nrpb2-2* plants. *LAT52::GUS* was expressed during pollen tube growth in the absence of the functional allele of a catalytic subunit of Pol II. A considerable number of *nrpb2-2* pollen tubes (blue stained) was present in the top portions of the pistils. Note that a tube from *nrpb2-2* pollen grew into ~2.0 mm in length (N).

nrpb2-2. Whereas *GUS*-stained *nrpb2-2* pollen tubes are rarely observed deeper than the top one-third of the pistil, *GUS*-stained control pollen tubes are easily detected throughout the top and middle one-thirds of the pistils and can be observed all the way to the base of the pistil (Figure 5C). Taken together, our results suggest that pollen germination, early pollen tube elongation, and fertilization are not severely affected by the lack of functional alleles for the RNA polymerase I, II, or III subunits. However, sustained pollen tube growth presumably requires *de novo* synthesis of essential RNA polymerase genes such that mutant pollen are at a competitive disadvantage compared to wild-type pollen, the phenomenon known as certation (HERIBERT-NILSSON 1920).

DISCUSSION

Genetic analyses have identified a large number of female gametophytic mutants in *Arabidopsis*, a signifi-

cant fraction of which correspond to mutant alleles of transcription factors (PAGNUSSAT *et al.* 2005). Our demonstration that mutations in RNA polymerases I, II, and III cause female gametophyte lethality are generally consistent with these findings and indicate that the female gametophyte is dependent on endogenous transcription machinery synthesized *de novo* during gametophyte development. In the absence of functional RNA polymerase subunits, female gametophytes can often progress to the two-nucleate stage, but typically arrest before, or shortly after, the second of the three mitotic divisions required for development of mature gametophytes. It is noteworthy that the SeedGenes Project database (<http://www.seedgenes.org/index.html>) (TZAFRIR *et al.* 2003, 2004) includes information for two T-DNA insertion alleles of *nrpb2*, named *emb 1989-1* and *emb 1989-2*. Embryos fail to develop in 90–94% of ovules bearing these mutant alleles, consistent with the female gametophytic lethal phenotype we describe in this article. However, 6–10% of *emb 1989-1* and *emb 1989-2* ovules are reported to arrest as preglobular embryos, indicating that

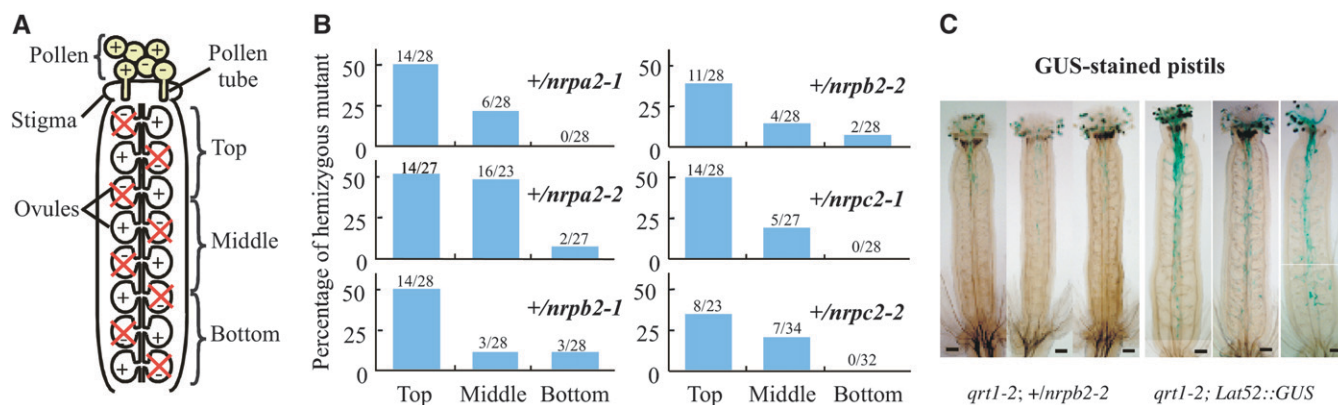


FIGURE 5.—Reduced paternal transmission of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles relative to wild-type alleles in self-fertilized hemizygotes is due to decreased, competitive fertilization of ovules farthest from the stigma. (A) A diagram of the female floral organ (the pistil), whose surface (the stigma) is the site where a pollen grain germinates and initiates formation of a pollen tube. Half of the pollen of a hemizygote has wild-type (+) RNA polymerase alleles and half are mutant (−), but all develop and mature. Likewise, within the ovary of a hemizygote, half of the ovules are wild type and half are mutant with respect to the RNA polymerase alleles. However, the latter fail to develop (denoted with an “X”) such that mutant alleles in fertilized ovules and seeds are derived from the male gametophyte. (B) Seeds collected from the top, middle, and bottom portions of siliques of the hemizygotes were germinated and resultant plants were genotyped. The numbers of plants of each genotype are indicated. Note that mutant alleles are more abundant in seeds developing nearest the stigma, at the top of the siliques, where the shortest pollen tubes would be needed to reach the ovules. (C) Self-pollinated pistils from *qrt1-2; +nrpb2-2* plants and transformants hemizygously carrying a *LAT52::GUS* reporter gene(s) inserted in an intergenic region (*qrt1-2; LAT52::GUS*). Pollen tubes from *qrt1-2; nrpb2-2* pollen (blue stained) were present in top portions of the pistils, while control pollen tubes (*qrt1-2; LAT52::GUS*) were observed all the way from the tops to the bottoms of the pistils.

the female gametophytes in these cases had completed development and had been fertilized, but produced embryos that were then unable to complete development. Cloning and sequencing of the region that defines the junction between the *NRPB2* gene and the T-DNA revealed that the T-DNA in *emb 1989-1* inserted 34 nucleotides upstream from the translation start site (Y. ONODERA, data not shown). Because the protein coding region is not disrupted, it is possible that the *emb 1989-1* allele is partially functional, which may explain how development can sometimes proceed to stages beyond what we have observed for the *nrpb2-1* and *nrpb2-2* alleles. We currently lack analogous data concerning the precise location of the T-DNA in the *emb 1989-2* allele.

A recent study of developing and mature pollen showed that 61.9% of all Arabidopsis genes are expressed during at least one stage of male gametophyte development, with 9.7% of the transcripts being pollen specific (HONYS and TWELL 2004). A large number of transcription factors are expressed during pollen development, suggesting that orchestrated waves of transcription are essential for pollen maturation. Mature pollen is also known to contain proteins, ribosomes, mRNAs, rRNAs, and tRNAs that are synthesized post-meiotically during pollen maturation or pollen tube growth (MASCARENHAS 1975, 1989). Therefore, we were surprised to find that functional alleles of RNA polymerases I, II, and III are not absolutely required in the haploid pollen genome to complete pollen development, germination, pollen tube growth, or fertilization. The simplest explanation is that transcription in pollen-

bearing defective polymerase alleles is conducted using RNA polymerases, or stored mRNAs encoding RNA polymerase subunits, that are synthesized premeiotically in the hemizygous microspore mother cell and are then partitioned into the microspores following meiosis. The one functional allele is apparently sufficient for microspore mother cells to load microspores with enough polymerase to support subsequent pollen development and postgermination pollen functions, including pollen tube growth and fertilization.

Transcript profiling using DNA microarray technology has shown that mRNAs encoding the core subunits for nuclear RNA polymerases are present within unicellular microspores at similar or greater abundance than in sporophytic tissues (HONYS and TWELL 2004). However, in mature pollen, mRNAs encoding transcription factors, RNA processing proteins, and translation machineries are less abundant than in vegetative tissues of the plant (HONYS and TWELL 2003; PINA *et al.* 2005; GRENNAN 2007). This holds true for transcripts encoding the core subunits for nuclear RNA polymerases I, II, and III, which either are not detected in mature pollen or are present at very low levels (HONYS and TWELL 2003; PINA *et al.* 2005). The idea that maternally derived polymerase subunit mRNAs are stored for translation late in pollen development is not readily supported by these observations, but the possibility cannot be ruled out. An alternative hypothesis is that polymerase proteins derived from the microspore mother cell, or translated from mRNAs partitioned into the unicellular microspores, persist in mature pollen. Plants hemizygous for a

single-copy transgene expressing a polymerase subunit-GFP fusion protein would be useful for testing this hypothesis. If the transgene were capable of rescuing plants that were homozygous for null alleles of the corresponding endogenous genes, one would expect the GFP marker to segregate 2:2 among the pollen. If GFP were observed in all pollen, this would indicate maternal loading of the polymerase subunit. Regardless of whether stored mRNA or stored protein is responsible for allowing the transmission of mutant polymerase alleles through the pollen, there are enough of the stored molecules to complete pollen development, germination, and fertilization. These developmental events are thought to span a period of at least 90 hr (BOWMAN 1994). However, additional *de novo* synthesis of Pol I, II, and III is apparently needed for full pollen vigor and for growth of pollen tubes long enough to reach the ovules farthest from the stigma.

Given the reduced fitness of mutant pollen relative to wild-type pollen, deleterious mutant polymerase alleles are unlikely to become widespread among a population. However, some gene evolution phenomena would seem to be favored by allowing mutant alleles to persist in the population for some period of time. For instance, a characteristic of the RNA polymerase I transcription system is that it evolves rapidly, such that the transcription machinery of one species cannot transcribe the rRNA genes of an unrelated species (GRUMMT *et al.* 1982; MIESFELD and ARNHEIM 1984; DOELLING and PIKAARD 1996). Species specificity appears to be explained by the rapid evolution of rRNA gene sequences and the corresponding coevolution of the transcription machinery, such that changes in gene sequences can be tolerated as a result of compensatory changes in the proteins that bind these sequences (or vice versa). Because haploid selection against defective alleles is less stringent in the male gametophyte than in the female gametophyte, at least for subunits of RNA polymerases I, II, and III, it is tempting to speculate that the male lineage could be the conduit for transmitting mutations that might initially be deleterious but could be tolerated if a compensatory mutation in an interacting protein or DNA sequence were to occur. Transmitting mutations at moderate frequency via the pollen would presumably buy time for such compensatory mutations to occur. However, the null hypothesis is that the capacity to transmit mutations in essential housekeeping genes such as RNA polymerases via pollen has no evolutionary advantage and is merely an unintended consequence of pollen development.

Research at Hokkaido University was supported in part by a grant in aid for scientific research from the Ministry of Education, Science and Culture and grants from the Akiyama Foundation and Grant Program of The Sumitomo Foundation. Pikaard lab research was supported by National Institutes of Health grants GM-060380 and GM-077590. The content of this article is solely the responsibility of the authors and does not necessarily reflect the views of the funding agencies that supported the work.

LITERATURE CITED

- ALONSO, J. M., A. N. STEPANOVA, T. J. LEISSE, C. J. KIM, H. CHEN *et al.*, 2003 Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**: 653–657.
- ARABIDOPSIS GENOME INITIATIVE, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- BOWMAN, J. (Editor), 1994 *Arabidopsis: An Atlas of Morphology and Development*. Springer-Verlag, New York.
- CHRISTENSEN, C., E. KING, J. R. JORDAN and G. N. DREWS, 1997 Megagametogenesis in *Arabidopsis* wild type and the Gf mutant. *Sex. Plant Reprod.* **10**: 49–64.
- DOELLING, J. H., and C. S. PIKAARD, 1996 Species-specificity of rRNA gene transcription in plants manifested as a switch in polymerase-specificity. *Nucleic Acids Res.* **24**: 4725–4732.
- DREWS, G., and R. YADEGARI, 2002 Development and function of the angiosperm female gametophyte. *Annu. Rev. Genet.* **36**: 99–124.
- EADY, C., K. LINDSEY and D. TWELL, 1994 Differential activation and conserved vegetative cell-specific activity of a late pollen promoter in species with bicellular and tricellular pollen. *Plant J.* **5**: 543–550.
- EARLEY, K., J. HAAG, O. PONTES, K. OPPER, T. JUEHNE *et al.*, 2006 Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**: 616–629.
- GRENNAN, A., 2007 An analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* **145**: 3–4.
- GROSSNIKLAUS, U., and K. SCHNEITZ, 1998 The molecular and genetic basis of ovule and megagametophyte development. *Semin. Cell Dev. Biol.* **9**: 227–238.
- GRUMMT, I., E. ROTH and M. R. PAULE, 1982 rRNA transcription *in vitro* is species-specific. *Nature* **296**: 173–174.
- HASHIDA, S., H. TAKAHASHI, M. KAWAI-YAMADA and H. UCHIMIYA, 2007 *Arabidopsis thaliana* nicotinate/nicotinamide mononucleotide adenylyltransferase (AtNMNAT) is required for pollen tube growth. *Plant J.* **49**: 694–703.
- HERIBERT-NILSSON, N., 1920 Zuwachsgeschwindigkeit der Pollenschläuche und gestörte Mendelzahlen bei *Oenothera Lamarckiana*. *Hereditas* **1**: 41–67.
- HIGASHIYAMA, T., 2002 The synergid cell: attractor and acceptor of the pollen tube for double fertilization. *J. Plant Res.* **115**: 149–160.
- HIGASHIYAMA, T., S. YABE, N. SASAKI, Y. NISHIMURA, S. MIYAGISHIMA *et al.*, 2001 Pollen tube attraction by the synergid cell. *Science* **293**: 1480–1483.
- HIGASHIYAMA, T., H. KUROIWA and T. KUROIWA, 2003 Pollen-tube guidance: beacons from the female gametophyte. *Curr. Opin. Plant Biol.* **6**: 36–41.
- HONYS, D., and D. TWELL, 2003 Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* **132**: 640–652.
- HONYS, D., and D. TWELL, 2004 Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol.* **5**: R85.
- JOHNSON, M., and D. PREUSS, 2002 Plotting a course: multiple signals guide pollen tubes to their targets. *Dev. Cell* **2**: 273–281.
- JOHNSTON, A., P. MEIER, J. GHEYSSELINCK, S. WUEST, M. FEDERER *et al.*, 2007 Genetic subtraction profiling identifies genes essential for *Arabidopsis* reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. *Genome Biol.* **8**: R204.
- JONES-RHOADES, M., J. BOREVITZ and D. PREUSS, 2007 Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins. *PLoS Genet.* **3**: 1848–1861.
- KASSAVETIS, G. A., C. BARDELEBEN, B. BARTHOLOMEW, B. R. BRAUN, C. A. P. JOAZEIRO *et al.*, 1994 Transcription by RNA polymerase III, pp. 107–126 in *Transcription Mechanisms and Regulation*, edited by R. C. CONAWAY and J. W. CONAWAY. Raven Press, New York.
- LARKIN, R., and T. GUILFOYLE, 1993 The second largest subunit of RNA polymerase II from *Arabidopsis thaliana*. *Nucleic Acids Res.* **21**: 1038.
- MASCARENHAS, J. P., 1975 The Biochemistry of angiosperm pollen development. *Bot. Rev.* **41**: 259–314.
- MASCARENHAS, J. P., 1989 The Male gametophyte of flowering plants. *Plant Cell* **1**: 657–664.

- MASCARENHAS, J. P., 1993 Molecular mechanisms of pollen tube growth and differentiation. *Plant Cell* **5**: 1303–1314.
- MCCORMICK, S., 1993 Male gametophyte development. *Plant Cell* **5**: 1265–1275.
- MCCORMICK, S., 2004 Control of male gametophyte development. *Plant Cell* **16**: S142–S153.
- MIESFELD, R., and N. ARNHEIM, 1984 Species-specific rDNA transcription is due to promoter-specific binding factors. *Mol. Cell. Biol.* **4**: 221–227.
- ONODERA, Y., J. R. HAAG, T. REAM, P. C. NUNES, O. PONTES *et al.*, 2005 Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**: 613–622.
- PAGNUSSAT, G., H. YU, Q. NGO, S. RAJANI, S. MAYALAGU *et al.*, 2005 Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development* **132**: 603–614.
- PAULE, M. R., and R. J. WHITE, 2000 Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.* **28**: 1283–1298.
- PINA, C., F. PINTO, J. FEIJÓ and J. BECKER, 2005 Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiol.* **138**: 744–756.
- PREUSS, D., S. RHEE and R. DAVIS, 1994 Tetrad analysis possible in *Arabidopsis* with mutation of the QUARTET (QRT) genes. *Science* **264**: 1458–1460.
- ROSSO, M., Y. LI, N. STRIZHOV, B. REISS, K. DEKKER *et al.*, 2003 An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol. Biol.* **53**: 247–259.
- RUSSELL, S., 1993 The egg cell: development and role in fertilization and early embryogenesis. *Plant Cell* **5**: 1349–1359.
- SCHNEITZ, K., M. HÜLSKAMP and R. E. PRUITT, 1995 Wild-type ovule development in *Arabidopsis thaliana*: a light microscope study of cleared whole-mount tissue. *Plant J.* **7**: 731–749.
- SPRUNCK, S., U. BAUMANN, K. EDWARDS, P. LANGRIDGE and T. DRESSELHAUS, 2005 The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). *Plant J.* **41**: 660–672.
- STEFFEN, J., I. KANG, J. MACFARLANE and G. DREWS, 2007 Identification of genes expressed in the *Arabidopsis* female gametophyte. *Plant J.* **51**: 281–292.
- TWELL, D., J. YAMAGUCHI and S. MCCORMICK, 1990 Pollen-specific gene-expression in transgenic plants: coordinate regulation of 2 different tomato gene promoters during microsporogenesis. *Development* **109**: 705.
- TZAFRII, I., A. DICKERMAN, O. BRAZHNİK, Q. NGUYEN, J. McELVER *et al.*, 2003 The *Arabidopsis* SeedGenes Project. *Nucleic Acids Res.* **31**: 90–93.
- TZAFRII, I., R. PENA-MURALLA, A. DICKERMAN, M. BERG, R. ROGERS *et al.*, 2004 Identification of genes required for embryo development in *Arabidopsis*. *Plant Physiol.* **135**: 1206–1220.
- YADEGARI, R., T. KINOSHITA, O. LOTAN, G. COHEN, A. KATZ *et al.*, 2000 Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. *Plant Cell* **12**: 2367–2382.
- YANG, H., N. KAUR, S. KIRIAKOPOLOS and S. MCCORMICK, 2006 EST generation and analyses towards identifying female gametophyte-specific genes in *Zea mays* L. *Planta* **224**: 1004–1014.
- YU, H., P. HOGAN and V. SUNDARESAN, 2005 Analysis of the female gametophyte transcriptome of *Arabidopsis* by comparative expression profiling. *Plant Physiol.* **139**: 1853–1869.

Communicating editor: J. A. BIRCHLER

Supplemental data

Table S1 methods:

NRBP12a (At5g41010) T-DNA insertion line SALK_049327 was obtained from ABRC. DNA was extracted from 1-3 leaves in microcentrifuge tubes using a modified version of a previously published protocol (2). Briefly, leaves were incubated 10 min in 300 ul of extraction buffer (200 mM Tris, pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) at 99°C. Cell debris was cleared by centrifugation at 14,000 x g, 8 min. The supernatant was transferred to a new tube containing an equal volume of isopropanol, mixed and incubated at room temperature for fifteen minutes. DNA was pelleted by centrifugation at 14,000 x g, 15 min. Pellets were washed once in 70% ethanol before resuspending in 100 ul of 1x TE buffer, pH 8.0. Debris was pelleted by centrifuging one minute at top speed in a microcentrifuge. 2 ul of DNA was used in a 20 ul PCR reaction with GoTaqGreen (Promega) and appropriate primers. The wild-type *NRBP12a* gene was amplified using forward primer 5'-TTATAGCCAATCAAGGATTATAGCAATGTGAAC-3' and reverse primer 5'-GAAATCAAAGTTTTGTTAGTATCTGTAAAAGATTG-3'. The T-DNA inserted allele was detected using the reverse primer above in combination with the SALK line T-DNA Left border primer, LBa1: 5'-TGGTTCACGTAGTGGGCCATCG-3'.

Figure S1. Developmentally arrested mutant female gametophytes within pistils just prior to anthesis, visualized by confocal fluorescence microscopy. **(A and B)** *nrapa2-2* female gametophytes arrested at the two-nucleate stage **(C)** A *nrbp2-2* female gametophyte arrested at the two-nucleate stage **(D)** A *nrbp2-2* female gametophyte

arrested at the four-nucleate stage. (E) A *nrpc2-2* female gametophyte arrested at the two-nucleate stage and displaying a prominent vacuole. Abbreviations: Mp, micropylar pole; Ch, chalazal pole; CPN, chalazal pole nucleus; MPN, micropylar nucleus; V, vacuole; Nu, nucleus. Scale bars = 10 μ m.

References for Supplemental Data

1. Alonso, J., Stepanova, A., Leisse, T., Kim, C., Chen, H., Shinn, P., Stevenson, D., Zimmerman, J., Barajas, P., Cheuk, R. et al. (2003). Genome-wide Insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653-657.
2. Herr, A., Molnar, A., Jones, A. and Baulcombe, D. (2006). Defective RNA processing enhances RNA silencing and influences flowering of *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 103, 14994-15001.

Table S1. Male-specific transmission of *RPB12a* mutant alleles.

Parental genotype		Genotypes of progeny	
Female parent	Male parent	homozygous wt (+/+)	hemizygous (+/-)
+/ <i>nrbp12a</i>	+/+	100% (20/20)	0% (0/20)
+/+	+/ <i>nrbp12a</i>	40% (24/60)	60% (36/60)

Table S2. Transgene rescue allows maternal transmission of mutant alleles

Parental genotype		Genotypes of progeny	
Female parent	Male parent	Homozygous wt (+/+)	Hemizygous (+/-)
<i>+nrpa2-1, NRPA2</i> transgenic #109	+/+	67% (12/18)	33% (6/18)
<i>+nrpa2-1, NRPA2</i> transgenic #110	+/+	62% (16/26)	38% (10/26)
<i>+nrpb2-1, NRPB2</i> transgenic #148	+/+	50% (9/18)	50% (9/18)
<i>+nrpb2-1, NRPB2</i> transgenic #149	+/+	55% (11/20)	45% (9/20)
<i>+nrpc2-1, NRPC2</i> transgenic #669	+/+	74% (31/42)	26% (11/42)

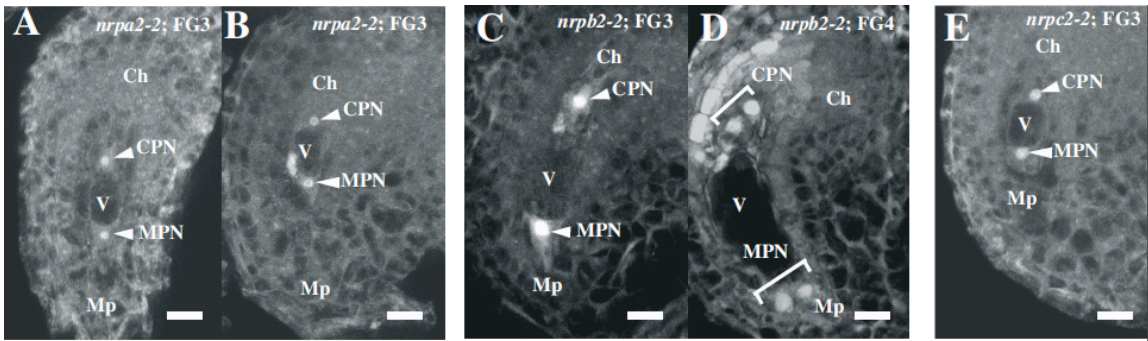


Figure S1

APPENDIX G

RNA POLYMERASE I: A MULTIFUNCTIONAL MOLECULAR MACHINE

A review published in *Cell* (2007), 137 (7): 1224-1225.

My contributions to this work:

Craig Pikaard and I reviewed and were later asked to write a *Cell* preview article for a research article out of Patrick Cramer's lab describing the structure and functional architecture of yeast RNA Polymerase I (Kuhn et al, 2007). I wrote the initial draft of the preview article and helped brainstorm the figure content.

RNA Polymerase I: A Multifunctional Molecular Machine

Jeremy R. Haag¹ and Craig S. Pikaard^{1,*}

¹Department of Biology, Washington University, 1 Brookings Drive, St. Louis, MO, USA

*Correspondence: pikaard@biology2.wustl.edu

DOI 10.1016/j.cell.2007.12.005

In this issue, Kuhn et al. (2007) report the complete structure of the 14-subunit yeast RNA polymerase (Pol) I enzyme at 12 Å resolution using cryo-electron microscopy (cryo-EM). Their study reveals that three subunits of Pol I perform functions in transcription elongation that are outsourced to the transcription factors TFIIF and TFIIIS in the analogous Pol II transcription system.

Bacteria and Archaea decode their genomes using a single DNA-dependent RNA polymerase, whereas eukaryotes have evolved at least three (Pol I, II, and III, plus IVa and IVb in plants). Furthermore, whereas the RNA polymerase of *Escherichia coli* is composed of only four different proteins, yeast RNA Pol I, II, and III are far more complicated, consisting of 14, 12, and 17 subunits, respectively (Werner, 2007). Among these are subunits that are orthologous to the bacterial polymerase subunits. Five additional subunits of Pol I, II, and III are identical and are encoded by the same genes. The remaining subunits are unique to Pol I, Pol II, or Pol III and are thought to mediate their distinct functions: Pol II mostly transcribes protein-coding genes and regulatory RNA genes (Hahn, 2004); Pol I transcribes genes encoding the 18S, 5.8S, and 25–28S rRNAs that form the catalytic core of ribosomes (White, 2005); Pol III primarily transcribes tRNA genes and 5S rRNA genes (White, 2005); and in plants, Pol IVa and Pol IVb function in a pathway generating short-interfering RNAs that direct DNA methylation (Pikaard, 2006).

Understanding the functions of the various eukaryotic polymerase subunits is a major challenge in which structural biology is playing a critical role. The high resolution (2.8–3.3 Å) crystal structures of bacterial RNA polymerase and yeast RNA Pol II (Cramer et al., 2001; Gnatt et al.,

2001; Zhang et al., 1999) revealed a remarkable conservation of structure at the core of these enzymes. Now, Kuhn et al. (2007) provide the most detailed and complete view of the Pol I enzyme to date. By combining structural analyses with manipulations of subunit compositions and biochemical assays, their study is a tour-de-force that reveals functions conserved among Pol I, II, and III as well as aspects of Pol I functional specialization.

As the starting point for their current work, Kuhn et al. (2007) derived a cryo-EM density map based on the analysis of ~40,000 purified Pol I

molecules and looked for correspondence between the density map and the Pol II crystal structure (Cramer et al., 2001). The Pol II structure fit perfectly onto the Pol I EM density map in the regions corresponding to the five subunits that are common to Pol I, II, and III. Highly conserved domains within paralogous catalytic subunits also fit nicely, including the active center and bridge helix that spans the template cleft. Interestingly, some domains of Pol II that lack obvious Pol I counterparts based on sequence comparisons, such as the jaw and lobe domains, are nonetheless apparent in the Pol I

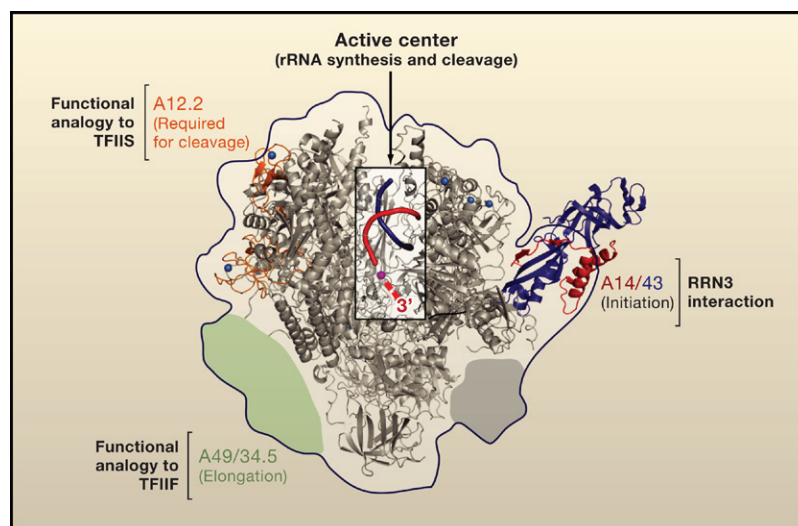


Figure 1. RNA Polymerase I

Annotated overview of the 12 Å RNA polymerase I structure highlighting the positions of functional subdomains. Figure adapted from Kuhn et al. (2007).

structure and presumably carry out analogous functions—a hypothesis that can now be tested based on the structural insight.

Regions displaying distinct structural variation between Pol I and Pol II are candidates for polymerase-specific functions. One such region of Pol I includes the A14/A43 subunit heterodimer, which has weak homology to the Rpb4/Rpb7 and C17/C25 heterodimers of Pol II and Pol III, respectively, but insufficient similarity to allow homology modeling based on the Pol II crystal structure. Kuhn et al. (2007) determined the crystal structure of the A14/A43 heterodimer at 3.1 Å resolution and fit the structure unambiguously into the EM density map. A43, in turn, is known to interact with Rrn3 (TIF-IA in mammals), an essential transcription factor that regulates Pol I activity in response to growth status and the cellular need for ribosomes and protein synthesis (Peyroche et al., 2000). Collectively, the new structural data indicate that Rrn3 interacts with Pol I on an upstream surface relative to the direction of transcription (Figure 1), an important new piece of the puzzle for understanding Pol I transcriptional activation.

One of the most interesting aspects of the study by Kuhn et al. (2007) involves the function of the Pol I-specific subunits A49 and A34.5. By determining the cryo-EM structures of Pol I with or without these subunits, the precise position of the A49/34.5 subcomplex was defined. The authors recognized that the A49 and A34.5 subunits have weak sequence and structural homology to the RAP74 and RAP30 subunits of transcription factor TFIIF, a factor needed for Pol II promoter clearance and transcript elongation. Indeed, data from *in vitro* and *in vivo* assays indicate that Pol I lacking the A49 and A34.5 subunits has impaired

transcription elongation activity that can be rescued by exogenously supplied A49/34.5 heterodimers. Collectively, the data suggest that the A49/34.5 subcomplex fulfills an elongation function accomplished by TFIIF in the context of Pol II transcription (Figure 1). The authors further suggest that the weakly homologous C37/C53 subcomplex is likely to carry out this same function in Pol III. Interestingly, RAP30 and RAP74 got their names as RNA polymerase II-associating proteins (Sopta et al., 1985). The fact that these proteins do not stably associate with Pol II, unlike the functionally analogous Pol I and Pol III subunits, provides one potential explanation for why Pol II has fewer subunits than Pol I and Pol III.

An important biochemical insight provided by Kuhn et al. (2007) is that Pol I has a strong 3'-end RNA cleavage activity *in vitro*. A similar RNA cleavage activity for Pol III is attributable to the C11 subunit, which shares sequence similarity with the Pol I subunit A12.2 (Figure 1). Indeed, Pol I missing the C-terminal domain of A12.2 is unable to cleave RNA. This domain also shows homology to TFIIS, a Pol II elongation factor that works with the Rpb9 subunit to stimulate RNA cleavage when Pol II encounters a roadblock to elongation and backtracks to extricate itself, yielding a 3' end that can be elongated in a second attempt to read through the problematic region. Ribosomal RNA gene primary transcripts are approximately 5 kb, so a similar activity may be necessary for Pol I to maintain its processivity. Importantly, the A12.2 subunit is required for Pol I termination (Prescott et al., 2004), suggesting that RNA cleavage may be part of the Pol I termination process as is the case for Pol II termination following the cutting of nascent Pol II transcripts at Poly(A)

cleavage sites. A third potential role of the RNA cleavage activity is in the proofreading of nascent transcripts and correction of misincorporated nucleotides in order to prevent non-functional or potentially deleterious RNAs from being incorporated into ribosomes.

The paper by Kuhn et al. (2007) is yet another clear example of how structure can illuminate function, and no doubt numerous follow-up studies will be spurred by their observations and speculations. Breakthrough papers always provide food for thought, and Kuhn, Cramer, and their colleagues have served up a feast with this exciting new study.

REFERENCES

- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). *Science* 292, 1863–1876.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). *Science* 292, 1876–1882.
- Hahn, S. (2004). *Nat. Struct. Mol. Biol.* 11, 394–403.
- Kuhn, C.-D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. (2007). *Cell*, this issue.
- Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C., and Riva, M. (2000). *EMBO J.* 19, 5473–5482.
- Pikaard, C.S. (2006). *Cold Spring Harb. Symp. Quant. Biol.* 71, 473–480.
- Prescott, E.M., Osheim, Y.N., Jones, H.S., Alen, C.M., Roan, J.G., Reeder, R.H., Beyer, A.L., and Proudfoot, N.J. (2004). *Proc. Natl. Acad. Sci. USA* 101, 6068–6073.
- Sopta, M., Carthew, R.W., and Greenblatt, J. (1985). *J. Biol. Chem.* 260, 10353–10360.
- Werner, F. (2007). *Mol. Microbiol.* 65, 1395–1404.
- White, R.J. (2005). *Nat. Rev. Mol. Cell Biol.* 6, 69–78.
- Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). *Cell* 98, 811–824.

APPENDIX H
CURRICULUM VITAE

JEREMY R. HAAG

160 Willow Court
Bloomington, IN 47401
haagj@indiana.edu
314-482-9181 (cell)
812-855-2659 (work)
812-855-6082 (fax)
Date of birth: February 18, 1979

Washington University in St. Louis
Plant Biology Graduate Program
Division of Biology and Biomedical Science
Campus Box 1137, 1 Brookings Drive
St. Louis, MO 63130

EDUCATION

Ph.D. **Washington University in St. Louis**, St. Louis, MO
Plant Biology, August 31, 2009

B.S. **Texas A&M University**, College Station, TX
Genetics (Cum Laude), December 2001

Ph.D. ABSTRACT

RNA Polymerases IV and V (Pol IV and Pol V) are plant-specific enzyme complexes with subunit homology to RNA Polymerase II (Pol II). The largest subunits in Pol IV and Pol V, NRPD1 and NRPE1 respectively, share a second largest subunit, NRPD2/NRPE2. The evolutionarily conserved Metal A and Metal B binding sites are required for Pol IV and V *in vivo* function fitting the prediction that these are functional polymerases. The Defective Chloroplast and Leaves-like (DeCL) domain at the C-terminus of both NRPD1 and NRPE1 is also required for complementation but other domains in the NRPE1 CTD are largely dispensable. Biochemical analysis reveals Pol IV to be a DNA-dependent RNA polymerase capable of producing RNA from a tripartite template that mimics an open transcription bubble. The Metal A binding site is required for Pol IV *in vitro* transcription while the enzyme is resistant to alpha-amanitin, a potent Pol II inhibitor. Pol IV has also been found to bind and co-localize with RNA-dependent RNA Polymerase 2 (RDR2) *in vivo* providing an explanation for how Pol IV RNA products are channeled specifically to RDR2 for the production of double-stranded RNA, which in turn are substrates for dicer cleavage and small-interfering RNA production. Biochemical analysis has also revealed that RDR2 is capable of transcribing both RNA and single-stranded DNA *in vitro*, consistent with previously analyzed RNA-dependent RNA polymerases from plants and other organisms.

RESEARCH EXPERIENCE

2002-Present **Ph.D. Thesis.** "Genetic and Biochemical Properties of *Arabidopsis* RNA Polymerases IV and V" supervised by Craig Pikaard. Identified Pol IV *in*

in vitro activity defining it as a DNA-dependent RNA polymerase. Also identified the CTD requirements of NRPD1 and NRPE1 proteins.

- Fall 2003 **Ph.D. Rotation.** Sigma-Aldrich (St. Louis), Plant Biotechnology Research & Development Team supervised by Keming Song. Sigma's plate-based high throughput co-IP system was tested for the analysis of plant protein-protein interactions and a series of epitope tags were tested for plant protein expression, purification and detection.
- 2002 **Research Assistant.** Texas A&M University (College Station, TX) supervised by Rodolfo Aramayo. Constructed a series of *his-3* integration vectors for the transformation of *Neurospora crassa*.
- Summer 2001 **Internship.** Dow AgroSciences (Indianapolis, IN), Trait Development Group supervised by Xueyi Hu and Siva Kumpatla. Performed extensive genotyping on a recombinant-inbred mapping population and constructed a SSR molecular marker linkage map for *Helianthus annuus* (sunflower).

RESEARCH SKILLS

Proficient in independent experimental design.

Experience with standard molecular biology techniques (PCR, RT-PCR, site-directed mutagenesis, cloning, bacteria transformation, DNA sequencing, Southern, Westerns, small RNA Northern, etc.).

Knowledgeable in the art of antigen design (peptides and recombinant proteins), antibody production, and affinity purification.

Extensive experience with protein purification from plants and bacteria for protein analysis and *in vitro* activity assays.

Skilled at transforming and screening *Arabidopsis thaliana* plants for the expression of epitope-tagged proteins and *in vivo* complementation assays.

Experience with basic phylogenetic analysis, gene mining and protein domain identification.

TEACHING

- Spring 2004 **Teaching Assistant.** Principles of Biology III (Bio3050). A lecture-based course focused on biochemistry and physiology. Led three discussion sections for the biochemistry module of this course, offered one-on-one help during advising hours, and assisted in weekly quiz design and grading.

PUBLICATIONS

1. Lee, D.W., **Haag, J.R.**, and Aramayo, R. (2003) Construction of strains for rapid homokaryon purification after integration of constructs at the histidine-3 (*his-3*) locus of *Neurospora crassa*. *Current Genetics* 43: 17-23.
2. Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., **Haag, J.R.**, Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L., and Richards, E.J. (2003) Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* 163: 1109-22.
3. **Haag, J.R.**, Lee, D.W., and Aramayo, R. (2003) A GATEWAY destination vector for high-throughput construction of *Neurospora crassa* histidine-3 gene replacement plasmids. *Fungal Genetics Newsletter* 50: 6-8.
4. Onodera, Y.*, **Haag, J.R.***, Ream, T.*, Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005) Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation. *Cell* 120: 613-622. * co-first authors
5. Earley, K.W., **Haag, J.R.**, Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant Journal* 45: 616-629.
6. Pontes, O., Li, C.F., Nunes, P.C., **Haag, J.**, Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006) The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. *Cell* 126: 79-92.
7. **Haag, J.** and Pikaard, C.S. (2007) RNA Polymerase I: a multifunctional molecular machine. *Cell* 131: 1224-1225 (Preview)
8. Pikaard, C.S., **Haag, J.R.**, Ream, T., and Wierzbicki, A. (2008) Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci.* 13: 390-397. (Review)
9. Onodera, Y., Nakagawa, K., **Haag, J.R.**, Pikaard, D.J., Mikami, T., Ream, T., Ito, Y., and Pikaard, C.S. (2008) Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. *Genetics* 180: 207-218.
10. Wierzbicki, A.T., **Haag, J.R.**, and Pikaard C.S. (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135: 635-648.
11. Ream, T.S., **Haag, J.R.**, Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. (2009) Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol. Cell* 33: 192-203.

12. **Haag, J.R.**, Pontes, O., and Pikaard, C.S. (2009) Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS ONE* 4(1): e4110. doi:10.1371/journal.pone.0004110.
13. Wierzbicki, A.T., Ream, T.S., **Haag, J.R.**, and Pikaard, C.S. (2009) RNA Polymerase V transcription guides ARGONAUTE4 to chromatin. *Nature Genetics* 41: 630-634.
14. Ream, T.S., Pontvianne, F., Nicora, C.D., Norbeck, A.D., **Haag, J.R.**, Pasa-Tolic, L., and Pikaard, C.S. (2009) Subunit compositions of Arabidopsis RNA polymerases I and III reveal insights into the evolution, functional diversification and redundancy of subunits among all five DNA-dependent RNA polymerases. In preparation.
15. **Haag, J.R.**, Ream, T.S., Pontes, O., Nicora, C.D., Norbeck, A.D., Pasa-Tolic, L., and Pikaard, C.S. (2009) DNA-DEPENDENT RNA POLYMERASE IV and RNA-DEPENDENT RNA POLYMERASE 2 are physically coupled to produce siRNA precursors. In preparation.
16. **Haag, J.R.**, Gu, J., Pontes, O., Tan, E.H., and Pikaard, C.S. (2009) Functional analysis of NRPD1 and NRPE1 C-terminal domains required for RNA-directed DNA methylation. In preparation.

CONFERENCES & SEMINARS

Invited Speaker:

- 2007 Biology Forum Departmental Seminar (Washington University in St. Louis)
“Genetic Screen to Identify RNA Polymerase IV Domains Required for siRNA-Dependent Gene Silencing”
- 2007 Plant Biology Departmental Retreat (St. Louis, MO)
“Structure-Function Analysis of RNA Polymerase IV Largest Subunits”
- 2005 Plant Lunch (Washington University in St. Louis)
“Plant RNA Polymerase IV: Role in siRNA-Directed Cytosine Methylation and Heterochromatin Dynamics”
- 2004 Plant Lunch (Washington University in St. Louis)
“Evaluation of Epitope Tags for Recombinant Protein Detection in Plants”
- 2003 Sigma Day (Donald Danforth Plant Science Center, St. Louis, MO)
“Evaluation of Epitope Tags in Plants”

First author posters have been presented at:

- 2008 The 6th International Conference on RNA Polymerases I and III (Quebec, Canada)

“Conserved Catalytic Active Site Residues Are Required for RNA Polymerase IV Function in Gene Silencing”

- 2008 The MSU Summer Symposium on Transcriptional Regulation and Systems Biology (East Lansing, MI)
“Conserved Catalytic Active Site Residues and a Unique C-terminal Domain Are Required for RNA Polymerase IV and V Function in Gene Silencing”
- 2007 24th Symposium in Plant Biology (Riverside, CA)
“RNA Polymerase IV active site requirements for siRNA production and RNA-directed DNA methylation”
- 2006 Midwest Meeting on Chromatin, Transcription and Nuclear Dynamics (Iowa City, IA) “RNA Polymerase IV subunit interactions and active site requirements for RNA-directed DNA methylation”
- 2005 16th International Conference on Arabidopsis Research (Madison, WI)
“Role of RNA Polymerase IV in siRNA-mediated DNA methylation and heterochromatin formation”

HONORS & AWARDS

- 2009 Plant Biology Program co-nominee with Thomas Ream for the DBBS Spencer T. and Ann W. Olin Biomedical Science Fellowship reflecting our teamwork in researching Pol IV/Pol V and the RNA-directed DNA methylation pathway.