Ghent University
Faculty of Sciences
Department of Plant Biotechnology and Genetics

Transcriptional and post-transcriptional regulation of leaf development in *Arabidopsis thaliana*
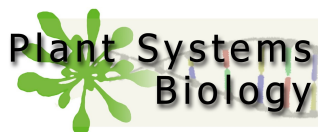
Frederik Coppens

thesis submitted as partial fulfillment of the requirements
for the degree of Doctor (Ph.D.) in Sciences: Biotechnology
Academic year 2010–2011

Promotor:
Prof. Dr. Dirk Inzé

Co-Promotors:
Prof. Dr. Gerrit T.S. Beemster
Prof. Dr. James C. Carrington

VIB - Plant Systems Biology
Systems Biology of Yield Group
Technologiepark 927, 9052 Ghent, Belgium

Promotor

Prof. Dr. Dirk Inzé
Department of Plant Biotechnology and Genetics
Ghent University

Co-promotors

Prof. Dr. Gerrit T.S. Beemster
Laboratory for Molecular Plant Physiology and Biotechnology
University of Antwerpen

Prof. Dr. James C. Carrington
Center for Genome Research and Biocomputing
Oregon State University, Oregon, USA

Promotion commission

Prof. Dr. Ann Depicker (chair)
Department of Plant Biotechnology and Genetics
Ghent University

Prof. Dr. Lieven De Veylder
Department of Plant Biotechnology and Genetics
Ghent University

Dr. Fabio Fiorani
Institut Phytosphäre (ICG 3)
Forschungszentrum Jülich, Germany

Prof. Dr. Godelieve Gheysen
Department of Molecular Biotechnology
Ghent University

Dr. Arp Schnittger
Institut de Biologie Moléculaire des Plantes
Université de Strasbourg, France

Prof. Dr. Yves Van de Peer
Department of Plant Biotechnology and Genetics
Ghent University

Prof. Dr. Mieke Van Lijsebettens
Department of Plant Biotechnology and Genetics
Ghent University

The Road goes ever on and on
down from the door where it began.
Now far ahead the Road has gone,
and I must follow, if I can,
pursuing it with eager feet,
until it joins some larger way
where many paths and errands meet.
And whither then? I cannot say.

J.R.R. Tolkien

Table of Contents

List of Acronyms

AGO	ARGONAUTE
amiRNA	artificial microRNA
ARF	AUXIN RESPONSIVE FACTOR
BLAST	Basic Local Alignment Search Tool
CDR	Cell Division Rate
CMT3	CHROMOMETHYLTRANSFERASE3
DCL	DICER-LIKE
DDL	DAWDLE
DDM1	DECREASED DNA METHYLATION1
DMS3	DEFECTIVE IN MERISTEM SILENCING 3
DRB	Double-stranded RNA Binding
DRD1	DEFECTIVE IN RNA-DIRECTED DNA METHYLATION1
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE2
dsRNA	double-stranded RNA
GFP	Green Fluorescent Protein
GRF	GROWTH-REGULATING FACTOR
GUS	β-glucuronidase
HEN1	HUA ENHANCER1
HMG	High Mobility Group
hc-siRNA	heterochromatic short interfering RNA
KYP	KRYPTONITE
KTF1	KOW DOMAIN CONTAINING TRANSCRIPTION FACTOR 1
MBD	Methylcytosine Binding-Domain
miRNA	microRNA
MIR	microRNA encoding gene
MET1	METHYLTRANSFERASE1
MOM1	MORPHEUS MOLECULE 1
nat-siRNA	natural antisense short interfering RNA
NAT	natural antisense transcript
NLS	Nuclear Localization Signal
PAZ	PIWI/ARGONAUTE/ZWILLE
PCR	Polymerase Chain Reaction
PIWI	P ELEMENT-INDUCED WIMPY TESTIS
PPR	pentatricopeptide repeat

pre-miRNA	microRNA precursor transcript
pri-miRNA	primary microRNA transcript
PTGS	Post-Transcriptional Gene Silencing
qPCR	quantitative Polymerase Chain Reaction
RDR	RNA-dependent RNA polymerase
RdDM	RNA-dependent DNA methylation
RDM1	RNA-DIRECTED DNA METHYLATION1
RISC	RNA-Induced Silencing Complex
RLER	Relative Leaf Expansion Rate
RNAi	RNA interference
ROS1	REPRESSOR OF SILENCING 1
rRNA	ribosomal RNA
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SDE5	SILENCING DEFECTIVE5
SE	SERRATE
SGS	SUPPRESSOR OF GENE SILENCING
sRNA	small RNA
SI	Stomatal Index
siRNA	short interfering RNA
ssRNA	single-stranded RNA
SI	Stomatal Index
S-PTGS	Sense-transgene Post-Transcriptional Gene Silencing
SPL	SQUAMOSA PROMOTER BINDING PROTEIN-LIKE
SUVH4	SU(VAR)3-9 HOMOLOG4
ta-siRNA	*trans*-acting short interfering RNA
TCP	TEOSINTE BRANCHED/CYCLOIDEA/PCF
TF	Transcription Factor
TGS	Transcriptional Gene Silencing
TOE	TARGET OF EAT
UBC24	UBIQUITIN-CONJUGATING ENZYME24
UTR	UnTranslated Region
WEX	WERNER SYNDROME-LIKE EXONUCLEASE
XRN	RNA exoribonucleases

Summary

Plant growth follows a strict developmental program but needs to incorporate also environmental cues to adapt to the encountered conditions. This requires a complex regulatory network to ensure an appropriate response to changing conditions. We used the first leaf pair of *Arabidopsis thaliana* as a model system to study the regulation of organ development. Leaf growth can be divided in subsequent phases according to the major process driving it. In a young leaf primordium cells divide continuously and cell size homeostasis is ensured by matching rates of cell expansion. Next, cell division ceases and cell expansion becomes the driving force for growth. When the leaf has attained its final size, maturity is reached.

In this thesis, I studied the regulation of leaf development at two regulatory levels. At the gene level, we analyzed the function of the *CYCA2* core cell cycle regulatory gene family. We also studied the function of two new proliferation specific gene families putatively involved in cell cycle regulation. On the other hand, we profiled small RNA sequences during development and linked this with the occurrence of DNA methylation.

The core machinery of the cell cycle in plants has been thoroughly studied, but our knowledge on how developmental and environmental signals impinge on cell division is still limited. CYCA2s are known core cell cycle regulators, involved in G2-to-M transition. Here, we studied the functional requirement of this gene family and showed that transcriptional repression is required for specific differentiation processes. Members of the CYCA2 protein family function in vascular development and differentiation of guard cells. For the latter process, we demonstrated that FOUR LIPS and MYB88, two transcription factors involved in stomatal development, directly repress *CYCA2;3* expression, thus ensuring correct guard cell differentiation.

Next to known 'core' cell cycle regulating genes, we also selected proliferation specific genes with unknown function, assuming them to be involved in the cell division process. We focused on two small gene families: three genes with four transmembrane domains (*4TMs*) and two genes containing three High Mobility Group (HMG) domains (*3xHMG-box*). Expression analysis and localization of transcriptional fusions with a fluorescent marker confirmed for both gene families the highly proliferation-specific expression pattern. Moreover, both families are highly induced in the M-phase of the cell cycle in synchronized cell cultures. The 4TMs localize to the cell plate during mitosis and we observed defects in cell plate forma-

tion upon overexpression and depletion of these genes. Therefore, we hypothesize that the *4TM* genes are involved in formation of the cell plate.

Profiling of small RNAs (sRNAs) in plants has thusfar mainly been focused on inflorescence tissue or whole seedlings. Here, we studied sRNAs during the different phases of development. Early in development, microRNAs implicated in nutrient stress response are upregulated, suggesting that at this phase nutrient availability is limiting for growth. We showed that specifically 24-nt sRNAs increase in expression during development. This class of sRNAs is known to be involved in RNA-dependent DNA methylation (RdDM) and can thus silence both transposons and genes. In general, the expression of sRNAs matching the coding sequences of protein-coding genes is positively correlated to the mRNA expression of this gene. We specifically selected genes that do not show this correlation, which were highly enriched in two categories: targets of microRNAs and *trans*-acting siRNAs, which generate phased sRNAs upon cleavage, and genes for which the sRNA profile is enriched for 24-nt sRNAs. This latter category is likely regulated through RdDM as this subset of genes shows increased DNA methylation in the gene body. This suggests that sRNA regulation could play an important role in regulating the leaf developmental process not only by preserving genome integrity by repressing transposon activity but also through silencing of protein-coding genes.

Samenvatting

Groei bij planten volgt enerzijds een strikt programma gedurende de ontwikkeling maar moet ook groei en ontwikkeling afstemmen op signalen uit de omgeving. Dit vergt een complex netwerk van regulators om ervoor te zorgen dat de plant gepast kan reageren op veranderende omgevingsomstandigheden. We hebben het eerste bladpaar van *Arabidopsis thaliana* gebruikt als modelsysteem om de regulatie van de ontwikkeling te bestuderen. Op basis van de belangrijkste groeiprocessen kan bladgroei opgedeeld worden in drie fasen. In een jong blad primordium delen cellen continu en de homeostase van de celgrootte wordt verzekerd door een aangepaste celexpansie. Vervolgens neemt celdeling af en wordt celexpansie het belangrijkste proces voor groei. De mature fase wordt bereikt wanneer het blad haar finale grootte bereikt heeft.

In deze thesis heb ik de regulatie van bladontwikkeling op twee regulatorische niveaus bestudeerd. Op het genniveau hebben we de functie van gekende celcyclusregulatoren, de *CYCA2* genfamilie, bestudeerd. We hebben ook de functie van twee nieuwe genfamilies bestudeerd die mogelijks betrokken zijn bij de regulatie van de celcyclus. Anderzijds hebben we het expressieprofiel van korte RNA sequenties geanalyseerd gedurende bladontwikkeling en dit gecorreleerd met DNA methylatie.

De belangrijkste componenten betrokken in de celcyclus van planten werden reeds grondig bestudeerd. Niettemin is onze kennis over hoe signalen omtrent ontwikkeling en invloeden vanuit de omgeving celdeling beïnvloeden nog steeds beperkt. CYCA2s zijn gekende celcyclusregulatoren, betrokken in G2-naar-M transitie. We hebben de functie van deze genfamilie bestudeerd en aangetoond dat repressie op transcriptieniveau nodig is voor specifieke differentiatieprocessen. Deze genfamilie reguleert ontwikkeling van de vasculatuur en differentiatie van sluitcellen van huidmondjes. Bij deze differentiatie zorgen FOUR LIPS en MYB88, twee transcriptiefactoren betrokken bij de ontwikkeling van huidmondjes, ervoor dat de expressie van *CYCA2;3* geïnhibeerd wordt. Dit is nodig voor de correcte differentie tot sluitcellen.

Naast gekende celcyclusregulatoren, hebben we ook genen geselecteerd die specifiek tijdens proliferatie tot expressie komen. Hierbij veronderstellen we dat deze betrokken zijn bij celdeling. We hebben twee kleine genfamilies meer in detail bestudeerd: drie genen met vier transmembranaire domeinen (*4TM*s) en twee genen met drie 'High Mobility Group' (HMG) domeinen (*3xHMG-box*). Analyse van de

expressie gedurende bladontwikkeling en de lokalisatie van transcriptionele fusies met een fluorescente merker bevestigen voor beide families de specificiteit voor delend weefsel. Bovendien worden beide families sterk geïnduceerd tijdens mitose in een gesynchroniseerde celcultuur. De 4TM eiwitten bevinden zich ter hoogte van de celplaat tijdens mitose en defecten bij de vorming van de celplaat werden geobserveerd bij zowel ectopische expressie als neerregulatie van deze genen. Daarom stellen we voorop dat deze genen betrokken zijn bij de vorming van de celplaat.

De analyse van korte RNA sequenties in planten heeft zich voornamelijk toegespitst op bloeiwijzen en volledige zaailingen. We hebben deze korte RNAs bestudeerd tijdens de bladontwikkeling. MicroRNAs die betrokken zijn bij de regulatie van nutriëntenstress zijn vroeg in ontwikkeling opgereguleerd. Dit suggereert dat tijdens proliferatie de beschikbaarheid van nutriënten limiterend is. We hebben aangetoond dat korte RNAs van 24 nucleotiden toenemen in expressie gedurende de ontwikkeling. Deze categorie van korte RNAs is betrokken bij RNA-afhankelijke DNA methylatie en kan zowel transposons als genen uitschakelen. Over het algemeen is de expressie van een proteïne-coderend gen positief gecorreleerd met de expressie van korte RNA sequenties waarvoor de sequentie overeenkomt. We hebben specifiek genen geselecteerd waarbij deze correlatie niet opgaat. Deze genen konden opgedeeld worden in twee categorieën: enerzijds genen die gereguleerd worden door microRNAs of *trans*-acting siRNAs die op hun beurt korte RNAs produceren en anderzijds genen die voornamelijk geassocieerd zijn met korte RNAs van 24 nucleotiden. Deze laatste categorie wordt wellicht gereguleerd door DNA methylatie aangezien in de coderende sequentie een verhoogd niveau van methylatie gedetecteerd werd. Dit suggereert dat regulatie door korte RNAs een belangrijke rol zou kunnen spelen in bladontwikkeling, zowel bij het beschermen van de integriteit van het genoom door de activiteit van transposons te verhinderen als door de regulatie van proteïne-coderende genen.

Research objectives

Plants have an indeterminate growth pattern which yields the necessary plasticity to adapt to changes in the environment. This requires an intricate regulatory network controlling growth and development. Plant growth can be reduced to two major processes: increasing the number and the size of cells, cell division and cell expansion, respectively. In early stages of development, cell proliferation is the driving force for growth. This role is gradually taken over by cell expansion. Transition from proliferation to expansion phase coincides with major transcriptional reprogramming, as we have shown in previous work.

Over the last decades, the core components of the cell cycle machinery have been identified and the major molecular mechanisms driving cell division have been unraveled. Research in our department has contributed to this field. However, the regulation of these components and its integration in the developmental program are largely unknown. We are only beginning to understand how environmental cues such as availability of nutrients and water impinge on this program. Expanding our knowledge on the basic regulatory mechanisms behind these processes is instrumental to dissect the complex regulatory network. This will ultimately enable further improvement of crops in terms of both yield as well as stress resistance.

In this research project we set out to deepen our knowledge of the regulation of cell division during development using the first leaf pair of *Arabidopsis thaliana* as a model system. In the first part of this thesis, we analyzed proliferation specific genes. Chapter 1 aims to introduce leaf development and the molecular components known to be involved in the regulation of cell proliferation and expansion during leaf growth. In Chapter 2, we set out to investigate the role of the *CYCA2* core cell cycle family in development. Next to the analysis of known cell cycle regulators, we tried to identify new putative regulators of development with unknown function based on their expression pattern during development by focussing on proliferation specific genes. This selection and the functional analysis of two gene families are described in Chapter 3.

The expression level of a gene is only one regulatory layer controlling growth and development. The regulatory potential of small RNA molecules has been established in recent years. The analysis of small RNAs makes up the second part of this thesis. In Chapter 4, I have reviewed the state-of-the-art of small RNAs in *Arabidopsis*. Profiling of small RNAs has been mainly focused on inflorescence tissue. Therefore, we profiled small RNAs during three different time points in leaf

xiv

development of *Arabidopsis* to assess their regulatory role (Chapter 5).

In the last part of this thesis, the protocol standardization and automation I performed are described. While, a manual approach for kinematic analysis in plants is explained in Chapter 6, the steps taken towards automation of this analysis are covered in Chapter 7. Expression profiling has become essential in molecular biology. While quantitative PCR has become the gold standard to determine gene expression, the methodology used to analyze the data is often not up to standard. In Chapter 8 I list the requirements for a good experimental setup and describe an algorithm that allows fast and accurate data analysis.

Part I

Proliferation Specific Genes in *Arabidopsis thaliana*

1

Leaf development in *Arabidopsis thaliana*

Frederik Coppens[1,2], Dirk Inzé[1,2], and Gerrit T.S. Beemster[1,2,3]

[1]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), 9052 Ghent, Belgium
[2]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Ghent, Belgium
[3]Department of Biology, University of Antwerp, 2020 Antwerp, Belgium

Multicellular organisms require a tightly controlled developmental program. Plants, in contrast to animals, are build in a modular way. During the embryonic development of animals, the whole body plan is at least rudimentary constructed and post-embryonic development consists mainly of enlargement of the body. In plants, the embryo essentially provides two meristems, regions of continuous cell division: the shoot apical meristem (SAM) and root apical meristem (RAM). These meristematic regions are a source of new cells that differentiate near the periphery and give rise to the formation of new 'modules'. The basic unit for the shoot, the phytomer, consists of a leaf, internode and axillary meristem that can develop a lateral shoot (1).

1.1   The leaf as a model for organ development

In this thesis, the first leaf pair of *Arabidopsis thaliana* ecotype Columbia is used as an experimental model system to study the molecular regulation of organ growth. Leaves are formed through differentiation near the flanks of the dome-like SAM from as little as four cells with an auxin-regulated spacing (2, 3). Already early in development, a dorsoventral symmetry is adopted so that the flat structure of the leaf blade can be formed. The development of this leaf primordium can be subdivided in three temporally separated phases. At first cell division is the predominant process accompanied by cell expansion to maintain a more or less constant cell size (of about 100 μm^2 for cells on the abaxial epidermis of *Arabidopsis* leaves (4)). In a second phase, cell expansion drives leaf growth. At this point cell division is limited to dispersed meristemoids that are associated with the formation of stomatal complexes. The final phase, maturity, is reached when the leaf has obtained its final size and both cell division and expansion cease (4, 5, 6, Andriankaja, Dhondt and Inzé, unpublished data).

The transition from proliferation to expansion occurs in a tip-to-base gradient (5, 7). Recently we have shown, for the third leaf, that this gradient is rapidly established (within 24 hours) with roughly the bottom half of the leaf proliferating and the top half expanding. Next, over a time span of about two days, the proportion of cells that are proliferating decreases (Andriankaja, Dhondt and Inzé, unpublished results). The transition to cell expansion is accompanied by a marked change in cell shape. Proliferating cells are rather polygonal while at the onset of expansion the characteristic undulations of the cell wall are formed. These undulations give rise to puzzle shaped cells that are characteristic for epidermal cells of dicots. The different tissue layers in the leaf show a slightly different timing of this gradient, with cell division stopping earlier in the epidermis compared to the inner tissue (5). The onset of cell expansion is accompanied by differentiation and endoreduplication, a modified cell cycle where mitosis is skipped resulting in a doubling of the DNA content of the cell. It is hypothesized that an increase in genomic material is needed to support larger cell sizes, particularly in fast growing species with a relatively small genome (8).

The same developmental stages – proliferation, expansion and mature – can be distinguished in root development. There however, they are spatially separated instead of temporally. In the root tip, adjacent to the root apical meristem, cells are dividing. Next, they undergo very rapid expansion to reach mature cell size (9, 10). A similar spatial gradient is present in monocot leaves (11, 12, 13).

1.2 Transcriptional reprogramming associated with transitions in leaf development

The temporal succession of the different developmental stages was used to study the transcriptional reprogramming during leaf development. Leaves were harvested at the three phases (proliferation, expansion and mature) and microarrays performed to assess the changes in expression on a genome-wide level (4). The equivalent data for the root was obtained by the Benfey lab (14). In the root, this was taken a step further using cell-file-specific marker lines to separate different cell types. These data are available at the ArexDB (www.arexdb.org, 15). A similar approach in the context of the developing leaf is technically much more challenging as good marker lines are not readily available.

Despite the tissue heterogeneity of the leaf material, large differences between the developmental stages were present in the transcription data. These microarrays revealed, as expected, that most known cell cycle related genes are highly expressed in young leaf material and are downregulated during leaf development. The core cell cycle machinery has been thoroughly studied so that we now have a good view of the mechanistic processes occurring during cell division (reviewed in Inzé and De Veylder (16) and Francis (17)). During the cell cycle, DNA is first replicated (S-phase) and the chromosomes are separated during mitosis (M-phase), after which the actual cell division (cytokinesis) takes place. Cells prepare for S and M during G1- and G2-phase, respectively. The transition from G1-to-S and G2-to-M are major checkpoints in cell cycle progression. At the heart of cell cycle regulation are Cyclin Dependent Kinases (CDKs) which depend on binding with a Cyclin (CYC) for activation and substrate specificity. CDKA is required at both G1-to-S and G2-to-M while CDKBs function mainly at progression through mitosis and cell cycle exit. The number of *CYC* genes is largely expanded in plants compared to other organisms, with 10 *CYCA*s, 11 *CYCB*s and 10 *CYCD*s in *Arabidopsis* (18, 19). Generally, D-type CYCs regulate G1-to-S transition integrating environmental cues and hormonal signals, A-type CYCs control S-to-M and B-type CYCs regulate G2-to-M and progression through mitosis (16). Both activating and inhibitory phosphorylation, as well as interactions with other proteins modulate the activity of these CDK-CYC complexes. The unidirectionality of the cell cycle is ensured by proteolysis of target proteins. In a multicellular organism, the spatial and temporal control of cell division is crucial. While the components of this upstream regulatory layer are still not fully unraveled, many of the components have been identified. These will be discussed below.

Remarkably, in the microarray data of the leaf developmental stages, few genes were found that are specifically expressed during the expansion phase of leaf development. This suggests that the switch between cell proliferation and cell expansion during early leaf development is a gradual process with a shutdown of proliferative

stimulators and activation of genes involved in cell cycle exit and differentiation. Recent detailed expression profiling using a time course experiment supports this view (Andriankaja, Dhondt and Inzé, unpublished data).

1.3   Molecular regulation of leaf development

The final shape and size of a leaf is characteristic for each plant species. The number and size of cells mainly determine the size of an organ, through the rate and duration of cell proliferation and the extent of cell expansion, respectively. Additionally, the number of founder cells can affect the size of an organ, but in dicotyledonous plants this has not yet been characterized extensively (20). Cell division and expansion are not merely cell-autonomous processes but need to be regulated at the level of the organ e.g. to coordinate growth of different tissue layers and regions to obtain a flat leaf. The growth-related molecular mechanisms have been studied extensively in recent years in the model system *Arabidopsis thaliana* and we are gaining insight in the regulatory mechanisms underlying growth and development. The identified genes belong to diverse and independent pathways, highlighting the complex integration of growth in the developmental program of plants (21, 22).

Here, I briefly review the major pathways involved in leaf growth, focussing on mechanisms influencing cell division and cell expansion. The regulatory pathways controlling leaf shape and patterning are outside the scope of this research project and discussed elsewhere (23, 24, 25, 26).

1.3.1   AINTEGUMENTA-related pathways

As in other developmental processes, transcription factors (TFs) play a crucial role in the regulation of growth. AINTEGUMENTA (ANT) is part of the APETALA2 (AP2) subfamily of TFs and is involved in maintaining meristem competence in leaf development. Overexpression of *ANT* results an enlarged organ size due to an increase in cell number while *ant* mutants have smaller organs composed of fewer cells (27, 28). Also other members of this AP2 subfamily, called AINTEGUMENTA-LIKE (AIL) have similar effects as ANT and can act redundantly. This has been demonstrated for AIL5 and AIL6 in flower development (29, 30). AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE (ARGOS) is a plant-specific activator of ANT. Ectopic expression of ARGOS results in prolonged expression of ANT and CYCD3;1 causing an increase in cell number in an ANT-dependent manner (31). The closest homolog of ARGOS, ARGOS-LIKE (ARL) also promotes growth but acts through cell expansion instead. While ARGOS is induced by auxin, ARL is stimulated by brassinosteroids.

AUXIN RESPONSE FACTOR 2 (ARF2) has been identified as a repressor of cell division and expansion (32, 33). ARF2 is a transcription factor that represses ANT

and CYCD3;1 expression, suggesting it has an antagonistic function to ARGOS in growth regulation (20). ARF2, implicated in auxin-mediated regulation, could serve as a point of integration of auxin and brassinosteroid (BR) signaling. The kinase BRASSINOSTEROID-INSENSITIVE 2 (BIN2) can, in the presence of BR, phosphorylate ARF2 inhibiting its DNA binding ability. This allows positive regulators of the ARF TF family to bind these promoter elements and stimulate expression of previously repressed genes (34).

1.3.2 Pathways involving GROWTH-REGULATING FACTOR

Another important family of TFs involved in growth are GROWTH-REGULATING FACTORs (GRFs). This plant-specific gene family is represented by nine members in *Arabidopsis*. Overexpression of *GRF1* and *GRF2* leads to enlarged leaves and cotyledons. Single mutants exhibit no phenotypic changes while a triple mutant in *grf1 grf2 grf3* has smaller organs, suggesting redundancy in the GRFs (35). These phenotypes are caused by changes in cell size. Similarly, overexpression of *GRF5* increases leaf area but this is due to an increase in cell number (36, 37). Several of the *GRF*s are regulated by miR396. The expression of miR396 increases during development and it has been shown that the balance of GRFs and miR396 quantitatively regulates cell number (38). The role of miR396 in leaf development will be described in more detail in Chapter 4.4.1.

GRFs bind to GRF1-INTERACTING FACTORs (GIFs), transcriptional coactivators. Three *GIF* encoding genes are present in the *Arabidopsis* genome. These act redundantly in the regulation of growth (36, 39). Mutants in *ANGUSTIFOLIA3* (*AN3*)/*GIF1* have narrow leaves while overexpression resulted in larger leaves with more cells. In a triple *gif1 gif2 gif3* mutant the final leaf size is reduced and leaf outgrowth is delayed (36, 39, 40). These phenotypes are caused by a shortening of the proliferation phase as well as a reduction in maximal cell division rate (40).

1.3.3 Translation and protein production

Growth is positively correlated to the expression level of *TARGET OF RAPAMYCIN* (*TOR*) kinase. TOR activity stimulates cell expansion and this is related to the levels of translationally active polyribosomes and *ErbB-3 BINDING PROTEIN1* (*EBP1*) expression (41). The human ortholog of EBP1 is part of ribonucleoprotein complexes and its activity regulates the production and assembly of the translational machinery. EBP1 regulates both cell number and cell size in developing leaves and its activity is necessary for the expression of *CYCD3;1* and *CYCB1;1* (42). A member of the TEOSINTE BRANCHED 1, CYCLOIDEA, PCF (TCP) transcription factor family, TCP20, is required for high levels of CYCB1;1. The motif recognized by TCP20 is also present in the promoter of many ribosomal proteins (43). This provides a mechanism to simultaneously regulate cell division and cell growth.

BIG BROTHER (BB) / ENHANCER OF DA1 (EOD) has been identified as a repressor of growth (44). BB is a E3 ubiquitin ligase and it likely functions by marking critical growth regulators for degradation. The abundance of BB correlates with cell proliferation activity and changes in the levels affects the duration of the proliferative growth phase during development. Also *DA1*, encoding a ubiquitin receptor, is implicated in the restriction of the period of proliferation. A dominant negative *da1* mutant allele affects both DA1 and DA1-RELATED1 (DAR1) causing increased leaf area through an extended proliferation phase. Double mutants in *DA1* and *BB* strongly enhances the *bb* phenotype, indicating that these genes function in parallel pathways and possibly regulate a common target (45).

1.3.4   Acting in concert

The formation of an organ requires coordination of growth. The cytochrome P450 KLUH (KLU)/CYP78A5 has been shown to affect the duration of proliferative growth and acts non-cell-autonomously. Therefore, this gene could be implicated in the coordination of growth. In young primordia KLU is ubiquitously expressed while later KLU is restricted to the perimeter of a developing organ. However, both *klu* mutations and ectopic expression of *KLU* affect cell division in a broader region and KLU could thus be part of an organ-size sensing mechanism based on the decreasing concentration of an unknown mobile signal (46). This hypothesis is supported by mathematical modeling of the transition of proliferation to expansion phase (47). In flower development, the abundance of the mobile signal seems to be integrated across the inflorescence, thus beyond the individual organ level (48). Also leaf initiation rate seems to be controlled by a pathway involving KLU (49).

To establish a flat leaf, growth between different regions and layers also needs to be coordinated tightly. Loss-of-function of TCP transcription factors (TF) results in an enlarged, wavy leaf. A prolongation of cell proliferation at the margin of the leaf in these mutants results in an excess of tissue at the perimeter and causes buckling of the leaf. Several *TCP* genes (*TCP2*, *TCP3*, *TCP4*, *TCP10* and *TCP24*) are regulated by miR319a. Overexpression of this microRNA in the *jaw-D* mutant, downregulates these *TCP*s, yielding a strong wavy phenotype (21, 50). Single mutants in the *TCP*s are less affected, indicating at least partial redundancy among the *TCP*s. TCP4 has been shown to regulate *GRF*s through upregulation of miR396, linking this TF to cell division regulation (38).

Also the *peopod* (*ppd*) mutant, in which two homologous DNA binding genes (*PPD1* and *PPD2*) are deleted, fails to make a flat leaf. In this case the central region of the leaf exhibits prolonged proliferation of specific cell types (e.g. meristemoids in the stomatal lineage), resulting in a bell-shaped leaf (51).

1.3.5   Pathways affecting cell expansion

The majority of the described pathways affect cell proliferation while cell expansion is also an important process contributing to growth. However, much less is known about pathways affecting cell elongation (20).

To be able to increase the size of cells, new cell wall material needs to be synthesized and the existing wall has to be loosened to allow it to stretch. Expansins mediate cell wall loosening and their activity is required for cell expansion in growing organs which has been demonstrated for EXP10. Ectopic expression of *EXP10* leads to larger leaves and longer petioles due to an increase in cell size while knockdown had the inverse effect (52, 53).

DELLA proteins are transcription factors that are involved in numerous developmental processes and growth responses (54). DELLAs act as repressors and their protein levels are controlled by gibberellin (GA) and auxin by inducing degradation while ethylene stabilizes these proteins (55, 56). Their role has been established in elongation growth through cell expansion but recently also a function in cell proliferation was discovered (57). Indeed, increasing the levels of GA through overexpression of an enzyme involved in GA synthesis, gibberellin 20-oxidase (GA20ox), resulted in larger leaves due to an increase in both cell number and cell size (37).

1.3.6   Conclusion

To be able to fulfill the growing demand for food, feed and bioenergy, biotechnology is expected to further optimize plants for human use. In this respect, growth is an important process, but surprisingly little is known of the underlying molecular network. Here, I gave an overview of the major pathways that are implicated in the determination of leaf size. The majority of these processes impinge on cell proliferation and more specifically the timing of exit of mitosis and transition from proliferation to expansion phase in leaf development. The deployed mechanisms are diverse and deeply intertwined, adding to the complexity to study the molecular nature of growth.

1.4   Outline

In the first part of my thesis, we studied two aspects of the regulation of cell division during proliferation using two approaches: (A) functional analysis of *CYCA2*s, a known core cell cycle family and (B) the functional characterization of two families of highly differential, proliferation specific genes with unknown function.

From the known core cell cycle genes, *CYCA2*s were chosen because *CYCA2;4* is highly proliferation specific during leaf development and has also been identified as a key activator of the cell cycle in a lateral root initiation system (58). This project

was performed in collaboration with the Root Development group (Tom Beeckman) of the Department of Plant Systems Biology (PSB) and this work has been submitted for publication to Developmental Cell.

Next to known cell cycle genes, we identified genes with unknown function that are specifically expressed during proliferation in both leaf and root development (4, 14). We selected two families for further analysis and made a start to unravel the function of these genes in regulation of proliferation. The first family consists of three genes that are among the most differential, proliferation specific genes. These genes are highly homologous and share the presence of four transmembrane domains.

A subfamily of the High Mobility Group (HMG) genes was also studied in more detail in collaboration with Dr. Klaus Grasser (Cell Biology & Plant Biochemistry, Regensburg University, Regensburg, Germany). The two selected genes are characterized by the presence of three HMG domains, a unique configuration found only in plants (59).

References

[1] T. A. Steeves, I. M. Sussex. *Patterns in Plant Development* (Cambridge University Press, Cambridge, UK, 1989).

[2] P. Piazza, S. Jasinski, M. Tsiantis. *New Phytol* **167**, 693 (2005).

[3] N. Carraro, A. Peaucelle, P. Laufs, J. Traas. *Plant Mol Biol* **60**, 811 (2006).

[4] G. T. S. Beemster, *et al. Plant Physiol* **138**, 734 (2005).

[5] P. M. Donnelly, *et al. Dev Biol* **215**, 407 (1999).

[6] H. Tsukaya. *Annu Rev Plant Biol* **57**, 477 (2006).

[7] K. A. Pyke, J. L. Marrison, A. M. Leech. *Journal of Experimental Botany* **42**, 1407 (1991).

[8] J. E. Melaragno, B. Mehrotra, A. W. Coleman. *Plant Cell* **5**, 1661 (1993).

[9] T. I. Baskin, A. Cork, R. E. Williamson, J. R. Gorst. *Plant Physiol* **107**, 233 (1995).

[10] G. T. S. Beemster, T. I. Baskin. *Plant Physiol* **116**, 1515 (1998).

[11] C. Granier, D. Inzé, F. Tardieu. *Plant Physiol* **124**, 1393 (2000).

[12] F. Tardieu, *et al. J Exp Bot* **51**, 1505 (2000).

[13] F. Tardieu, C. Granier. *Plant Mol Biol* **43**, 555 (2000).

[14] K. Birnbaum, *et al. Science* **302**, 1956 (2003).

[15] S. M. Brady, *et al. Science* **318**, 801 (2007).

[16] D. Inzé, L. De Veylder. *Annu Rev Genet* **40**, 77 (2006).

[17] D. Francis. *New Phytol* **174**, 261 (2007).

[18] K. Vandepoele, *et al. Plant Cell* **14**, 903 (2002).

[19] G. Wang, *et al. Plant Physiol* **135**, 1084 (2004).

[20] H. Breuninger, M. Lenhard. *Curr Top Dev Biol* **91**, 185 (2010).

[21] N. Gonzalez, G. T. S. Beemster, D. Inzé. *Curr Opin Plant Biol* **12**, 157 (2009).

[22] B. A. Krizek. *Curr Opin Plant Biol* **12**, 17 (2009).

[23] D. Szakonyi, A. Moschopoulos, M. E. Byrne. *J Plant Res* **123**, 281 (2010).

[24] J. L. Bowman, S. K. Floyd. *Annu Rev Plant Biol* **59**, 67 (2008).

[25] H. Tsukaya. *Int J Dev Biol* **49**, 547 (2005).

[26] M. Van Lijsebettens, M. Van Montagu. *Int J Dev Biol* **49**, 453 (2005).

[27] Y. Mizukami, R. L. Fischer. *Proc Natl Acad Sci U S A* **97**, 942 (2000).

[28] G. Horiguchi, *et al. Plant J* **60**, 122 (2009).

[29] S. Nole-Wilson, T. L. Tranby, B. A. Krizek. *Plant Mol Biol* **57**, 613 (2005).

[30] B. Krizek. *Plant Physiol* **150**, 1916 (2009).

[31] Y. Hu, Q. Xie, N.-H. Chua. *Plant Cell* **15**, 1951 (2003).

[32] Y. Okushima, I. Mitina, H. L. Quach, A. Theologis. *Plant J* **43**, 29 (2005).

[33] M. C. Schruff, *et al. Development* **133**, 251 (2006).

[34] G. Vert, C. L. Walcher, J. Chory, J. L. Nemhauser. *Proc Natl Acad Sci U S A* **105**, 9829 (2008).

[35] J. H. Kim, D. Choi, H. Kende. *Plant J* **36**, 94 (2003).

[36] G. Horiguchi, G.-T. Kim, H. Tsukaya. *Plant J* **43**, 68 (2005).

[37] N. Gonzalez, *et al. Plant Physiol* **153**, 1261 (2010).

[38] R. E. Rodriguez, *et al. Development* **137**, 103 (2010).

[39] J. H. Kim, H. Kende. *Proc Natl Acad Sci U S A* **101**, 13374 (2004).

[40] B. H. Lee, *et al. Plant Physiol* **151**, 655 (2009).

[41] D. Deprost, *et al. EMBO Rep* **8**, 864 (2007).

[42] B. M. Horváth, *et al. EMBO J* **25**, 4909 (2006).

[43] C. Li, *et al. Proc Natl Acad Sci U S A* **102**, 12978 (2005).

[44] S. Disch, *et al. Curr Biol* **16**, 272 (2006).

[45] Y. Li, *et al. Genes Dev* **22**, 1331 (2008).

[46] E. Anastasiou, *et al. Dev Cell* **13**, 843 (2007).

[47] T. Kazama, Y. Ichihashi, S. Murata, H. Tsukaya. *Plant Cell Physiol* **51**, 1046 (2010).

[48] S. Eriksson, *et al. Curr Biol* **20**, 527 (2010).

[49] J.-W. Wang, *et al. Plant Cell* **20**, 1231 (2008).

[50] J. F. Palatnik, *et al. Nature* **425**, 257 (2003).

[51] D. W. R. White. *Proc Natl Acad Sci U S A* **103**, 13238 (2006).

[52] H. T. Cho, D. J. Cosgrove. *Proc Natl Acad Sci U S A* **97**, 9783 (2000).

[53] D. J. Cosgrove. *Nat Rev Mol Cell Biol* **6**, 850 (2005).

[54] C. Schwechheimer, B. C. Willige. *Curr Opin Plant Biol* **12**, 57 (2009).

[55] P. Achard, W. H. Vriezen, D. Van Der Straeten, N. P. Harberd. *Plant Cell* **15**, 2816 (2003).

[56] X. Fu, N. P. Harberd. *Nature* **421**, 740 (2003).

[57] P. Achard, *et al. Curr Biol* **19**, 1188 (2009).

[58] S. Vanneste, *et al. Plant Cell* **17**, 3035 (2005).

[59] M. Stros, D. Launholt, K. D. Grasser. *Cell Mol Life Sci* **64**, 2590 (2007).

<div style="text-align: right; font-size: 3em;">2</div>

Plant *CYCA2*s are G2/M regulators that are transcriptionally repressed during differentiation

Steffen Vanneste[1,2,8], Frederik Coppens[1,2,8], EunKyoung Lee[3], Tyler Donner[4], Zidian Xie[5], Gert Van Isterdael[1,2], Stijn Dhondt[1,2], Freya De Winter[1,2], Bert De Rybel[1,2,7], Lieven De Veylder[1,2], Jiří Friml[1,2], Dirk Inzé[1,2], Erich Grotewold[5], Enrico Scarpella[4], Fred Sack[3], Gerrit T.S. Beemster[1,2,6] and Tom Beeckman[1,2,*]

[1]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), 9052 Ghent, Belgium

[2]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Ghent, Belgium

[3]Department of Botany, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

[4]Department of Biological Sciences, University of Alberta, Edmonton, AB, T6G 2E9, Canada

[5]Department of Plant Cellular & Molecular Biology and Plant Biotechnology Center, The Ohio State University, Columbus, OH, 43210, USA

[6]Department of Biology, University of Antwerp, 2020 Antwerp, Belgium

[7]Present address: Laboratory of Biochemistry, Wageningen University, Wageningen, The Netherlands

[8]These authors contributes equally to this work

[*]Corresponding author

Summary

In multicellular organisms, morphogenesis relies on a strict coordination of cell proliferation and differentiation in time and space. In contrast to animals, plant development is characterized by continuous organ formation and adaptive growth responses during the entire life-span. How developmental and environmental signals interact with the plant cell cycle machinery is largely unknown. Here, we show that plant A2-type cyclins are G2-to-M regulators acting prior to activation of mitotic genes, such as B-type cyclins. Through tissue-specific expression, members of this small gene family contribute to fine-tuning of local proliferation during morphogenesis. Moreover, we found that temporal expression of *CYCA2;3* in guard cells is restricted through direct control by the stomatal transcription factors FOUR LIPS/MYB124 and MYB88, providing a direct link between developmental programming and cell cycle exit in plants. Transcriptional downregulation of *CYCA2* expression might represent an important mechanism to coordinate proliferation in a developmental context in plants.

2.1   Introduction

After fertilization, plant zygotes undergo a series of patterned cell divisions during the establishment of its body axis. After germination, this minimal body plan is elaborated by the iterative development of new organs that shape the adult plant. Each new organ is formed according to a predictable pattern, which reflects a complex interplay between plant hormones and developmental programs. One of the targets of morphogenetic cues is the modulation of local cell proliferation and differentiation. Because plant cells cannot move within the plant body due to their rigid cell walls, cell proliferation needs to be controlled in time and space. While insights into the mechanisms of plant development and cell cycle regulation are expanding, only a few direct connections between these processes have been identified at the molecular level (1, 2).

Cell proliferation is characterized by consecutive cycles of DNA replication (Synthesis; S-phase) and cell division (Mitosis; M-phase). The S-phase is preceded by a G1-phase, in which cells prepare for DNA synthesis, and the M-phase is preceded by a G2-phase, in which cells prepare to divide. Transition from one phase to another in an orderly fashion largely depends on the periodic activity of Cyclin Dependent Kinases (CDKs) that require binding to a cyclin (CYC) for activity (reviewed by Satyanarayana and Kaldis (3) and De Veylder *et al.* (4)). Further modulation of CYC/CDK activity can be achieved by phoshorylation, dephosphorylation and binding to several co-factors. Such elaborate control mechanisms allow a strict control over cell proliferation.

Based on sequence homology and conserved motifs, many core cell cycle regu-

lators have been annotated in plant genomes (5). Interestingly, plants have tremendously elaborated their repertoire of cyclins compared to animals. As an example, the *Arabidopsis* genome encodes 10 A-type, 11 B-type and 10 D-type cyclins, while animal genomes usually encode 1 or 2 of each type. This expanded number of cyclins in plants could reflect a means to integrate a broader range of signals into control of proliferation.

In plants, components of the G1-to-S transition control cell proliferation and differentiation events in shoots (6, 7) and roots (1, 8, 9), highlighting the key role of this transition in the decision of the cell to exit cell cycle and activate differentiation. However, differentiation does not preclude an active G1-to-S pathway, as some differentiating cell types are known to go through multiple rounds of DNA duplication without mitosis (endoreduplication, 10). An active G1-to-S pathway in these cells argues for the G2-to-M transition as an additional target for developmental regulation of proliferation.

Although much progress in plant cell cycle regulation has been made, there remains a big gap in our understanding of G2/M transition. Here, we address the functional requirement of the subfamily of plant A2-type cyclins in plant cell cycle regulation and highlight their transcriptional regulation during terminal differentiation of guard cells.

2.2   Results

Sequence similarity (5), co-regulation during the cell (11), subcellular co-localization (12), common interaction partners (13) and mild phenotypes in single mutants (14, 15), together suggest redundancy among the four CYCA2s in the *Arabidopsis* genome and obscure functional analysis of this gene family. To address their function in plant cell cycle regulation and development, we aimed to overcome redundancy through mutants in which several *cyca2* genes were knocked-out.

2.2.1   CYCA2s regulate G2/M transition in the root

As *CYCA2* expression is strongly associated with proliferative tissues, such as root meristems (Supplemental Figure 2.1), we probed the impact of their loss-of-function on root growth. Consistent with redundancy, we did not find strong phenotypes in single and double mutants (data not shown). Only when 3 out of 4 *CYCA2*s were mutated, growth defects could be detected easily. Triple mutants had shorter roots and were impaired in their capacity to form lateral roots (Supplemental Figure 2.2). To address if these growth defects were due to a defect in proliferation, we analyzed root meristems. Primary root meristems were smaller and contained fewer dividing cells than WT (Figure 2.1A, B). Similarly, developing lateral root primordia of *cyca2;234* were composed out of fewer cells compared to WT (Figure 2.1C),

suggesting that a defect in proliferative capacity is causal to the observed growth defects.

To determine at which cell cycle stage CYCA2s act, we compared cell cycle progression during synchronized lateral root initiation in WT versus *cyca2;234* triple mutants (Figure 2.1D). In WT, expression of auxin signaling genes and G1-to-S regulatory genes precedes expression of G2-to-M regulatory genes (16, 17, 18). In contrast, mitotic regulators, such as B-type cyclins, were not induced within the same time-course in *cyca2;234* mutants, whereas expression of auxin signaling genes and G1-to-S regulators was not affected. The delay in activation of mitotic regulators shows that plant CYCA2s function early in the G2/M transition, as was predicted based on sequence homology (5) and expression in synchronized cell suspensions (11).
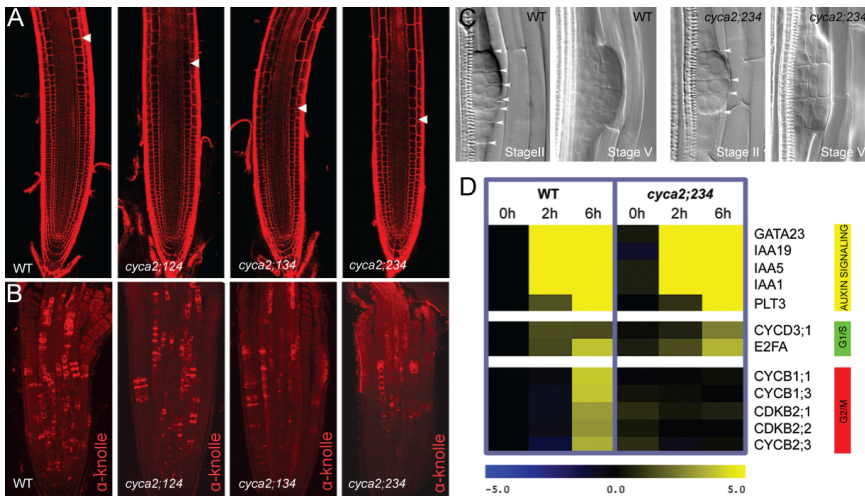

Figure 2.1: *cyca2* triple mutants have defects in cell cycle progression. (A) Propidium iodine stained root meristems of WT, *cyca2;124*, *cyca2;134* and *cyca2;234* 10 days after germination. Arrowheads indicate end of the meristem, defined as the position where cells start elongating. (B) Immunolocalization of the mitosis-specific syntaxin KNOLLE, labeling actively dividing cells in roots of 7 day old WT, *cyca2;124*, *cyca2;134* and *cyca2;234*. (C) Stage II and stage V lateral root primordia of WT and *cyca2;234* cleared with chloral hydrate. Lateral root primordia of *cyca2;234* are composed of fewer cells than WT. Arrowheads indicate periclinal cell walls. Stages as defined by Malamy and Benfey (19). (D) Transcriptional responses of auxin signaling genes, G1/S and G2/M regulators in WT and *cyca2;234* root segments during auxin-induced lateral root initiation. 0h, 2h and 6h correspond to time of auxin treatment (10 μM) after being germinated in presence of the auxin transport inhibitor NPA (10 μM). Range indicator from blue to yellow represent expression levels on a log2 scale relative to NPA germinated WT (0h).

2.2.2 CYCA2s drive proliferation in leaves, while repressing endo-reduplication

To obtain the characteristic final size and shape, morphogenesis of the leaf depends on a tight coordination between cell proliferation, cell cycle exit and differentiation. Early leaf development displays high cell division activity that is followed by a gradual tip-to-base deceleration of proliferation and the start of differentiation-associated endoreduplication and cell expansion (20, 21). We found dramatic increases in ploidy levels and cell sizes in the mature first true leaves of *cyca2* triple mutants (Figure 2.2A, B). We analyzed *cyca2;234* leaf development in greater detail, to address the mechanism of the enhanced ploidy levels and cell sizes. Kinematic analysis of leaf growth showed a slow-down of cell division rate in *cyca2;234* leaves (Figure 2.2C). Moreover, as soon as the first leaf pair became macroscopically apparent (after 8 days of growth, Stage 1.02, 22), DNA content was already dramatically higher than the WT (Figure 2.2D; 2C, 4C and 8C). Moreover, ploidy levels continued to rise in *cyca2;234* (after 14 days of growth), while endoreduplication had already stopped in WT (Figure 2.2D; 16C and 32C). Thus, enhanced ploidy levels in *cyca2;234* are the combined result of an early onset and extended duration of endoreduplication. Collectively, our phenotypic and molecular analyses in roots and shoots of *cyca2* triple mutants demonstrate that plant CYCA2s are fundamental elements of the cell cycle, exerting their function in early G2-to-M transition, just like their animal counterparts.

2.2.3 Tissue-specific *CYCA2* expression contributes to vascular proliferation near hydathodes

In addition to their expression in meristems, the expression pattern of *CYCA2;1* and *CYCA2;4* remarkably mimicked the reticulate venation pattern in the leaf (Figure 2.3A). The promoter activities of these two genes in the leaf further overlapped with one of the earliest hallmarks of the vascular progenitor cell state (pep-rocambial, Figure 2.3B), namely the promoter activity of the homeodomain-leucine zipper (HD-Zip) III gene *ATHB8* (23, 24). These tissue-specific expression patterns are suggestive for a specific function in leaf vascular development. Indeed, the number of vascular hypertrophy zones was reduced in *cyca2;234* leaves compared to the WT (Figure 2.3C). However, this was associated with a change in leaf shape in *cyca2;234*, as fewer serration tips were formed in these mutants (Figure 2.3D). Introgression of a mutation in the vascular-specific *CYCA2;1* into a *cyca2;234* background reduced vascular cell proliferation even more without further affecting the number of serration tips (Figure 2.3C, D), which supports that vascular proliferation defects in *cyca2* mutants are not secondary consequences of leaf growth defects but rather due to tissue-specific modulation of *CYCA2* levels.
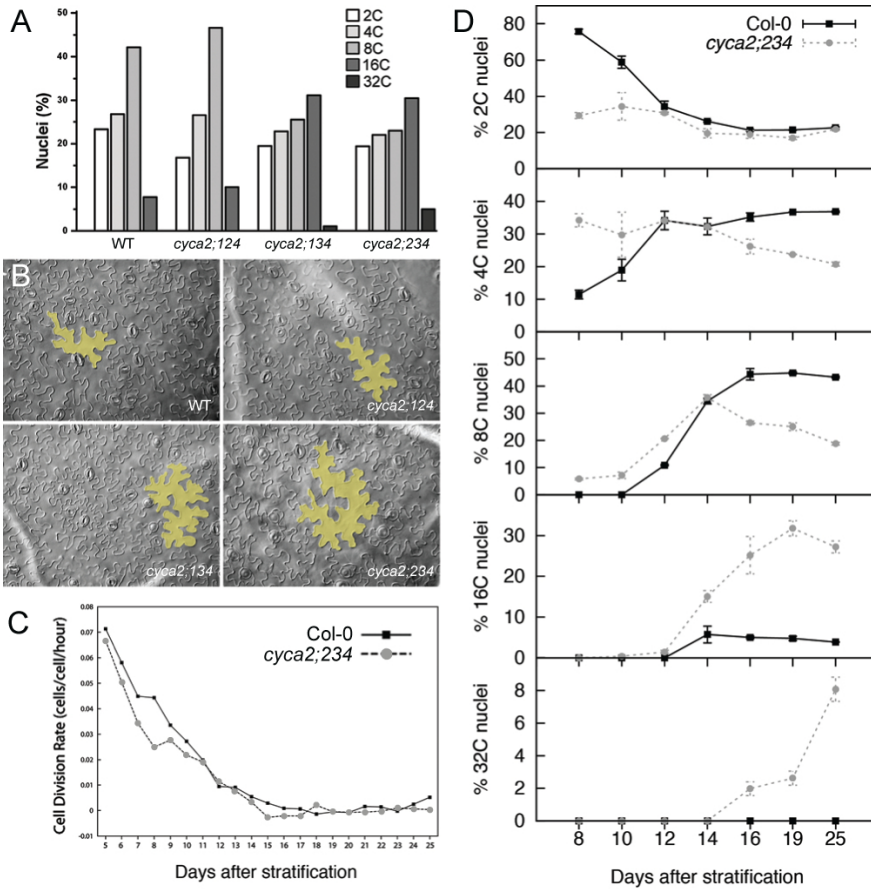
Figure 2.2: Leaf development shows enhanced endoreduplication and slowed down cell cycle progression in *cyca2* triple mutants. (A) Distribution of nuclear ploidy in mature primary leaves of WT, *cyca2;124*, *cyca2;134* and *cyca2;234*. Triple mutants *cyca2;134* and *cyca2;234* show highest ploidy levels (B) Pavement cell size in mature primary leaves of WT, *cyca2;124*, *cyca2;134* and *cyca2;234*. Yellow overlays highlight representative cells. (C) Kinematic analysis reveals a slow-down in cell division rates in developing primary leaves of *cyca2;234* compared to WT. (D) Evolution of ploidy levels during the development of WT and *cyca2;234* primary leaves. In early stages, WT has predominantly 2C nuclei and a low 4C fraction. Later, the 2C fraction drops rapidly, while higher ploidy fractions increase until ± 16 days after stratification. In *cyca2;234*, the 2C fraction is already low at the earliest stage analyzed, while the 4C fraction is already high and even a small fraction 8C nuclei can be detected. At later stages higher ploidy fractions continue to increase, and do not saturate within the time-frame of our analysis. Data are represented as mean ± SEM.
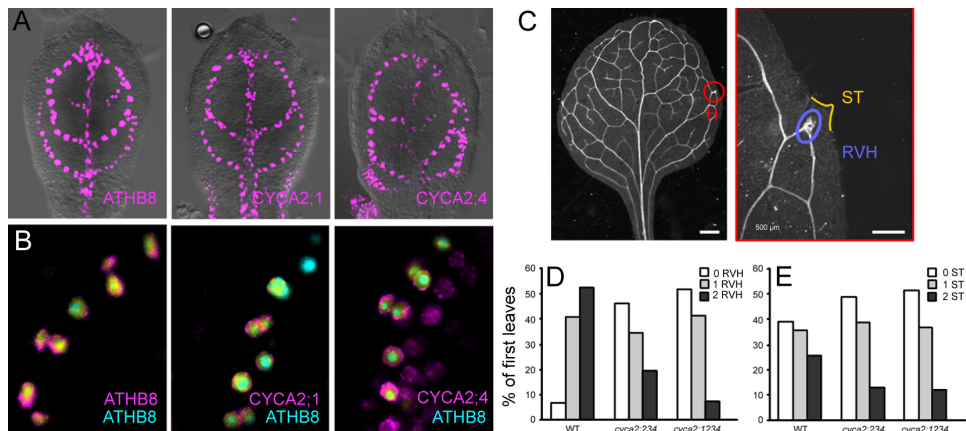
Figure 2.3: Tissue-specific expression of CYCA2s is required for vascular cell proliferation. (A) Expression pattern of pCYCA2;1::HTA6:eYFP and pCYCA2;4::HTA6:eYFP in 4 days after germination (DAG) leaves resemble that of pATHB8::HTA6:eYFP, which is an early hallmark of vascular development. (B) Co-expression of pATHB8::HTA6:eYFP, pCYCA2;1::HTA6:eYFP and pCYCA2;4::HTA6:eYFP with pATHB8::eCFP:NLS in 4 DAG leaves. *Note*: *CYCA2;4* expression is initiated slightly earlier than *ATHB8* and in wider expression domains that over time narrow to single cell files. In contrast, *CYCA2;1* expression is initiated slightly later than *ATHB8*, but its expression is always confined to single cell files. Images color-coded with a dual-channel look-up table (LUT) from cyan to magenta through green, yellow and red (25). Preponderance of cyan signal over co-localized magenta signal is encoded in green, opposite in red, and co-localized cyan and magenta signals of equal intensity in yellow. (C) Overview of cleared mature primary WT leaf and a detail of a hydathode (H) that shows vascular hypertrophy (RVH) and a serration tip (ST). (D) Frequency of regions of vascular hypertrophy (RVH) in mature primary leaves. RVH frequency gradually reduces between WT, *cyca2;234* and *cyca2;1234*. (E) Frequency of serration tips (ST) per mature primary leaves. ST frequency shows a reduction of ST number between WT and *cyca2;234* and *cyca2;1234*, but not between *cyca2;234* and *cyca2;1234*.

2.2.4 Guard mother cell division requires CYCA2 activity

Besides vascular-specific expression of *CYCA2;1* and *CYCA2;4*, we found that *CYCA2;2* and *CYCA2;3* expression patterns showed remarkable association to the stomatal lineage (Figure 2.4A). These expression patterns suggest a role for these CYCA2s in modulating cell divisions in the stomatal lineage.

Stomatal development requires a sequence of asymmetric cell divisions, and finally a single, symmetric division of the guard mother cell (GMC) to give rise to a pair of guard cells that form the stomate (26). Inspection of the *cyca2* triple mutant
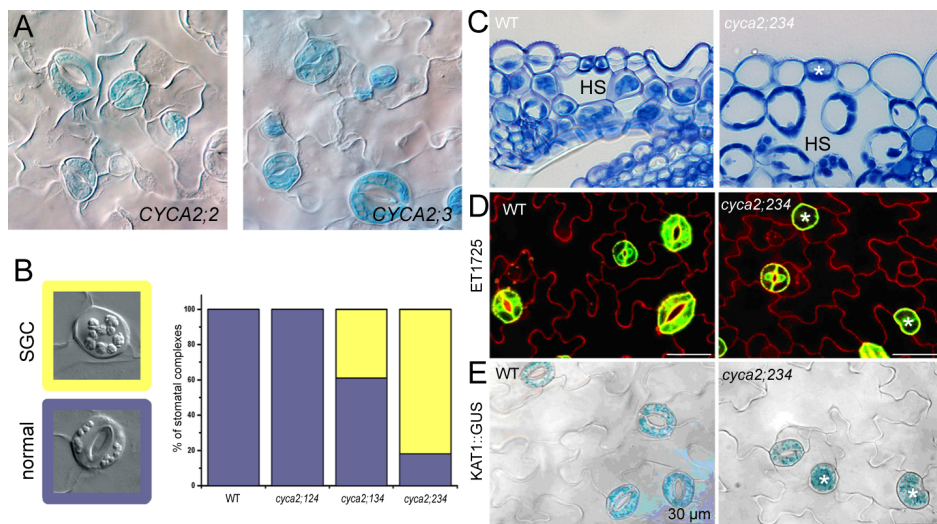
Figure 2.4: Stomatal expression of *CYCA2*s is required for guard mother cell division. (A) Stomatal expression of pCYCA2;2::GUS:GFP and pCYCA2;3::GUS:GFP (B) Stomatal phenotypes (left) of WT and representative triple mutant. Graph: quantification of stomatal phenotypes. The *cyca2;234* triple mutant displays the highest frequency of single guard cells (SGC). Blue = normal stoma. Yellow = SGC. Triple mutants shown: *cyca2;124*, *cyca2;134* and *cyca2;234*. (C) Anatomical section through a WT stomatal complex and a *cyca2;234* SGC showing correct placement of abnormal SGC (asterisk) over a hypostomatal space (HS). (D-E) Expression of mature guard cell identity markers, (D) ET1728:GFP and (E) pKAT1::GUS in WT and *cyca2;234* (asterisks indicate SGCs).

leaf epidermis revealed unpaired oval cells displaying cell wall thickenings and plastid accumulations reminiscent of guard cells (Figure 2.4B). Similar to normal stomatal guard cells, these cells were positioned above large intercellular spaces in the subjacent mesophyll (Figure 2.4C). They also expressed mature guard cell identity markers, pKAT1::GUS (27) and ET1728 (Figure 2.4D, 28), indicating that they correspond to aberrant, single guard cells (SGCs) at positions where a stoma would be expected. These SGCs have twice the nuclear-DNA content (4C) of normal guard cells (2C) (Supplemental figure 2.3), implying that a round of DNA replication occurred without cell division. Yet, the aberrant cells attained a guard cell identity and formed single guard cells instead of a pair of guard cells. Consistent with specific expression patterns of *CYCA2;2* and *CYCA2;3*, frequencies of SGCs in mutants correlated with the respective mutations (Supplemental Table 2.2). Thus, tissue-specific *CYCA2* expression is required to couple cell division to differentiation in dividing guard mother cells.

2.2.5 CYCA2s and CDKB1s act redundantly during guard mother cell division

SGCs were earlier reported in transgenic plants harboring the *CDKB1;1-N161* dominant negative construct (29) and *cdkb1;1 cdkb1;2* double mutants (2). Moreover, CDKB1;1 has been shown to form a functional complex with CYCA2;3 (30) and *CDKB1;1* is also expressed in the stomatal lineage (29), suggesting that CYCA2s and CDKB1s form functional complexes during guard mother cell division. Therefore, we generated a *cyca2;234 cdkb1;1/1;2* quintuple mutant which showed an even higher frequency of SGCs (Figure 2.5). Thus, all five genes act redundantly to promote symmetric division and stomatal morphogenesis.
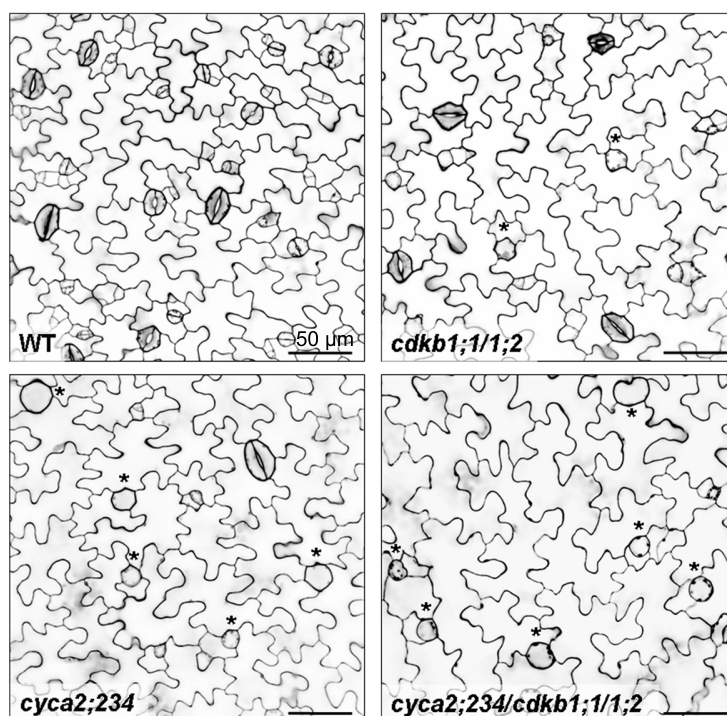
Figure 2.5: *CYCA2;2, CYCA2;3, CYCA2;4* and *CDKB1;1* and *CDKB1;2* genes act redundantly in promoting guard mother cell symmetric division. About 94% of stomata are undivided in the *cyca2;234* triple loss-of-function mutant while 98-99% of stomata are undivided in the quintuple *cdkb1;1 cdkb1;2 cyca2;234* mutant. Cell walls in 4-day-old developing cotyledons were visualized using propidium iodide and laser scanning confocal microscopy. Asterisks indicate abnormal single guard cells.

2.2.6   FLP/MYB88 regulates timely repression of *CYCA2;3* during terminal differentiation of guard cells

These tissue-specific requirements for *CYCA2* expression in vascular tissues and stomatal lineage, demonstrate that multiple CYCA2s allow a more subtle regulation of the plant cell cycle across its tissues and outside the meristems. Furthermore, the expression of *CYCA2*s is not only subject to spatial control in the stomatal cell lineage, it is also restricted in time. Together with *CDKB1;1*, stomatal *CYCA2;2* and *CYCA2;3* expression was induced in late guard mother cells (GMCs) and remained high in young guard cells, while their expression was strongly reduced in mature stomata (Figure 6A), suggesting their expression might be actively repressed. Potential candidate *CYCA2* repressors, transcription factors FOUR LIPS (FLP/MYB124), its closest ortholog MYB88 (31) and basic helix-loop-helix transcription factor FAMA (32), are expressed at comparable stages of stomatal development as *CYCA2;2*, and *CYCA2;3* (Figure 2.6A). Moreover, loss-of-function mutations in *FOUR LIPS* (*FLP/MYB124*) and *MYB88*, induce clusters of four or more guard cells in direct contact (31), whereas mutants in *FAMA* have similar clusters of cells that never acquire guard cell identity (32). To address the functional requirement of *CYCA2* expression for guard cell hyperproliferation in these mutants, we generated *cyca2;234 fama* quadruple and *cyca2;234 flp myb88* quintuple mutants. The *cyca2;234 fama* plants did not show any SGCs but did form cell clusters that lack guard cell fate, although the clusters had fewer cells than *fama* single mutants (Supplemental Figure 2.4). By contrast, the formation of stomatal clusters in a *cyca2;234 flp-7 myb88* background was completely suppressed (Figure 2.6B) demonstrating that *CYCA2* gene products are required for the *flp-7 myb88* stomatal phenotype. As FLP and MYB88 are transcription factors, we analyzed *CYCA2* expression in a *flp myb88* background. Ten days after sowing, cotyledons of *flp-7 myb88* seedlings showed 5-fold higher *CYCA2;3* expression than the WT (Figure 2.6C). Moreover, in *flp-1 myb88* stomata, *CYCA2;3* promoter activity remained high after the GMC divided (Figure 2.6D) suggesting FLP and MYB88 are repressors of *CYCA2;3* promoter activity. To test if this was a direct effect, we performed ChIP-PCR using polyclonal antibodies raised against FLP and MYB88 (33). In WT, *CYCA2;3* promoter chromatin fragments were enriched after ChIP, while these were lost in *flp-1 myb88* mutants (Figure 2.6E), demonstrating a specific, direct interaction of FLP and MYB88 with *CYCA2;3* chromatin. Thus FLP and MYB88 restrict *CYCA2;3* transcription after GMC division through direct interaction with its promoter.

2.3   Discussion

In the first rough sketch of the plant cell cycle, A-type cyclins were assumed to function in S-phase and G2-to-M transition. This was in part justified by their expression pattern in synchronized suspension cells (11, 34, 35), the rescue of growth of yeast cyclin mutants (36) and their ability to induce maturation of *Xenopus* oocytes (37). Also in plants they were able to drive cells into mitosis upon ectopic overexpression (14).

Loss-of-function had previously revealed that single mutants in *cyca2;1* and *cyca2;3* had increased levels of endoreduplication (14, 15), suggesting they are negative regulators of endoreduplication, presumably by stimulating G2-to-M transition. Consistent with these data, we found that *cyca2* triple mutants had enhanced endoreduplication levels. Moreover, we found that cell proliferation rates in developing leaves were reduced in these mutants. All these data only indirectly implicate CYCA2s in G2-to-M regulation.

More conclusive evidence came from analysis of stomatal development in triple *cyca2* mutants. Normally, a guard mother cell divides symmetrically to give rise to two guard cells that make up the stomatal pore. After this division, the daughter cells exit mitosis and acquire guard cell identity. In *cyca2* triple mutants, single guard cells occur frequently. Interestingly, these cells with guard cell identity have 4C nuclei, which is the double of normal guard cells. Therefore, these aberrant cells are presumably the product of an undivided guard mother cell that had completed S-phase but failed to undergo mitosis.

To address the role of CYCA2s in G2-to-M transition in greater detail, we used *in planta* cell cycle synchronization to compare cell cycle progression between WT and mutant. While G1-to-S regulatory genes were not affected, B-type cyclin expression was delayed in these mutants. These data imply that plant CYCA2s act upstream of B-type cyclins during G2-to-M transition.

Proliferation and differentiation are mostly mutually exclusive processes. Some cells exit the cell cycle after mitosis and remain in G1-phase, while other differentiating cells undergo several rounds of a modified cell cycle in which G2-to-M transition is omitted and only DNA synthesis occurs (endoreduplication). In animals, some developmental programs coordinate cell cycle exit during differentiation through transcription factor activity (reviewed by Myster and Duronio (38), Buttitta and Edgar (39)). One strategy is to induce CDK inhibitory proteins, while another is to repress cell cycle activating proteins. Interestingly, transcription of A-type cyclins is often actively repressed during differentiation processes (40, 41, 42, 43, 44). In plants, it is not known how developmental signals can modulate the switch between a full cell cycle to the endocycle or cell cycle exit during differentiation. Here, we found that FLP/MYB88 repress *CYCA2;3* expression during cell cycle exit in differentiating guard cells. This mechanism resembles the PROSPERO-dependent

mechanism in *Drosophila* that links neuronal lineage development with the transcriptional regulation of cell cycle regulatory genes (40).

Mutants in *cyca2* function have higher ploidy levels (14, 15, our data), while overexpression strongly suppresses endoreduplication (14, 30), suggesting that CYCA2s are prime negative determinants of endoreduplication in leaves. Early stages of leaf development are highly proliferative in nature, while later stages switch gradually to differentiation-associated endoreduplication and cell expansion (20, 21). Interestingly, *CYCA2;3* expression is rapidly repressed during the switch from proliferation to endoreduplication in differentiating leaves (14). However, it remains to be seen if this effect is directly mediated by differentiation-induced transcription factors.

Stomatal development ends by a single symmetric division of a guard mother cell, whose daughter cells undergo terminal differentiation into individual guard

---

Figure 2.6 *(on the next page)*: **FLP/MYB124 and MYB88 are direct repressors of *CYCA2;3* expression during guard mother cell division.** (A) Expression analysis of transcriptional promoter:reporters (except for pTMM:TMM:GFP translational fusion). The expression patterns of the three transcription factors (*FLP*, *MYB88* and *FAMA*) are similar i.e. present in late GMCs, symmetric divisions and young guard cells. TOO MANY MOUTHS (TMM) is an early marker. *CYCA2;2*, *CYCA2;3* and *CDKB1;1* are roughly expressed during the same phases as the transcription factors. pCYCA2;2::GUS:GFP, pCYCA2;3::GUS:GFP, pCDKB1;1::GFP, pFLP::GUS:GFP, pMYB88::GUS:GFP, pFAMA::GFP and pTMM::TMM:GFP during stomatal development and especially during guard mother cell division. A meristemoid (M) develops into a guard mother cell (GMC). Late GMCs have thickened cell walls which are usually bisected by the symmetric division (SD) that produces two young guard cells (GC). The latter undergo further morphogenesis including stomatal pore formation. (B) Frequency of different stomatal phenotypes in WT, *flp-7 myb88*, *cyca2;234* and *cyca2;234 flp-7 myb88*. In WT, all stomata are normal (type I), while in *flp-7 myb88* and in *cyca2;234* the majority are stomatal clusters (type II) and single guard cells (type III), respectively. In *cyca2;234 flp-7 myb88* quintuple mutant, type II are not formed and most of guard cells are type III. Note that this mutant combination also displays a 'fusion' phenotype in type IVa-c. (C) Relative expression levels of *CYCA2;1*, *CYCA2;2*, *CYCA2;3*, *CYCA2;4* in cotyledons of WT and *flp-7 myb88* 10 days after germination as determined by qPCR. *CYCA2;3* was expressed in *flp-7 myb88* to markedly higher levels than WT. WT levels were used as reference. Data are represented as mean $\pm$ SEM. (D) Expression of CYCA2;3::GUS-GFP in WT and *flp-1 myb88*. While expression of pCYCA2;3::GUS:GFP is absent from mature guard cells in WT, it is strongly expressed in *flp-1 myb88* stomatal clusters. (E) ChIP-PCR on 4 fragments upstream (-1.4 kb) of the translational start of *CYCA2;3* (P1-P4). For each fragment an input control, a negative control (IgG) and specific polyclonal against FLP/MYB88 was loaded. Comparisons between WT and *flp-1 myb88* ChIP-PCR data demonstrate antibody-specificity. Strongest, specific binding was observed for P2.
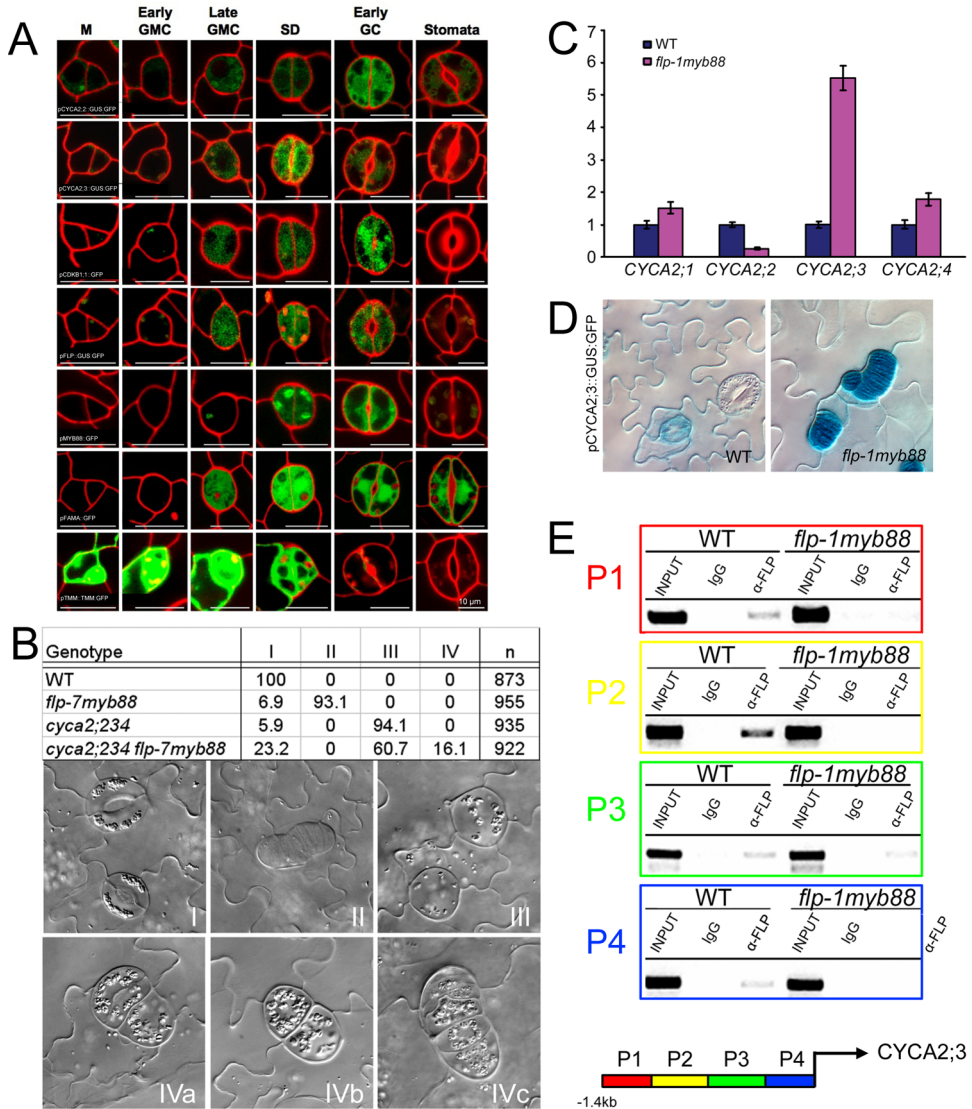
cells (reviewed by Bergmann and Sack (26)). Mutants in stomatal transcription factors *FLP* and *MYB88* do not stop dividing after the guard mother cell has divided, even though guard cell identity markers are expressed (31). We found that down-regulation of *CYCA2;3* after the first guard mother cell division did not occur in *flp myb88*. Direct interaction with *CYCA2;3* promoter chromatin corroborates that FLP and MYB88 are direct repressors of *CYCA2;3* expression in guard cells. Similarly, expression of an interacting CDK (30), *CDKB1;1* is also a directly repressed by FLP and MYB88 (33). These data are consistent with a model in which FLP and MYB88 enforce cell cycle exit during terminal differentiation of guard cells by direct repression of CYCA2/CDKB1;1 kinase complexes.

In summary, we found evidence that transcriptional repression of *CYCA2*s is part of the mechanism that coordinates the switch between proliferation and differentiation.

2.4 Experimental procedures

2.4.1 Plant material and growth conditions

We used *Arabidopsis thaliana* seedling of the ecotype Col-0 and L*er* and mutants for the various A2-type cyclins from publicly available collections (SALK (45), GABI-KAT (46) and EXOn Trapping Insert Consortium (EXOTIC, www.jic.bbsrc.ac.uk)), and stomatal lineage mutant alleles *flp-1 myb88*, *flp-7 myb88* (31) and *fama-1* (32). The cyclin mutant alleles used are *cyca2;1-1* (SALK_121077), *cyca2;1-2* (SALK_123348), *cyca2;2-1* (GABI_120D03), *cyca2;3-1* (SALK_092515), *cyca2;3-2* (SALK_086463), *cyca2;3-3* (SALK_043246), *cyca2;4-1* (SALK_070301) and *cyca2;4-2* (GAT_5.10009). Promoter::reporter lines for *FLP* (31), *CDKB1;1* (33) and *CYCA2;1* (47) have been reported previously. For detection of T-DNA inserts we used primers specific to the left border of the T-DNAs used for mutagenesis (LBC1, LB_GABI and LB_EXOTIC) in combination with gene-specific primers (Supplemental table 2.2). The alleles *cyca2;1-1*, *cyca2;2-1*, *cyca2;3-1* and *cyca2;4-1* are representative knock-out alleles and have been used for generating triple mutants. After surface sterilization, seeds were sown on 0.5x Murashige and Skoog (MS, 48) medium supplemented with 1% sucrose and 0.8% agar. After stratification, plates were moved to cooled benches in a growth chamber (temperature: 22 °C; irradiation: 65 μE·m^{-2}·sec^{-1} photosynthetically active radiation; photoperiod: 16 hr light/8 hr dark or continuous light).

2.4.2 Immunofluorescence localization

One week old seedlings, grown on 0.5x MS medium under continuous illumination, were fixed in paraformaldehyde (Merck, Darmstadt, Germany). Immunolocaliza-

tion was performed as (49). The rabbit anti-KNOLLE antibody (1:2000) (50), and the fluorochrome-conjugated secondary antibody anti-rabbit-Cy3 (1:600) (Dianova, Hamburg, Germany) were used. Fluorescence detection was done with a confocal laser-scanning microscope Zeiss 710 (Carl Zeiss, Zaventem, Belgium).

2.4.3 Cloning

Promoter::GUS:GFP fusions of *MYB88*, *CYCA2;2*, *CYCA2;3* and *CYCA2;4* were generated through Gateway cloning of promoter fragments into pKGWFS7 (51). PCR fragments of *CYCA2;2*, *CYCA2;3* and *CYCA2;4* promoters were described previously (52). To generate the *CYCA2;1* and *CYCA2;4* transcriptional fusions (pCYCA2;1::HTA6:EYFP and pCYCA2;4::HTA6:EYFP, respectively), 1808 bp upstream of the *CYCA2;1* start codon and 1963 bp upstream of the *CYCA2;4* start codon were amplified from *Arabidopsis thaliana* ecotype Col-0 genomic DNA and cloned into the Gateway-adapted pFYTAG binary vector, which contains a translational fusion between the coding region of *histone 2A* (*HTA6*, *AT5G59870*) and that of the enhanced YFP (EYFP, 53).

2.4.4 Vascular expression analysis

The origin of the pATHB8::HTA6:EYFP and the pATHB8::ECFP:Nuc have been described (54, 55). Seeds were sterilized and germinated, and seedlings and plants were grown, transformed and selected as described (54, 55). The progeny of 8 independent transgenic lines of pCYCA2;1::HTA6:EYFP and pCYCA2;4::HTA6:EYFP were inspected, to identify the most representative expression pattern. We define 'days after germination' (DAG) as days following exposure of imbibed seeds to light. Dissected seedling organs were mounted and imaged as described (23, 54, 55). Brightness and contrast were adjusted through linear stretching of the histogram in ImageJ (National Institutes of Health, rsb.info.nih.gov/ij). Signal levels and colocalization were visualized as described (55).

2.4.5 Quantitative PCR

RNA was extracted with the RNeasy kit (Qiagen, Venlo, The Netherlands). Poly(dT) cDNA was prepared from 1 µg total RNA with the Superscript III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) and quantified on an iCycler apparatus (BioRad, Hercules, CA, USA) with the Platinum SYBR Green qPCR Supermix-UDG kit (Invitrogen, Merelbeke, Belgium). PCR was carried out in 96-well optical reaction plates heated for 10 minutes to 50 °C to allow UNG activity, followed by 10 minutes of 95 °C to activate hot start Taq DNA polymerase and 40 cycles of denaturation for 20 seconds at 95 °C and annealing-extension for 20 seconds at 58 °C. Target quantifications were performed with specific primer pairs designed using

Beacon Designer 4.0 (Premier Biosoft International, Palo Alto, CA, USA). Expression levels were normalized to *AT5G25760* (Q_PEX4) and *AT4G16100* (Q_UNKN1), which showed constitutive expression across samples. All qPCR experiments were performed in triplicates and the data was processed using qBase (56).

Histochemical staining and anatomical analysis. The β-glucuronidase (GUS) assays were performed as described earlier (57). For microscopic analysis, chlorophyll was removed by EtOH treatment and further cleared by mounting in 90% lactic acid (Acros Organics, Brussels, Belgium). All samples were analyzed by differential interference contrast microscopy.

For anatomical sections, samples were fixed overnight in 1% glutaraldehyde (Agar Scientific Ltd, Stansted, Essex, UK) and 4% paraformaldehyde (Merck, Darmstadt, Germany) in 50 mM phosphate buffer (pH7.0). Samples were dehydrated and embedded in Technovit 7100 resin (Heraeus Kulzer, Wehrheim, Germany) according to the manufacturers protocol. Sections of 5 μm were cut with a microtome (Minot 1212; Leitz, Wetzlar, Germany), dried on Vectabond coated object glasses (Vector Laboratories, Burlingame, CA, USA), stained with Toluidine Blue for 8 minutes (Fluka Chemica, Buchs, Switzerland) and rinsed in tap water for 30 seconds. After drying, the sections were mounted in DePex medium (British Drug House, Poole, UK).

2.4.6   Flow cytometry

Primary leaves of 3-week old seedlings were chopped with a razor blade in 300 μl of buffer (45 mM $MgCl_2$, 30 mM sodium citrate, 20 mM 3-[N-morpholino]propane-sulfonic acid, pH7.0, and 1% Triton X-100). To the supernatants, 1 μl of 4′,6-diamidino-2-phenylindole (DAPI, Sigma, St. Louis, MO, USA) from a stock of 1 mg/ml was added, which was filtered over a 30 μm mesh. The nuclei were analyzed with a CyFlow©ML (Partec, Görlitz, Germany) flow cytometer.

2.4.7   Guard cell nuclear content measurement

Nuclei were stained fluorescently by fixing 3-week-old cotyledons in a mixture of 9:1 ethanol:acetic acid (v/v). After the samples had been rinsed, they were stained for 24 h with 0.1 μg/ml of DAPI, mounted in Vectashield mounting medium (Vector Laboratories, Burlingame, CA, USA) and observed under a 63x oil immersion objective on a Zeiss Axioskop equipped with an Axiocam CCD camera (Carl Zeiss, Zaventem, Belgium). Images were obtained using the Axiovision software and were analyzed in grayscale with the image analysis program ImageJ (version 1.28; rsb.info.nih.gov/ij/). Relative fluorescence units were reported as integrated density, which is the product of the area and the average fluorescence of the selected nucleus.

2.4.8 Kinematic analysis of leaf development

Plants of the wild type and the *cyca2* triple mutants were sown in quarter sections of round 12 cm Petri dishes filled with 100 ml of 0.5 x MS medium (Duchefa, Haarlem, The Netherlands) and 0.6% plant tissue culture agar (Lab M Limited, Bury, Lancaster, UK). At relevant time points after sowing plants or primary leaves (depending on the size) of the respective genotypes were harvested. All healthy plants were placed in methanol overnight to remove chlorophyll and subsequently they were cleared and stored in lactic acid for microscopy.

For each time point five leaves were analyzed by drawing the abaxial epidermis at two positions in the leaf, as described previously (58). The following parameters were determined: total area of all cells in the drawing, total number of cells and number of guard cells. From these data, we calculated the average cell area. We estimated the total number of cells per leaf by dividing the leaf area by the average cell area (averaged between the apical and basal positions). Finally, average cell division rates for the whole leaf were determined as the slope of the log2-transformed number of cells per leaf, which was done using five-point differentiation formulas (59).

2.4.9 FLP/MYB88 ChIP experiment

Polyclonal antibodies against the FLP/MYB88 proteins were generated by inoculating rabbits with Ni-NTA-affinity purified NHis6-MYB88. ChIP experiments were performed essentially as before (33). In brief, ten-day old shoots of wild type, *flp-1 myb88* (200 mg fresh weight for each) were cross-linked in 1% formaldehyde for 20 minutes by vacuum filtration and the cross-linking reaction was stopped by the addition of 0.1 M glycine (final concentration) for additional 5 minutes. Tissues were ground to the fine powder using mortar and pestle in liquid nitrogen and then suspended in 300 µl of lysis buffer (50 mM HEPES, pH7.5; 150 mM NaCl; 1 mM EDTA, pH8.0; 1% Triton X-100; 0.1% sodium deoxycholate; 0.1% SDS; 1 mM phenylmethanesulphonylfluoride (PMSF); 10 mM sodium butyrate; 1x protein protease inhibitor, Sigma, St. Louis, MO, USA) and sonicated to achieve an average DNA size of 0.3 - 1 kb. The sonication conditions using the Bioruptor (Diagenode, Liège, Belgium) were as follows: at high power; 30 seconds of sonication followed by 30 seconds of break; change ice every ten minutes; 30 minutes in total. After clearing, using 30 µl salmon sperm DNA/protein-A agarose (Millipore, Brussels, Belgium) at 40 °C for at least one hour, the supernatant fractions were incubated respectively with 1 µl FLP/MYB88 rabbit polyclonal antibody or 1 µg rabbit IgG (Abcam, Cambridge, UK) at 40 °C overnight. At the same time, 10% of the supernatant was saved as the input fraction. The chromatin-antibody complex was incubated with salmon sperm DNA/protein-A agarose (Millipore, Brussels, Belgium) at 40 °C for at least 3 hours, washed with lysis buffer, LNDET buffer (0.25 M LiCl; 1% NP40; 1% sodium

deoxycholate and 1 mM EDTA, pH8.0) and TE buffer (10 mM Tris-HCl, pH7.5; 1 mM EDTA, pH8.0) twice respectively, and the complex was reverse cross-linked in elution buffer (1% SDS; 0.1 M NaHCO$_3$; 1 mg/ml proteinase K) overnight at 65 °C. DNA was extracted using the PCR Cleaning Kit (Qiagen, Venlo, The Netherlands). Used primer sequences are listed in Supplemental Table 2.2.

Acknowledgements

Contributions

Steffen Vanneste, Frederik Coppens, Gert Van Isterdael, Freya De Winter, Bert De Rybel, Jiří Friml, Gerrit T.S. Beemster and Tom Beeckman conceived the general idea, isolated higher order mutants and performed cell division related experiments. Stijn Dhondt, Lieven De Veylder and Dirk Inzé generated and provided promoter::GUS:GFP lines. Tyler Donner and Enrico Scarpella conceived and performed vascular-related experiments. Steffen Vanneste, EunKyoung Lee, Fred Sack and Tom Beeckman conceived and performed the stomata related experiments. Zidian Xie and Erich Grotewold conceived and performed ChIP-PCR. Steffen Vanneste, Frederik Coppens and Tom Beeckman wrote the paper with input from all authors.

References

[1] R. Sozzani, *et al. Nature* **466**, 128 (2010).

[2] Z. Xie, *et al. Plant Cell* **22**, 2306 (2010).

[3] A. Satyanarayana, P. Kaldis. *Oncogene* **28**, 2925 (2009).

[4] L. De Veylder, T. Beeckman, D. Inze. *Nat Rev Mol Cell Biol* **8**, 655 (2007).

[5] K. Vandepoele, *et al. Plant Cell* **14**, 903 (2002).

[6] W. Dewitte, *et al. Plant Cell* **15**, 79 (2003).

[7] W. Dewitte, *et al. Proc Natl Acad Sci U S A* **104**, 14537 (2007).

[8] M. Wildwater, *et al. Cell* **123**, 1337 (2005).

[9] E. Caro, M. M. Castellano, C. Gutierrez. *Nature* **447**, 213 (2007).

[10] J. E. Melaragno, B. Mehrotra, A. W. Coleman. *Plant Cell* **5**, 1661 (1993).

[11] M. Menges, S. M. de Jager, W. Gruissem, J. A. H. Murray. *Plant J* **41**, 546 (2005).

[12] J. Boruc, *et al. Plant Physiol* **152**, 553 (2010).

[13] J. Boruc, *et al. Plant Cell* **22**, 1264 (2010).

[14] K. K. Imai, *et al. Plant Cell* **18**, 382 (2006).

[15] T. Yoshizumi, *et al. Plant Cell* **18**, 2452 (2006).

[16] K. Himanen, *et al. Plant Cell* **14**, 2339 (2002).

[17] K. Himanen, *et al. Proc Natl Acad Sci U S A* **101**, 5146 (2004).

[18] S. Vanneste, *et al. Plant Cell* **17**, 3035 (2005).

[19] J. E. Malamy, P. N. Benfey. *Development* **124**, 33 (1997).

[20] P. M. Donnelly, *et al. Dev Biol* **215**, 407 (1999).

[21] G. T. S. Beemster, *et al. J Plant Res* **119**, 43 (2006).

[22] D. C. Boyes, *et al. Plant Cell* **13**, 1499 (2001).

[23] T. J. Donner, I. Sherr, E. Scarpella. *Development* **136**, 3235 (2009).

[24] E. Scarpella, P. Francis, T. Berleth. *Development* **131**, 3445 (2004).

[25] D. Demandolx, J. Davoust. *Journal of Microscopy* **185**, 21 (1997).

[26] D. C. Bergmann, F. D. Sack. *Annu Rev Plant Biol* **58**, 163 (2007).

[27] R. L. Nakamura, *et al. Plant Physiol* **109**, 371 (1995).

[28] M. J. Gardner, *et al. J Exp Bot* **60**, 213 (2009).

[29] V. Boudolf, *et al. Plant Cell* **16**, 945 (2004).

[30] V. Boudolf, *et al. Plant Physiol* **150**, 1482 (2009).

[31] L. B. Lai, *et al. Plant Cell* **17**, 2754 (2005).

[32] K. Ohashi-Ito, D. C. Bergmann. *Plant Cell* **18**, 2493 (2006).

[33] Z. Xie. *Methods Mol Biol* **592**, 1 (2010).

[34] J. P. Reichheld, *et al. Proc Natl Acad Sci U S A* **93**, 13819 (1996).

[35] O. Shaul, *et al. Proc Natl Acad Sci U S A* **93**, 4868 (1996).

[36] Y. Y. Setiady, *et al. Plant J* **8**, 949 (1995).

[37] J. P. Renaudin, *et al. Proc Natl Acad Sci U S A* **91**, 7375 (1994).

[38] D. L. Myster, R. J. Duronio. *Curr Biol* **10**, R302 (2000).

[39] L. A. Buttitta, B. A. Edgar. *Curr Opin Cell Biol* **19**, 697 (2007).

[40] L. Li, H. Vaessin. *Genes Dev* **14**, 147 (2000).

[41] A.-M. Martinez, *et al. Genes Dev* **20**, 501 (2006).

[42] C. G. James, A. Woods, T. M. Underhill, F. Beier. *BMC Mol Biol* **7**, 30 (2006).

[43] S. Sebastian, *et al. Proc Natl Acad Sci U S A* **106**, 4719 (2009).

[44] J. Pan, *et al. Oncogene* **29**, 6245 (2010).

[45] J. M. Alonso, *et al. Science* **301**, 653 (2003).

[46] M. G. Rosso, *et al. Plant Mol Biol* **53**, 247 (2003).

[47] S. Burssens, *et al. Planta* **211**, 623 (2000).

[48] T. Murashige, F. Skoog. *Physiologia Plantarum* **15**, 473 (1962).

[49] M. Sauer, T. Paciorek, E. Benková, J. Friml. *Nat Protoc* **1**, 98 (2006).

[50] M. H. Lauber, *et al. J Cell Biol* **139**, 1485 (1997).

[51] M. Karimi, D. Inze, A. Depicker. *Trends Plant Sci* **7**, 193 (2002).

[52] M. Benhamed, *et al. Plant J* **56**, 493 (2008).

[53] C. Zhang, F. C. Gong, G. M. Lambert, D. W. Galbraith. *Plant Methods* **1**, 7 (2005).

[54] M. G. Sawchuk, P. Head, T. J. Donner, E. Scarpella. *New Phytol* **176**, 560 (2007).

[55] M. G. Sawchuk, T. J. Donner, P. Head, E. Scarpella. *Plant Physiol* **148**, 1908 (2008).

[56] J. Hellemans, *et al. Genome Biol* **8**, R19 (2007).

[57] T. Beeckman, G. Engler. *Plant Mol Biol Rep* **12**, 37 (1994).

[58] L. De Veylder, *et al. Plant Cell* **13**, 1653 (2001).

[59] R. O. Erickson. *Annual Review of Plant Physiology* **27**, 407 (1976).

2.5 Supplemental data
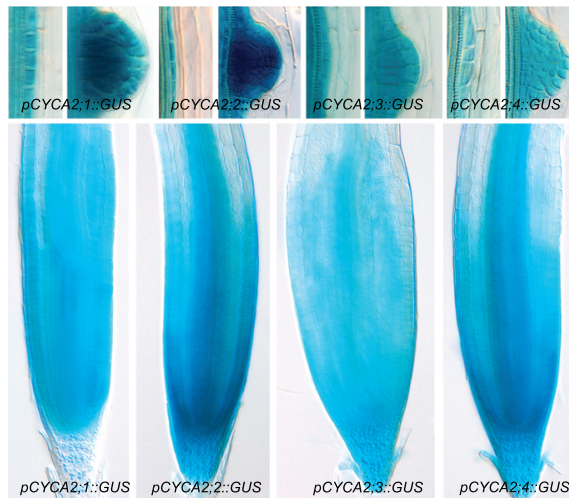
Supplemental Figure 2.1: relates to Figure 1 | Expression patterns of *CYCA2*s in the root. Expression analysis of pCYCA2;1::GUS, pCYCA2;2::GUS:GFP, pCYCA2;3::GUS:GFP and pCYCA2;4::GUS:GFP in shoots, developing lateral roots and root apical meristems.
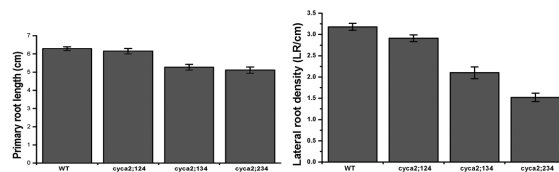
Supplemental Figure 2.2: relates to Figure 1 | Root phenotypes of *cyca2* triple mutants. Primary root lengths and lateral root densities of 10 day old seedlings of WT, *cyca2;124*, *cyca2;134* and *cyca2;234*. Data are presented as mean ± SEM.
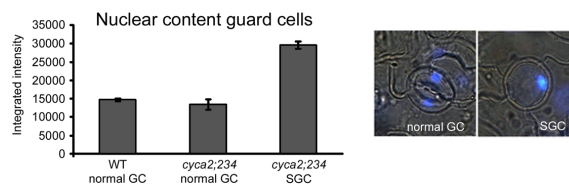
Supplemental Figure 2.3: relates to Figure 4 | Nuclear contents of guard cells. Nuclear content of guard cells (GC) and single guard cells (SGC) in epidermal peels of WT and *cyc2;234*, estimated by integrated intensity of DAPI fluorescence. SGC have double nuclear contents than normal guard cells both in WT as in *cyca2;234*. Data are presented as mean ± SEM.
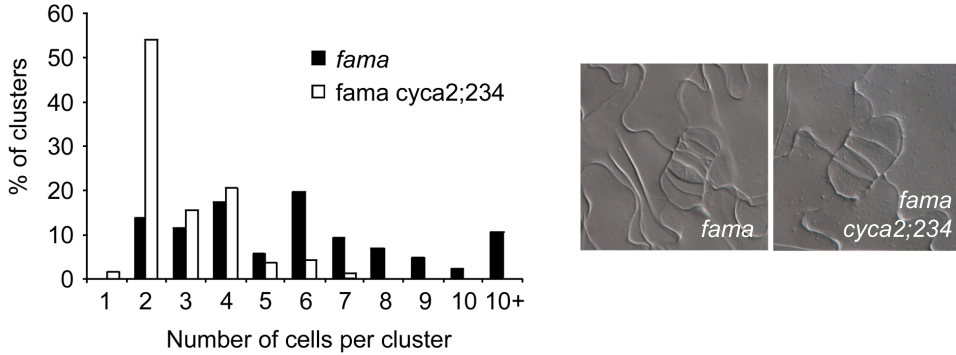
Supplemental Figure 2.4: relates to Figure 6 | Genetic interaction between *fama* and *cyca2;234*. Frequency distribution of number of cells per cluster in *fama* and *fama cyca2;234* in mature cotyledons. The number of cells per cluster is dramatically reduced by *cyc2;234* mutation, but does not give rise to *cyca2;234* single guard cells. Note that these cells do not have any stomatal identity since *FAMA* is required for a guard cell fate.

Supplemental Table 2.1: relates to Figure 4 | Analysis of stomatal phenotypes of various mutant combinations.

Genotype	Counts	Normal	SGC	Normal (%)	SGC (%)
Col-8	501	501	0	100	0
cyca2;1-1 (SALK_121077)	501	501	0	100	0
cyca2;1-2 (SALK_123348)	478	478	0	100	0
cyca2;2-1 (GABI_120D03)	965	965	0	100	0
cyca2;3-1 (SALK_092515)	845	830	15	98	2
cyca2;3-2 (SALK_086463)	960	939	21	98	2
cyca2;3-3 (SALK_043246)	988	975	13	99	1
cyca2;4-1 (SALK_070301)	805	805	0	100	0
cyca2;4-2 (GAT_5.10009)	824	824	0	100	0
cyca2;12 (*cyca2;1-1cyca2;2-1*)	546	546	0	100	0
cyca2;14 (*cyca2;1-1cyca2;4-1*)	464	464	0	100	0
cyca2;24 (*cyca2;2-1cyca2;4-1*)	474	474	0	100	0
cyca2;23 (*cyca2;2-1cyca2;3-1*)	831	699	132	84	16
cyca2;34 (*cyca2;3-1cyca2;4-1*)	1149	913	236	79	21
cyca2;3-2cyca2;4-2	1286	899	387	70	30
cyca2;3-3cyca2;4-2	964	681	283	71	29
cyca2;124 (*cyca2;1-1cyca2;2-1cyca24-1*)	529	529	0	100	0
cyca2;134 (*cyca2;1-1cyca2;3-1cyca2;4-1*)	744	457	287	61	39
cyca2;234 (*cyca2;2-1cyca2;3-1cyca2;4-1*)	734	42	692	6	94

Supplemental Table 2.2: Primer sequences used for genotyping, cloning, ChIP-PCR and qPCR.

Name	Sequence	
Genotyping	**LP**	**RP**
CYCA2;1_1	CAATGATTTGATCACGGAAAC	TTTGAGAAACAAACACTCTGG
CYCA2;1_2	GCTTGTTGAATCAGTGGAGTG	TGTATTCCTACCTCCACGAGC
CYCA2;2_2	ATCCGTGTATGCGTGTGATTC	AGGAGCTCTAATGGCGGTAAG
CYCA2;3_1	AACATTTTTATATGCATGGTTGC	TCGATGTCTACAAATTTTGGG
LB_GABI-KATa	CCCATTTGGACGTGAATGTAGACAC	
LB_SALK	GTGGACCGCTTGCTGCAACTCTCT	
LB_EXOTIC	CCGTCCCGCAAGTTAAATATG	
Cloning (HTA6:EYFP)	**FW**	**RW**
pCYCA2;1	GGGGACAAGTTTGTACAAAAAAGCAGGC	GGGGACCACTTTGTACAAGAAAGCTGGG
	TGGAAGTGACTAGCAGGATTCG	TCACTCCACTGATTCAACAAG
pCYCA2;4	GGGGACAAGTTTGTACAAAAAAGCAGGC	GGGGACCACTTTGTACAAGAAAGCTGGG
	TAAAGACCACGGGAGCGTCGT	TGATTGAAACCCTTTGACAC
pMYB88	GGGGACAAGTTTGTACAAAAAAGCAGGC	GGGGACCACTTTGTACAAGAAAGCTGGG
	TAAATTTCTTCTAACTTGGCCTCTGATA	TTAAAAAGTTTTGGCCTTTCTCTCTCTC
ChIP_PCR	**FW**	**RW**
CYCA2;3_P1	CTCACGGGTTGATCTCCC	GAGACACGCACACACTAGTGTG
CYCA2;3_P2	GCTGGACAAGATCTTCTGC	GAGAATCTGGGGTTATTATTTGTCTAC
CYCA2;3_P3	CGTCCATCTCTCGTTAGTATTTTAGAG	CCAGCTTCGTTCTCTTTGC
CYCA2;3_P4	TACTCACGTGATTGTACTGTAG	CTCTAAAATACTAACGAGAGAATGGAC
qPCR	**FW**	**RW**
Q_CYCA2;1	GAGTCACTAGATCAAGAGCTAAGGC	TAGTTTCTTTACCACCTCGCTTGC
Q_CYCA2;2	GGAAGGTGCTAACAGTGCCATG	ACCTGATCCATCTTCTTGTTGCTC
Q_CYCA2;3	GTTCCCTTGCCTCTGCTTTGC	GCTCGCTTCTTCCTCTTGTTGG
Q_CYCA2;4	TTCAGTGTTGCGGTGGAGAATATC	TTGTGAAGTAGCCGAAGAAGATGC
Q_GATA23	AGTGAGAATGAAAGAAGAGAAGGG	CGTGGCTGGAATAATATGAAATACC
Q_IAA19	GTGGTGACGCTGAGAAGGTT	CGTGGTCGAAGCTTCCTTAC
Q_IAA5	TCCAAGGAAACATTTCCCAAG	CCGGAGAAAGAAACAGTCTCG
Q_IAA1	ACCGACCAACATCCAATCTC	TGGACGGAGCTCCATATCTC
Q_PLT3	CGGCGAAATGCAGCTTCTGACTC	GGTGCCATAAGTCCCATTGTTCCC
Q_CYCD3;1	TTCGTTCGTAGCACCATTATCAGG	CGGAGATTACGAGCAGGAGGAGAC
Q_E2Fa	GGTGTTGTAGATCGGGAGGAAG	CTAGGAGTTTGACAGGGACTGC
Q_CYCB1;1	CCTGGTGGAGTCGGTTGATTGATG	CGACATGAGAAGAGCACTGAGAC
Q_CYCB1;3	TTATTGGAGTCGAAGGAGGAAAGC	CTTTAGCAAACCTCGTTTGTATCAGG
Q_CYCB2;3	AGTTTCAAGTCGAAAGAACCAGAGACG	CAACCGACAGCCAAAGGATTATTC
Q_CDKB2;1	CGAGTGTGGGGATTTATGTC	CGCTGGGGAGAGTGAATG
Q_CDKB2;2	CATCAGAAGCTTTCGTCAAG	TTTGAGCGTCATTGTCTTCC
Q_PEX4	GCGACTCAGGGAATCTTCTAAG	AGCAAGAGGATCATCATCATCATC
Q_UNKN_1	GGAGATTAAGCAACCTGAGGAGTG	GTGGTGGTGTGGAGGAGAC

3

Functional analysis of new putative cell proliferation regulators

Frederik Coppens[1,2], Daniël Van Damme[1,2], Steffen Vanneste[1,2], Dirk Inzé[1,2], and Gerrit T.S. Beemster[1,2,3]

[1]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), 9052 Ghent, Belgium
[2]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Ghent, Belgium
[3]Department of Biology, University of Antwerp, 2020 Antwerp, Belgium

Summary

Growth and development of plants is driven by two processes: cell division and cell expansion. Plants go through a strict developmental program in which the precise regulation of proliferation and expansion is crucial. Additionally, environmental cues need to be incorporated to adapt growth and development to changing conditions. While the primary components of the core cell cycle machinery are well established, our knowledge of the upstream regulators is still limited.

In this research project, we identified proliferation specific genes, based on microarray data of both leaf and root. We specifically selected genes with unknown function. For all genes, the proliferation specific expression pattern could be validated by qPCR. We focused on two small gene families in *Arabidopsis*.

The first family consists of four plant-specific genes that code for proteins with four transmembrane domains, hence named 4TM. We show that these genes are mitotically expressed and localize to the forming cell plate, vesicles and the plasma membrane. They co-localize with the syntaxin KNOLLE, which functions in vesicle fusion. Upon downregulation of the 4TM genes through an artificial microRNA, aberrant cell divisions occur and unfinished, wavy cell plates are formed. Therefore, we hypothesize that this novel, proliferation specific gene family is involved in the formation of the cell plate during cytokinesis.

Two genes with 3 High Mobility Group (HMG) boxes make up the second gene family we studied. This is a subfamily of the HMGB family and the only HMG genes in *Arabidopsis* that contain 3 HMG boxes that function in DNA binding. Orthologues of these genes are only found in plants. Also for this gene family, we showed a mitosis specific expression but their function in proliferation remains elusive.

Using microarray data as a starting point, we successfully identified unknown proliferation specific genes. The initial functional analysis confirmed that these genes are likely important for the regulation of specific processes during mitosis of plant cells.

3.1 Identification of putative new cell cycle regulators

3.1.1 Selection of proliferative genes

To select proliferation specific genes, we analyzed genome-wide expression data from samples of the three defined stages (proliferation, expansion and mature) during leaf (1) and root development (2). To select for genes that are specifically expressed during proliferation and that decrease rapidly in subsequent phases, we applied a number of criteria: (A) expression above the background in proliferating tissue, (B) differential expression (FDR < 0.05), (C) maximum expression during proliferation, (D) at least 1.5 fold change between the minimum and maximum expression, (E) the fold change between proliferation and expansion must be higher than the fold change from expansion to mature. These criteria were applied to both the leaf and root dataset and only genes that have a proliferation specific expression pattern in both organs were selected. This analysis yielded a long-list of 391 genes which were ranked according to fold change during leaf development. To verify the specificity of the criteria we assessed the presence of known core cell cycle genes in our selection (3). On the ATH1 microarray, the Affymetrix CDF identifies 22592 genes of which 55 are core cell cycle genes (0.24%). The list of genes we selected contained 13 of these core cell cycle genes while at random only a single one was expected (χ^2 test p-value 4e-35). This confirms that we identified genes that are specifically expressed during proliferation. Further selection for functional analysis

was performed based upon the availability of T-DNA insertion lines and on prior knowledge on the function of the gene, based on The *Arabidopsis* Information Resource (TAIR, www.arabidopsis.org), selecting for genes of unknown function (4, 5). An initial selection of 10 genes was made (Table 3.1).

Table 3.1: Selected genes that are specifically expressed during proliferation and have unknown function. The 'Name' was used internally, the name according to TAIR is provided when applicable. The log2-expression at the three developmental stages (proliferation (Prol), expansion (Exp) and mature (Mat)) in the leaf as well as the linear fold change (FC) between minimum and maximum values are listed. If the gene was selected in the initial selection of 10 genes, this is marked by an asterisk in the Selection (S) column.

AGI	Name	TAIR	Prol	Exp	Mat	FC	S
AT5G16250	4TM1		8.51	4.53	3.84	25.38	*
AT3G02640	4TM2		8.23	4.06	3.25	31.45	*
AT5G36710	4TM3a		7.76	4.20	3.73	16.37	*
AT5G36800	4TM3b						
AT4G11080	3xHMG-box1		5.90	3.30	3.55	6.03	
AT4G23800	3xHMG-box2		7.53	2.91	2.75	27.43	*
AT5G23420		HMGB6	5.60	3.62	2.87	6.63	
AT2G25060		ENODL14	8.62	4.85	4.06	23.64	*
AT4G31840		ENODL15	8.58	5.53	4.38	18.45	*
AT1G09450	HASP		4.81	3.02	2.86	3.87	*
AT2G32590	BAR		5.62	3.12	3.19	5.66	*
AT4G05520		ATEHD2	6.90	4.06	4.59	7.13	*
AT3G42660	WD40		5.85	5.00	4.12	3.33	*

Within the top-ranked genes, three homologous genes were present: *AT5G16250*, *AT3G02640* and *AT5G36710*. One additional homologue is present in *Arabidopsis*, *AT5G36800*, which is identical to *AT5G36710* and this gene was therefore added to the selection. These genes have four transmembrane domains and are hence referred to as *4TM1* (*AT5G16250*), *4TM2* (*AT3G02640*), *4TM3a* (*AT5G36710*) and *4TM3b* (*AT5G36800*).

Our initial selection contained a gene encoding for a High Mobility Group (HMG) gene of the HMG-box (HMGB) subfamily. In the long-list of 391 proliferative genes, two additional HMGB genes are present and these genes were therefore added to the selection. *AT5G23420* contains one HMG-box and is known as *HMGB6*. The other two genes, *AT4G11080* and *AT4G23800*, contain three HMG-boxes and we propose to name them *3xHMG-box1* and *2*, respectively, conforming with previously used terminology (6).

The selection of proliferation specific genes also contained two *EARLY NODULIN-LIKE PROTEIN*s (*ENODL14* and *ENODL15*). These plastocyanin-like domain-containing proteins belong to a family with 22 members in *Arabidopsis* (7). Both genes contain a glycosylphosphatidylinositol (GPI) anchor, indicating they are associated with a membrane (8). *HASP* is a homolog of *haploid germ cell specific nuclear protein* kinase (*haspin*), a histone H3 kinase that acts upon Thr3 (9, 10, 11). *BAR*, named after the *Drosophila* condensin I subunit BARREN/CAP-H protein which is required for sister-chromatid segregation in mitosis. In *Arabidopsis* it localizes to the cytoplasm during interphase and to chromosomes from prometaphase to cytokinesis (12). *ATEHD2* contains an Eps15 homology (EH) domain, known to be involved in protein-protein interactions, a calcium binding and GTPase domain. The last gene that was selected contains several WD40 or beta-transducin repeats, 40 amino acid motifs often terminated by a Trp-Asp (W-D) dipeptide. WD40-repeat containing proteins belong to a large family found in all eukaryotes and function e.g. in signal transduction, cell cycle control and apoptosis. The WD40-repeats are important for protein-protein interactions and the assembly of protein complexes.

3.1.2   Validation of proliferation specific expression profile

The expression profile during leaf development of the selected genes (Table 3.1) was verified with quantitative PCR (qPCR) in a leaf developmental series that included 7 rather than the three data points that were used for the microarray analysis, allowing for a much better resolution. For this, *Arabidopsis* plants were grown *in vitro* and the first leaf pair was harvested 8, 10, 12, 14, 16, 19 and 24 days after stratification. This time series encompasses proliferation (day 8), the transition from proliferation to expansion (day 10), expansion (day 14 and 16) and maturity (day 19 and 24). For all genes analyzed, the proliferation specific expression pattern could be confirmed (Figure 3.1).

All *4TM* members showed a sharply declining expression profile with less than 10% of the expression level remaining after proliferation (Figure 3.1A). The coding sequence for *4TM3a* and *4TM3b* is identical, which prevents the development of primers specific for either. Therefore, expression can only be assessed for both genes together. However, these two genes are part of a mistakenly annotated tandem duplication and actually correspond to one single gene (see 3.2.4). For the two *3xHMG-box* genes the expression profile is rapidly declining (similar to the *4TMs*), while *HMGB6* levels decrease more gradually (Figure 3.1B). A highly proliferation specific pattern is also observed for *ENODL14*, *ENODL15* and *HASP* (Figure 3.1C) while the expression of *BAR*, *ATEHD2* and *WD40* decreases more gradually (Figure 3.1D). For *WD40* even a slight peak at transition from proliferation to expansion (day 10) can be observed. All expression patterns confirm that our criteria allowed selecting for highly proliferation specific genes.
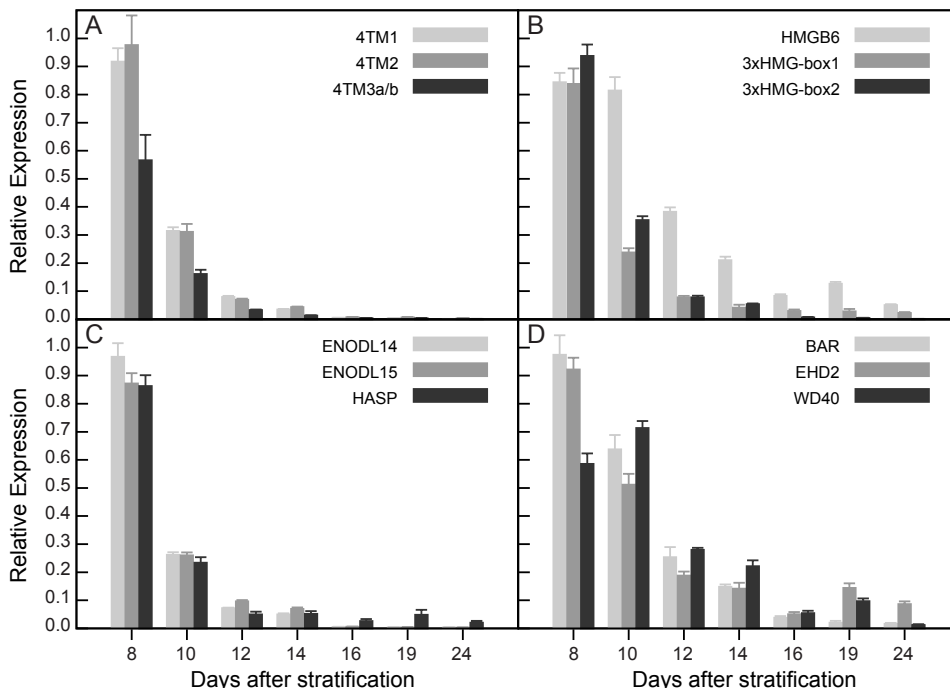
Figure 3.1: Expression profile during leaf development of selected proliferation specific genes. (A) *4TM* family (B) *HMGB* family (C) *ENODL14*, *ENODL15* and *HASP* (D) *BAR*, *EHD2*, *WD40*. Data are represented as mean relative expression for two biological replicates ± SEM, rescaled to a maximum of 1.

3.1.3 Proliferative genes: description and final selection

The gene with the highest fold change that fulfilled all criteria was *4TM2*, an unknown protein-coding gene. A homology search using the basic local alignment search tool (BLAST, 13), revealed that this protein has high similarity to three other genes in *Arabidopsis*: *4TM1* (*AT5G16250*), *4TM3a* (*AT5G36710*) and *4TM3b* (*AT5G36800*). *4TM1* and *4TM3a* were also present in our selection and ranked third and tenth according to fold change in all proliferative genes (391 in total). *4TM3b* is fully identical on the nucleotide level to *4TM3a* and consequently it is impossible to make a distinction between both genes based on probe hybridization in the microarray data. These two genes are part of a larger, 59.3 kb tandem duplication on chromosome 5. Towards functionality, the only recognizable feature of this gene family is the presence of four transmembrane domains, therefore the genes were named *4TM* genes.

The tetraspanin family also contains four transmembrane domains and one of its members, *TORNADO2*, is implicated in patterning during early leaf development (14). A tetraspanin gene contains four transmembrane domains, a short first extracellular loop and a longer second extracellular loop with a conserved CCG protein motif (15). This second extracellular loop is not present in the *4TM* family, therefore this family is distinct from the tetraspanins.

Due to the extremely high fold change of this whole family in the proliferation phase and their completely unknown function, we decided to study the function in more detail. As a first step, we profiled the expression in a synchronized cell culture and developed and studied transcriptional and translational Green Fluorescent Protein (GFP) fusions. The functional analysis was initiated using loss-of-function and gain-of-function mutations of the genes (see Chapter 3.2).

Ranked second, based on fold change, was *3xHMG-box2* (*AT4G23800*). This gene belongs to the High Mobility Group (HMG) family of transcription factors, more specifically to the HMG-box (HMGB) family and contains three HMG-boxes (6). The *Arabidopsis* genome encodes a second HMGB-protein with three HMG-boxes, 3xHMG-box1 (AT4G11080), which was also present in the long-list of proliferative genes. In collaboration with Dr. Klaus Grasser we investigated the function of the two *3xHMG-box* genes. In this context, we performed expression analysis in a synchronized cell culture, developed a transcriptional fusion with GFP and analyzed the spatiotemporal expression. We specifically selected these two genes because, despite the fact that they belong to the large HMGB gene family, they are distinct from other members of this family due to the presence of 3 (instead of 1) HMG-boxes.

The remaining genes were not retained for further research due to the lack of a functional knockout and/or obvious phenotypic differences compared to the wild type. While also the genes we did select do not fully comply to these criteria, their novel characteristics described above were key to their selection. I will briefly discuss each of the genes we did not retain here.

In the long-list, a third HMGB gene, *HMGB6* (*AT5G23420*), was also present. For *HMGB6* we analyzed three SALK lines (16): SALK_044697, SALK_044693 and SALK_138632. In the first line we could not detect a T-DNA insertion in the genome but in the others the insertion was validated and localized to the first (200 bp downstream of the start codon) and third intron (536 bp downstream of the start codon), respectively. Using qPCR we established that homozygous plant of both these lines do not express *HMGB6* above the detection limit. However, no visible phenotypic changes could be observed and therefore this gene was not studied further.

For the *ENODL* genes, four T-DNA insertion lines were analyzed: SALK_019896 and SALK_025894 for *ENODL14* and SALK_036851 and SALK_093027 for *ENODL15*. Based on PCR analysis, no T-DNA insert could be found in the proximity of *ENODL15* for SALK_036851. The insertion could be mapped to the 5′ UTR for SALK_019896 and SALK_025894 (73 and 410 bp upstream of the start codon, respec-

tively) and to just after the stop codon for SALK_093027. Among these lines, only SALK_019896 had a decreased expression of *ENODL14* but no statistically significant phenotypic changes were observed for parameters such as leaf size, cell size and endoreduplication at maturity. Therefore, these genes were not investigated further.

In *HASP* only one SALK line was available and we could confirm that a T-DNA is inserted in the second intron (282 bp downstream of ATG). The mRNA level was however not decreased and it could be shown that the portion of the gene, containing the kinase domain, downstream of the insertion is constitutively expressed likely driven by the *Npt* promoter in the T-DNA insert (data not shown). Additionally, two GABI-KAT lines were phenotypically analyzed (GABI_320A05 and GABI_858F01), but also here no obvious phenotypic changes were observed and therefore this gene was not investigated further. Recently, it has been shown that the phosphorylation of histone H3 at Thr3 by Haspin is required for accumulation of Aurora B at centromeres in human cell lines (17). Aurora B is a component of the complex required for correct spindle-kinetochore attachment during chromosome segregation and for cytokinesis (18, 19). This confirms that our selection procedure has identified important cell cycle regulators.

From the three T-DNA insertion lines available for *BAR* (SALK_017766, SALK_013559, SALK_072400), line SALK_013559 had a severe phenotype, reminiscent of the constitutive ethylene response in *ctr1* mutants (20), while the others had no visible phenotype. However, there was no correlation between the phenotype and the expression level of *BAR*, indicating that the phenotype is caused by an independent mutation. Therefore, we did not retain this gene for further analysis.

For *ATEHD2*, none of the T-DNA insertion lines (SALK_046090, SALK_056796, SALK_046090, SALK_096593) showed phenotypic alterations, hence no further research was conducted. Recently, this gene has been implicated in endocytosis in *Arabidopsis* and reorganization of actin (21, 22). A link with the cell cycle was not addressed yet, but both processes are involved in cytokinesis (23, 24).

For *WD40*, we mapped the T-DNA insertion to an exon of this gene (966 bp downstream from the start codon). This insertion resulted in a strong decrease of the expression (40-fold). Mutant plants showed a highly variable leaf area between different experiments (data not shown), probably due to susceptibility to environmental factors. As this was the only mutant line available and such variability complicates research, we decided to focus on other genes.

3.2 Transmembrane proteins

3.2.1 4TMs are expressed in M-phase

With the validation of the microarray results by qPCR, we established that the selected genes of the 4TM family are indeed specifically expressed during proliferation (see 3.1.2). In order to assess if the expression changes during the cell cycle, we analyzed the transcript levels in an aphidicolin synchronized *Arabidopsis* cell culture. Addition of aphidicolin causes a block in early S-phase and upon release the majority of the cells goes through the cell cycle simultaneously (25). Collection of cells every two hours allows for profiling of the expression of genes during the different phases of the cell cycle (Figure 3.2). As controls *Histon H4* and *KNOLLE* (*KN*) were used as markers for G1/S-phase and M-phase, respectively (26, 27, 28). The expression of all *4TM* genes is very similar to that of *KN*, indicating a mitosis specific expression, with a 50-fold increase from G2 to M.
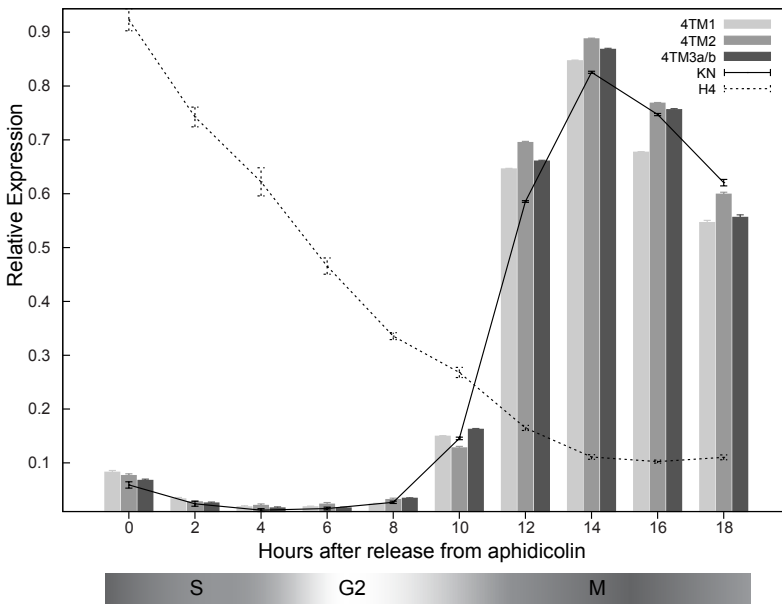
Figure 3.2: Mitosis specific expression of *4TM* genes during the cell cycle. The expression was profiled using an aphidicolin synchronized cell culture from release from aphidicolin for 18h. *Histone H4* (*H4*) and *KNOLLE* (*KN*) were used as markers for G1/S and M, respectively. The timing of the phases of the cell cycle is indicated below the graph. Data are represented as mean relative expression for two biological replicates $\pm$ SEM, rescaled to a maximum of 1.

3.2.2 *4TM*s are plant specific and conserved in all angiosperms

Using Position-Specific Iterated BLAST (PSI-BLAST) (29), we searched for homologues of the 4TM proteins in other species. Remarkably, we only identified homologues in other plants. For in-depth analysis of the homologues, we used PLAZA, a comparative genomics resource that integrates structural and functional annotation of 23 plants: 11 dicots, 5 monocots, 2 mosses and 5 algae (bioinformatics.psb.ugent.be/plaza/, 30). In all angiosperms included in PLAZA, homologues were found but not in evolutionary older plants. Both in dicotyledonous species and monocots homologues were found, but the phylogenetic tree reveals that monocots make up a separate clade most related to 4TM3a/b (Figure 3.3).
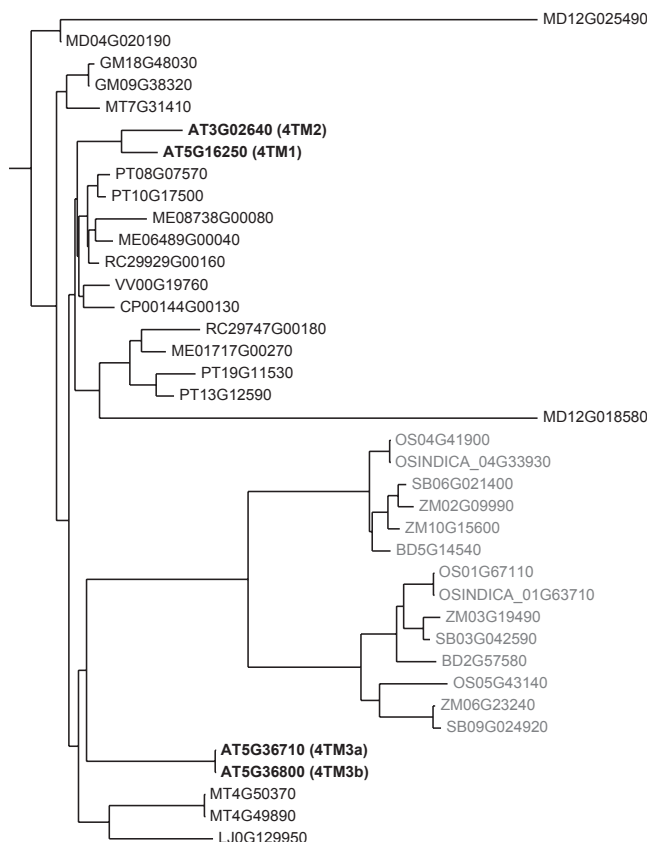
Figure 3.3: Phylogenetic tree of the homologous gene family of 4TM. Included species: *Malus domestica* (MD), *Glycine max* (GM), *Medicago truncatula* (MT), *Arabidopsis thaliana* (AT), *Populus trichocarpa* (PT), *Manihot esculenta* (ME), *Ricinus communis* (RC), *Vitis vinifera* (VV), *Carica papaya* (CP), *Oryza sativa* ssp. japonica (OS), *Oryza sativa* ssp. indica (OSINDICA), *Sorghum bicolor* (SB), *Zea mays* (ZM), *Brachypodium distachyon* (BD) and *Lotus japonicus* (LJ). Monocot species are indicated in grey, the selected genes of *Arabidopsis* in bold (based upon PLAZA, bioinformatics.psb.ugent.be/plaza/, 30).

3.2.3   Four transmembrane domains is the only known feature for the 4TM family

The *4TM* genes are 552 (*4TM1*, *4TM3a/b*) or 558 (*4TM2*) nucleotides in length and do not contain introns. They encode for proteins containing 183 (*4TM1*, *4TM3a/b*) or 185 (*4TM2*) amino acids. Protein sequence alignment using Clustal W (31) showed that 4TM1 and 4TM2 are 79% identical while the identity with 4TM3a/b is 60% and 57%, respectively (Figure 3.4). As indicated previously, 4TM3a and 4TM3b are identical. As a starting point for the functional analysis we searched for known domains in the protein sequences using InterProScan (32). This identified 4 transmembrane domains in both 4TM1 and 4TM2. For 4TM1 a signal-peptide was recognized in the first 63 amino acids, while 4TM3 contained a prokaryotic lipoprotein in the first 28 amino acids. Remarkably, 4TM3 did not yield transmembrane domains while the sequence identity is highest in these specific regions (Figure 5). Using algorithms specifically designed to predict transmembrane domains, four such domains were found in all *4TM* genes (TMHMM (33), HMMTOP (34), TMpred (35), TopPred (36)). Based on the sequence characteristics, we conclude that all 4TM family members contain 4 transmembrane domains and are therefore likely located in a membrane. As the signal-peptide and lipoprotein could only be identified in one of the members of the 4TM family, we believe this is an artifact of the prediction algorithms.

```
CLUSTAL 2.0.12 multiple sequence alignment

4TM3a   MGMS--------KSKGNTHNIFLLCNYILLGSASSCIFLTISLRLFPSLSGLSLIFLYTL 52
4TM3b   MGMS--------KSKGNTHNIFLLCNYILLGSASSCIFLTISLRLFPSLSGLSLIFLYTL 52
4TM2    MGLIPQPQESIQESHYYTHKLFLTANYVLLGASSSCIFLTLSLRLIPSLCGFFLILLHAT 60
4TM1    MGFISSSS-PVEESHYHTHKIFLFSNYILLGAASSCIFLTLSLRLIPSICGFLLILLHAT 59
        **:.          :*:  **::**  .**:***::*******:****:**:.*: **:*::

4TM3a   TIATAVSGCSIFASSTSATASDRLYGSHMVATVLTAIFQGAVSVLIFTRTGDFLRFLKSY 112
4TM3b   TIATAVSGCSIFASSTSATASDRLYGSHMVATVLTAIFQGAVSVLIFTRTGDFLRFLKSY 112
4TM2    TIAAAVSGC-----AAASYGKNRWYAAHMIATVLTAIFQGSVSVLIFTNTSNFLESLNSY 115
4TM1    TIAAAVSGC-----AAASCGRNRWYAAHMVATVLTAIFQGSVSVLIFTNTSKFLGSLKSY 114
        ***:*****       ::::  . :*  *.:**:**********:*******.*..**   *:**

4TM3a   VREEDGEVILKLSGGLCVLMFCLEWIVLVLAFLLKYSDYLDE---SVVDDDDFKVRRQEE 169
4TM3b   VREEDGEVILKLSGGLCVLMFCLEWIVLVLAFLLKYSDYLDE---SVVDDDDFKVRRQEE 169
4TM2    VREKEASMILKLAGGLCVVIFCLEWIVLVLAFFLKYYAYVDGDNNGVAMKRTGKVQ-SEE 174
4TM1    VREEDAAVILKLGGGLCIVIFCLDWIVLVCAFFLKYYAYVDG-GDGVAMKRTGKVQ-SEE 172
        ***::. .****.****:::***:***** **:***  *:*    .*. .    **: .**

4TM3a   DLKDWPSYPFQLKI 183
4TM3b   DLKDWPSYPFQLKI 183
4TM2    TLKNSP-WAFQV-- 185
4TM1    NPKDWP-WPFQV-- 183
         *: *  :.**:
```

Figure 3.4: Sequence alignment of 4TM family genes. The amino acid residues that are identical in all 4 sequences are indicated with an asterisk, a colon and dot denote conserved and semi-conserved residues, respectively. The grey bars indicate the transmembrane domains as predicted by TMHMM.

3.2.4 *4TM3a* and *4TM3b* are erroneously annotated as part of a tandem duplication

Two of the members of this gene family, *4TM3a* and *4TM3b*, are identical in sequence. The genes are located in a 59.3 kb tandem duplication, which was already observed with the first publication of the sequence of chromosome 5 (37). The two copies of the tandem duplication differ by one base, an insertion of a G at position 15403 of the second copy. While a nearly identical duplication is possible, it is highly unlikely. Therefore, we investigated this region in more detail.

First, we verified the alignment of the bacterial artificial chromosome (BAC) clones that were used to assemble this region. According to the data available on TAIR, three BAC clones were used: F24C7, MPK17 and F5H8. These are schematically represented in Figure 3.5A, the symbols represent 100 bp at the termini. Next, we matched the sequences of the ends on all BACs (Figure 3.5B). This shows that the end of F24C7 is also present internally in this BAC clone, as well as in MPK17. We verified that the sequence delineated by the same symbols (e.g. open circle and filled square) is identical in all instances. The only exception to this is the insertion of G, marked by the 'x'. The two 4TM genes are located between the open triangle and filled circle. In Figure 3.5C the alignment that was made in the *Arabidopsis* genome is shown. An alternative alignment is constructed in Figure 3.5D which does not contain the duplication. This assumes that the BAC clone F24C7 is a chimera. This does however not provide evidence for the absence of the duplication. The only available BAC clone that encompasses the junction of the two tandem duplications is F24C7, identification of this junction in other, not publicly available, BAC clones could confirm the presence of the duplication.

As an alternative approach, we tried to amplify the region of the junction of the duplication (Figure 3.5E). Only the combination of primers 1-2 and 5-6 yielded a PCR product (data not shown, list of primers in Supplemental table 3.1). These fragments are located just before and after the junction, while all other combinations contain the junction. While the absence of PCR product does not prove the absence of the duplication, it is consistent with the alternative assembly suggested in Figure 3.5D.

Previously, genome deletions in the *grandifolia-D* mutants were mapped using tiling arrays on genomic DNA (38). We revisited the transcriptome data obtained for the WT (Figure 3.5F) which shows a clear decrease in the signal at chromosome 5 at the position of the duplication (dark grey boxes). If the duplication is an assembly artifact as suggested in Figure 3.5D, the corresponding DNA signal from this specific region is spread over the two copies of the probes on the array because they are based on the official assembly (Figure 3.5C). As a consequence this would yield a 50% lower signal in this region. At this position in the chromosome, a decreased signal is indeed observed (Figure 3.5F). The low signal just before the duplication can be explained by the presence of a heterochromatic knob (37).

Finally, we analyzed T-DNA insertion lines (see further) in 4TM3a/b. Because both genes are identical, it cannot be determined in which of the two genes the T-DNA is inserted. For SALK line SALK_101711, we could determine the insert site at 275 bp downstream of the stop codon. In the segregating population, wild type plants as well as mutants with a heterozygous and homozygous insertion could be identified by PCR on genomic DNA. This confirms that the insertion is at a locus for which there is no duplicate. If the duplication would be present, 'homozygous mutants' would not exist as this would require the insertion of the T-DNA in the exact same place in both copies of the duplication. However, the identification of homozygous mutants is based on the absence of PCR product for the WT region (16). Although the genotyping was confirmed in an independent experiment, it cannot be excluded that, for unknown reasons, this PCR reaction failed repeatedly. With the data that is publicly available we cannot exclude the presence of the duplication in the genome, there are however several lines of evidence that suggest this duplication is an artifact of the BAC alignment. As we can anyhow not distinguish between *4TM3a* and *4TM3b*, we will assume that there are only 3 *4TM* genes and use '*4TM3*' instead of *4TM3a/b*.

### 3.2.5   4TM localizes to proliferative tissue

To validate the expression of the *4TM* genes, transcriptional fusions with green fluorescent protein (GFP) and β-glucuronidase (GUS) were made. To enhance the fluorescent signal, a nuclear localization signal (NLS) was added (p4TM::NLS:GFP:GUS). The promoter sequences were obtained from the SAP *Arabidopsis* Promoterome Database (39). For *4TM1* and *4TM2*, 391 and 649 bp upstream of the start codon were identified as the promoter and successfully cloned using Gateway constructs (40). For *4TM3a* and *4TM3b* the predicted promoter was much longer: 3618 and 1997 bp, respectively. Why a different length was identified for

---

Figure 3.5 *(on the next page)*: **Evidence for the absence of the 59.3 kb tandem duplication on chromosome 5.** (A) Bacterial artificial chromosome (BAC) clones used as annotation units for this region. The symbols represent 100 bp at the termini of the BAC clones (the illustrations are not to scale). (B) Localization of the termini of the BAC clones in the entire sequence of all BAC clones. The sequence in between the same two symbols is always also identical, with the exception of the insertion of a single G marked by the 'x'. (C) The alignment of the BAC clones in the genome of *Arabidopsis*, the dashed line indicates the location of the duplication. (D) An alternative alignment without the duplication assuming that BAC clone F24C7 is a chimera. (E) Illustration of the PCR reactions that were performed to try to amplify the junction of the tandem duplication. (F) Tiling array data for genomic DNA of chromosome 5, the dark grey boxes indicate the location of the duplication.
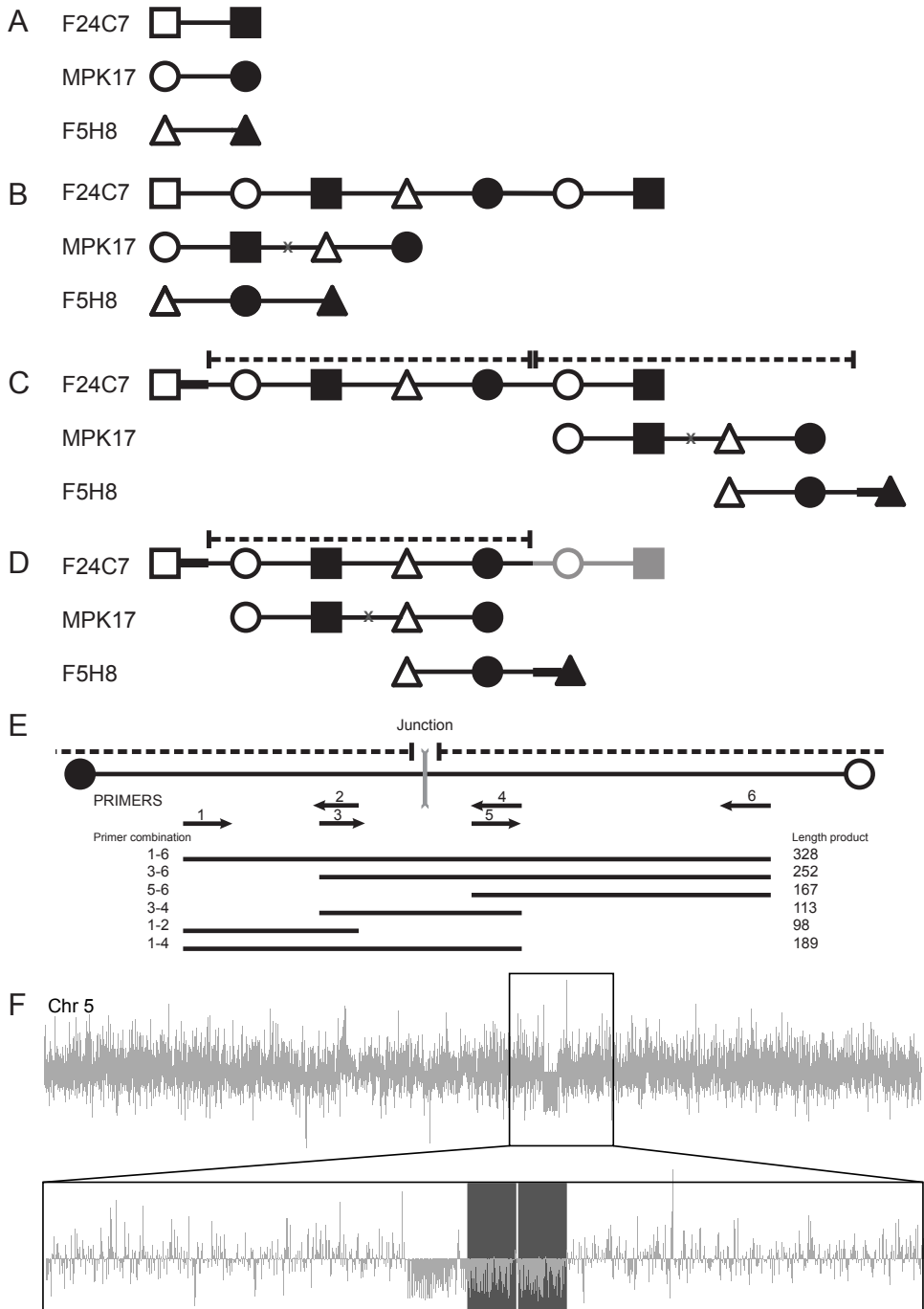
these two genes is unclear. However, for both sequences the promoter could not be amplified and therefore their cloning was abandoned and we continued with the promoters of *4TM1* and *4TM2*. For both construct, three independent transformants carrying a single locus insertion in homozygous state were analyzed. All analyzed plants had GFP expression specifically in proliferating tissue such as the root tip, lateral root primordia and young developing leaves (Figure 3.6A-C). Later in leaf development, expression is limited to the guard cells and small cells of the stomatal lineage (Figure 3.6D). This demonstrates that the promoter sequences that were used are indeed conferring proliferation specificity to the *4TM* genes. The transcriptional fusions had high fluorescence levels, suggesting very high expression levels during mitosis. Moreover, analysis of GUS staining was not possible due to nearly instantaneous over-staining of the tissue (within seconds).
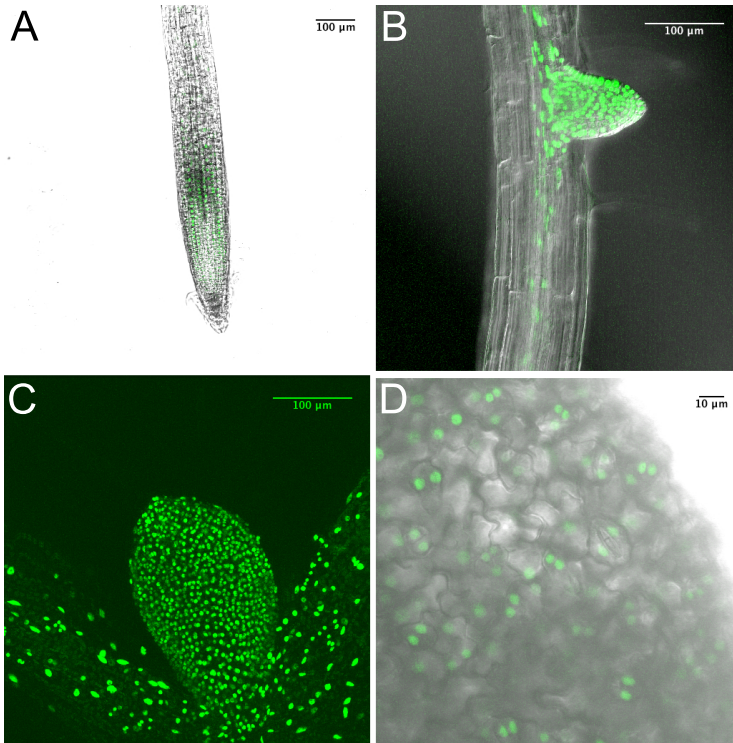
Figure 3.6: Transcriptional fusion of 4TM with nuclear localized GFP and GUS: p4TM1::NLS:GFP:GUS (B, C), p4TM2::NLS:GFP:GUS (A, D). (A) Fluorescent signal is observed in the root tip, (B) lateral root primordium, (C) leaf primordium (leaf 1 and 2, 5 days after stratification) and (D) guard cells of older leaves (leaf 1 and 2, 12 days after stratification).

3.2.6 4TM localizes to the newly forming cell wall after mitosis

Next, we wanted to assess where the 4TM proteins localize subcellularly. Translational fusions of the coding sequence of *4TM1* and *4TM2* with GFP, both N- and C-terminal were expressed in *Arabidopsis* using the 35S constitutive promoter (Figure 3.7A, B). For *4TM3* cloning of the coding sequence did not succeed, despite several attempts, therefore we continued with 4TM1 and 4TM2. For each construct, at least three independent, homozygous transformants with a single insert site were analyzed. The C-terminal fusion of both 4TM1 and 4TM2 had GFP signal at the plasma membrane, the newly formed cell plate during mitosis and showed punctate expression in the cytoplasm (Figure 3.7A). Both N-terminal fusion also had fluorescent signal at the plasma membrane and spots in the cytoplasm, but GFP was not associated with the cell plate (Figure 3.7B). We hypothesize that localization of the protein is affected by the N-terminal fusion with GFP, likely because this disrupts the functionality of an N-terminal localization signal. Plants carrying the N-terminal fusions also showed spots with very intense fluorescence (Figure 3.7B), which we believe are an artifact of the mis-localization due to aggregation of the hydrophobic transmembrane domains in the proteins.

As we found that the native promoters of *4TM1* and *4TM2* were highly expressed, we made C-terminal fusions of both genes driven by their own promoter (p4TM1::4TM1:GFP and p4TM2::4TM2:GFP). The fluorescent signal is similar to the constitutively expressed C-terminal fusion, but more specific for dividing or recently divided cells. Also the signal-to-noise ratio is higher, due to the high expression levels brought about by the native promoters (Figure 3.7C, D). 4TM is localized to the forming cell plate and upon docking to the plasma membrane (PM), also the PM is fluorescently labeled. In the cytoplasm, a punctate pattern of fluorescence suggest that 4TM is also present in vesicles, which seem to fuse with the cell plate (inset in Figure 3.7C), alternatively these vesicles could also bud from the cell plate. In the young leaf primordium, the newly formed cell plate and, upon docking, the PM is labeled, corroborating the localization pattern of 4TM in dividing tissue (Figure 3.7D).

3.2.7 4TM co-localizes with KNOLLE

The subcellular expression pattern of the *4TM* genes is reminiscent of that of the mitotic syntaxin KNOLLE (KN) (27). We confirmed co-localization of 4TM with KN at the cell plate and in vesicles in the cytoplasm using a KN antibody and a secondary antibody conjugated to a fluorochrome (Cy3) (Figure 3.8A-C). In Figure 3.8C cell '1' is finalizing cell division, KN and 4TM are strongly present in the cell plate and in dispersed vesicles in the cytoplasm while 4TM also labels the plasma membrane (PM). In cells '2' and '3', progressively older cells, the signal for both 4TM and KN decreases. 4TM seems to be more persistent, but this could be an artifact of the
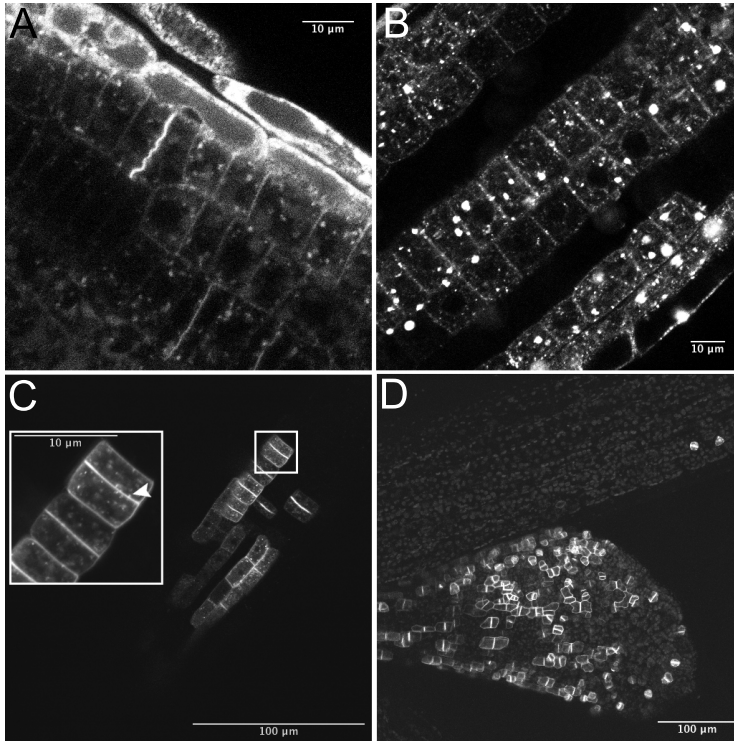
Figure 3.7: Subcellular localization of 4TM: p35S::4TM1:GFP (A), p35S::GFP:4TM2 (B), p4TM1::4TM1:GFP (C, D). (A) C-terminal translational fusion of 4TM1 and 4TM2 with GFP driven by the 35S promoter exhibits fluorescence at the plasma membrane (PM), the newly formed cell plate during mitosis and showed punctate expression in the cytoplasm. (B) The N-terminal fusion does not localize to the cell plate and contains spots with high fluorescence, which are likely artifacts of the mis-localization and overexpression due to aggregation of the hydrophobic transmembrane domains. (C, D) C-terminal translation fusion of 4TM1 driven by its native promoter in the root tip (C) and young leaf primordium (5 days after stratification) (D) marks proliferative cells in which the cell plate and vesicles in the cytoplasm are fluorescently labeled. The PM becomes labeled when the forming cell plate docks to the PM. The inset in C shows the fusion of vesicles with the cell plate or budding of vesicles from the cell plate (indicated by the arrowhead).

GFP fusion construct or due to a difference in signal strength of the GFP-tagged proteins and the antibody. For unknown reasons, the KN antibody binds to KN at the cell plate and dispersed vesicles (Figure 3.8B) while fluorescently labeled KN also is present at the plasma membrane upon fusion of the cell plate (Figure 3.8D) (41).
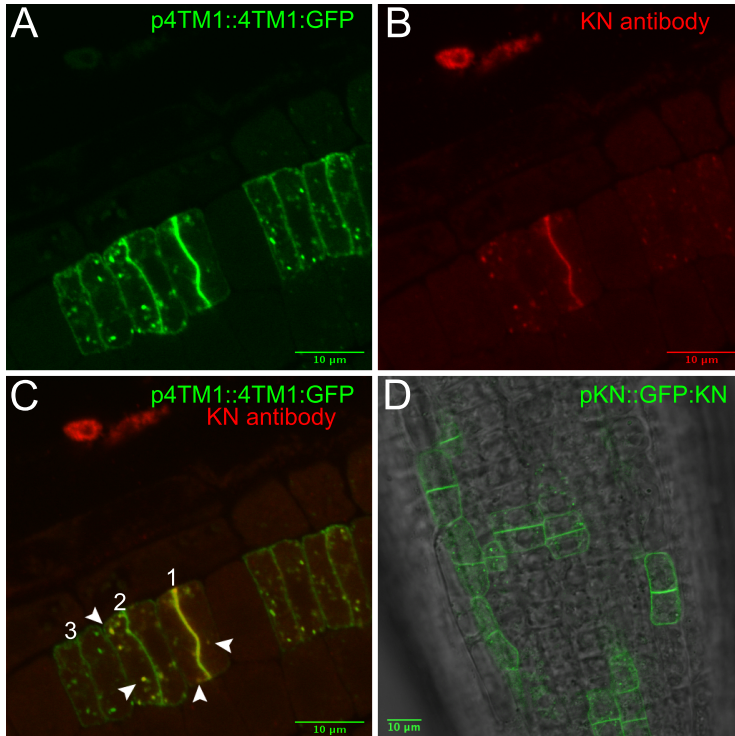
Figure 3.8: Co-localization of 4TM with KNOLLE (KN). (A) C-terminal GFP-labeled 4TM1 driven by its native promoter (p4TM1::4TM1:GFP), as described earlier (Figure 3.7C, D). (B) KN antibody visualized by a secondary antibody linked to a fluorochrome (Cy3), labeling the cell plate and transport vesicles. (C) Overlay of p4TM1::4TM1:GFP and KN antibody. The arrowheads indicate vesicles at which 4TM1 and KN co-localize. The cell indicate by '1' has recently divided and the cell plate has just fused with the plasma membrane (PM) but the signal in the PM is still weak. In cell '2' cell division has been finalized. While 4TM is present in the cell plate, PM and in dispersed vesicles, the KN antibody only labels vesicles. Later, the 4TM signal is still present in the PM and some vesicles while KN cannot be detected at this point (cell '3'). (D) Compared to the KN antibody, an N-terminal translational fusion with GFP (pKN::GFP:KN) labels additionally the PM upon fusion of the cell plate with the PM.

3.2.8　Aberrant cell divisions upon 4TM depletion

To obtain loss-of-function mutants, we analyzed 14 T-DNA insertion lines from the SALK collection (16). For 9 of the 14 lines, a T-DNA insertion could be mapped in the proximity of a gene of the *4TM* family (see 3.5). Only in *4TM1* we could identify a SALK line (SALK_031814) which leads to a full knockout of the gene due to the introduction of a premature stop codon. This mutant did however not show any phenotypic changes, probably due to redundancy among the *4TM* genes.

To overcome possible redundancy, artificial miRNAs (amiRNA) were made by replacing the mature miRNA in the miR319a precursor, by site-directed mutagenesis, with an engineered sequence targeting our genes of interest (42). Using the publicly available designer software (wmd3.weigelworld.org), we made four amiRNAs that target *4TM2* (ami4TM2), *4TM3* (ami4TM3), both *4TM2* and *4TM3* (ami4TM23) and all three *4TM* genes (ami4TM123) (Figure 3.9). We expressed these amiRNAs constitutively in *Arabidopsis* using the 35S promoter.

```
4TM2       361 GCGTCTATGATCTTGAAACTAGCTGGTGGGCTCTGTGTTGTAATCTTTTGCCTTGAGTGG 420
ami4TM2     21 -----------------------CCGCCCGAGACACAACATTAT--------------- 0
                                     ** ****************

4TM3       181 TGCTCGATCTTCGCCTCCTCCACATCCGCCACCGCGAGCGATAGATTATACGGTTCACAC 240
ami4TM3     21 ----------------------------TGGCGCTGGCTATCTAATATT--------- 0
                                           ******* ************

4TM2       241 GCAGCTCACATGATCGCAACTGTCCTTACCGCCATTTTCCAAGGCTCAGTCTCTGTTCTC 300
4TM3       232 GGTTCACACATGGTAGCCACAGTCCTCACGGCCATTTTCCAAGGCGCTGTCTCTGTTCTG 291
ami4TM23    21 ----------------------TGAGTGCCGGTAAAAGGTTCT---------------- 0
                                     ** ** *************

4TM1        79 ATCCTTCTCGGCGCTGCTTCAAGTTGTATCTTCCTCACACTCTCTCTCCGTCTAATCCCT 138
4TM2        82 GTCCTCCTCGGTGCATCGTCAAGCTGCATCTTCCTCACTCTCTCTCTCCGTCTAATCCCT 141
4TM3        58 ATCCTCTTAGGCTCAGCCTCAAGTTGCATCTTCCTCACAATCTCCCTCCGTCTCTTCCCA 117
ami4TM123   21 ------------------GTTCCACGTAGAAGGAGTGAT------------------- 0
                                 **** ** ***********
```

Figure 3.9: Alignment of artificial miRNA (amiRNA) sequence with 4TM mRNA. The asterisks indicate perfect matches between the amiRNA and the target. The position of the target site in the coding sequence is indicated (5′ to 3′). ami4TM2 targets *4TM2*, ami4TM3 *4TM3*, ami4TM23 both *4TM2* and *4TM3* and ami4TM123 all three *4TM* genes.

We analyzed the expression levels of the *4TM* genes in homozygous amiRNA transformants carrying a single insert to verify the functionality of the amiRNAs (Figure 3.10). The amiRNAs targeting a single gene lead to the highest reduction of expression, circa 75% for *4TM2* and 60% *4TM3*. Among the plants that express ami4TM23 only one line that showed reduced expression in any *4TM* gene could be identified. This transformant is only affected in *4TM3* which is expressed at 35% of the WT expression level. The amiRNA that targets all three *4TM* genes showed less reduction with the expression reduced to half for *4TM2* and *4TM3*, while *4TM1* remained at 60 to 75%.
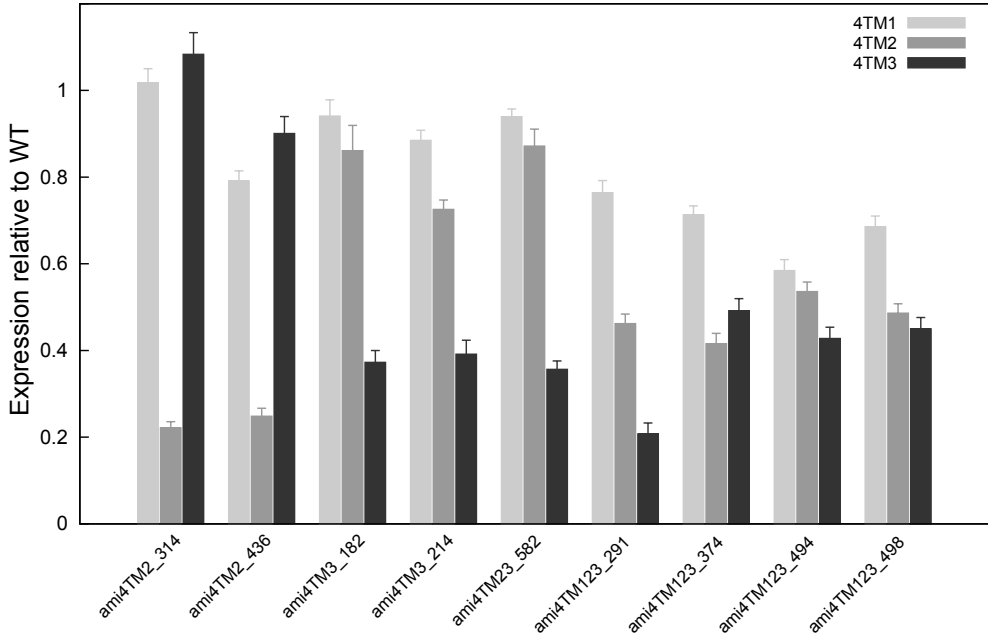
Figure 3.10: Expression of *4TM* genes in amiRNA transformants using qPCR.
ami4TM2 targets *4TM2*, ami4TM3 *4TM3*, ami4TM23 both *4TM2* and *4TM3* and
ami4TM123 all three *4TM* genes (the numbers indicate the transformation event).
The data are represented as mean ± SEM, relative to the wild type expression.

Macroscopically, none of the amiRNA transformants showed phenotypic differences compared to the wild type. The main root of *Arabidopsis* is strictly organized and was used as a model system to study defects at the cellular level. We focused on the ami4TM123 transformants as these are most likely to overcome redundancy among *4TM* genes. The overall organization of the root is not disrupted, but in the endodermis aberrant cell divisions occur (Figure 3.11). In those cases the cell plate does not divide the cell evenly, but resembles an asymmetric division (Figure 3.11A, B). The presence of aberrant divisions was observed in 25% (12 of 47 roots) of ami4TM123 plants (line 498) but also in 20% of ami4TM3 (3 of 15 roots) and 10% of ami4TM23 (2 of 22 roots) while for wild type this was only the case in 3% of plants (1 of 34 roots). We also observed this phenotype in N-terminal translational GFP fusions of 4TM2. Additionally, we observed unfinished and wavy cell plates in cortex cells of roots with aberrant cell divisions (Figure 3.11C, D). These wavy cell plates did not expand further and remained stable as such for at least 30 minutes. Within this time frame, a normal cell plate would have expanded and docked to the plasma membrane.
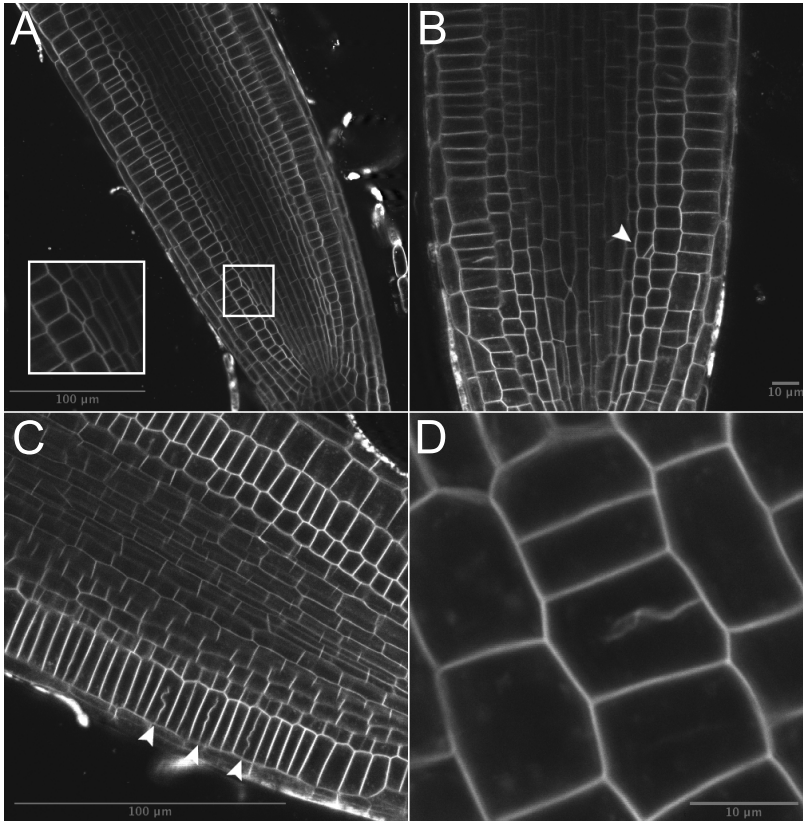
Figure 3.11: Aberrant cell divisions upon depletion of 4TM by artificial miRNAs.
(A, B) Aberrant cell divisions in the endodermis of the root, indicated by the inset
in A and arrowhead in B. (C, D) Unfinished and wavy cell plates in the cortex, indi-
cated by the arrowheads in C. Representative examples are shown for ami4TM123
line 498.

3.2.9   Discussion

Based on the expression profile during both leaf and root development, we have
identified proliferation specific genes. We focused on a small family of genes with
four transmembrane domains and hence named them *4TM* genes. All family mem-
bers were confirmed to be highly upregulated during mitosis using an aphidicolin
synchronized cell culture. The specificity for dividing cells could be confirmed by
both transcriptional and translational fusions with a fluorescent marker.

The *Arabidopsis* genome encodes for four genes of the *4TM* family. Two of the
genes, *4TM3a* and *4TM3b*, are located in a 59.3 kb tandem duplication on chro-

mosome 5. Both copies of the tandem duplication are identical with the exception of the insertion of a single G. We investigated this duplication in more detail and found only one BAC clone, F24C7, that confirmed its presence. Assuming that this BAC clone could be a chimera, we showed that an alternative alignment is possible such that this duplication is absent. The identification of homozygous T-DNA insertion lines in the sequence within the alleged duplication provides strong evidence that the duplication is not present. However, this is based on the absence of a PCR product and it can therefore not be excluded that this is due to a technical aspect of the experiment. Finally, tiling array data on DNA samples showed a lower signal specifically at the position of this duplication. Combining these results, we are convinced that the duplication is an artifact of the BAC alignment and is not present in the genome. Therefore, we assume that there are three members of the *4TM* family in *Arabidopsis*.

Phylogenetic analysis learned that genes homologous to *4TM*s are only found in angiosperms, suggesting that these proteins function at a point in the cell cycle specific to plants. While the key players in the cell cycle are conserved in the plant and animal kingdom, the main difference lies in the actual division into two daughter cells. In mammalia this is achieved by a contractile ring that constricts centripetally while plant cells form a cell plate that expands in a centrifugal way. The translational fusions localize the 4TM protein at the cell plate, suggesting they play a role in the plant-specific process of cell division after mitotic division. When the cell plate merges with the plasma membrane (PM), the fluorescent signal also spreads into the PM. In the cytoplasm, fluorescent foci are visible, likely corresponding to transport vesicles from the ER to the cell plate and/or from the cell plate to the trans-golgi network. This localization closely resembles KNOLLE (KN), a mitotic syntaxin known to be involved in cell plate formation. KN is, together with KEULE, implicated in vesicle fusion during somatic cytokinesis (43, 44).

To further investigate the function of the 4TM proteins, we ectopically expressed 4TM1 and 4TM2 in *Arabidopsis* plants. Despite a 20- to 30-fold overexpression in selected lines, no phenotypical changes were observed (data not shown). Next, we tried to knock out the *4TM* genes using T-DNA insertion lines. This yielded a mutant line with an insertion in *4TM1* introducing a premature stop codon. However, this did not lead to phenotypic changes in the plants probably due to the redundancy among *4TM* genes. To overcome this, we developed artificial miRNAs (amiRNA) against the *4TM* genes, resulting in a knockdown of all three *4TM* genes. It is possible that the selection of mutants resulted in a bias towards relatively weak mutant lines that still support plant growth and that a stronger downregulation of *4TM*s results in severe developmental defects. Macroscopically, the amiRNA overexpressing lines are indiscernible from wild type, but at the cellular level we observed aberrations. The root tip has a strict organization and is therefore suited to detect subtle changes in cell division patterns. In the endodermis, we observed

aberrant cell divisions, resembling asymmetric cell divisions, due to misorienta-
tion of the forming cell plate (Figure 3.11A, B). If the resulting small cells contain
a nucleus remains to be determined. In some cortex cells, a wavy cell plate that
is docked to one site to the plasma membrane could be observed (Figure 3.11C,
D). This cell plate seemed to have stopped expanding as it appeared as a stable
structure for at least 30 minutes while in normal cell plate development the divi-
sion should have been finalized in the mean time. In the translational fusions, fully
expanded cell plates were observed that also showed a wavy pattern compared to
the normal straight cell wall (Figure 3.7A). We hypothesize that the 4TM proteins
function in the establishment of the cell plate and the trafficking of material from
the ER to the forming cell plate. Overexpression of the 4TM proteins causes the
delivery of excess material to the forming cell plate, resulting in a wavy cell wall,
while depletion of 4TMs can lead to the inability to finish expansion of the cell
plate. In some cases, the cell plate cannot dock to the plasma membrane, suggest-
ing that 4TMs are also implicated in the control of cell plate orientation. While the
orientation is establishment before the presence of 4TM, already at the formation of
the preprophase band, these proteins could be involved in the guidance of the cell
plate towards these positions.

3.3   High Mobility Group genes

3.3.1   Introduction

High Mobility Group (HMG) genes are rich in charged amino acid residues causing high electrophoretic mobility in Triton-urea gels, hence the name (45). Both mammals and plants code for genes containing HMG domains. The HMG proteins are associated with eukaryotic chromatin. HMG genes are divided into three classes, based on structural differences: the HMG nucleosome binding family (HMGN), the HMG AT-hook family (HMGA) and the HMG-box family (HMGB). Only the latter two families have been identified in plants (46). While HMGA proteins interact with A/T-rich double-stranded DNA, HMGB proteins bind DNA non-sequence specifically. In *Arabidopsis*, one HMGA gene is present while it encodes 15 proteins with a HMG-box (6, 46). The two proliferation-specific genes we identified belong to the HMGB family.

Plant genes containing a HMG-box can be subdivided in four families: HMGB-type proteins with a single HMG-box domain flanked by a basic N-terminal and acidic C-terminal domain (8 members in *Arabidopsis*), structure-specific recognition protein 1 (SSRP1, 1 member in *Arabidopsis*), proteins containing both an AT-rich interaction domain (ARID) and a HMG box (4 members in *Arabidopsis*) and proteins containing three HMG-boxes (3xHMG-box, 2 member in *Arabidopsis*). These last two families are only found in plant species. The HMG-genes we selected make up the last group of 3xHMG-box genes. Next to the three HMG domains, they share an N-terminal domain that deviates from known protein domains. The role of these proteins has not yet been established (6).

The HMG-box domain mediates DNA binding and plant HMGB proteins can bind to promoter regions and bend DNA. Plant HMGB proteins have a higher affinity for DNA structures (e.g. DNA mini-circles, super-coiled DNA, four-way junctions) compared to linear DNA. They can also interact with nucleosomes and function in chromatin remodeling e.g. through competition with histone H1 (47). The HMG domain is also involved in protein-protein interactions, serving as a chaperone that facilitates nucleoprotein complex formation (6). Transcription factors regulate the transcription of genes through binding at *cis*-acting DNA target sequences. This often requires the assembly of multiprotein complexes and factors as HMGB proteins can modulate DNA binding or protein interactions (46, 47). Overexpression of the maize HMGB1 in tobacco caused a reduction of the cell size in the proliferation zone of the primary root, resulting in a reduced length of the primary root (48). In *Xenopus*, HMGB3 is involved in the regulation of cell proliferation (49) and in mice HMGB2 is implicated in germ cell differentiation (50).

3.3.2 Analysis of T-DNA insertion lines

To start the functional analyis of the two *Arabidopsis 3xHMG-box* genes (*3xHMG-box1* (*AT4G11080*) and *3xHMG-box2* (*AT4G23800*)), we analyzed the available SALK (16) and GABI-KAT lines (51). For *3xHMG-box2*, no insertion lines were available. In *3xHMG-box1* two SALK lines were analyzed but for SALK_087786 no T-DNA insertion could be detected in the proximity of this gene using any of the primers (Supplementary Table 3.3). For SALK_087783 the insertion could be verified 5 bp upstream of the start codon, but the insertion had no effect on the transcription level and hence no phenotypic changes could be observed in that line (data not shown). Of the two available GABI-KAT lines in *3xHMG-box1*, for GABI_151G12 no T-DNA insertion could be found and the expression level of *3xHMG-box1* was similar to wild type in GABI_171F06 and no phenotypic changes were observed.

3.3.3 *3xHMG-box* genes are conserved in plants

We identified homologs of *3xHMG-box* genes using PLAZA (30), yielding genes in both monocots and dicots as well as the lycophyte *Selaginella moellendorfii* and the moss *Physcomitrella patens* (Figure 3.12). From all land plants included in PLAZA2.0, only *Medicago truncatula* and *Manihot esculenta* did not contain a homolog. All monocot species contain only one *3xHMG-box* gene (6), while dicot species have either one or two copies. Both *S. moellendorfii* and *P. patens* contain two copies.

Figure 3.12: Phylogenetic tree of the homologous gene family of *3xHMG-box* genes. Included species (PLAZA 2.0, 30): *Selaginella moellendorfii* (SM), *Populus trichocarpa* (PT), *Ricinus communis* (RC), *Vitis vinifera* (VV), *Carica papaya* (CP), *Glycine max* (GM), *Lotus japonicus* (LJ), *Arabidopsis lyrata* (AL), *Arabidopsis thaliana* (AT), *Malus domestica* (MD), *Sorghum bicolor* (SB), *Zea mays* (ZM), *Oryza sativa* ssp. indica (OSINDICA), *Oryza sativa* ssp. japonica (OS), *Brachypodium distachyon* (BD) and *Physcomitrella patens* (PP). Monocot species are indicated in grey, the selected *Arabidopsis* genes in bold (based upon PLAZA, bioinformatics.psb.ugent.be/plaza/, 30).
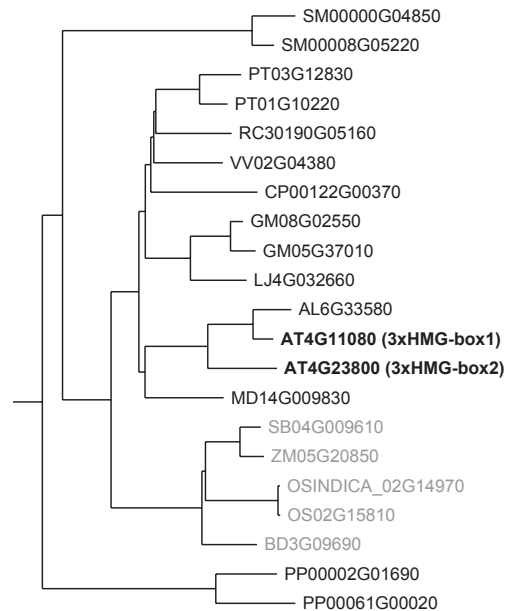
3.3.4 Proliferation specific expression of *3xHMG-box* genes

To further validate the expression pattern of the *3xHMG-box* genes, we made a transcriptional fusion of the promoter of *3xHMG-box2* with green fluorescent protein (GFP) and β-glucuronidase (GUS), including a nuclear localization signal (NLS) to concentrate the fluorescent signal to the nucleus (p3xHMG-box2::NLS:GFP:GUS, Figure 3.13). The fluorescent signal was observed in the root tip (Figure 3.13A-C) as well as lateral root primordia (Figure 3.13D-F). In young leaf primordia, small pavement cells and guard cells contain the fluorescent marker (Figure 3.14A-C), while cells that are already expanding do not express GFP (inset Figure 3.14C). Later in leaf development, the expression is limited to guard cells (Figure 3.13D-F).
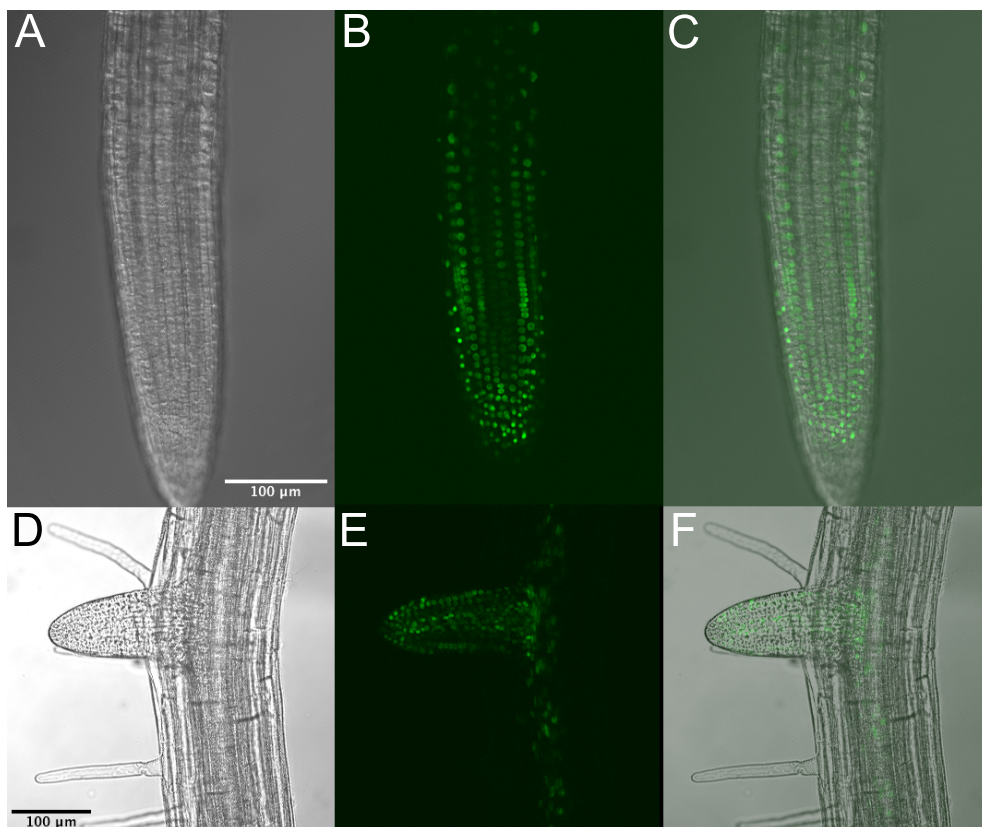
Figure 3.13: A p3xHMG-box2::NLS:GFP:GUS transcriptional fusion is expressed in proliferative tissue in the root. (A-C) Expression in the root tip (DIC (A), GFP (B), DIC+GFP (C)), (D-F) lateral root primordia and stele (DIC (D), GFP (E), DIC+GFP (F)).
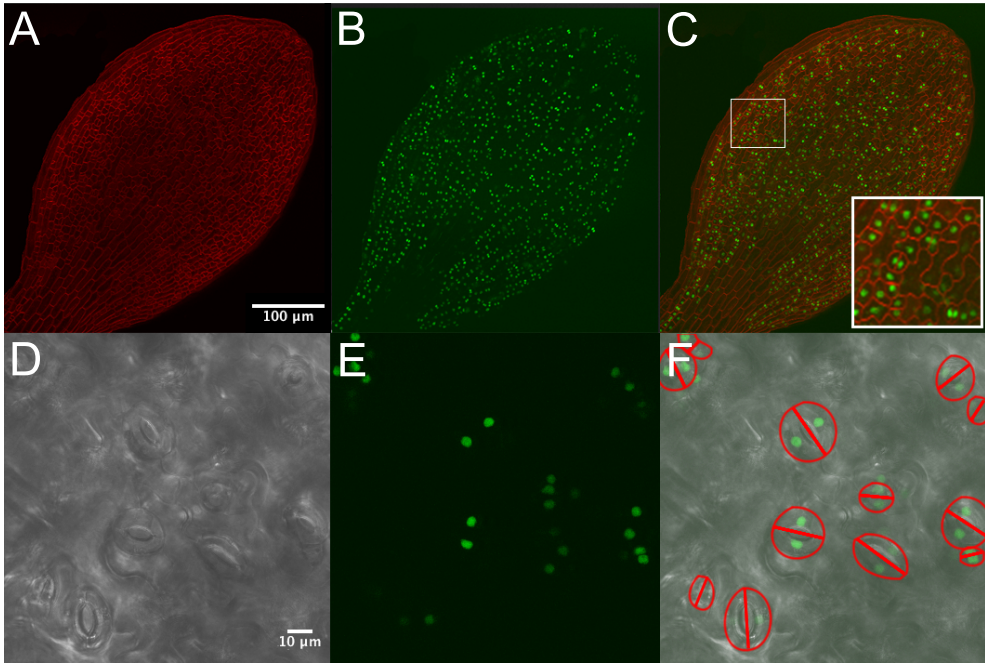
Figure 3.14: A p3xHMG-box2::NLS:GFP:GUS transcriptional fusion is expressed in proliferative tissue in the leaf. (A-C) Expression in recently divided pavement and guard cells in young leaf primordia (4 days after stratification) while cells that start to expand are devoid of expression (Inset C) (FM4-64 (A), GFP (B), FM4-64+GFP (C)). (D-F) In older leaves (12 days after stratification) the expression is limited to guard cells (DIC (D), GFP (E), DIC+GFP (F) with stomata marked in red).

3.3.5 *3xHMG-box* genes are mitosis specific

The validation of the selection for proliferation specific genes, confirmed that both *3xHMG-box* genes are indeed high expressed during the proliferation phase of leaf development and their expression decrease steeply upon the transition to expansion (Figure 3.1). This is confirmed by the transcriptional fusion of *p3xHMG-box2* with GFP. To establish if the *3xHMG-box* genes are differentially regulated during the cell cycle, we analyzed their expression in an aphidicolin synchronized cell culture using qPCR (Figure 3.15). Both genes show a very similar mitosis specific expression pattern, the fold change being substantially higher for *3xHMG-box2*. This corroborates the difference in fold change found in the microarray data (Table 3.1).
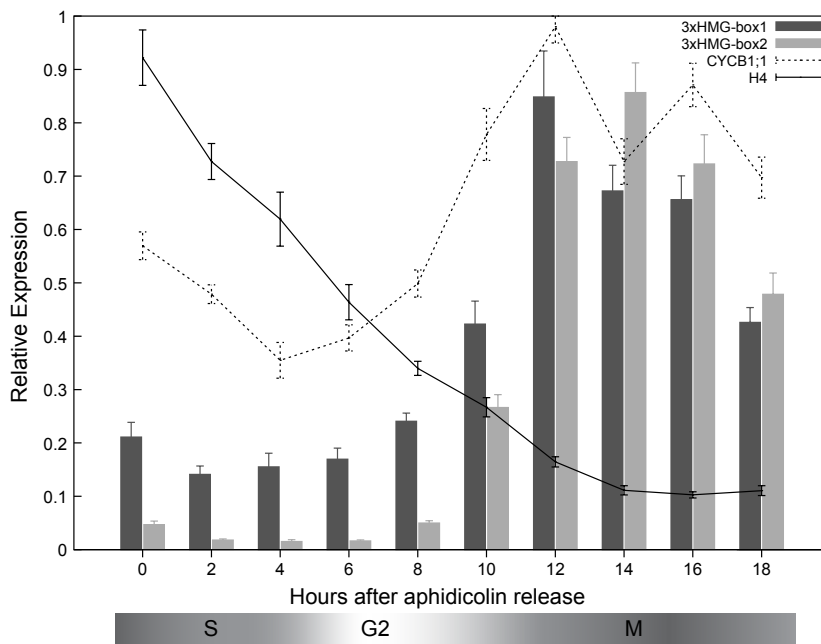
Figure 3.15: Expression of *3xHMG-box* genes in a synchronized cell culture occurs specifically in M-phase. For reference, *Histone H4* is an S-phase marker, *CYCB1;1* a mitotic marker. Note that the fold-change upregulation during M-phase exceeds that of the B-type cyclin. Data are represented as mean $\pm$ SEM for two biological repeats, rescaled to a maximum of 1.

3.3.6 Discussion

We have identified two genes from the HMG family that are specifically expressed in proliferation. These two genes contain 3 HMG-box domains and were therefore named *3xHMG-box1* and *3xHMG-box2* (*AT4G11080* and *AT4G23800*, respectively). They constitute a plant-specific sub-family of *HMGB* genes. We confirmed the microarray data and thus the proliferation specific nature of these genes with both qPCR and a translational GFP fusion. Furthermore, we demonstrated that these genes are differentially regulated during the cell cycle, peaking during mitosis.

The key question for future studies regards the biological function of the 3xHMG-box proteins. Our data suggest a function for the two proteins during mitosis. The localization studies have been performed using transcriptional marker fusions with a nuclear localization signal, limiting its use to the spatiotemporal expression pattern. The use of translational marker fusions driven by its own promoter would allow determining the subcellular localization of the 3xHMG-box proteins. The behavior of these proteins during mitotic division would be of particular interest.

While available T-DNA insertion lines did not yield a functional knockout, alternative approaches could be used to eliminate or reduce the expression of these genes. As redundancy between the two homologues is likely, the use of artificial microRNAs provides an elegant solution to achieve this.

Members of the HMG family are known for their ability to bind DNA and change its conformation. The HMGB subfamily have a higher affinity for DNA structures such as DNA mini-circles compared to linear DNA (47). DNA binding assays will reveal if this is also the case for the 3xHMG-box proteins. Furthermore, next to the full-length protein, the contribution of the individual HMG boxes could be assessed. The DNA bending activity could be assayed using a DNA ligation experiment. Short DNA fragments, which are unable to circularize in the absence of a DNA bending protein, are used in this assay to test the DNA bending activity of a protein of interest by assessing if the protein facilitates circularization in the presence of a DNA ligase (52).

The described follow-up experiments would help elucidate the biological function of the 3xHMG-box proteins. Other members of the HMG protein family are known to bind DNA in e.g. the promoter region but also interact with nucleosomes and function in chromatin remodeling (47). SSRP1 is part of the facilitates chromatin transcription (FACT) complex that destabilizes nucleosomes and thus assists in the progression of RNA polymerase on chromatin templates. Upon mutation of SSRP1, diverse defects in vegetative and reproductive development have been recently demonstrated (53). In maize, HMGB1 and HMGB5 have been shown to modulate the interaction of DOF transcription factors with nucleosomes (54, 55). Therefore, this plant-specific subfamily could be involved in transcriptional regulation of mitotic genes. Alternatively, these proteins could be involved in DNA/chromatin condensation or histone modification during mitosis. It is well established that the phosphorylation status of histone H3 differs in mitosis compared to interphase and the acetylation of histone H3 and H4 are reduced during M-phase (56, 57). The conservation of this small protein family throughout the plant kingdom suggests an important function but further research is required to unravel its biological significance.

3.4   Perspectives and conclusion

We have identified two new families of proliferation specific genes that function during mitosis of the cycle. The *4TM* family consists of three members that have 4 transmembrane domains and these genes are specifically expressed during mitosis. The 4TM proteins localize to the cell plate and likely play a role in its formation. The second family contains three HMG-boxes for which we have shown that they also are specifically induced during mitosis. The data on the *3xHMG-box* genes will be combined with a detailed phylogenetic analysis and DNA binding assays performed in the lab of Dr. Klaus Grasser (Cell Biology and Plant Biochemistry, Regensburg University, Germany) and submitted for publication.

To establish the function of 4TM proteins during the cell cycle further, co-localization with vesicle markers will be instrumental to determine the dynamics of the 4TM proteins. We have therefore initiated co-localization experiments with markers for the trans-golgi network, endosomes and multivesicular bodies. To determine the order in which 4TM and KNOLLE operate in the formation of the cell plate, a time-lapse experiment using fluorescently tagged 4TM and KNOLLE will be conducted.

Also, generation of a mutant with a stronger downregulation of the 4TM proteins could aid in unraveling the function. Therefore, crosses will be made of the ami4TM123 line with the T-DNA insertion line in *4TM1*. This will lead to a full knockout of *4TM1* combined with a potentially even stronger downregulation of the other two genes due to the absence of *4TM1* mRNA. Alternatively, the expression of the amiRNA driven by the 4TM native promoter could enhance the silencing capacity.

The identification of binding partners of 4TM proteins can be a valuable source of information concerning the function of these proteins in development in general and cell division more specifically. We performed tandem affinity purifications (TAP-TAG) (58) but the standard protocol uses a strong detergent while for the identification of complexes with transmembrane proteins a gentle extraction is required so that the complex is not destroyed. The extraction of intact complexes of membrane proteins can be improved by crosslinking the complex prior to extraction. The purification protocol for transmembrane proteins is now being optimized by the Functional Proteomics group of Geert De Jaeger (PSB) based on a new tag that is resilient to crosslinking (59).

In conclusion, we have selected proliferation specific genes based upon expression profiling during leaf and root development. This approach has identified promising candidates to further deepen our knowledge of the regulation of the cell cycle and growth and development of plants.

3.5   Experimental procedures

3.5.1   Plant material

Arabidopsis thaliana ecotype Columbia-0 was grown on round Integrid Petri dishes (Integrid Petri Dish, Beckon Dickinson Labware, Le Pont de Claix, France) on 0.5x MS medium (Duchefa Biochemie, Haarlem, The Netherlands) containing 1% sucrose, 0.5% MES (Duchefa Biochemie, Haarlem, The Netherlands) and 1% Plant Tissue Agar (Lab M Limited, Bury, Lancaster, UK). Seeds were stratified for 2 days. For the developmental time series, the first leaf pair was harvested in liquid nitrogen 8, 10, 12, 14, 16, 19 and 24 days after stratification. For two independent biological replicates, leaves were pooled from 256 plants for day 8, 64 plants for day 10 and 16 plants for all other time points.

T-DNA insertion lines were obtained from the NASC stock center, the position of the insertion relative to the start codon is indicated (NA signifies that no T-DNA insertion was detected in the proximity of the gene): for 4TM1 (SALK_028428, -344; SALK_066297, NA; SALK_031814, +250; SALK_032395, -146; SALK_039559, -453), 4TM2 (SALK_115974, +381; SALK_114432, NA; SALK_046332, NA; SALK_046334, NA; SALK_046343, +691; SALK_060336, -207; SALK_076440, NA) and 4TM3 (SALK_041170, -168; SALK_101711, +827). Detection of T-DNA inserts was done using primers specific to the left border of the T-DNAs used for mutagenesis (LBC1) in combination with gene-specific primers (Supplemental Table 3.3).

3.5.2   Synchronized cell culture

The PSB-D cell culture (58) was synchronized using 6.2 µg/ml aphidicolin for 21.5h. Cells were collected every two hours up to 18h after removal of aphidicolin for two biological replicates.

3.5.3   RNA extraction and cDNA conversion

RNA was extracted using TriReagent (Applied Biosystems, Lennik, Belgium) and cleaned with RNeasy Mini kit (Qiagen, Venlo, The Netherlands). RNA quality was verified using an ND-1000 Spectrophotometer (Thermo Scientific, Erembodegem, Belgium). 1 µg of total RNA was converted to cDNA with SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA, USA) using poly(A) primers according to manufacturers instructions.

3.5.4   Expression analysis

Quantitative PCR (qPCR) was performed on a LightCycler480 (Roche, Vilvoorde, Belgium) using LightCycler480 SYBR Green I Master (Roche, Vilvoorde, Belgium).

All pipetting was performed by a Xiril100 robot (Xiril, Hombrechtikon, Switzerland). The qPCR reaction was performed in 5 µl using 10 ng of cDNA and 200 ng/µl of forward and reverse primer. Primers were designed using Beacon Designer 4.0 (Premier Biosoft International, Palo Alto, CA, USA). Each qPCR reaction was performed in three technical replications. A list of all primers used is provided in Supplemental Table 3.2.

Data analysis was performed using the $\Delta\Delta C_T$ method (60), taking the primer efficiency into account. The data were normalized using three normalization genes (*CDKA;1*, *CKIIa*, *CBP20*) according to the GeNorm algorithm (61). The first time-point was used as calibrator. The result for two biological replicates was averaged and standard errors were determined.

3.5.5 Cloning

The overexpression lines were generated through Gateway cloning of the coding sequence in pK7FWG2 and pK7WGF2 (62). The coding sequence for *4TM1* and *4TM2* was amplified from *Arabidopsis thaliana* ecotype Columbia genomic DNA using Phusion High-Fidelity DNA polymerase (Finnzymes, Espoo, Finland). The promoters of *4TM1* and *4TM2* were described previously (39) and constructs were made through multisite Gateway cloning (40).

3.5.6 Confocal imaging

Plants for confocal imaging were grown vertically, using growth medium as described. Plant material was mounted in water and fluorescence detection was done with a confocal laser-scanning microscope Zeiss 510 or Zeiss 710 (Carl Zeiss, Zaventem, Belgium).

3.5.7 Immunolocalization

One-week-old seedlings, grown on 0.5x MS medium under continuous illumination, were fixed in paraformaldehyde (Merck, Darmstadt, Germany). Immunolocalization was performed as described (63). The rabbit anti-KNOLLE antibody (1:2000) (64) and the fluorochrome-conjugated secondary antibody anti-rabbit-Cy3 (1:600) (Dianova, Hamburg, Germany) were used. Fluorescence detection was done with a confocal laser-scanning microscope Zeiss 710 (Carl Zeiss, Zaventem, Belgium).

3.5.8 Artificial miRNAs

The hairpin structure for the artificial microRNAs were constructed as described (42). Cloning was done using Gateway technology (40), the multiple cloning site was replaced by attB1 and attB2 sequences by an additional

PCR reaction with primers GGGGACAAGTTTGTACAAAAAAGCAGGCTCCC-
CAAACACACGCTCGGA and GGGGACCACTTTGTACAAGAAAGCTGGGTCC-
CCATGGCGATGCCTTAAA.

Acknowledgments

Contributions

Frederik Coppens performed the selection and validation of new proliferative genes, all expression analyses, the phylogenetic analyses, all analyses related to the mistakenly annotated tandem duplication and the cloning and analysis of the 4TM, amiRNA and HMG constructs. The co-localization with KNOLLE was performed by Steffen Vanneste. Daniël Van Damme assisted for the cell biological analyses and helped shape the 4TM project through discussions.

References

[1] G. T. S. Beemster, *et al. Plant Physiol* **138**, 734 (2005).

[2] K. Birnbaum, *et al. Science* **302**, 1956 (2003).

[3] K. Vandepoele, *et al. Plant Cell* **14**, 903 (2002).

[4] D. Swarbreck, *et al. Nucleic Acids Res* **36**, D1009 (2008).

[5] P. Lamesch, *et al. Curr Protoc Bioinformatics* **Chapter 1**, Unit1.11 (2010).

[6] M. Stros, D. Launholt, K. D. Grasser. *Cell Mol Life Sci* **64**, 2590 (2007).

[7] K. Mashiguchi, T. Asami, Y. Suzuki. *Biosci Biotechnol Biochem* **73**, 2452 (2009).

[8] G. H. H. Borner, K. S. Lilley, T. J. Stevens, P. Dupree. *Plant Physiol* **132**, 568 (2003).

[9] H. Tanaka, *et al. FEBS Lett* **355**, 4 (1994).

[10] J. Dai, S. Sultan, S. S. Taylor, J. M. G. Higgins. *Genes Dev* **19**, 472 (2005).

[11] J. Dai, B. A. Sullivan, J. M. G. Higgins. *Dev Cell* **11**, 741 (2006).

[12] S. Fujimoto, *et al. Planta* **222**, 293 (2005).

[13] S. F. Altschul, *et al. J Mol Biol* **215**, 403 (1990).

[14] G. Cnops, *et al. Plant Cell* **18**, 852 (2006).

[15] M. E. Hemler. *Nat Rev Mol Cell Biol* **6**, 801 (2005).

[16] J. M. Alonso, *et al. Science* **301**, 653 (2003).

[17] F. Wang, *et al. Science* **330**, 231 (2010).

[18] D. Demidov, *et al. Plant Cell* **17**, 836 (2005).

[19] D. Francis. *New Phytol* **174**, 261 (2007).

[20] J. J. Kieber, *et al. Cell* **72**, 427 (1993).

[21] M. Bar, *et al. Plant J* **55**, 1025 (2008).

[22] M. Bar, M. Sharfman, S. Schuster, A. Avni. *PLoS One* **4**, e7973 (2009).

[23] Y. Mineyuki, K. W. Jeon. *The Preprophase Band of Microtubules: Its Function as a Cytokinetic Apparatus in Higher Plants*, vol. Volume 187, 1–49 (Academic Press, 1999).

[24] P. Dhonukshe, *et al. Dev Cell* **10**, 137 (2006).

[25] M. Menges, J. A. H. Murray. *Plant J* **30**, 203 (2002).

[26] G. S. Stein, J. L. Stein, A. J. van Wijnen, J. B. Lian. *Curr Opin Cell Biol* **4**, 166 (1992).

[27] W. Lukowitz, U. Mayer, G. Jürgens. *Cell* **84**, 61 (1996).

[28] M. Menges, S. M. de Jager, W. Gruissem, J. A. H. Murray. *Plant J* **41**, 546 (2005).

[29] S. F. Altschul, *et al. Nucleic Acids Res* **25**, 3389 (1997).

[30] S. Proost, *et al. Plant Cell* **21**, 3718 (2009).

[31] M. A. Larkin, *et al. Bioinformatics* **23**, 2947 (2007).

[32] E. M. Zdobnov, R. Apweiler. *Bioinformatics* **17**, 847 (2001).

[33] A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer. *J Mol Biol* **305**, 567 (2001).

[34] G. E. Tusnády, I. Simon. *J Mol Biol* **283**, 489 (1998).

[35] K. Hofmann, W. Stoffel. *Biol Chem Hoppe-Seyler* **374**, 166 (1993).

[36] M. G. Claros, G. von Heijne. *Comput Appl Biosci* **10**, 685 (1994).

[37] S. Tabata, *et al. Nature* **408**, 823 (2000).

[38] G. Horiguchi, *et al. Plant J* **60**, 122 (2009).

[39] M. Benhamed, *et al. Plant J* **56**, 493 (2008).

[40] M. Karimi, A. Bleys, R. Vanderhaeghen, P. Hilson. *Plant Physiol* **145**, 1183 (2007).

[41] I. Reichardt, *et al. Curr Biol* **17**, 2047 (2007).

[42] R. Schwab, *et al. Plant Cell* **18**, 1121 (2006).

[43] I. Waizenegger, *et al. Curr Biol* **10**, 1371 (2000).

[44] F. F. Assaad, Y. Huet, U. Mayer, G. Jürgens. *J Cell Biol* **152**, 531 (2001).

[45] G. H. Goodwin, C. Sanders, E. W. Johns. *Eur J Biochem* **38**, 14 (1973).

[46] K. D. Grasser. *Plant Mol Biol* **53**, 281 (2003).

[47] K. D. Grasser, D. Launholt, M. Grasser. *Biochim Biophys Acta* **1769**, 346 (2007).

[48] J. Lichota, C. Ritt, K. D. Grasser. *Biochem Biophys Res Commun* **318**, 317 (2004).

[49] K. Terada, *et al. Dev Biol* **291**, 398 (2006).

[50] L. Ronfani, *et al. Development* **128**, 1265 (2001).

[51] M. G. Rosso, *et al. Plant Mol Biol* **53**, 247 (2003).

[52] M. Grasser, *et al. J Mol Biol* **358**, 654 (2006).

[53] I. B. Lolas, *et al. Plant J* **61**, 686 (2010).

[54] M. Cavalar, *et al. Biochemistry* **42**, 2149 (2003).

[55] M. Grasser, J. M. Christensen, C. Peterhänsel, K. D. Grasser. *Biochemistry* **46**, 6375 (2007).

[56] A. Houben, *et al. Biochim Biophys Acta* **1769**, 308 (2007).

[57] M. d. l. P. Sanchez, *et al. Semin Cell Dev Biol* **19**, 537 (2008).

[58] J. Van Leene, *et al. Mol Cell Proteomics* **6**, 1226 (2007).

[59] C. Guerrero, C. Tagwerker, P. Kaiser, L. Huang. *Mol Cell Proteomics* **5**, 366 (2006).

[60] M. W. Pfaffl. *Nucleic Acids Res* **29**, e45 (2001).

[61] J. Vandesompele, *et al. Genome Biol* **3**, RESEARCH0034 (2002).

[62] M. Karimi, D. Inze, A. Depicker. *Trends Plant Sci* **7**, 193 (2002).

[63] M. Sauer, T. Paciorek, E. Benková, J. Friml. *Nat Protoc* **1**, 98 (2006).

[64] M. H. Lauber, *et al. J Cell Biol* **139**, 1485 (1997).

3.6 Supplemental data

Supplemental Table 3.1: PCR primers for verification of the junction of the two copies of the 59.3 kb duplication on chromosome 5

Name	Sequence
Duplication_1	AAACGAGGAGGTTTAGACGGATTC
Duplication_2	CGTCACCCACCCACTTACATTG
Duplication_3	CAATGTAAGTGGGTGGGTGACG
Duplication_4	ATTTGAGAAGTTGCATTTGGGTTCC
Duplication_5	CATGGAACCCAAATGCAACTTCTC
Duplication_6	TGCAACTCCTCCCATACCAAAATAG

Supplemental Table 3.2: Primers for qPCR

Name	AGI	Forward	Reverse
3xHMG-box1	AT4G11080	GCCTAACATGACTTTTGCTTTTAG	TTCCACTTAGCACCCAGAATG
3xHMG-box2	AT4G23800	AAACATCCCGTATCTGCCTTCC	CTCCTCTCCAGTGATCTTTGCG
CYCB1;1	AT4G37490	CCTGGTGGAGTGGTTGATTGATG	CGACATGAGAAGAGCACTGAGAC
KN	AT1G08560	AGTAAAGGTGTTCACAAGGCAGAG	GCCCATTAGTCACAGCGGTTC
H4	AT5G59970	ACCAAATTGCGTGTTTCCATTG	ATGTCTGGTCGTGGAAAGGGAG
CDKA	AT3G48750	ATTGCGTATTGCCACTCTCATAGG	TCCTGACAGGGATACCGAATGC
CKIIa	AT3G50000	GAGCATCTACAACGGTTTACATTGG	TAGGCCGGTCATCGAGAATAGTC
CBP20	AT5G44200	ACCACCATTAACGTGCGTCAAC	GATCTTGGCGAGAGAATCGGTATC
4TM1	AT5G16250	TCGTTCTTGTGTGTGCTTTC	TTCGGATTCTCCTCACTCTG
4TM2	AT3G02640	CTGCAAACTACGTCCTCCTC	TTCTTACCATAAGATGCGGC
4TM3a	AT5G36710	CGGTCTCTCCCTTATCTTCC	GAGGACTGTGGCTACCATGT
4TM3b	AT5G36800		
3xHMG-box1	AT4G11080	AGCCTAACATGACTTTTGCTTTTAG	GCCTGATACTTCTCCTCATAAGG
3xHMG-box2	AT4G23800	AAAACATCCCGTATCTGCCTTCC	CTCCAGTGATCTTTGCGACCTC
HMGB6	AT5G23420	TGATGGACCTAAGCCAAAGAGAC	GCTAACGAACCATTATGCTCTGAC
ENODL14	AT2G25060	CCCATTAGCGGCTCAGTTAGG	GAACCAAGCACATAGACCAAGAAC
ENODL15	AT4G31840	CTACTTCGTCAGTGGAACAGAGG	AATCCAAGAACAAGCCCAAATACG
HASP	AT1G09450	TCAGCCATACACCAGGAGTCAG	TGAAGTGTTGCCAGCAGAAGAC
BAR	AT2G32590	ACCTTCCTTTGATGCCCTTAACG	ACTTGACACCCTCCATAGACTCC
ATEHD2	AT4G05520	AAGCCTTTGGAAGTCACATATCGG	GCTCATCTGGTCCACTCATTGC
WD40	AT3G42660	GCTATCTCCTACAGTTGGTGTTCC	AGATGCCGAAGGTAAGCTACTCC

Supplemental Table 3.3: PCR primers for analysis of Salk lines

Salk-line	Forward primer	Reverse primer
SALK_028428	TGGATTTTGAGAAATGGGTTTCG	CAGCGCTTAATAGCTCGTATTTC
SALK_066297	TGGATTTTGAGAAATGGGTTTCG	CAGCGCTTAATAGCTCGTATTTC
SALK_031814	TTTCAAATTTGTTTCTTTTGTGTG	CGAAACCCATTTCTCAAAATCCA
SALK_032395	AGCAGCATCTTCCTCACGCAC	TCACCTGTCGATGATGGTGGA
SALK_039559	TGGATTTTGAGAAATGGGTTTCG	CAGCGCTTAATAGCTCGTATTTC
SALK_115974	CCCATGACGATGTTGGAGGTT	TCTCCATTTCGTGAAGCTCTGG
SALK_114432	CCCATGACGATGTTGGAGGTT	TCTCCATTTCGTGAAGCTCTGG
SALK_046334	CCATTTCACCAATGAAGGACCA	TGCAAACTACGTCCTCCTCGG
SALK_046343	CCATTTCACCAATGAAGGACCA	TGCAAACTACGTCCTCCTCGG
SALK_046332	CCATTTCACCAATGAAGGACCA	TGCAAACTACGTCCTCCTCGG
SALK_060336	CCCATGACGATGTTGGAGGTT	TCTCCATTTCGTGAAGCTCTGG
SALK_076440	CCCATGACGATGTTGGAGGTT	TCTCCATTTCGTGAAGCTCTGG
SALK_041170	ATGGGAAGAGACGGAGGGAGA	AAAAACTGTTTGTTACGTTTGATTG
SALK_101711	TGGACACACATTGTGGAAAACAAA	CAAGGATTGGCCTTCTTACCCA
SALK_087783	CCGAATTTCGCAGTCCTAATG	CATCTTCTCCAACATCGCTTG
SALK_087786	CCGAATTTCGCAGTCCTAATG	CATCTTCTCCAACATCGCTTG

Part II

Small RNAs in Leaf Development of *Arabidopsis thaliana*

4

Small RNAs in *Arabidopsis*

Frederik Coppens[1,2], Dirk Inzé[1,2], and Gerrit T.S. Beemster[1,2,3]

[1]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), 9052 Ghent, Belgium
[2]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Ghent, Belgium
[3]Department of Biology, University of Antwerp, 2020 Antwerp, Belgium

4.1   Introduction

Tight spatiotemporal regulation of gene expression is essential in multicellular organisms. To provide a robust system, several layers of control are in place, including promoter activity, chromatin status, compartmentalization, complex formation, phosphorylation, ubiquitination,. . . A novel mechanism of gene regulation was discovered in *Caenorhabditis elegans* in 1993 where a small, non-coding RNA regulated a protein-coding gene through complementarity in the 3′ untranslated region (1). The key discovery to unravel the mechanism of RNA interference (RNAi) was made by Craig Mello and Andrew Fire, who established the crucial role of double-stranded RNA (2), for which they received the Nobel Prize for Medicine in 2006. A mechanism similar to RNAi (3, 4, 5) was found as a defense against viruses and to silence transgenes in different plant species, named post-transcriptional gene silencing (6), as well as in fungi, where the term quelling was used (7). Since then, enormous progress has been made to elucidate the regulatory pathways involving small RNA

(sRNA) molecules (8). sRNAs are important regulators of gene expression and it has been established that they play key roles in development (9, 10, 11, 12), resistance against viruses (13, 14, 15, 16, 17) and genome defense against selfish DNA and mobile elements (18, 19, 20). Due to their potential in gene regulation, small RNA pathways have been exploited early on, before the underlying mechanisms were known, in plant biotechnological approaches to understand gene functions and to improve crops (reviewed by Frizzi and Huang (21)).

In the beginning of this century, sRNAs were first sequenced in animals (22, 23, 24) followed soon by plant species (25, 26). More recently next generation sequencing technologies revealed an enormous diversity of sRNAs in *Arabidopsis* (27, 28, 29, 30) but also in a growing number of other plant species such as cotton, rice, maize, soybean, *Medicago*, trifoliate orange, *Brachypodium*, wheat, moss, Norway spruce and poplar (31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45).

The plant pathways for biogenesis and function of sRNAs conform to the same general framework (Figure 4.1). In all the currently known pathways in plants, double-stranded RNA (dsRNA) is involved in the biogenesis of sRNAs: either generated through RNA-dependent RNA polymerases (46, 47, 48, 49, 50, 51), from transcribed inverted repeats or stem-loops (52, 53, 54), through overlapping transcripts (55, 56) or through addition of exogenous dsRNA (57, 58). This dsRNA is processed by RNaseIII-like enzymes (3, 59, 60) producing short 20-to-25-nucleotide (nt) sRNA/sRNA* duplexes (3, 59), where the 'starred' sRNA represents the antisense strand. Both strands are methylated at the 2-hydroxyl group of the $3'$ terminal nucleotides (61, 62, 63). One strand of this duplex, called 'guide strand', is loaded into an RNA-induced silencing complex (RISC, 4) and guides this to its RNA or DNA target for cleavage, translational inhibition or methylation (64, 65) while the antisense sRNA*, or 'passenger strand', is thought to be degraded (66). Two major size classes of sRNAs are formed in plants: 21- and 24-nt sRNAs. In general terms, the shorter sRNAs cause mRNA degradation or translational inhibition and the longer ones are active in DNA methylation (25, 64, 67).

RNA silencing has specialized in different pathways that, although they share a general mechanism, have developed unique triggers, components and functions (for recent reviews see Vazquez *et al.* (65), Chen (64), Voinnet (53) and Carthew and Sontheimer (68)). Small RNAs (sRNAs) are typically classified according to their origin (Figure 4.2). MicroRNAs are generated from a single-stranded RNA molecule, encoded in the genome, that folds into an imperfect stem-loop structure. All other sRNAs are grouped under the term short interfering RNAs (siRNAs, 69). These are subdivided into *trans*-acting siRNAs (ta-siRNA) that stem from specific *TAS*-loci, natural antisense transcript siRNAs (nat-siRNAs) originate from convergent transcription and heterochromatic siRNAs (hc-siRNAs) are associated with transposons and repeats.
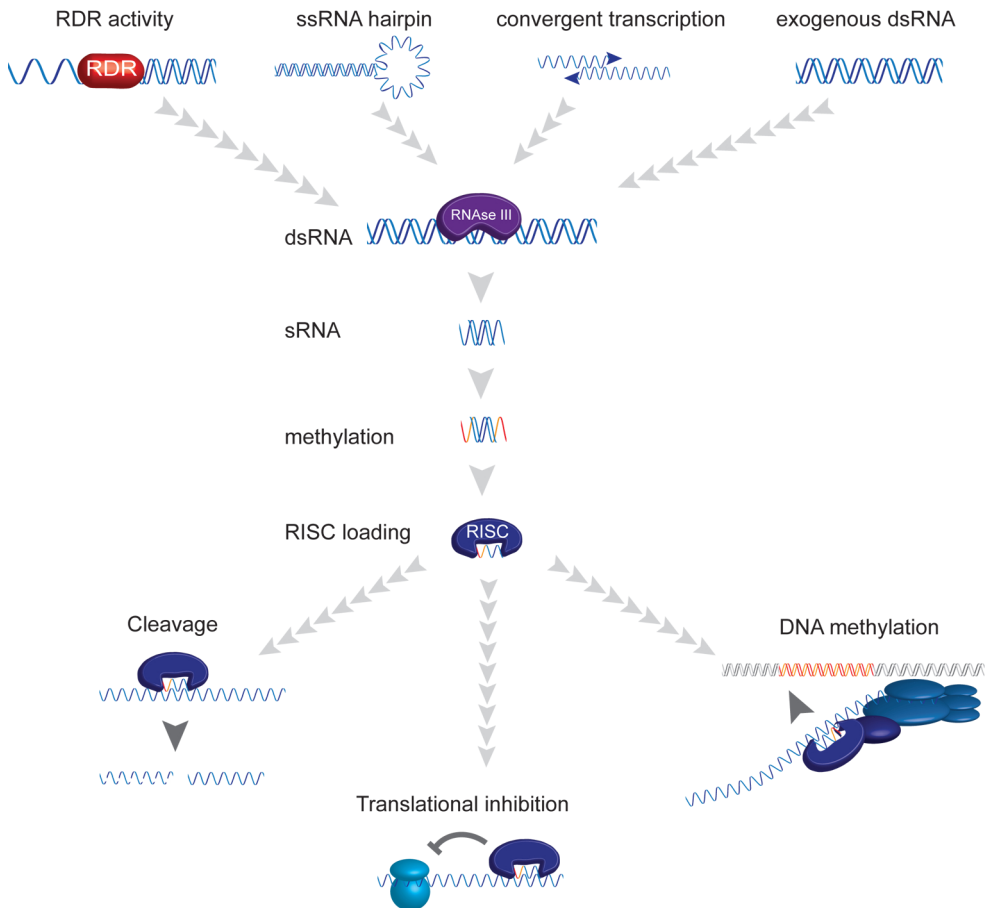
Figure 4.1: General mechanism of RNA silencing. Double-stranded RNA (dsRNA) can be formed through different ways: RNA-dependent RNA polymerase (RDR) activity, folding of single-stranded RNA (ssRNA) in a hairpin structure, convergent transcription or addition of exogenous dsRNA. This dsRNA is recognized by RNaseIII-like enzymes that cleave it in short 21- to 24-nt siRNA duplexes with a 2-nt 3′ overhang and the 2-hydroxyl group of the 3′ termini is methylated (in red). This duplex is loaded into an RNA-induced silencing complex (RISC), containing an ARGONAUTE protein as catalytic core, and one strand of the duplex is discarded. The remaining guide strand targets the RISC to complementary targets and causes cleavage, translation inhibition or DNA methylation at these loci.
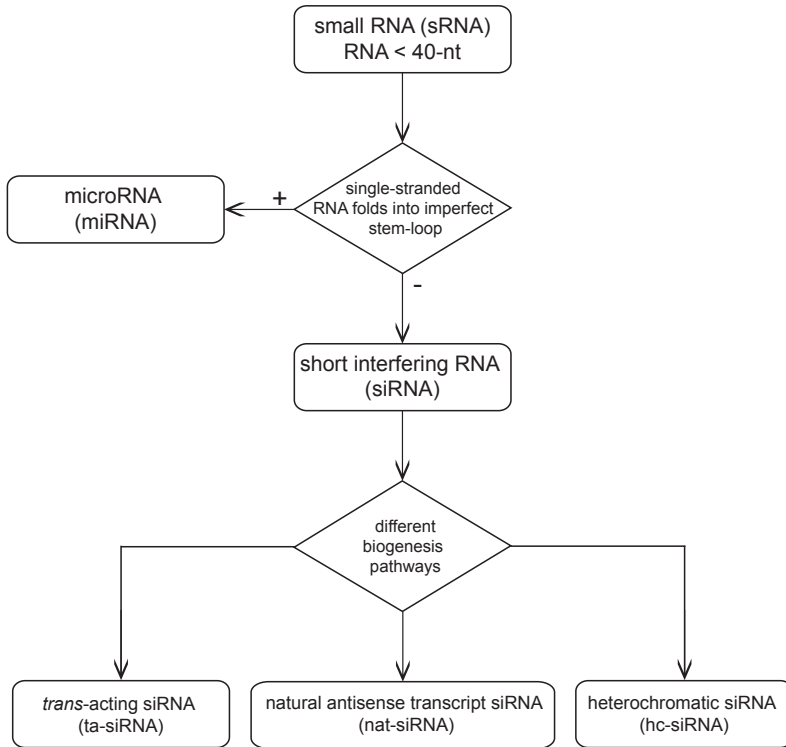
Figure 4.2: Nomenclature for small RNAs that is used in this review.

MicroRNAs and ta-siRNAs have important functions in gene regulation, especially in development (70, 71). nat-siRNAs have been found in specific inducing conditions, such as salt stress (55). Small RNAs are also important to protect genome integrity through silencing of transposable elements (20) and as defense against particular viruses (17). As a side-effect, introduced transgenes can also be targeted (72). In this introduction I will give an overview of the different silencing pathways in plants and outline for each the biogenesis pathway and known targets. First, the major gene families involved in the different biogenesis pathways will be covered.

4.2   Molecular components

A limited number of protein families are involved in most of the small RNA biogenesis pathways discovered to date. These common components will be introduced here, while pathway specific proteins are described as part of the different small RNA biogenesis pathways.

4.2.1   Dicer-like enzymes

Arabidopsis encodes four Dicer-like (DCL) proteins, double-stranded RNA (dsRNA)-specific RNaseIII ribonucleases, named DCL1 to DCL4 (Figure 4.3, 30, 60, 73). They contain N-terminal DexH-box RNA-helicase-C motifs, a Domain of Unknown Function 283 (DUF283), a PIWI/ARGONAUTE/ZWILLE (PAZ) domain, two tandem ribonuclease-III (RNaseIII) motifs and one or two C-terminal dsRNA-binding domains (74). DCL1, the founding member of this family in *Arabidopsis*, has been identified in several mutant screens and therefore has been known under different names (EMBRYO DEFECTIVE76 (EMB76, 75), SHORT INTEGUMENTS (SIN1, 76), SUSPENSOR1 (SUS1, 77) and CARPEL FACTORY (CAF, 78)). Based on its similarity with DICER in *Drosophila*, it was renamed *DICER-LIKE1* (*DCL1*) (74). In mammals, *Caenorhabditis elegans*, *Chlamydomonas reinhardtii* and fission yeast *Schizosaccharomyces pombe*, the Dicer-like family is represented by a single gene, while *Drosophila* and the fungi *Neurospora* and *Mucor* have two. In plants, however, the *DCL* family has expanded to 4 members in *Arabidopsis* and *Physcomitrella patens*, while 6 genes have been identified in rice (74, 79, 80, 81).
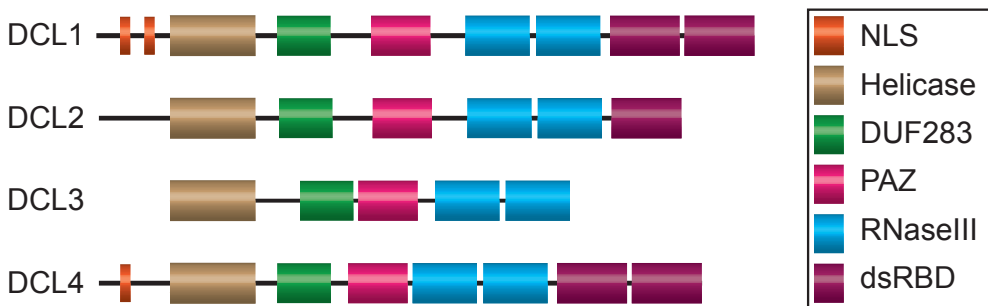
Figure 4.3: The domain structure of the *Arabidopsis DCL* family: Nuclear Localization Signal (NLS), DexH-box RNA-helicase-C motif (Helicase), Domain of Unknown Function 283 (DUF283), PIWI/ARGONAUTE/ZWILLE domain (PAZ), two ribonuclease-III motifs in tandem (RNaseIII) and one or two C-terminal dsRNA-binding domains (dsRBD) (73, 82).

DCLs cleave dsRNA into 21-to-24-nt sRNA duplexes with a 3′ 2-nt overhang (3, 69, 83, 84, 85) and 2′,3′-hydroxyl and 5′-monophosphate termini (86). Cleavage activity of DCL1 requires a divalent cation and ATP (87). DCL1 and DCL4 produce 21-nt, DCL2 22-nt and DCL3 24-nt sRNAs (30, 73, 88, 89, 90). The crystal structure of the protein, more specifically the distance between the PAZ and RNAseIII domains, can explain these different size classes (91, 92), but different binding partners could also play a role in modulating this characteristic cutting distance.

DCL1 and DCL4 contain a Nuclear Localization Signal (NLS) but also DCL3 and DCL2 reside in the nucleus (74, 93, 94, 95). The mammalian Dicer, Drosha, requires

phosphorylation at either one of two serine residues in the N-terminus for nuclear localization (96).

DCL1 is the most important family member for miRNA biogenesis (26, 90), explaining the severe developmental defects that have been observed in different mutants (reviewed in Schauer *et al.* (74)). DCL2 processes exogenous elements and some nat-siRNA loci (55). DCL4 works in a hierarchical way with DCL2 to produce sRNAs from viruses and transgenes (14, 97, 98). DCL3 is implicated in siRNA biogenesis from heterochromatic loci (94). Although each DCL has its preferred substrates, there is also functional redundancy (89, 99, 100). dsRNA binding (DRB) proteins that interact with DCLs, have limited redundancy. DCL1 and HYPONAS-TIC LEAVES 1 (HYL1)/DRB1 interact and also DCL4 and DRB4 form a complex (95, 101). DCL2 and DCL3 seem to operate without such a binding partner. For the other DRB homologs (DRB2, DRB3, DRB5), thus far no function has been uncovered (101).

4.2.2   RNA-dependent RNA polymerases

The *Arabidopsis* genome contains six genes that encode RNA-dependent RNA (RDR) polymerases (50). Three of them (RDR1, RDR2 and RDR6) have established roles in small RNA silencing. *In vitro*, RDR2 can synthesize dsRNA from a single-stranded RNA (ssRNA) template both primer-independent and primer-dependent (86). RDR6 is able to use both ssRNA and ssDNA as a template to generate *de novo* a complementary RNA strand. Primer-dependent activity was not detected. The activity of RDR1 has not been demonstrated yet, but is likely similar to RDR2 and RDR6 (48, 86). The exact mechanisms of target recognition for RDRs are not yet known. Aberrant features, such as the lack of a poly-A tail or 5′ cap, could attract RDR activity. A cytoplasmic 5′-3′ RNA exoribonuclease, EXORIBONUCLEASE 4 (XRN4), degrades decapped mRNAs in yeast and antagonizes silencing in *Arabidopsis*, confirming this hypothesis (102, 103). RDR6 cannot distinguish aberrant RNA from RNA that is properly capped and has a poly-A tail, therefore probably other proteins are required to recognize (the lack of) these features (104).

All three RNA silencing RDRs are involved in siRNA biogenesis from plant viruses (105, 106, 107). For RDR1, viral defense seems to be the primary role (51, 104, 108, 109, 110), while RDR2 mainly functions in the heterochromatic pathway (28, 29, 94). RDR6 has a broader functional diversification. It targets exogenous sequences, both from viral and transgene origin (51, 104, 110, 111, 112, 113, 114), it is indispensable for endogenous loci such as ta-siRNA generating loci (51, 71, 110, 115) and for the biogenesis of natural antisense transcript siRNAs (55, 56). RDR6 is also required for transitivity, a process where secondary siRNAs are produced from regions flanking the primary targeted site, and for long-range spread of transgene silencing (116, 117, 118, 119).

4.2.3 Methyl transferase

Contrary to mammals, all sRNAs that have been found in plants are methylated. The sRNA:sRNA* duplex is methylated at the 2-hydroxyl group of the 3′ ends by the methyltransferase HUA ENHANCER1 (HEN1) (61, 62, 63, 120, 121). HEN1 methylates individual strands successively and does not require additional proteins for its activity (122). The addition of a methyl group protects the 3′ end from uridylation and exonucleolytic RNA degradation (62). This modification could potentially affect interactions of siRNAs, but at this point no experimental evidence supports this hypothesis (121).

4.2.4 RISC and ARGONAUTES

Small RNA function is mediated by an RNA-induced Silencing Complex (RISC), containing an sRNA guide and an ARGONAUTE (AGO) protein as catalytic center (4, 123, 124, 125). The AGO protein, guided by the loaded sRNA, regulates gene expression through cleavage, translational inhibition or DNA methylation and chromatin remodeling (126). Upon target cleavage, the 3′ cleavage fragment is degraded by 5′-3′ RNA exoribonucleases (XRN), primarily XRN4, assisted by the nucleotidase/phosphatase FIERY1 (FRY1) (103, 127, 128, 129). 5′ fragments are probably degraded by the exosome (48, 130, 131). The 3′-5′ exoribonuclease WERNER SYNDROME-LIKE EXONUCLEASE (WEX) is also involved in post-transcriptional silencing but its precise function remains elusive (132).

The mechanisms behind loading an sRNA in RISC are not yet known (53). Recent studies of human RISC loading suggested a Dicer-mediated siRNA transfer to AGO (133). The sRNA duplex is loaded into RISC in an ATP-dependent process (134, 135, 136, 137), in association with the Hsc70/Hsp90 chaperone machinery (138, 139). The chaperone machinery could mediate a dynamic conformational change in the AGO protein allowing the duplex to enter (139). Subsequent release of the tension could help drive strand separation. This loading mechanism is referred to as the 'rubber band' model (139, 140). The requirement to open the AGO protein matches the fact that the resolved structure of the protein does not accommodate enough space for the sRNA duplex (133, 141, 142). To generate an active RISC complex, one of the strands of the siRNA duplex needs to be removed. Reports in the animal field have shown cleavage of the passenger strand upon loading of the duplex into RISC (134, 136, 143), but an alternative mechanism based on thermal destabilization and unwinding has also been demonstrated (137, 142, 144). The same two mechanisms, cleavage and cleavage-independent unwinding are present in plants, an siRNA required endonuclease activity for removal of the passenger strand but the two investigated miRNA* did not. Both mechanisms required ATP hydrolysis (139). Cleavage of targets does not require ATP, but removal of the cleaved target from the RISC complex does (145). If these *modi operandi* can be generalized in

plants for siRNAs or miRNAs remains to be investigated. The Hsc70/Hsp90 chaperone machinery does not mediate removal of the cleavage fragment, but which additional factors it requires is still unknown (138).

The founding member of the ARGONAUTE family, AGO1, was discovered in plants based on developmental defects in a null mutation, with leaves that resembled a small squid (146). AGO proteins are evolutionary ancient, as homologues are found in eubacteria and archaea. They can be divided into three subfamilies. In plants, only one clade is present, but similar to DCLs, this has expanded to 6 members in *P. patens*, 10 in *Arabidopsis* and 18 in rice (80, 147, 148).

At the C-terminus, all AGOs contain three conserved domains: P ELEMENT-INDUCED WIMPY TESTIS (PIWI)/ARGONAUTE/ZWILLE (PAZ), middle (MID) and PIWI (126). The N-terminal domain is variable. The PAZ domain is also found in DCLs. The MID and PAZ domain bind the 5′ phosphate and 3′ end of sRNAs, respectively (141, 149, 150, 151). The strand of the siRNA duplex that is loaded into RISC is called 'guide', the other strand is denominated 'passenger'. The distinction is made based upon the relative stability of the 5′ ends (152, 153). The PIWI domain adopts an RNAseH fold and confers RNA endonuclease activity to the proteins (123, 124, 154, 155, 156). In *Arabidopsis*, slicer activity has been shown for a subset of AGO proteins: AGO1, AGO4 and AGO7 (88, 125, 126, 157, 158). The target RNA is cleaved at the phosphodiester bond between the 10th and 11th nucleotide complementary to the guide strand (69, 85, 159). The 5′ nucleotide of the guide strand proved to be important for AGO loading (158, 160, 161, 162) and this mechanism appears to be conserved in all angiosperms (163). This preference for a specific 5′ nucleotide has been attributed in human AGO2 to interaction of the 5′ terminus with a rigid loop in the MID domain (164). Upon duplex formation with the target, the 3′ end of the sRNA is released but the 5′ end stays anchored to the MID domain, which could be important for proper target processing (156, 164). Additionally, the small RNA biogenesis pathway can exhibit preference for a specific AGO as demonstrated in rice (162).

AGO1 is the most important member of the AGO family for miRNAs and ta-siRNA processing, which contain predominantly a 5′ uridine (88, 125, 158, 160). AGO7 is specialized in loading miR390 to form ta-siRNAs from the *TAS3* locus. The close association of AGO1 with miRNAs explains the dramatic developmental defects *ago1* mutants exhibit, especially compared to the other members of the AGO family that show little (*ago7* and *ago10*) or no obvious phenotype (*ago2*, *3*, *4*, *5*, *6* and *9*) (126). AGO10, also known as PINHEAD (PNH) or ZWILLE (ZLL) (165), is at least partially functionally redundant with AGO1 (166). Based on discrepancies between mRNA levels and protein levels, Brodersen *et al.* (167) demonstrated that AGO1 and AGO10 may function through translational inhibition for certain miRNAs. Remarkably, the phenotypes of *ago1* and *ago10* are more severe in a L*er* background than a Col background, suggesting the presence of ecotype specific modifiers (126).

AGO4, AGO6, AGO8 and AGO9 are closely related. *AGO8* and *AGO9* appear to be the result of a recent duplication. The low expression and splicing-induced frameshift suggest that *AGO8* could be a pseudogene. AGO4, AGO6 and AGO9 preferentially load 24-nt sRNAs with a 5′ adenosine (157, 160, 168). 24-nt small RNAs from the heterochromatic pathway are loaded into AGO4. AGO6 acts in heterochromatic siRNA, DNA methylation and transcriptional gene silencing pathways, showing partial redundancy with AGO4 (169). AGO9 is crucial in specification of the cell fate of the *Arabidopsis* ovule (168, 170).

Mutants in the remaining AGO members, *AGO2*, *AGO3* and *AGO5*, do not show obvious developmental phenotypes (126, 161, 171). AGO2 prefers a 5′ adenosine while AGO5 favors a cytosine at the 5′ end but their biological function remains to be elucidated (158, 160, 161).

4.3   Small RNA biogenesis pathways

Small RNA (sRNA) biosynthesis pathways are categorized based on the origin of the double-stranded RNA (dsRNA) that triggers sRNA production. Best studied are microRNAs (miRNAs), which originate from imperfect foldback structures. Another endogenous pathway produces *trans*-acting short interfering RNAs (tasiRNAs) that are generated starting from specific *TAS* loci. Both pathways regulate their targets post-transcriptionally through RNA cleavage and/or translational inhibition (53, 64, 65). Transcriptional gene silencing occurs for mobile elements such as transposons in order to maintain genome integrity, but also promoters can be targeted by the silencing machinery, thus regulating gene expression (64). The RNA silencing machinery also plays a role in plant defense against viruses, deploying strategies depending on the virus (17). Recently, a new mode of action was identified inhibiting transcription elongation in *C. elegans*, termed co-transcriptional silencing (172). Here, I will describe the state-of-the-art of the small RNA biogenesis pathways in plants.

4.3.1   microRNAs

MicroRNAs (miRNAs) are post-transcriptional regulators of gene expression with a remarkable bias towards development in plants, including the unicellular green alga *Chlamydomonas reinhardtii*, as well as animals (9, 10, 53, 70, 80). miRNAs are generally 20- to 22-nt sRNAs generated from imperfect foldback hairpin structures (9, 10, 53, 70). The first miRNAs were all found in mutants with developmental defects and have roles in e.g. organ identity, organ morphogenesis, developmental transitions and leaf and floral morphogenesis (9, 173).

Although the biogenesis and structure of miRNAs in plants and animals is quite similar, they exhibit different regulatory approaches. While metazoan miRNAs typ-

ically target multiple sites in the 3′ untranslated region (UTR), plant miRNAs usually recognize a single target site within the coding region. Also the base-pairing rules for target recognition seem to differ, where the plant mechanism is more stringent and requires high complementarity. Here, I will focus on the mechanisms in plants. For the biogenesis and function of miRNAs in animals I refer to recent reviews (68, 174, 175, 176, 177).

4.3.1.1   Origin & evolution

Genes encoding miRNAs have been found in higher plants, animals and the alga *Chlamydomonas reinhardtii*, but not in unicellular choanoflagellates and fungi, suggesting that the miRNA pathway was established before the split of plants and animals. Conserved miRNAs present in both plants and animals have not been identified, indicating that miRNA encoding (*MIR*) genes have evolved separately in both kingdoms (81, 178, 179, 180). In plants, conservation of 'ancient' miRNAs dates back to a common ancestor of angiosperms and mosses, but after separation of algae (181, 182, 183, 184). Only few miRNAs (e.g. miR156/157, miR160, miR165/166, miR170/171, miR159, miR319, miR390, miR396, miR408) are found that are in common between mosses, lycopods and angiosperms. The majority of these miRNAs is involved in developmental regulation and retain their homologous targets (80). Twenty-one miRNA families have been found in all angiosperms and thus date back from before the split of monocots and dicots (miR156/157, miR159, miR160, miR162, miR164, miR165/166, miR167, miR168, miR169, miR170/171, miR172, miR319, miR390, miR393, miR394, miR395, miR396, miR397, miR398, miR399 and miR408) (181). An exhaustive study for conservation of miRNAs in 155 plant species was performed by Sunkar and Jagadeeswaran (184). While in most angiosperms miRNAs are encoded from independent loci, in *P. patens* and *C. reinhardtii* polycistronic precursors give rise to miRNAs for 48 out of 205 predicted miRNAs (80, 185). These loci produced paralogs and not unrelated miRNAs as for animal miRNA clusters.

   With the use of deep sequencing techniques, it became clear that, next to the already known and generally conserved miRNAs, a large number of species-specific miRNAs is present in plants (27, 39, 40, 42, 179, 183, 186, 187, 188, 189, 190, 191, 192). About 200 miRNAs have been identified in *Arabidopsis* of which the most conserved have important functions in patterning, development and cell identity (9, 53). More recently evolved miRNAs make up the majority of miRNAs and are predicted to cover a broader range of target processes (27, 179, 183). These lineage specific ones are usually expressed at lower levels. The identification of miRNAs in very specific conditions or cell-type specific expression patterns furthermore suggests that not all miRNAs have been identified at present. An exhaustive overview of currently known miRNAs can be found at miRBase (www.mirbase.org, 193) and a recently developed plant-specific database Plant microRNA Database

(bioinformatics.cau.edu.cn/PMRD, 194).

The emergence of miRNAs at different time points during evolution, can provide some insight into the origin of miRNAs and help understand exceptions on the canonical biogenesis pathway. Currently, it is believed that *MIR* genes originate from inverted repeats, possibly through an inverted duplication (27, 195, 196) but also through coincidental trapping of potential regulatory sequences in miniature inverted-repeat transposable elements (MITE, 196, 197) or in foldback structures (178, 198) present in the genome. Over a thousand inverted repeat loci have been found in *Arabidopsis* (29). During evolution these potential regulatory sites can be selected for and used to regulate its nascent target or a new unrelated target can be adopted. Young miRNAs are more divergent compared to conserved miRNAs and seem to be under less evolutionary constraints. Also the targets of young miRNAs show more divergence, as observed in the closely related *Arabidopsis thaliana* and *A. lyrata* (196, 199). During evolution, *MIR* genes gradually adopt a short dsRNA hairpin structure and the different intermediary forms are related to the biogenesis pathway used to produce mature miRNAs. The canonical miRNAs are processed by DCL1. Long inverted repeats are substrates for DCL2, while DCL4 is the primary DCL for an intermediary length (27, 195). The switch to DCL1 might ensure better processing accuracy (27). Based on these new insights, it has been suggested to use the term 'miRNA-like siRNAs' for sRNAs from long hairpins that are not *MIR* genes (200).

4.3.1.2 Biogenesis

Most *MIR* genes are transcribed by RNA polymerase II (PolII), similar to protein-coding genes (Figure 4.4, 201, 202, 203, 204, 205). Therefore, the primary miRNA transcript (pri-miRNA) has a 5′ 7-methyl guanosine cap and is 3′ poly-adenylated (205, 206) and makes them prone to the same regulatory mechanism. Their expression can be spatially and temporally restricted as well as induced by environmental clues or as a response to stress.

This pri-miRNA adopts an imperfectly matching foldback structure and is further processed into a stem-loop precursor (pre-miRNA) through cleavage by DICER-LIKE1 (DCL1). This takes place in nuclear processing bodies, named Dicing-bodies (D-bodies, 207, 208, 209) and is promoted by interaction with DAWDLE (DDL), an RNA binding protein, that probably facilitates binding of DCL1 and stabilizes the pri-miRNA (210). Also CAP BINDING PROTEIN 20 (CBP20) and CBP80, subunits of the nuclear cap-binding complex (CBC) are required (95, 129, 211). A second cleavage action of DCL1 produces a miRNA/miRNA* duplex with a 2-nt 3′ overhang from the pre-miRNA (53, 69). The excised miRNA loop is degraded by the partially redundant exoribonucleases XRN2 and XRN3 (127). Both the dsRNA binding protein HYPONASTIC LEAVES 1 (HYL1) and the C2H2 zinc-finger protein SERRATE (SE) are required for efficient and accurate processing of the pri- and pre-

miRNA (87, 171, 212, 213, 214, 215, 216, 217). A HYL1 homodimer probably binds the miRNA/miRNA* region of the precursor (218).

The secondary structure of the stem-loop varies widely in plants (219). Recently, determinants for correct processing of the pri- and pre-miRNA have been elucidated. The lower stem region is important and the initial, loop-distal cleavage occurs at about 15 nt from an unpaired region in the lower stem. The upper stem region is more tolerant to mutations, but a defined structure at the second cleavage site is needed as well as a minimum stem length between the miRNAs site and the loop region (220, 221, 222, 223). Alternative processing is possible as has been demonstrated for miR159 and miR319, which both contain the miRNA site at the proximal base of the foldback and are loop-to-base processed (224, 225).

The miRNA/miRNA* duplex is 2′O-methylated at the 3′ termini by the S-adenosyl methionine-dependent methyltransferase HUA ENHANCER 1 (HEN1) (61, 62, 63, 226, 227, 228). Through this addition of methyl groups, the small RNA duplex is protected from uridylation and subsequent exonucleolytic degradation (62). Transport from the nucleus to the cytoplasm occurs either during or following the processing of the pre-miRNA to a miRNA/miRNA* duplex. Some miRNAs depend on HASTY (HST), an exportin-5 ortholog (229, 230, 231, 232, 233), for their accumulation in both nucleus and cytoplasm (234). This suggests a role for HST in export from the nucleus, but it is unclear at this point, where in the biogenesis pathway this exactly occurs (53).

AGO1 is the major effector of miRNAs, but also AGO7 and AGO10 have been shown to use miRNAs as guide (235, 236). AGO1 activity is promoted by SQUINT (SQN), a cyclophilin40 homolog (237, 238) and HSP90.2 (139, 238). While the miRNA strand is loaded into the RISC complex, the miRNA* is thought to be degraded (239). Recent evidence however demonstrated incorporation of the passenger strand in AGOs and target cleavage (161, 240, 241). The exact localization of RISC loading and degradation is not yet known (53).

Figure 4.4 *(on the next page)*: **miRNA biogenesis pathway.** *MIR* genes are transcribed by PolII and adopt an imperfect foldback structure that is stabilized by DAWDLE (DDL). The primary miRNA transcript (pri-miRNA) is processed to a miRNA precursor (pre-miRNA) through cleavage by DCL1, aided by HYPONASTIC LEAVES 1 (HYL1) and SERRATE (SE). Also the nuclear cap binding complex (CBC) is important for this processing. A second cleavage by DCL1 releases the miRNA/miRNA* duplex which is 3′ methylated by HEN1. HASTY (HST) is involved in the transport from the nucleus to the cytoplasm. Next, the guide strand is loaded into AGO1, assisted by SQUINT (SQN) and HSP90.2. This forms an active RISC complex that regulates targets by cleavage or translational inhibition. (DNA grey/yellow, RNA blue, methylated RNA red)
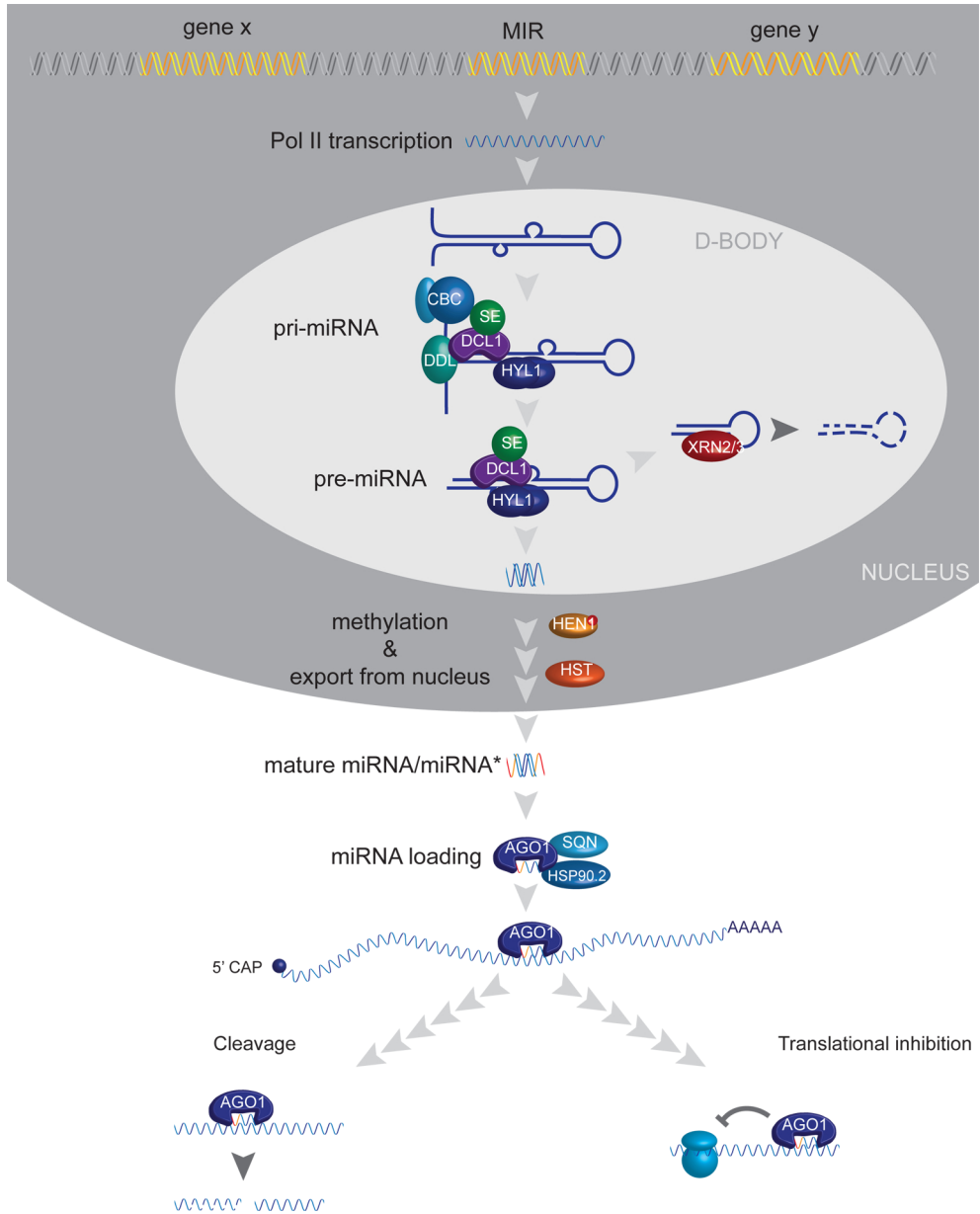
4.3.1.3 Alternative processing

Although redundancy within the DCL family has been observed (89, 99), miRNA biosynthesis almost exclusively depends on DCL1 (90). Most miRNA stem-loops are processed by DCL1, but some miRNAs depend on DCL4 (miR822, miR839). Their precursors are evolutionary young and do not (yet) conform to the canonical miRNA stem-loop structure but are long hairpins with extensive complementarity (26, 27).

For miRNA accumulation different dependencies were found in leaves and flowers. A *dcl2* mutant allele and to a lesser extent also a *dcl4* mutant, caused accumulation of a subset of miRNAs in leaves but not in flowers, likely due to reduced competition for co-factors such as HYL1 (212). When *dcl1* and *dcl4* are mutated, an additional *DCL2* mutation reduces mortality, indicating that DCL2 somehow interferes with development (90). DCL1 activity inhibits *DCL4* and *DCL3* expression (107).

Also DCL3 can process hairpins, yielding long-miRNAs which guide methylation through AGO4 instead of AGO1 for the other DCLs (27, 195). This could be attributed to the transposon-derived origin of some miRNAs (53).

4.3.1.4 Regulation of miRNA biosynthesis

Transcriptional regulation
The miRNA regulatory system is itself also highly regulated. As most *MIR* genes are transcribed by PolII, their promoters are under control of transcription factors. Indeed, *MIR* loci could be identified in which known promoter motifs are overrepresented (204, 242, 243, 244). Their expression can also be affected by chromatin remodeling such as histone modifications and DNA methylation. Recently, histone acetylation has been shown to regulate miRNA accumulation (245). Analogous to protein-coding genes, miRNAs can be induced by specific conditions and have precise spatiotemporal expression patterns. Examples will be addressed at the end of this review.

Post-transcriptional regulation
The availability of the necessary components also impinges on miRNA biogenesis. Short interspaced elements (SINE) can form stem-loops and bind to HYL1, thus mimicking miRNA-deficient mutants. Competition for HYL1 could therefore control miRNA production (246). HEN1 seems to be a limiting factor in L*er* background, as siRNAs and miRNAs compete for methylation by this methyltransferase. Surprisingly, this was not the case in Col background, which might be attributed to a negative modulator of HEN1 activity in Col (247). These cross-regulations between RNA silencing pathways are not uncommon. In *rdr2* mutant background all heterochromatic siRNAs are lost (see further), leading to an upregulation of

miRNAs, possibly due to the increased availability of limiting factors. The same pattern is seen in *dcl2dcl3dcl4* triple mutants (30).

Feedback loops

Control of miRNAs by their targets is well known in the animal field and also found in plants (248, 249, 250, 251). *DCL1* is targeted by miR162 (252) and an intron of *DCL1* encodes miR838 (27). The abundance of *DCL1* transcripts could thus be controlled by a regulatory feedback mechanism. Similarly to the arrangement in *Arabidopsis* (27), also in *P. patens* a miRNA has been found in an intron of *DCL1* (miR1047), but this is most likely independently evolved as this could not be found in rice (80).

Also the major effector *AGO1* is itself targeted by a miRNA, miR168 (235, 253, 254). miR168 is thought to be inefficiently loaded into AGO1, explaining that in *dcl1* mutants miR168 (and miR165/166) is only slightly affected, compared to the strong reduction of most miRNAs. This mild effect is hypothesized to be caused by enhanced loading into AGO1 due to less competition (253). A similar feedback loop was found in *P. patens* with miR904, which could have evolved from a common ancestor (80). The presence of feedback loops in both angiosperms and bryophytes indicates the importance of these regulatory controls. AGO10 negatively regulates AGO1 at the protein level but the interplay between these proteins is not yet understood (255).

Arabidopsis hypomorphic *dcl1* mutants accumulate *DCL3* and *DCL4* transcripts and *DCL1* upregulation reduced *DCL3* and *DCL4* expression levels, suggesting normal DCL1 activity negatively regulates these *DCL*s, through as yet unknown mechanisms. *DCL2* on the other hand, seems to be only slightly affected by DCL1 (107, 256).

Target mimicry

In animals, the effectiveness of miRNAs or siRNAs has been shown to depend on the abundance of its targets. This could affect target repression, e.g. a miRNA with few targets could downregulate sequences with non-optimal target sites while the presence of many or high-abundant targets could cause diminished repression ability (257). In plants, a regulatory mechanism based on this principle was identified and termed 'target mimicry'. INDUCED BY PHOSPHATE STARVATION1 (IPS1) is a non-protein-coding gene targeted by miR399. However, the target site contains mismatches at positions 10-11, impeding cleavage and effectively sequestering miR399. Upon low phosphate levels, IPS1 accumulation deviates miR399 from its target PHOSPHATE2 (PHO2)/UBIQUITIN-CONJUGATING ENZYME24 (UBC24), causing a reduction in phosphate content of the shoot (258).

The study of ta-siRNA production at the *TAS3* locus in *Arabidopsis* revealed that a non-cleavable target site competes more efficiently with cleavable sites (44). Also in human cell lines this non-cleavage efficiency gain holds true and has been applied by introducing constructs containing binding sites in tandem, 'microRNA sponges'

(259, 260). An example of an expressed pseudogene serving as decoy for miRNA targeting of its protein-coding homologue has recently been identified in human (261). This provides new insights in the potential regulatory role of pseudogenes in the genome.

Degradation of small RNAs

The steady-state level of mature miRNAs is regulated through the balance of biogenesis and degradation. SMALL RNA DEGRADING NUCLEASE1 (SDN1), and likely also its homologs SDN2 and SDN3, degrade single-stranded short RNAs (262). Loss of SDN activity results in increased mature miRNA levels and causes pleiotropic developmental phenotypes, illustrating the importance of miRNA turnover in development. 2′O-methylation of the sRNA inhibited degradation by SDN1. Remarkably, also uridylation of sRNAs in the absence of methylation provided protection to exonuclease activity (262).

Different miRNAs can have different intrinsic stability. In human this has been investigated for an unstable miRNA where seven 3′ terminus nucleotides could confer increased stability when mutated (131).

In *C. elegans* and *S. pombe* Enhanced RNAi-1 (Eri-1) exonucleases degrade siRNA duplexes with a 2-nt 3′ overhang (263, 264). In worms, this is tissue specifically expressed in gonads and certain neurons, providing a protection against RNA silencing in these cells (263). Also in *C. elegans*, the 5′-3′ exoribonuclease XRN-2 has been identified as a mediator of miRNA homeostasis through release from the RISC complex and subsequent degradation of mature miRNAs. This turnover was repressed by target annealing *in vitro*, providing possibly an additional layer of regulation. Not all miRNAs seem to be affected by XRN-2, allowing the intriguing possibility of modulation of specific miRNAs dependent on certain conditions or developmental stages (265).

4.3.1.5   Mode of action: cleavage or translational inhibition

In plants, cleavage was long thought to be the major mode of action of miRNAs. In animals it was already clear that translational inhibition is an important way of silencing through small RNAs (266), but also in plants the importance of this mechanism is now recognized. Lowering target site complementarity shifted miR398 activity from cleavage to translational inhibition of *COPPER/ZINC SUPEROXIDE DISMUTASE1* (*CSD1*) (267). This shift resulted in a less strict regulation accompanied by variable penetrance, suggesting that target recognition could depend much less on complementarity than initially believed (268).

Through the use of antibodies against the product of target genes and demonstration of association of miRNA-loaded AGO1 and AGO10 with polysomes, the widespread nature of translational inhibition became clear (167, 269). It was demonstrated that miR156/157 can repress *SQUAMOSA PROMOTER BINDING PROTEIN-*

LIKE 3 (*SPL3*) through translational inhibition (270). The point mutation in *ago1-27* (147) appears to uncouple cleavage from inhibition, disrupting the latter, probably through disabling protein interactions with additional factors needed for translational inhibition. Mutations in *AGO10* cause similar phenotypes as *ago1* but specifically suppress translational inhibition by miRNAs (165, 166, 167).

Microtubuli appear to be involved in translational inhibition as a mutant in the P60 subunit of KATANIN (KTN1), an ATP-dependent microtubule-severing enzyme (271), is deficient in translational inhibition. For the cleavage mode of action, KTN1 is not necessary. Dependency on the microtubular cytoskeleton has also been reported in worms, flies and yeast (167). P-body component VARICOSE (VCS) (272, 273) was also found to be indispensible for translational inhibition. This protein is an ortholog of Ge-1, which is required for decapping of miRNA targets in flies (274), linking translational inhibition and mRNA decay (167).

Brodersen *et al.* (167) illustrated with different classes of *miRNA-deficient* (*mad*) mutants that most miRNAs act in two modes: target degradation and translation inhibition. What determines the ratio between the two modes of action is not clear, as it seems to be independent of complementarity and its localization within the target sequence. Moreover, translational inhibition can be uncoupled from target slicing and could be the primary mode of action for certain (near-) perfect complementary miRNAs. This clearly demonstrates that translational inhibition is not a mere side effect and offers an explanation for discrepancies seen with overexpression of miRNAs and artificial microRNAs (see further) that phenocopy known mutants while mRNA levels of the target genes are not significantly affected (275, 276, 277, 278). The precise mechanism of translational inhibition and at which step in translation it functions, is not yet known.

The two modes of action of miRNAs, cleavage and translational inhibition, both resulting in a reduction of the gene product, differ mainly in the reversibility. The cleavage mode of action regulates irreversible switches such as needed in cell fate changes during development. Also fine-tuning of spatiotemporal regulation fits this mode. In other cases, reversibility is required. Stress responses for example, should be rapidly suppressed when the stress is gone. Translational inhibition serves this purpose, as translation can continue immediately (53). Thus, miRNAs can be strict regulators of spatiotemporal gene expression, but also dampen expression or serve as a backup system or safeguard to ensure robustness of regulation.

4.3.1.6   Methylation

Evidence is emerging that miRNA target loci can be silenced in a dual mode, both through cleavage and methylation. Bao *et al.* (279) reported for miR165/166 induction of DNA methylation downstream of its targets *PHABULOSA* (*PHB*) and *PHAVOLUTA* (*PHV*), but this seemed not to be generally the case for miRNAs or ta-siRNAs (280). Remarkably, these *MIR* genes also generate 23- to 26-nt sRNAs

associating with AGO4 and AGO7, next to *bona fide* 20- to 22-nt miRNAs (195). Recently, new *MIR* genes have been identified that also produce both 20- to 22-nt miRNAs dependent on DCL1 and AGO1 and 23- to 27-nt dependent on the heterochromatic siRNA pathway (see further) (281). Also for known miRNAs it has been shown that many can produce 23- to 26-nt sRNAs from the same strand and region as the miRNA. How siRNA generation is limited to the miRNA-generating region, is not known. The hairpin structure could play a role, but also the miRNA itself might be involved. The longer siRNAs methylate the miRNA target genes probably through AGO4, but the generating locus is not affected. Especially asymmetric CHH (where H is A, C or T) sites are methylated, a known hallmark of the heterochromatic siRNA silencing pathway (281).

Also in *Physcomitrella patens* and rice, siRNAs matching miRNA loci were identified. In rice these siRNAs were more abundant than the miRNAs for about half of these sites (281). A similar miRNA action was found by Wu *et al.* (162). Three classes of miRNA sites were identified that generate 24-nt sRNAs guiding DNA methylation both in *cis* and in *trans*. Two of them generate 24-nt from the same site as 21-nt, one requiring both DCL1 and DCL3, the other only DCL3. RDR2 is not required, but other RDRs could act redundantly in rice (162, 200, 282). Upon abscisic acid treatment in *P. patens*, DNA methylation is induced by miRNAs. Additionally, a mutant in *DCL1b* was identified that causes DNA methylation and abolishes cleavage activity of miRNAs. DNA methylation was also observed in WT plants expressing an artificial miRNA (amiRNA, see further), where its target was methylated when amiRNA expression was high, probably triggered by the miRNA:mRNA duplex (283).

4.3.1.7   Computational prediction

In the last years, numerous algorithms to predict miRNA sequences as well as their targets have been developed. Here, I want to touch upon the general methodology behind these approaches.

miRNA prediction
Discovery of miRNAs in the genome is usually based on the identification of imperfect inverted repeats and their potential to form a hairpin containing a miRNA/miRNA* duplex. The presence of bulges or unpaired nucleotides and the number of consecutive mismatches at the duplex site serve as additional criteria. Conservation between *Arabidopsis* and rice was used to filter the candidate sequences, requiring higher sequence homology at the miRNA and miRNA* site compared to the loop region (206, 284, 285). Also conservation of the secondary structure of the stem-loop was used as a criterium (284).

Evolutionary conservation was included in some prediction algorithms (71, 206). Several approaches specifically use homologs and secondary structures to find

miRNAs (31, 43, 286, 287, 288, 289). Deep sequencing results have shown that evolutionary conservation might not be the best way to restrict miRNA discovery, as species-specific miRNAs have been found and validated (29). A computational approach without genome comparison yielded 592 new miRNAs, many not conserved in other plant species (290). Intragenomic matching of potential miRNAs with targets avoids the need for inter-species comparison (291) and was more recently combined with machine learning techniques (292). With the advent of deep sequencing technology, discovery of small RNAs in a wide range of species became possible and facilitates species specific miRNA discovery (33, 34, 35, 37, 38, 39, 40, 42, 43, 45). Due to the realization that some criteria did not only apply for miRNAs but also for tasiRNAs (see further) additional rules were added such as precise excision from the stem-loop hairpin and presence of both miRNA and miRNA* (27, 179). The current knowledge of biogenesis pathways can be incorporated and e.g. sRNAs dependent on RDRs or PolIV/PolV, a hallmark for heterochromatic siRNAs, could be excluded (see further). However, indirect regulation could lead to false negatives (293).

Target prediction
The first target prediction algorithms in plants relied heavily on near-perfect complementarity between the miRNA and its target sequence, allowing no gaps and a maximum of three mismatches (268, 291). Perfect target complementarity to the target proved not to be necessary (294) and based upon experimental evidence, less stringent scoring rubrics were applied (206). Extensive pairing at the $5'$ terminus was favored (277), whereby only one mismatch was allowed from positions 2 to 12 (71, 295). Additionally, sequences with mismatches at positions 10 and 11, flanking the cleavage site, were discarded (295). Consecutive mismatches need to be limited to two nucleotides over the whole length. Also, the free energy of the miRNA/target alignment was used as a cutoff (at least 72% of a perfect match) (71, 295, 296, 297, 298). Recently, a user-friendly web-based target prediction tool was created within our research departement (http://bioinformatics.psb.ugent.be/webtools/tapir, 299). Also for targets, homology has been used as a first approach to limit potential candidates (206, 300). Alternative approaches start from EST databases allowing a broad range of plant species to be analyzed (300, 301). Recently, a new approach based on sequencing of $5'$ ends of polyadenylated mRNAs, termed 'degradome sequencing', revealed cleavage products of miRNAs and allowed validation of previously predicted targets (189, 302, 303, 304).

A better understanding of the biochemical requirements of RNA silencing machinery will help improve target prediction algorithms (305). This knowledge will also increase the performance of prediction methods, helping to build accurate and sensitive methods that do not rely on conservation. Integration of all available information, including e.g. expression data will further optimize both miRNA and target prediction algorithms (292).

4.3.2 *Trans*-acting short interfering RNAs

A second class of plant small RNAs is known as *trans*-acting short interfering RNAs (ta-siRNAs). Similar to miRNAs, these endogenous siRNAs regulate target genes that are different from the nascent locus, also referred to as heterosilencing. In *Arabidopsis*, 4 ta-siRNA generating (*TAS*) families have been identified (*TAS1-4*), each of these loci contains at least one miRNA target site, which is essential for the biogenesis of ta-siRNAs (Figure 4.5). This cleavage site serves as a starting point for a phased production of mainly 21-nt ta-siRNAs. Repeated DCL activity cleaves in every cycle a 21-nt siRNA/siRNA* duplex from the end of the target dsRNA, leading to a population of ta-siRNAs in a 21-nt phase with the miRNA cleavage site. ta-siRNAs play an essential role in development, with *ARF3* and *ARF4* as best studied targets. These phased siRNAs have not only been found in *Arabidopsis* (44, 236, 306) but also in rice (36, 200, 307), poplar (45), pine (308), *P. patens* (309) and a green alga (185). They arise from non-coding as well as protein-coding transcripts and long imperfect double-stranded RNA (dsRNA).

4.3.2.1 Biogenesis

TAS loci are transcribed by PolII and thus have a $5'$ cap and $3'$ poly(A)-tail (115). The transcript is recognized by an AGO-containing RISC complex loaded with a miRNA that subsequently cleaves the non-coding RNA. This miRNA-guided cleavage is indispensable for the onset of ta-siRNA biogenesis (71). Therefore, ta-siRNA biogenesis depends on all components from the miRNA pathway (28, 110). SUPPRESSOR OF GENE SILENCING3 (SGS3) homodimers stabilize the primary *TAS* RNA and interact with RDR6 that generates dsRNA using one of the two cleavage products as template (51, 71, 110, 115, 310). As both SGS3 and RDR6 localize to the cytoplasm (310, 311), transport from the nucleus to the cytoplasm is necessary[1]. SILENCING DEFECTIVE5 (SDE5) is required in ta-siRNA biosynthesis and probably assists RDR6 in dsRNA generation. A shared domain with a human mRNA export factor suggests a role for SDE5 in RNA trafficking from the nucleus to the cytoplasm (315). The dsRNA is recognized by a DCL4-DRB4 complex, that cuts it in a phased way, starting from the miRNA cleavage site, yielding 21-nt ta-siRNAs with $3'$ overhang (71, 73, 89, 95, 110, 115, 316, 317). The 21-nt window can be attributed to the structure of DCL (91). If this RDR6-produced dsRNA has blunt ends or if it is optimized for use by DCLs with a $3'$ overhang is not known. The processing of the dsRNA takes place in the nucleus, thus requiring again transportation through the nuclear membrane (95, 311, 318). The resulting mature ta-siRNA duplex is $3'$ $2'$O-methylated by HEN1, protecting it from degradation (61, 62, 63). One strand of the

---

[1]After finalizing this review, nuclear localization of RDR6, has been first suggested (312) but also demonstrated using antibodies (313). While nuclear localization of RDR6 was already shown previously (314), this was only seen in *rdr6* mutants. If RDR6 is functional in the nucleus remains to be determined.
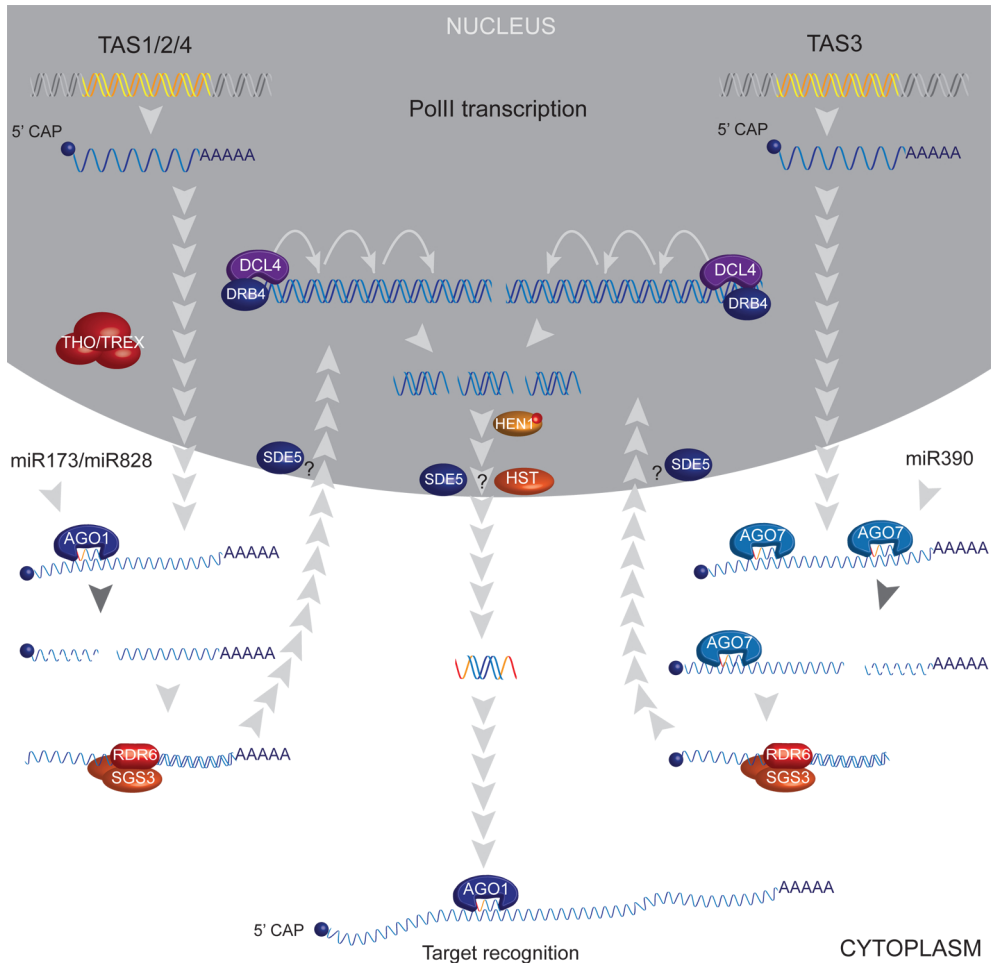
Figure 4.5: ta-siRNA biogenesis pathway. All *TAS* loci are transcribed by PolII and targeted by a miRNA. *TAS1* and *TAS2* contain a miR173 target site, *TAS4* a miR828 target site, while *TAS3* has two miR390 target site of which only the 3′ proximal one is cleavable. For *TAS1* and *TAS2* the THO/TREX complex could be involved in nuclear export. Cleavage of the *TAS* transcripts triggers dsRNA generation by RDR6, assisted by SUPPRESSOR OF GENE SILENCING3 (SGS3). Upon import in the nucleus, possibly mediated by SILENCING DEFECTIVE5 (SDE5), DCL4-DRB4 processes the dsRNA in multiple cycles, leading to phased 21-nt ta-siRNAs. These are methylated by HUA ENHANCER1 (HEN1) and exported from the nucleus. This transport could be controlled by SDE5 and/or HST. The ta-siRNA is loaded into a RISC complex containing predominantly AGO1 as effector (73, 82).

resulting mature ta-siRNA duplex is loaded into a RISC complex with AGO1 as
main component (88, 158, 236, 319), guiding cleavage of near-perfect complemen-
tary targets (51, 71, 88, 110, 115, 125). Rajagopalan *et al.* (27) found that *TAS* loci
produce predominantly 21-nt sRNAs that begin with uridine. The starting base is
more important for ta-siRNA sorting into AGO proteins than the asymmetry rules
that apply for miRNAs (27, 236).

Target recognition takes place in the cytoplasm, requiring again transport, which
might involve HASTY (HST) (234). Recently AtTEX1 has been identified as neces-
sary for ta-siRNA production of *TAS1* and *TAS2*, but not *TAS3*. It encodes a ho-
molog of the TEX1 subunit of the THO/TREX complex that is involved in mRNA
export in yeast and animals (320, 321). Also siRNA production from transgenes
and some endogenous inverted repeats requires AtTEX1 (322). Other subunits of
an *Arabidopsis* THO/TREX complex could be identified and T-DNA insertions in
two of them confirmed their role in ta-siRNA biogenesis (322, 323). This mRNA
transport complex could transport an siRNA precursor to the site of processing.

4.3.2.2 ta-siRNA generating loci in *Arabidopsis*

In *Arabidopsis* 4 families of *TAS* loci have been identified, where *TAS1* and *TAS3* are
represented by three members, resulting in a total of 8 loci (Table 4.1). The first
two families, *TAS1* and *TAS2*, contain a single miR173 target site and the 3′ cleavage
product serves as a template for RDR6 to make dsRNA (71). This target site diverges
from the canonical rules, as it has mismatches at the cleavage position (44, 71, 324).
This imperfect-matching target site of miR173 is necessary and sufficient for ta-
siRNA production (236, 325). The 3′ fragment of the *TAS4* transcript is processed in
the ta-siRNA pathway after cleavage by miR828. *TAS1*- and *TAS2*-derived siRNAs
mainly target *PPR* genes (71, 115), while *TAS4* controls *MYB* transcription factors
(27).

Table 4.1: *TAS* **loci in** *Arabidopsis* **and the miRNA that cleaves the native tran-
script.** Also synonymous names are indicated.

TAS	locus	miRNA	Synonym	References
TAS1a	AT2G27400	miR173	Cluster 64	(51, 71, 110)
TAS1b	AT1G50055	miR173	Cluster 28	(51, 71)
TAS1c	AT2G39675	miR173	Cluster 68	(51, 71)
TAS2	AT2G39681	miR173		(71)
TAS3a	AT3G17185	miR390		(71, 158, 326)
TAS3b	AT5G49615	miR390		(29, 306)
TAS3c	AT5G57735	miR390		(306)
TAS4	AT3G25795	miR828		(27)

Finally, *TAS3* has somewhat different characteristics compared to the other loci. *TAS3* has two miR390 target sites, both of which are bound by miR390-AGO7 complexes, but only the 3′ target site, which has the highest complementarity, is cleaved (44, 158). The 5′ target site has strict dependencies on targeting by miR390-AGO7 (158). The combination of the 5′ target site with the 3′ cleavage site is sufficient to trigger ta-siRNA generation from the sequence in between (325). This suggests that the binding of a miR390-AGO7 RISC complex is instrumental in guiding the *TAS3* loci to the ta-siRNA pathway. The 5′ cleavage product is converted subsequently by RDR6 to dsRNA, which again differs as for the other loci the 3′ fragment is further processed. ta-siRNAs originating from *TAS3* target *AUXIN RESPONSE FACTOR2* (*ARF2*), *ARF3* and *ARF4*, transcription factors involved in auxin signaling, and therefore this locus plays an important role in development (71, 115, 326).

4.3.2.3 Origin and Evolution

The three *TAS1* loci and *TAS2* seem to be paralogs (115). The genome organization suggest that *TAS2* and *TAS1c* arose from a direct duplication and that *TAS1a* and *TAS1b* arose from *TAS1c* after it diverged from *TAS2*. *TAS2* could be related to pentatricopeptide repeat (*PPR*) genes, targets of *TAS1*- and *TAS2*-derived ta-siRNAs, as it has complementarity with these genes (115, 327, 328). It has also been suggested that *TAS* loci could originate from miRNA-targeted protein-coding genes (306). While *TAS1*, *TAS2* and *TAS4* are only present in *Arabidopsis* and its close relatives, *TAS3*-like loci are also in this respect divergent, since they are evolutionary conserved in seed plants, with e.g. 4 members in *P. patens* (27, 44, 71, 110, 306, 309, 326, 329).

4.3.2.4 Specificity

Amongst miRNA-targeted sequences, *TAS* loci constitute only a small fraction. What determines if a cleaved transcript is targeted for degradation or inhibited in translation, or if it is processed in the ta-siRNA pathway? There must be a mechanism that targets RDR6 activity specifically to *TAS* loci. For *TAS1*, it has been shown that cleavage is necessary (71) and sufficient (236) for entering the ta-siRNA pathway. A single miR173 target site is sufficient to initiate phased siRNA production. The complex that performs the cleavage is crucial, which suggests that miR173-AGO1 has unique properties that allow entry into the ta-siRNA pathway, as this behavior could not be mimicked by other miRNAs (44, 236, 306, 325). Therefore, the RISC-miRNA complex may interact with RDR6, directly or through its associated factors. Another possibility is the recruitment through aberrant features of the mRNA such as the lack of 5′ cap or 3′ poly-A tail (103). The non-coding nature of the transcript could also be an important hallmark, as has been shown in mammals (330). miR828 (*TAS4*), miR168 (*AGO1*), miR393 (various targets) and

miR472 (CC-NBS-LRR domain) (306) show a similar phasing behavior to miR173, although at lower levels (27, 44).

For the *TAS3* locus, the mechanism for entry in the ta-siRNA pathway seems to be different. The region yielding phased ta-siRNAs lies between two miR390 target sites. Only the 3' site is cleaved while 5' site recognition is necessary, but no cleavage occurs there. Both sites are necessary for the production of ta-siRNAs from these loci and their respective behavior (cleavage for 3' and no cleavage for 5') is important for the correct processing of these loci. Moreover, this dual targeting seems to be evolutionary conserved (44). Other *Arabidopsis* genes that contain two or more small RNA complementary sites have also been found to produce siRNAs in a phased manner from the intermittent region. This is the case for *ARF3* and *ARF4* (targeted by two ta-siRNAs) and 15 *PPR* genes (miR161, miR400 and *TAS1b*, *TAS2* ta-siRNAs). For genes with only one site, this was not found with the exception of *AGO1* (miR168) and targets from miR393 (44). Again, this could be due to the loss of the hallmark signatures of correctly processed mRNA, as described earlier. However, dual targeting is not sufficient for entry into the ta-siRNA pathway as has been demonstrated with two miR159 sites (325).

Although there are distinct mechanisms at play for *TAS1*, *TAS2* and *TAS4* on the one hand and *TAS3* on the other, there are also common themes. All loci associate with a RISC complex at the 5' end of the *TAS* locus, which is predicted to be relatively stable, either due to the presence of mismatches at the cleavage site, or a non-cleaved target site. This prolonged stability could be one of the determining factors for entry in the ta-siRNA pathway. Several *PPR* genes that generate ta-siRNA-like small RNAs (29, 44, 306, 319) are targeted by miR161 for cleavage and are also predicted to be targets of miR400, for which no cleavage has been shown (306). Again, a stable RISC-target complex could be formed, which subsequently recruits RDR6 and engages into the ta-siRNA pathway (158).

Several loci with internal foldback structures that produce phased siRNAs have been found that do not require miRNA initiation (36, 200, 307), because of this lack, they are seen as evolutionary intermediates of miRNAs instead of genuine *TAS* loci (324).

4.3.2.5   Alternative Processing

The predominantly phased production of ta-siRNAs is important. This methodology ensures a structured, predictable siRNA production from the *TAS* loci. Random siRNAs from these loci would not have sufficient homology to perform the regulatory role of the phased siRNAs. This phasing in register with the miRNA cleavage site, is however not absolute. Forward frame shifts are detected and their likelihood increases with the number of cycles the *TAS*-dsRNA is processed, thus with the distance from the cleavage site (306). *TAS3* might be controlled by its own ta-siRNAs that arise from the antisense strand. What the importance is, is not

known, but it could also set the register for some of the ta-siRNAs produced (71). Next to out-of-phase 21-nt ta-siRNAs, also other size classes, such as 22- and 24-nt sRNAs matching the *TAS* loci are found. As these are also detected in wild type plants, this is a secondary, perhaps redundant, pathway (30, 73, 89, 115). Using mutant analysis, it has been established that DCL4 is necessary for the production of canonical 21-nt ta-siRNAs (89), but also DCL2 and DCL3 can use *TAS* loci as template yielding 22- and 24-nt sRNAs, respectively. Alternative species of 24-nt depend on DCL3 and RDR2 but not RDR6, SGS3 and DCL1, suggesting that the heterochromatic siRNA pathway can use the native *TAS* transcript as a template (see further, 71, 115). Surprisingly, double mutants in *dcl2 dcl4* generated more 21-nt ta-siRNAs, which could be due to decreased competition for a binding factor such as HYL1 (89). DCL1-mediated 21-nt ta-siRNA production was also observed in *dcl2 dcl3 dcl4* triple mutants with a strong *dcl2* allele, but not in a weak allele which is dysfunctional but probably capable of competition with DCL1 (30, 90).

For *TAS2*, the 3′ cleavage product yields both 21-nt ta-siRNAs as alternative species of 24-nt. This in contrast to the 5′ cleavage product, for which only 24-nt siRNAs are found, confirming that only the 3′ fragment is recognized by RDR6 and subsequently goes through the ta-siRNA biogenesis pathway, while the whole *TAS* locus can be used as a template for the heterochromatic siRNA pathway (115).

4.3.2.6   Why ta-siRNAs?

Why has evolution selected such a complicated biogenesis pathway to control gene expression? The development of ta-siRNAs has several advantages over miRNAs. By generating a multitude of siRNAs, *TAS* loci allow for the simultaneous regulation of a larger number of targets. Due to its targeting of gene families with several similar ta-siRNAs, there is an amplification of the signaling. It has also been demonstrated that ta-siRNAs can initiate a cascade of secondary siRNAs, similar to transcription factor cascades (306, 319). Additionally, there is evidence that ta-siRNAs can function non-cell-autonomously (see further, 158, 331), while miRNAs generally cannot move from cell-to-cell.

4.3.3   Heterochromatic siRNAs & RNA-directed DNA methylation

Plants contain a large and diverse set of small RNAs, predominantly 24-nt in length, that mediate heterochromatin modification of their target sequences through transcriptional silencing involving DNA methylation and chromatin remodeling (28, 65, 332, 333). This RNA silencing mechanism is called RNA-directed DNA methylation (RdDM) (334). The 24-nt siRNAs arise mainly from pericentromeric regions, heterochromatic DNA, repeat elements, retroelements and some methylated DNA regions. They ensure *de novo* DNA methylation and maintenance of a silent, heterochromatic state (67, 94, 335, 336, 337, 338). They are therefore refered to as

heterochromatic siRNAs (hc-siRNAs), in literature also the term repeat-associated siRNAs (ra-siRNAs) is used. This class of small RNAs is produced exclusively by DCL3 in *Arabidopsis*. In maize also a class of 22-nt siRNAs match repeat sequences (339). In conifers the regulating function might be taken over by 21-nt siRNAs as they lack a DCL3 homolog (340). Also *C. reinhardtii* lack 24-nt siRNAs and a clear DCL3 homolog (182, 185, 341).

Most hc-siRNAs match transposons and repeats, but also development, plant defense and possibly stress are under this epigenetic control mechanism (178, 335, 336, 342, 343). Transposable elements have the capability to disrupt genomes, and therefore need to be controlled by limiting their transpositional activity (94, 344, 345). This pathway can also regulate transposon-derived promoters and enhancers, prevent spurious transcription of intergenic regions or antisense transcription in adjacent genes (25, 94, 346). Well studied examples of gene regulation driven by DNA methylation are *FLOWERING LOCUS C* (*FLC*), *FLOWERING WAGENINGEN* (*FWA*), *SUPERMAN* (*SUP*) and *SUPPRESSOR OF drm1 drm2 cmt3* (*SDC*) (345, 347, 348, 349, 350, 351, 352). Also at telomeres in *Arabidopsis*, the heterochromatic siRNA pathway helps in maintaining heterochromatic state, but other independent pathways are also at play (353). An RdDM mechanism does not seem to exist in mammalian somatic cells (54), but PIWI-interacting RNAs, a class of sRNAs that is not represented in plants, could fulfill this role in the male germline (354, 355).

4.3.3.1   Biogenesis

Recent research has revealed new insights in the heterochromatic siRNA-mediated DNA methylation pathway. It was hypothesized that RNA-directed DNA methylation requires the convergence of two independent pathways: siRNA biogenesis and transcription at heterochromatic loci. The intersection point seems to be AGO4 and both pathways are necessary to form heterochromatin and silence target loci (356, 357). This addresses the paradox that transcriptional silencing requires transcription of the target sequence (344, 358).

In this biogenesis pathway, two plant-specific, atypical DNA-dependent RNA polymerases are central: PolIV is required for siRNA production and PolV facilitates siRNA-directed DNA methylation. *Arabidopsis* encodes five RNA polymerase (359). These are multi-protein enzyme complexes where the two largest subunits form the DNA entry and RNA exit channels as well as the RNA synthesis catalytic center. Other subunits are important for regulation and assembly (335, 360). PolII transcribes most genes, such as protein-coding genes and miRNAs (361). PolI transcribes ribosomal RNA (rRNA) genes (362), PolIII short structural RNAs, e.g. tRNAs and 5S rRNA (363). PolIV and PolV (initially called PolIVa and PolIVb, respectively) function specifically in the RdDM pathway and are composed of a dedicated large subunit (NUCLEAR RNA POLYMERASE D1 (NRPD1) and NUCLEAR RNA POLYMERASE E1 (NRPE1), respectively) and a common smaller subunit

NRPD2/NRPE2. They share several of the smaller subunits of PolII (364, 365, 366). While the N-terminus of the largest subunit of PolIV and PolV is similar to that of PolII, the C-terminus has diversified, allowing recruitment of different factors. Both contain a DeCL motif, which is required for 4.5S rRNA processing in chloroplasts (335, 367, 368), while only NRPE1 has WG/GW repeats important for protein-protein interactions (369, 370). Metal A and B binding sites, crucial for activity of PolII, are needed for their function in RdDM (356, 364, 371). The presence of all these features suggests RNA polymerase activity for PolIV and PolV (336). This has only been experimentally validated for PolV that can transcribe RdDM target loci (356). Currently, these polymerases are studied by several research groups and new subunits have been discovered (364, 372, 373). Recently, a function for PolV in heterchromatin organization during interphase, independent of siRNA-production has been found (374). In the near future, further details of these molecular components and their function will undoubtedly be unraveled. Here, I present the current view on this biogenesis pathway (Figure 4.6).

siRNA biogenesis
PolIV is thought to make a transcript that can be used as a substrate by RNA-DEPENDENT RNA POLYMERASE2 (RDR2) to make dsRNA (335). The mechanism by which PolIV works, is not yet known. One possibility is that PolIV can transcribe, contrary to PolII, methylated DNA (94, 376). However, PolIV polymerase activity has not yet been demonstrated. PolIV localizes to siRNA source and target loci in the nucleoplasm and PolII appears to play a role in its recruitment (377, 378). For PolIV activity, CLASSY1 (CLSY1), a putative chromatin remodeling protein from a plant-specific SNF2 subfamily, is required. CLSY1 is found mainly around the inner periphery of the nucleolus, where RDR2 also localizes (379). CLSY1 probably resides at the interface of PolIV and RDR2 (335). PolIV and RDR2 activity are required for dsRNA production and to subsequently generate and amplify heterochromatic siRNAs (94, 333, 360, 371, 373, 376, 379, 380, 381, 382). The ssRNA produced by PolIV is transported to the nucleolus by an unknown mechanism and becomes a substrate for RDR2. The resulting dsRNA is processed by DCL3 into 24-nt siRNA duplexes that are 3' methylated by HEN1 and loaded into AGO4 (357). RDR2, DCL3 and AGO4 all localize to Cajal bodies within the nucleolus, which are known as ribonucleoprotein assembly centers (369, 377, 383). While DCL3 is most important, in its absence DCL2 and DCL4 are partially redundant and the resulting 22- and 21-nt can establish and maintain methylation (30). DCL1 cannot produce siRNAs from most RDR2-dependent loci (90). The stability of AGO4 depends on siRNA production, suggesting that the loading of siRNAs in AGO4 could be required for structural integrity. This could be a protection mechanism against irrelevant siRNAs (369). Also constituents of the Cajal body contribute to the stability of AGO4 (384). At least for some loci, AGO6 acts redundantly with AGO4 (169). AGO9 likely is the effector in the ovule (168, 170).
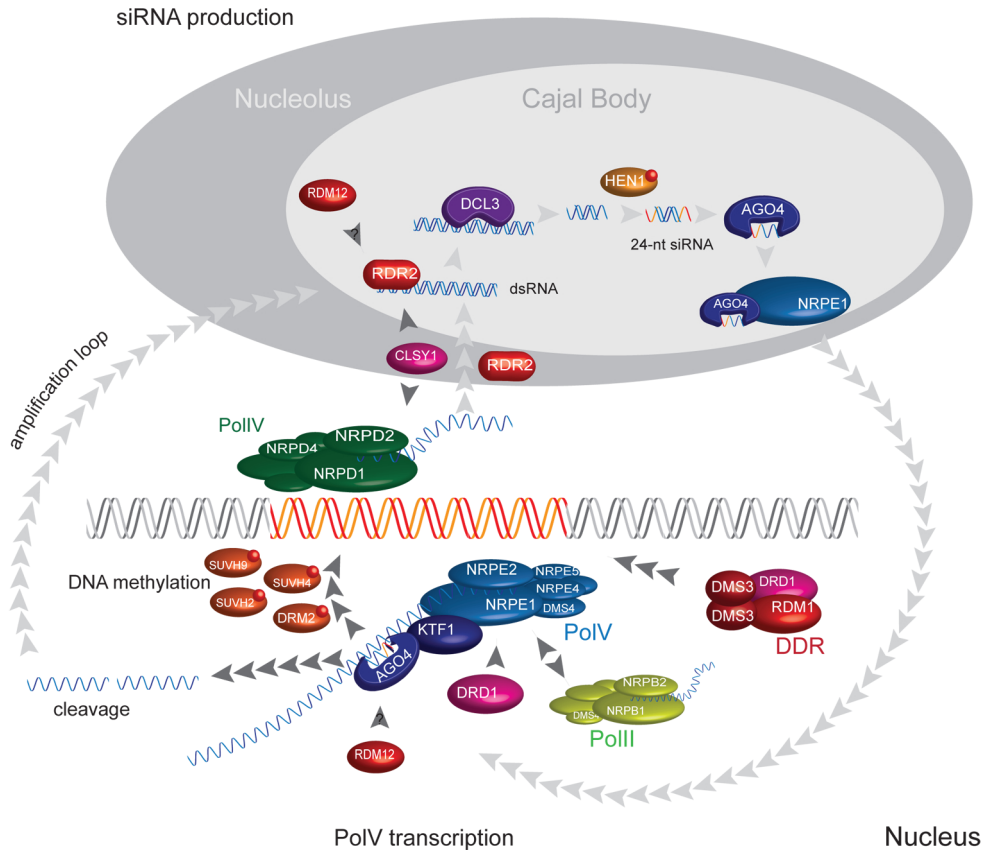
Figure 4.6: Heterochromatic siRNA biogenesis pathway. Methylated DNA regions (indicated in red) are probably transcribed by PolIV, assisted by the chromatin remodeling protein CLASSY1 (CLSY1). This transcript is converted to dsRNA by RDR2 and processed to a 24-nt siRNA/siRNA* duplex by DCL3. After HEN1 mediated methylation, the guide strand is loaded into AGO4. AGO4 probably binds to NRPE1, the largest subunit of PolV in the Cajal bodies. PolV transcribes siRNA target loci, assisted by the DRD1-RDM1-DMS3 (DDR) complex. AGO4 binds to PolV transcripts guided by the loaded siRNA, KOW DOMAIN CONTAINING TRANSCRIPTION FACTOR 1 (KTF1) could act as a scaffold. Chromatin remodeling by DRD1 could enable PolV activity. PolII replaces PolV at certain loci. AGO4, possibly through KTF1, recruits the *de novo* DNA methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE (DRM2) and histone modifiers such as SUVH4, SUVH2 and SUVH9. Alternatively, AGO4 could cleave the transcript and initiate an amplification loop (adapted from Law and Jacobsen (375)).

Transcription of heterochromatic loci

PolV has been shown to transcribe RdDM target loci (356). It requires the plant-specific proteins DEFECTIVE IN RNA-DIRECTED DNA METHYLATION1 (DRD1), DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), and RNA-DIRECTED DNA METHYLATION1 (RDM1) for its transcription activity (356, 357, 385). DRD1 is an RNA-binding protein from the same SNF2-family as CLSY1 (386). DMS3 is similar to a structural-maintenance-of-chromosomes hinge domain, containing short coiled-coil regions on either site of the hinge. DMS3 probably dimerizes and binds DNA (387). The function of RDM1 is unknown. It contains a unique fold that has no structural homology to other known proteins and that may be crucial for its function (385, 388). These three proteins form a DRD1-DMS3-RDM1 complex, dubbed DDR, which probably mediates chromatin association or activation of PolV (346, 357, 380, 385, 386, 387). Additionally, DRD1 could remodel chromatin to enable PolV activity (356). All of the proteins involved in this part of the pathway, have been shown to co-localize with chromosomal loci, that are both source and target of siRNAs, or bind to proteins that localize at these loci (377, 385).

Integration and feedback

AGO4, loaded with a hc-siRNA, interacts with PolV transcripts through siRNA-mediated guidance (357). AGO4 can interact directly with PolV through WG/GW repeats in its C-terminus (369, 370) at the Cajal bodies (369, 377) and the target loci (384), but this interaction is reported to be weak or transient (357). Moreover, active transcription is required for interaction of AGO4 with chromatin, indicating that AGO4 physically interacts with PolV transcripts. siRNAs themselves are not needed for guidance of PolV to its target loci or for transcriptional activity of PolV (356, 357). A recently identified RdDM effector, KOW DOMAIN CONTAINING TRANSCRIPTION FACTOR 1 (KTF1)/Spt5-like (SPT5L), was shown to bind AGO4 in the nucleoplasm through WG/GW repeats in the C-terminus of KTF1 and interact with PolV RNA transcripts (357, 365, 389, 390). This protein could act as a scaffold to recruit AGO4 to PolV transcripts (389, 390) and/or attract the DNA methylation effectors (mainly DOMAINS REARRANGED METHYLTRANSFERASE (DRM2)) and histone modifiers (KRYPTONITE (KYP)/SU(VAR)3-9 HOMOLOG4 (SUVH4), SUVH2, SUVH9) to these loci (54, 336, 357, 369, 375, 384). AGO4 has been shown to function in a non-catalytic mode to recruit these downstream effectors. But also a catalytic-dependent AGO4 function was revealed, which could initiate generation of secondary hc-siRNAs through a reinforcement loop (157).

Two recent screens identified an additional player in RdDM: RDM12 / INVOLVED IN DE NOVO 2 (IDN2). This protein, similar to SGS3, was found to be crucial for *de novo* methylation in RdDM. RDM12/IDN2 is hypothesized to facilitate targeting of downstream factors (391) but could also associate with RDR2, similar to the role of SGS3 for RDR6 (392). DAWDLE (DDL), necessary for stabilization of pri-miRNAs, could also play a role as hc-siRNAs are also affected in mutants

(210). Also for HASTY (HST), implicated in miRNA export from the nucleus, a role in 24-nt siRNAs was recently observed (52).

A role for RNA Polymerase II

Heterochromatic loci have been subdivided in two groups. Type I loci are high-copy-number repeats or transposons for which siRNA production depends on both PolIV and PolV. Low-copy-number repeat and intergenic sequences make up type II loci, that depend on PolII instead of PolV (333, 376, 378, 380, 381). This could explain the lack of co-localization of PolV and AGO4 observed outside the Cajal bodies (369, 377) while PolII and AGO4 co-localize in the nucleoplasm (393). Therefore, PolII could be important at target sites in the nucleoplasm. PolII transcription could also function through recruiting PolV to certain loci (378). Additionally, a newly identified RdDM component, DMS4/RDM4, interacts with both PolII and PolV (373, 394). The methylated state or, alternatively, the transcription by PolV or PolII, attracts PolIV and thus feeds positively back into the siRNA production. This feedback loop reinforces silencing at these loci (360, 377, 378, 395, 396, 397, 398).

4.3.3.2   DNA methylation and heterochromatin in *Arabidopsis*

Deep sequencing techniques, especially 454 technology, allowed mapping of DNA methylation with base-pair resolution (397, 399). DNA methylation in plants is found at cytosine residues in all sequence contexts: symmetric CG and CHG (where H is A, C or T) and asymmetric CHH. Genome-wide on average 24% CG, 6.7% CHG, 1.7% CHH sites were methylated (399). This genome-wide analysis revealed that about 30% of the methylation is siRNA mediated (397, 399). Cytosine methylation and histone H3K9 and H3K27 methylation are hallmarks of heterochromatin and transcriptionally silent DNA in plants (94, 280, 400, 401).

   DOMAINS REARRANGED METHYLTRANSFERASE (DRM) is responsible for siRNA-directed *de novo* DNA methylation in all sequence contexts in *Arabidopsis*. DRM2 is most important while DRM1 has a minor role (402, 403). CG and CHG are maintained by METHYLTRANSFERASE1 (MET1) and CHROMOMETHYL-TRANSFERASE3 (CMT3), respectively (404). However, at many loci CMT3 and DRM2/DRM1 can control CHG and CHH methylation redundantly (399). CHH methylation is most sensitive to loss of methylation as this type does not have a dedicated maintenance methyltransferase and thus requires continuous *de novo* methylation (18, 404). Small RNAs guide DRM2 to maintain DNA methylation in all sequence contexts (345, 405, 406). Silencing through CHH methylation can indeed be reactivated when the (transgenic) RNA trigger is removed (346, 380, 386). CG methylation can be lost passively, due to lack of maintenance in dividing cells, or actively through e.g. REPRESSOR OF SILENCING 1 (ROS1), a DNA glycosylase/lyase working through a base excision and repair mechanism. *ROS1* expres-

sion was found to be linked to RdDM (346, 390, 407, 408). DRD1, necessary for PolV function, was also found to be required for full erasure of CG methylation at 5S rDNA (409), which could be due to the downregulation of *ROS1* in a *drd1* mutant (346). Active demethylation by ROS1 might be guided by siRNAs bound to ROS3, thus dynamically regulating methylation and demethylation which could be important for an efficient response to e.g. environmental changes (410, 411).

DRM2-dependent *de novo* methylation is linked to CG and CHG methylation pathways involving MET1, DECREASED DNA METHYLATION1 (DDM1), and CMT3 (395, 399, 412), possibly by recruiting PolIV to these methylated sites. Genome-wide analysis of sRNAs and DNA methylation in mutants in the latter process, suggested that sRNAs can control different hierarchies of DNA methylation establishment in each context and showed complex interplay between the different methylation pathways (397). H3K4 demethylation at target loci is required for recruitment of DRM2 by AGO4 and hc-siRNAs (19). LYSINE-SPECIFIC DEMETHYLASE1-LIKE1 (LDL1) and LDL2 play an important role in *de novo* DNA methylation and heterochromatic silencing of *FWA* and likely mediate this through removal of H3K4 methylation (413). Methylcytosine binding-domain proteins (MBD6 and MBD10) act together with RdDM to effect large-scale rDNA silencing (414). DDM1, a chromatin remodeling helicase, is required for maintenance of DNA methylation in all sequence contexts (415). Other chromatin remodeling factors, DRD1 and CLSY1, play a role in the biogenesis pathway as described. REQUIRED TO MAINTAIN REPRESSION1 (RMR1) is similar to DRD1 and CLSY1 and is required for RdDM in maize (416).

KRYPTONITE (KYP)/SU(VAR)3-9 HOMOLOG4 (SUVH4) dimethylates lysine 9 of Histone 3 (H3K9me2) with a preference for CHG methylated sites (352, 417, 418), while SUVH5 and SUVH6 are required at certain loci (419, 420). SUVH4 is required for CMT3-dependent DNA methylation (406) and both H3K9me2 and CHG methylation are tightly correlated throughout the genome (421). The H3K9 demethylase INCREASE IN BONSAI METHYLATION1 (IBM1) and DDM1 are important to prevent ectopic non-CG methylation in genic regions and thus to target both H3K9 methylation as non-CG DNA methylation to repeats and transposons (422, 423). SUVH2 and SUVH9 are required for DRM2 activity (424). SUVH2 binds to methylated CG residues, while SUVH9 shows specificity for CHH methylation. These could function through recruitment or retainment of components of the DRM2 pathway to siRNA targeted loci. SUVH2 has been reported to have H3K9 methylation activity, but this is still under debate (401, 424). Also histone H2B de-ubiquitination by UBIQUITIN-SPECIFIC PROTEASE 26 (UBP26) is required for heterochromatic H3K9me2 and siRNA-directed DNA methylation (425). Recently, ARABIDOPSIS TRITHORAX-RELATED PROTEIN5 (ATXR5) and ATXR6 were identified as H3K27 monomethyltransferases and H3K27me deposition was found to be independent of both DNA methylation and H3K9me2. Both modifications (H3K27me

and H3K9me2/DNA methylation) are required for silencing of heterochromatic regions (426).

MORPHEUS MOLECULE 1 (MOM1) shares target loci with the RdDM pathway. It is required for the heterochromatic mark H3K9me2, but not to maintain the associated DNA methylation. This suggests that MOM1 is needed for the transmission of DNA methylation to histone modification, specifically at sites where non-CG methylation is important for maintenance of cytosine methylation (427). MOM1 works in parallel to the PolV pathway to silence a subset of loci, for other loci it seems to work in the same pathway but also counteracting of silencing has been observed (428, 429).

HISTONE DEACETYLASE6 (HDA6) removes acetyl groups from multiple core histone lysines. HDA6 activity is required for transcription suppression downstream of DNA methylation. They accumulate histone markers typical of active genes and associate with PolI and PolII. Symmetrical CG and CHG methylation depends on HDA6 functionality. Recruitment of HDA6 by DNA methylation probably is part of a reinforcement mechanism (430).

RdDM might be a reversible method of silencing while other additional layers, such as heterochromatic marks or CG methylation could provide a more stable, but less dynamic repression (346). The association of RdDM with H3K27me histone modifications, considered to be more dynamic compared to H3K9me2, confirms this (54, 346).

The different specificities of the proteins involved in RdDM and heterochromatin formation as well as the complex interplay between all these components can explain the different dependencies that have been reported for the variety of loci that has been studied (94, 345, 360, 380, 381, 386, 395, 396, 405).

4.3.3.3   A role in germplasm

PolIV-dependent siRNAs play a role in the developing seed and thus the germ line and embryo. A picture is emerging, suggesting that accessory cells sacrifice their genome integrity to ensure immobilization of transposable elements in the embryo, thus protecting its progeny (431, 432, 433, 434, 435, 436). A subset of PolIV-dependent siRNAs is specifically expressed in a burst in these accessory cells during seed development and is only maternally expressed. Maternal activators or paternal inhibitors involved in this process are not yet known (435).

4.3.4   Natural antisense transcript short interfering RNAs

Transcription of complementary DNA strands can lead to formation of dsRNA. Such natural antisense transcripts (*NATs*) are ubiquitous in eukaryotes (437, 438, 439, 440, 441, 442, 443, 444, 445) and can be divided into two classes: *cis-NATs* and *trans-NATs*. The former are transcribed from opposite strands of the same genomic

locus, the latter from distinct loci. Antisense transcription has been implicated in post-transcriptional gene silencing (55, 56, 446) and transcriptional silencing (447) as well as alternative splicing, polyadenylation (439) and RNA maturation or stability (448). These regulatory units are difficult to study as they are expressed in specific tissues and conditions and do not seem to be evolutionary conserved (449, 450). The presence of loci that combine *cis-NAT*s, *trans-NAT*s and miRNA targeting, illustrates the vast regulatory potential at the RNA level (449). Here, I briefly introduce the state-of-the-art in plants.

4.3.4.1   *cis*-natural antisense pairs

In *Arabidopsis*, several studies identified over a thousand potential *cis-NAT* gene pairs (438, 439, 450, 451). However, only a subset of these genes seem to be anti-correlated and only in specific conditions or developmental stages. Also, no clear correlation with small RNA production was found for genes belonging to a *cis-NAT* pair. However, within the *cis-NAT*s, the overlapping region was enriched in sRNAs compared to the whole transcripts (450, 451). The first example of a functional *cis-NAT* pair was found under salt-stress conditions from the overlapping genes Δ^1-*pyroline-5-carbyxolyate dehydrogenase* (*P5CDH*) and *SIMILAR TO RCD ONE 5* (*SRO5*) (55). *SRO5* is induced by salt stress, while *P5CDH* is present in plants grown under normal conditions and its expression is reduced by NaCl treatment. *Pseudomonas* infection induces a GTP-binding protein, *ATGB2*, which causes repression of its antisense partner *PPRL*, a pentatricopeptide repeats (PPR) protein-like gene (56). This type of regulatory mechanism might be very important under stress (452). Next to these stress-induced examples, also normal development is regulated by *cis*-nat-siRNAs as illustrated by the requirement of a *NAT*-pair for double fertilization in *Arabidopsis* (453). The biogenesis pathways are slightly different between the different cases, but they share the requirement for an amplification loop through PolIV and an RNA-dependent RNA polymerase (RDR), reminiscent of the heterochromatic biogenesis pathway (64, 65).

Stress-induced nat-siRNAs
Upon salt-stress *SRO5* is induced, producing a DCL2-dependent 24-nt nat-siRNA from the overlapping region with *P5CDH*, which represses the latter. P5CDH catabolizes P5C, an intermediate in proline biosynthesis, and functions thus in a key pathway for stress response. The requirement for RDR6, SGS3 and PolIV suggests that a primary siRNA is formed which triggers an amplification loop producing secondary siRNAs. Phased, DCL1-dependent 21-nt siRNAs in register with the 24-nt siRNA cleavage site were found, but these are not necessary for repression of *P5CDH* (55). In the case of the *ATGB2-PPRL NAT*-pair, a 22-nt nat-siRNA is formed requiring DCL1 and HYL1 as well as the amplification loop of PolIV-RDR6-SGS3. Additionally, RESISTANT TO P. SYRINGAE (RPS2) and NON-RACE

SPECIFIC DISEASE RESISTANCE1 (NDR1), components of the resistance signaling pathway, are necessary for the formation of the nat-siRNA (56).

Long siRNAs of 30- to 40-nt have been found from which several are generated from *NAT* pairs, induced in specific developmental growth conditions or upon bacterial infection (446). In this study, a cluster of heterogeneous 39- to 41-nt siRNAs induced by *Pseudomonas* infection was studied and found to be dependent on DCL1 and DCL4. A secondary amplification loop involving RDR6, PolIV and also PolV is implicated in its biogenesis. The production of these nat-siRNAs contributes to bacterial resistance by silencing of a repressor of plant defense through mRNA degradation by XRN4 (446).

nat-siRNAs in development

Specific aspects of development can also be regulated by *cis*-nat-siRNAs. Recently, *cis*-NATs that produce phased nat-siRNAs have been identified that could play a role in organ-specific regulation (163). A functional pair has been demonstrated to control reproduction, facilitating gametophyte formation and double fertilization (453). *KOKOPELLI* (*KPL*) and *ARIADNE14* (*ARI14*), a putative E3 ubiquitin ligase, form a *cis*-nat-siRNA pair. In *kpl* mutants, single fertilization often occurs, leading to seed abortion. *ARI14* overexpression resulted in similar phenotypes. While *ARI14* is present in both the vegetative cell and sperm in the male gametophyte, *KPL* is sperm-specific and leads to downregulation of *ARI14* in sperm cells. ARI14 is likely a non-active E3 ubiquitin ligase hypothesized to bind substrates, thus preventing degradation by active E3 ubiquitin ligases. The biogenesis pathway of siRNAs from this pair has similar dependencies as the stress-related nat-siRNAs. Production occurs through DCL1 and HYL1, stabilized by HEN1 and amplified by an RDR2-SGS3-PolIV mechanism (453).

4.3.4.2 *trans*-natural antisense pairs

Trans-NAT gene pairs have been identified in *Arabidopsis* and rice (449, 450). Often both *cis*- and *trans*-antisense transcripts are found for the same locus, suggesting complex interplay of regulatory networks. In *Arabidopsis*, UDP-glucosyl transferase genes were found to be enriched in *trans-NAT*s as well as *cis-NAT*s (449). *Trans-NAT*s were found to be hotspots for sRNA biogenesis in angiosperms, this in contrast to *cis-NAT*s where only a small proportion is associated with small RNA production (163, 450). These *trans*-pairs are formed in specific stress or development conditions and are rarely evolutionary conserved (449, 450). While *cis-NAT*s have been validated experimentally, *trans-NAT* research remained *in silico*. This leaves a large part of the complex regulatory network on the RNA level to be unraveled.

4.3.5 Exogenous introduction of RNA or DNA

When RNA or DNA, either single-stranded or double-stranded, is introduced in plants, silencing of these sequences is often observed. This phenomenon is mediated by usage of the small RNA pathways that have been described above. However, depending on the characteristics of the introduced genetic material, different mechanisms are at play. Here, I want to briefly touch upon the most important findings.

It is already known for a long time that transgenes can cause silencing based on homology (454, 455, 456, 457, 458, 459). Immunity against viruses can be achieved by expressing transgenic constructs containing viral sequences (460, 461, 462, 463). Vice versa, infection of a virus can induce silencing of homologous sequences (464, 465). This indicates that similar mechanism are at work for both defense against viruses and transgene silencing. At least for some viruses, the natural resistance is based upon RNA degradation, providing another link with transgene silencing (466, 467). Also the mode of action has similarities: silencing of transgenes spreads from a initial focal point systemically through the plant, similar to a virus infection (468) and the spreading signal is also graft transmissable (469). Silencing-deficient mutants are hypersensitive to virus infection, confirming that similar siRNA pathways are used for silencing transgene and viral RNA (90, 112, 126, 147, 228).

4.3.5.1 Virus defense through RNA silencing

Viruses require the machinery of their host to replicate their genetic information. Therefore, DNA or RNA is introduced into host cells. Double-stranded RNA (dsRNA), required for small RNA production, can be obtained from viruses through foldback structures in single-stranded RNA (ssRNA) viral transcripts, transcription intermediates from viral replicases or convergent transcript for DNA geminiviruses (17). Also conversion of ssRNA to dsRNA by RNA-dependent RNA polymerases is possible (48). Between viruses and plants a continuous arms race is ongoing. Viruses produce viral suppressors of RNA silencing (VSR) to overcome the silencing machinery (for review see Li and Ding (470), Ding and Voinnet (471) and Ruiz-Ferrer and Voinnet (17)). The mode of action of these VSRs is varied: some inhibit components of the sRNA machinery such as DCL4 (14, 90) or AGO1 (472, 473, 474), while others sequester small RNA duplexes (475) or prevent RNA-dependent DNA methylation by limiting methyl group donor availability (476, 477). The expansion and redundancy of the DCL family, could work to the advantage of the plant as inhibiting one DCL member is not necessarily detrimental for the silencing system (14, 90, 478). All four DCLs have been implicated in virus defense (107, 478). DCL4 is the major antiviral DCL, often hierarchical with DCL2 (14), but DCL3 can also process viral dsRNA (14, 17) while DCL1 has both been reported to be facilitating as inhibiting other DCLs (107, 478). AGO1 and AGO7 play a role as the major effectors

(107, 147). Amplification and subsequent generation of secondary siRNAs is important for building up resistance against viruses. All three RNA-dependent RNA polymerases (RDRs) involved in RNA silencing (RDR1, RDR2, RDR6) play a role in defense against certain viruses (46, 106, 112, 113). Especially RDR1 (47, 108, 109) and RDR6 with its binding partner SGS3 (47, 112, 114, 479, 480) have been found to be important in viral defense.

4.3.5.2   Transgenes

Transcriptional and post-transcriptional gene silencing (TGS/PTGS) have been observed for transgenes in plants (481, 482, 483, 484). Silencing of sense-transcripts requires RDR6 for dsRNA formation assisted by the dsRNA-binding protein SGS3, the RNA helicase SDE3 and export factor SDE5, and also the exonuclease WEX and methylase HEN1 are involved in this pathway (111, 112, 228, 315, 485, 486, 487). Induction of silencing by inverted repeat constructs, does not require dsRNA formation due to the formation of a hairpin loop (479). The repression is often associated with sequence specific RNA-directed DNA methylation (RdDM, 488, 489, 490) and is released in chromatin mutants such as *ddm1* and *met1* (491).

4.3.5.3   Transitivity

Spreading of silencing allows for silencing outside the targeted region, this process is called transitivity. siRNAs matching the sequence that triggers silencing are called 'primary siRNAs', while 'secondary siRNAs' originate from the flanking regions. This spreading requires RDR activity and serves as an amplification mechanism (48, 118, 398, 492). Transitivity in both 5′ and 3′ direction has been observed (48, 50, 117, 118, 350, 493). While 5′ spreading could be primer-dependent, this is not possible for the 3′ direction. Aberrant features of the RNA probably attract RDR6. While spreading of silencing is often found for transgenes, this has only been rarely reported for endogenous genes. Spreading of methylation from a *LINE* element into the adjacent *BONSAI* locus has been reported (494). The *SDC* locus shows extensive spreading dependent on CG methylation by MET1 (350). For unknown reasons, transitivity does not always occur for transgenes (380, 386, 495). For transgenes, a stepwise mechanism for spreading of methylation has been described, involving PolIV to transcribe methylated DNA in a direct or indirect way, thus producing RNA for secondary siRNA biogenesis (398). DCL2 and DCL4 have been identified as the main DCLs involved in transitivity (98, 496).

4.3.5.4   Endogenous RNAi pathway

The term RNAi is used for the introduction of a construct that generates a long dsRNA foldback. Recently, Dunoyer *et al.* (52) showed that also endogenous in-

verted repeats (IR) can be used as a substrate for siRNA generation, as found in flies (497). This kind of loci, producing symmetrically organized and heterogeneous siRNA populations, are frequent in the *Arabidopsis* genome (292). In contrast to DCL1, the other DCLs can use long, near-perfect dsRNA molecules and generate 21-, 22- and 24-nt siRNAs through DCL4, DCL2 and DCL3 activity, respectively. DCL2 is usually seen as a surrogate for DCL4 (14, 498), in this pathway however DCL2 is, together with DCL3, the primary source of siRNAs while DCL4 is secondary in the hierarchy. DCL1 does not contribute to siRNA production at these IR loci. This pattern is similar to a transgenic approach with IR, but the hierarchy of DCL4 and DCL2 is reversed (52, 498). Although DCL1 does not play a significant role in generation of siRNAs directly, it does have a facilitating role, possibly through cleavage of the primary transcript, allowing other DCLs easier access to the stem-loop. Conform all identified plant siRNAs, methylation through HEN1 occurs. Surprisingly, a decrease of 24-nt siRNA accumulation was seen in *hst* mutants, which was also observed for 24-nt heterochromatic and virus-derived siRNAs (52). This was not found in previous research, but indicates a role for HST beyond transport of miRNAs (234). DLC3-derived AGO4-loaded siRNAs mediate DNA methylation at these loci. For the 22-nt siRNAs post-transcriptional silencing capacity was demonstrated. The RdDM pathway is however not involved in siRNA production from these loci itself. 24-nt duplexes where shown to be non-cell-autonomous and functional, similar to what was shown for 21-nt duplexes (499).

This endogenous RNAi pathway could be a rapidly evolving system that produces a wide range of siRNAs available for selection by evolution, thus providing a molecular sensing of the environment and stress conditions.

4.3.5.5 Small RNAs as tools

The small RNA pathway has been used, without knowledge of its molecular background, to silence genes for research and biotechnological purposes. 'Flavr Savr' tomatoes are the best-known commercial product that was released. An antisense transgene was introduced that decreases the production of polygalacturonase, a cell wall degrading enzyme, and delays the softening of the tomato. Downregulation through gene silencing has been triggered initially by expressing sense or antisense constructs, later inverted repeats were found to be more efficient (500, 501). Examples of the use of RNA silencing in plants can be found in Frizzi and Huang (21). Another approach is virus-induced gene silencing (VIGS), where a part of the gene of interest is expressed in a viral vector (465). This proved to be a fast and effective technology to knockdown genes and study their function without the need for stable plant transformation (502, 503). This has recently been adapted to express *MIR* genes, called 'MIR VIGS', providing a tool to study miRNAs (504). Also an RdDM approach to silence a *MIR* promoter has been deployed *in planta* to study miRNAs (505).

Through the formation of dsRNA, a wide variety of siRNA species can be formed, possibly causing so-called off-target effects (506, 507, 508). Artificial miRNAs (amiRNAs) were successfully designed according to the miRNA-target prediction rules to target specific genes or groups of genes. By replacing the miRNA and miRNA* sequences in a hairpin-backbone a single, predictable amiRNA is produced (276, 509, 510, 511). In plants first implemented in *Arabidopsis*, this approach has been expanded to rice (512), tomato (513), *Physcomitrella patens* (514, 515) and *Chlamydomonas reinhardtii* (516). amiRNAs have rapidly become a valuable tool to study gene function in a wide variety of species. A web-based tool to design amiR-NAs for more than 90 plant species is available at wmd3.weigelworld.org.

Also the ta-siRNA pathway has been utilized as a means of gene silencing. Similar to amiRNAs, the sequence of *TAS* loci can be adapted to target genes of interest for silencing (158, 517).

4.3.6   Mobility

Systemic silencing of transgenes and virus-derived sequences has already been observed over a decade ago (518) and well documented over the years (97, 498, 519). Sequence specificity of the silencing suggested a nucleic acid as signaling molecule (518). Short-distance movement of silencing from cell-to-cell occurs through plasmodesmata with a range of 10-15 cells due to dilution of the signal (116). Viral-encoded movement proteins (MP) have been shown to influence spreading of the silencing signal. This depends on their RNA binding ability and potentially also the capacity to change the size exclusion limit of plasmodesmata (520, 521). Long range cell-to-cell movement occurs through reiteration of the signal in recipient cells by RDR6-dependent secondary siRNA formation (480). In *Solanaceae* movement of a silencing signal through the phloem was demonstrated through grafting experiments (117, 469, 518), recently in *Arabidopsis* this process was also described (52, 522, 523).

From specific miRNAs that were studied, these small RNAs are thought to act mostly cell-autonomously, but exceptions have been described (276, 509, 510, 524). miR165/166 has been demonstrated to act non-cell-autonomously in the root, moving from the endodermis to the stele (525). Also miR390 possibly moves to adjacent cells (526). The presence of miRNAs, as well as other 21- and 24-nt small RNAs, in the phloem has been demonstrated (527, 528, 529, 530, 531, 532). For miR399, induced by phosphate deficiency, a role in long-distance signaling has been well characterized (523, 531, 532, 533, 534) and this could also be the case for miR395, involved in sulfate accumulation and allocation (532). For ta-siRNAs produced by *TAS3*, intercellular movement from the adaxial side of the leaf, where they are produced, to the abaxial side was shown (331). This creates a gradient over the leaf regulating *AUXIN RESPONSIVE FACTOR*s (*ARF*s) which determine abaxial identity.

Voinnet and Dunoyer recently showed that double-stranded sRNA duplexes constitute the mobile signal and both 21-nt (499) and 24-nt sRNAs (52) are mobile. The exact mechanism behind the movement is not yet known.

DCL4 is a key component in cell-to-cell movement from inverted-repeat transgenic constructs (97). Further spreading requires amplification of the signal through RDR6 (116). This RNA polymerase activity is not required for the generation of the silencing signal, but only for the perception in recipient cells (480). Next to RDR6, also PolIV, CLSY1 and RDR2 have been found to be required for signal reception of long-range silencing through the phloem (52, 379, 498, 535). Other components of the heterochromatic pathway (DCL3, AGO4, DRD1 and PolV) do not seem to be involved (379). At the level of DCLs and AGOs there is redundancy (379, 498). This suggests that both RDR6 as well as part of the heterochromatic pathway could be co-opted as an amplification mechanism.

21-nt siRNAs produced by DCL4 were found to be necessary and sufficient for silencing upon spreading (499). DCL4 is therefore the primary endonuclease implicated in spreading of silencing, acting hierarchically with DCL2 as has been observed in virus resistance (14, 498, 535). Remarkably, ta-siRNAs, which have been shown to be mobile, are also produced by DCL4 (331, 536). At high doses of sRNAs, also DCL1 and DCL3 could act redundantly in this process (498).

Using a transgenic inverted repeat, Molnar *et al.* (522) showed the mobility of 23- and 24-nt sRNAs, produced in a PolIV-dependent pathway, that were capable of triggering RNA-dependent DNA methylation in receiving cells. Buhtz *et al.* (531) found 24-nt sRNAs in phloem sap, consistent with these data. The mobility of heterochromatic siRNAs could allow for spreading of epigenetic modifications and adaptation in progeny (65).

The different (transgenic) systems that were deployed to determine the molecular basis for small RNA mobility differ in the required components (52, 379, 498). This has been attributed to the region where the constructs are incorporated (52), but also our limited knowledge of substrate specificity for RDR2 and RDR6, targeting of PolIV and redundancy among DCLs and AGOs complicate this research. The rapidly increasing knowledge on small RNA pathways in general, will help to resolve outstanding questions in the coming years.

4.4   RNA silencing in development

The importance of small RNA pathways in development was early on clear as several mutants, especially in miRNA biogenesis, exhibited developmental phenotypes (73, 89, 94, 112, 115, 213, 235, 537). A large series of *dcl1* mutants was identified ranging from embryo-lethality to pleiotropic developmental defects in floral development, timing of flowering and leaf morphology (74). Also other components of the miRNA biogenesis pathway (AGO1, HEN1, HYL1, HST, SE, SQN) show

similar phenotypes (146, 171, 229, 237, 538, 539). The first predicted targets for miRNAs were biased towards transcription factors and F-box proteins functioning in developmental patterning and cell differentiation (206, 268). Some miRNAs and ta-siRNAs are highly conserved in land plants regulating growth, development and stress responses (70, 540).

The ability to handle a variety of stress conditions is crucial for the survival of plants. Also in these signaling pathways, miRNAs take a prominent role. Large scale studies for stress responsive miRNAs have been performed in *Arabidopsis* (541, 542, 543, 544, 545, 546), but also other plant species such as rice (34, 547, 548, 549) and maize (550). For detailed information I refer to recent reviews (17, 551, 552, 553, 554, 555). The role of miRNAs in development has been reviewed regularly (9, 11, 64, 70, 556). Here, I will focus on the role of miRNAs and *trans*-acting short interfering RNAs (ta-siRNAs) in leaf development.

4.4.1   Morphogenesis & pattern formation

4.4.1.1   Adaxial – abaxial polarity in the leaf

Proper differentiation of abaxial and adaxial cell fate is required for blade out-growth and formation of a planar leaf (557). Class III HomeoDomain leucine zip-per transcription factors (HD-ZIPIII) *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*) and *REVOLUTA* (*REV*) regulate adaxial characteristics. All HD-ZIPIII members contain a miR165/166 target site and loss of regulation by these miRNAs leads to adaxi-alization of the leaf (86, 277, 510, 537, 558, 559, 560, 561, 562, 563). miR165/166 is expressed at the abaxial side of the leaf restricting *PHB*, *PHV* and *REV* to the adaxial side (527, 560, 564, 565, 566, 567). The abaxial expression of these miRNAs is reg-ulated by *cis*-acting elements (567). Both in maize and rice, miR165/166 regulation of leaf polarity through HD-ZIPIII genes is conserved (568, 569).

The ta-siRNA pathway also affects leaf patterning (Figure 4.7). ta-siRNAs pro-duced by the *TAS3* locus, dependent on AGO7 and miR390, target the abaxial determinants *AUXIN RESPONSE FACTOR 3* (*ARF3*) and *ARF4* (44, 71, 316, 326). The ta-siRNA biogenesis pathway also downregulates miR165/166 (570). In *Ara-bidopsis*, miR390 is expressed throughout the leaf (331). *TAS3a*, the most important *TAS3* locus in developing leaves, and AGO7 are expressed in the adaxial domain (331, 561, 563, 571). Co-localization of all necessary components to generate *TAS3* ta-siRNAs occurs thus only adaxially. However, these ta-siRNAs form a dorso-ventral gradient over the leaf indicating that these small RNAs, or their precursor, are mobile (331, 536). Recently, it has been shown that probably the sRNA du-plex moves from cell-to-cell (499). Counter-intuitively, an sRNA gradient can create sharp expression boundaries of target genes (572).

In maize and rice, *TAS3* ta-siRNAs also regulate homologs of *ARF3* (189, 573) and downregulate miR166 (568, 573, 574). Although similar principles are used,

finetuning of the regulatory mechanism is different as demonstrated by the adaxial accumulation of miR390 (526). The mechanism by which ta-siRNAs regulate miR166 and the role of ARFs remains to be elucidated (575). The role of *TAS3* ta-siRNAs is not limited to leaf polarity, but also vegetative phase change (see further) and lateral root development are under its control (251, 576).

4.4.1.2   Regulation of cell proliferation

CUP-SHAPED COTYLEDON 1 (*CUC1*) and *CUC2*, two NAC transcription factors (TF), are targeted by miR164 (268, 577, 578). The miR164 family has three members that are at least partially redundant (579, 580). *cuc1 cuc2* double mutants fail to maintain the apical meristem and to establish lateral organ boundaries (577, 581). Upregulation of *CUC2*, either through decreased miR164 levels or usage of a miRNA-resistant *cuc2* allele, causes enlarged organs by extending proliferation phase (582, 583). Upregulation of *CUC2* leads to increased serrations, probably due to modulation of cell proliferation during outgrowth of serrations (584). Leaf shape in tomatoes is also controlled by miR164 and the *CUC2*-homolog *GOBLET* (*GOB*). Both loss-of-function *gob* mutants and gain-of-function *GOB* mutants have simple leaves instead of compound leaves because they fail to initiate higher order leaflets or through ectopic proliferation between these leaflets, respectively (585, 586). The miR164-*CUC* regulatory module is also involved in flower development (580), phyllotaxy (587) and axillary meristem formation (588). Another member of the NAC TF family, *NAC1*, is involved in transduction of auxin signaling for lateral root emergence. miR164 is induced by auxin and mediates *NAC1* homeostasis to regulate auxin-signaling (589). *ORESARA1* (*ORE1*), a NAC TF, is also targeted by miR164. ETHYLENE INSENSITIVE 2 (EIN2) mediates age-dependent downregulation of miR164, upregulating *ORE1* with ageing. EIN2 additionally induces *ORE1* independent of miR164 (Figure 4.7A, 590, 591).

A *jaw-D* mutant, overexpressing miR319a, causes repression of five TEOSINTE BRANCHED/CYCLOIDEA/PCF (TCP) family transcription factors (*TCP2*, *TCP3*, *TCP4*, *TCP10* and *TCP24*) (592). These TCPs promote differentiation and downregulation leads to a delay in leaf maturation and thus prolonged proliferation phase. This results in an increase in leaf lamina growth. One of the targets, *TCP4*, promotes expression of *LIPOXYGENASE2* (*LOX2*), which is crucial in jasmonic acid (JA) biosynthesis (593, 594). JA promotes plant senescence and inhibits cell cycle progression (594, 595, 596). The link between miR319 and JA biosynthesis thus helps explain the effect on plant growth and leaf senescence (Figure 4.7B, 556). Ectopic expression of a chimeric *TCP3* repressor leads to downregulation of miR164 and upregulation of *CUC* genes and overexpression of a miR319-resistant *TCP3* resulted in downregulation of *CUC1* and *CUC2*. Also indirect TCP3-mediated regulation of *CUC* genes independent of miR164 could be involved (597). This link with miR164-*CUC* regulation could contribute to the growth phenotype in the *jaw-*

D mutant. Also in tomato miR319 is implicated in leaf development by regulating *LANCEOLATE* (*LA*), a TCP TF. Compound leaves are converted into simple leaves in a miRNA-resistant *la* allele mutant due to precocious differentiation and down-regulation of *LA* results in prolonged proliferation (598).

miR396 is represented by two loci and regulates GROWTH-REGULATING FAC-TOR (GRF) TFs involved in leaf growth (*GRF1*, *GRF2*, *GRF3*, *GFR7*, *GRF8*, *GRF9*) (70, 599, 600). Leaf development can be subdivided in three phases: proliferation (cell division and cell size homeostasis), expansion (cell expansion and onset of endoreduplication, a modified mitotic cell cycle) and finally maturity. The transition from proliferation to expansion occurs in a tip-to-base gradient (601). miR396 expression increases during development similar to the mitotic gradient over a developing leaf (601, 602). Overexpression of miR396 causes reduced expression levels of mitotic genes, suggesting it is involved in establishment of the mitotic gradient during leaf development (602). miR396 overexpressing lines are more tolerant to drought, possibly through the lower density of stomata (600). Ectopic expression of miR396 reduces cell proliferation resulting in narrow leaves due to less cells. Cell

---

Figure 4.7 *(on the next page)***: Role of small RNAs in leaf development.** (A) Regulation of dorsoventral polarity in the leaf by small RNAs in *Arabidopsis*. Adaxial co-localization of ARGONAUTE7 (AGO7), *TAS3a* and miR390 generates ta-siRNAs that downregulate *AUXIN RESPONSE FACTOR 3* (*ARF3*) and *ARF4*. Repression of miR165/166, through an unknown mechanism involving these ta-siRNAs, causes a high expression level of the adaxial determinants *PHABULOSA* (*PHB*), *PHAVO-LUTA* (*PHV*) and *REVOLUTA* (*REV*) (grey font designates lower expression). The ta-siRNAs move through the leaf from cell-to-cell forming a gradient (represented by the background gradient). In the abaxial domain of the leaf downregulation of *PHB*, *PHV* and *REV* contributes to abaxial identity. (B) Regulation of cell proliferation and leaf senescence during leaf development in *Arabidopsis*. Amongst other *TEOSINTE BRANCHED/CYCLOIDEA/PCF* (*TCP*) TFs, *TCP3* and *TCP4* are targeted by miR319. TCP3 inhibits *CUP-SHAPED COTYLEDON* (*CUC*) genes both directly and indirectly through miR164. *CUC* genes regulate the extent of the proliferation phase in leaf development. miR164 also represses *ORESARA1* (*ORE1*), a promoter of leaf senescence. ETHYLENE INSENSITIVE 2 (EIN2) promotes leaf senescence by inhibiting miR164 and activating *ORE1*. TCP4 upregulates jasmonic acid (JA) biosynthesis through upregulation of *LIPOXYGENASE2* (*LOX2*). JA promotes leaf senescence and inhibits the mitotic cell cycle. miR396 downregulates *GROWTH REGULATING FACTOR*s (*GRFs*), thus inhibiting cell proliferation. In *Taxus*, miR164 and miR396 are regulated by JA. (C) Regulation of vegetative phase transition (250). miR156 is a master regulator of juvenile-to-adult transition by targeting *SPL* TFs that promote adult traits. A subset of these *SPL*s regulate miR172b. This miRNA represses *TARGET OF EAT1* (*TOE1*) and *TOE2*, two *AP2* TFs that promote juvenile characteristics. Negative feedback loops likely stabilize the regulatory system.
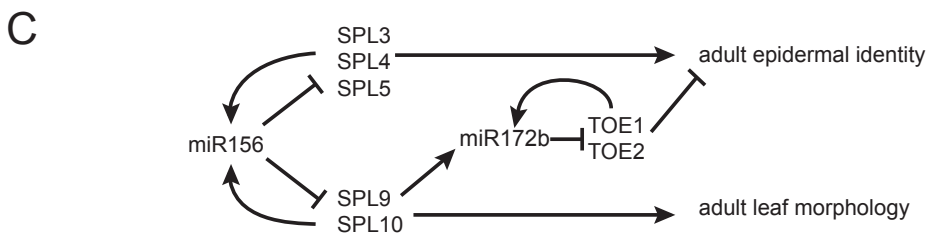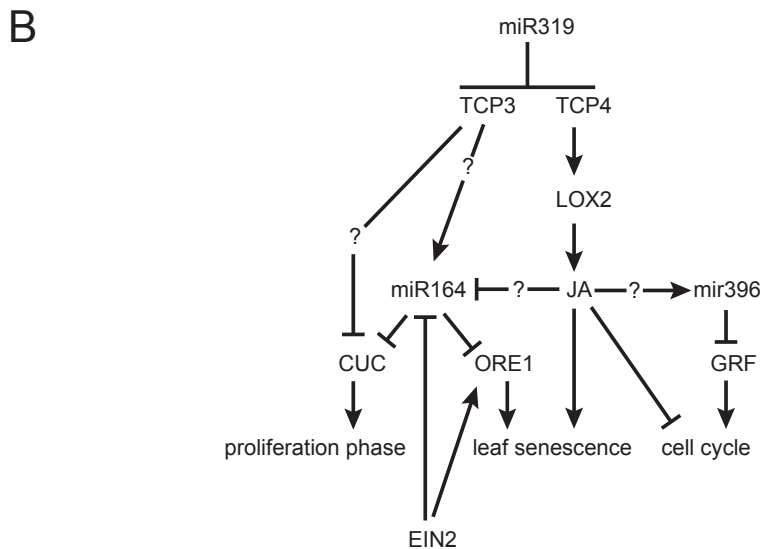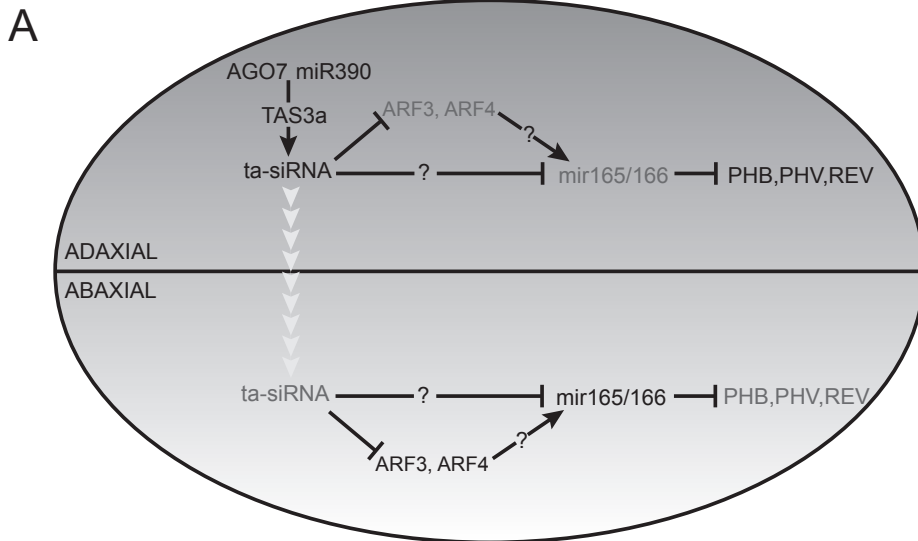
number is quantitatively regulated by the balance between miR396 and *GRF*s (602). *GRF*s are known regulators of cell number (599, 603), but the underlying mechanism is not yet known. *GRF*s that are not targeted by miR396 act redundantly with those that are under its regulatory control. GRF-INTERACTING FACTOR1 (GIF1)/ANGUSTIFOLIA3 (AN3) acts together with GRFs and is also repressed indirectly by miR396 through unknown mechanisms (600). miR396 is positively regulated by TCP4, itself under control of miR319 (602) embedding its regulatory role in a larger network that controls proliferation and senescence of leaf development (Figure 4.7B). Adding to the interconnection of the network, in *Taxus chinensis* jasmonic acid (JA) downregulates miR164 and upregulates miR396.

Finally, it was recently shown that miR159 represses *MYB33* and *MYB65*, two MYB TFs, in vegetative tissues. Failure to inhibit these TFs leads to a reduction of cell proliferation (604).

4.4.2   Phase transition

The development of a plant can be divided into several phases: juvenile, adult and generative (605). The transition from juvenile-to-adult is characterized by differences in morphology such as leaf shape and size, number of hydathodes, trichome distribution and the sensitivity to floral stimuli (606, 607, 608). The first leaf pair is small, flat and round, has a smooth margin and only trichomes on the adaxial surface. Subsequent leaves become progressively larger, are oval, curl downwards, show serrations, have short petiols and can produce both abaxial and adaxial trichomes. These morphological changes in subsequent leaves, are referred to as heteroblasty.

Small RNAs play key roles in the regulation of this transition in *Arabidopsis* and this is also conserved in maize and rice (608, 609, 610, 611). Mutants in *HASTY* (*HST*), implicated in miRNA biogenesis, have accelerated vegetative phase change (229, 230, 234). SQUINT (SQN) is important for the activity of miR156 and *sqn* mutations also lead to precocious adult traits (237, 238). *SERRATE* (*SE*) mutants also show this phenotype (612, 613).

miR156 was identified as necessary and sufficient for juvenile characteristics (250, 608). This miRNA has higher expression levels in juvenile leaves compared to adult leaves (614). Several *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) transcription factors contain a target site for miR156 (270, 608, 614). SPL3, SPL4 and SPL5 promote epidermal differentiation with adult characteristics (295, 614), while SPL9, SPL15, SPL10 and SPL11 stimulate all aspects of adult leaf morphology (250, 615, 616). The different SPL TFs have overlapping but distinct functions in leaf development (250). All these TFs are targeted by miR156 which serves as a master regulator (608).

A different miRNA, miR172, promotes adult traits and flowering (617, 618, 619). SPL9 and SPL10, itself regulated by miR156, regulate transcription of miR172b (250).

Indeed, miR172 has an opposite expression pattern as miR156, with expression levels increasing during shoot development (250, 617, 619, 620). Next to miR172, SPL9 has additional, but currently unknown, targets that mediate differentiation of the epidermis. The effect of miR172 on phase change is mediated by its targets *TARGET OF EAT1* (*TOE1*) and *TOE2*, two *APETALA2* (*AP2*) TFs that promote juvenile identity, contributing to the effect of miR156 on development. Negative feedback loops involving both miRNAs and their targets likely stabilize this regulatory system (Figure 4.7C). The downstream effector genes are currently unknown (250).

Next to the miRNA biogenesis components HST, SQN and SE, a second set of sRNA related proteins affects juvenile-to-adult transition: AGO7, RDR6, SGS3 and DCL4; all involved in processing *TAS3* transcripts (51, 73, 115, 316, 621). There is however a difference between the two gene sets: the miRNA-related set accelerates the rate of transition from juvenile-to-adult and results in rosettes with fewer leaves, while the *TAS3*-set changes the timing of the first appearance of adult traits, but does not change transition or total number of leaves (622). Additionally, *sqn* mutants and a miR156-resistant *spl15* allele resulted in an increase of cell number in the leaves as well as a decrease in cell size, this was not the case for *rdr6*, *sgs3*, *ago7* or *arf3* and *arf4* (51, 115, 316, 621, 623). Mutant alleles of *ARF3* that are not repressed by *TAS3*-derived ta-siRNAs accelerate phase change (624) and a screen for suppressors of the precocious transition in *ago7*, resulted in *arf3* and *arf4* alleles (622). Thus next to organ polarity, *TAS3* ta-siRNAs and its targets *ARF3* and *ARF4* have additional roles in leaf development. The expression of *ARF3* and *ARF4* as well as the abundance of *TAS3* ta-siRNAs does not change during vegetative development, indicating that additional input is required to add a temporal component to this regulatory mechanism (622).

4.4.3   Hormone biosynthesis and signaling

Small RNAs have important roles in hormone biosynthesis and signaling. The involvement of miR319 with jasmonic acid (JA) biosynthesis through *TCP4* and *LOX2* was described before. Abscisic acid (ABA) induces miR159 during seed germination. miR159 targets *MYB33* and *MYB101*, two MYB TF that are positive regulators of ABA responses, generate a regulatory feedback loop (625).

AUXIN RESPONSE FACTORs (ARFs) are key regulators in auxin-signaling. From the 23 described *ARF*s in *Arabidopsis* (626), about one third is controlled by miRNAs or ta-siRNAs. miR160 regulates *ARF10*, *ARF16*, *ARF17* (627, 628). Proper regulation of *ARF17* is required for normal development and it is a negative regulator of adventitious root initiation (249, 627). *ARF10* and *ARF16* regulate root cap formation (628). Downregulation of *ARF10* by miR160 is important for seed germination and post-embryonic development and modulates ABA sensitivity during germination (629). miR167 is involved in auxin signaling through repression of *ARF6* and *ARF8*, required for ovule and pollen development (630, 631) and posi-

tive regulators of adventitious root initiation (249). The regulation of *TAS3*-derived ta-siRNAs of *ARF2*, *ARF3* and *ARF4* has been described earlier (316, 571, 622, 624).

NAC1, a NAC TF targeted by miR164, transduces the auxin signal for lateral root development downstream of TRANSPORT INHIBITOR RESPONSE 1 (TIR1) (632). Auxin induces miR164 creating a negative feedback loop that attenuates auxin signaling (589). *AtHB15*, a HD-ZIPIII TF targeted by miR165/166, is important for vascular development and has been linked to auxin signaling (633).

Additionally, the auxin receptors *TIR1*, *AUXIN-SIGNALING F-BOX PROTEIN 2* (*AFB2*) and *AFB3* are targeted by miR393 (206, 541, 634). Bacterial infection can induce miR393 and thus repress auxin signaling, which contributes to basal defense (634, 635). Also miR160 and miR167 are upregulated upon *Pseudomonas syringae* infection, involving additional components of auxin signaling in bacterial defense (179, 636).

4.5   Conclusion

One of the most prominent observations about small RNA silencing is, despite the tremendous progress that has been made the last years, that our knowledge of this molecular regulation mechanism is still limited. The different biogenesis pathways are deeply intertwined and auto-regulatory feedback loops add to the complexity. Understanding the specificity of components such as RDRs, DCLs and AGOs will be important in elucidating the regulatory pathways that impinge on specific loci. Profiling of small RNAs in specific conditions and cell types still has a high potential in discovering new small RNA regulatory modules.

Further progress will likely require a systems biology approach to connect the different interacting components to functional outputs. A first effort has been made to model small RNAs and their targets into a network (637). Both RNA molecules that generate small RNAs as those that are targeted were defined as nodes. Predicted interactions between source and target make up the edges. The majority of its nodes have very low degree of connectivity: only few edges come in or go out. Hub nodes of very high degree are present, ensuring robustness and reducing the distance 'messages' have to travel. The pattern of hubs is dissortative, they are not preferentially connected to each other, a characteristic which could serve as a safeguard against failure as this inhibits propagation from one hub to another. There is also a high degree of clustering. A large number of isolated 'islands' exists, but about 85% of the nodes is interconnected. These are all known properties of biological networks (637). This first attempt to model small RNA targeting and interactions illustrates the presence of such a highly complex network in *Arabidopsis* and this could direct wet-lab research to further piece the genetic puzzle together.

References

[1] R. C. Lee, R. L. Feinbaum, V. Ambros. *Cell* **75**, 843 (1993).

[2] A. Fire, *et al. Nature* **391**, 806 (1998).

[3] P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel. *Cell* **101**, 25 (2000).

[4] S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon. *Nature* **404**, 293 (2000).

[5] D. Yang, H. Lu, J. W. Erickson. *Curr Biol* **10**, 1191 (2000).

[6] A. J. Hamilton, D. C. Baulcombe. *Science* **286**, 950 (1999).

[7] C. Cogoni, *et al. EMBO J* **15**, 3153 (1996).

[8] X. Chen. *Plant J* **61**, 941 (2010).

[9] J. C. Carrington, V. Ambros. *Science* **301**, 336 (2003).

[10] A. C. Mallory, H. Vaucheret. *Nat Genet* **38 Suppl**, S31 (2006).

[11] G. Chuck, H. Candela, S. Hake. *Curr Opin Plant Biol* **12**, 81 (2009).

[12] G. Stefani, F. J. Slack. *Nat Rev Mol Cell Biol* **9**, 219 (2008).

[13] O. Voinnet. *Trends Genet* **17**, 449 (2001).

[14] A. Deleris, *et al. Science* **313**, 68 (2006).

[15] P. M. Waterhouse, A. F. Fusaro. *Science* **313**, 54 (2006).

[16] S. Mlotshwa, G. J. Pruss, V. Vance. *Trends Plant Sci* **13**, 375 (2008).

[17] V. Ruiz-Ferrer, O. Voinnet. *Annu Rev Plant Biol* **60**, 485 (2009).

[18] S. W.-L. Chan, I. R. Henderson, S. E. Jacobsen. *Nat Rev Genet* **6**, 351 (2005).

[19] S. W.-L. Chan. *Trends Plant Sci* **13**, 383 (2008).

[20] A. Girard, G. J. Hannon. *Trends Cell Biol* **18**, 136 (2008).

[21] A. Frizzi, S. Huang. *Plant Biotechnol J* **8**, 655 (2010).

[22] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl. *Science* **294**, 853 (2001).

[23] N. C. Lau, L. P. Lim, E. G. Weinstein, D. P. Bartel. *Science* **294**, 858 (2001).

[24] R. C. Lee, V. Ambros. *Science* **294**, 862 (2001).

[25] C. Llave, K. D. Kasschau, M. A. Rector, J. C. Carrington. *Plant Cell* **14**, 1605 (2002).

[26] B. J. Reinhart, *et al. Genes Dev* **16**, 1616 (2002).

[27] R. Rajagopalan, H. Vaucheret, J. Trejo, D. P. Bartel. *Genes Dev* **20**, 3407 (2006).

[28] K. D. Kasschau, *et al. PLoS Biol* **5**, e57 (2007).

[29] C. Lu, *et al. Genome Res* **16**, 1276 (2006).

[30] I. R. Henderson, *et al. Nat Genet* **38**, 721 (2006).

[31] B. Zhang, *et al. Gene* **397**, 26 (2007).

[32] M.-B. Ruan, *et al. Genomics* **94**, 263 (2009).

[33] K. Nobuta, *et al. Nat Biotechnol* **25**, 473 (2007).

[34] R. Sunkar, *et al. BMC Plant Biol* **8**, 25 (2008).

[35] C. Lu, *et al. Proc Natl Acad Sci U S A* **105**, 4951 (2008).

[36] C. Johnson, *et al. Genome Res* **19**, 1429 (2009).

[37] C. Johnson, *et al. Nucleic Acids Res* **35**, D829 (2007).

[38] S. Subramanian, *et al. BMC Genomics* **9**, 160 (2008).

[39] G. Szittya, *et al. BMC Genomics* **9**, 593 (2008).

[40] C. Song, *et al. BMC Genomics* **11**, 431 (2010).

[41] J. Zhang, Y. Xu, Q. Huan, K. Chong. *BMC Genomics* **10**, 449 (2009).

[42] B. Wei, *et al. Funct Integr Genomics* **9**, 499 (2009).

[43] I. A. Yakovlev, C. G. Fossdal, O. Johnsen. *New Phytol* **187**, 1154 (2010).

[44] M. J. Axtell, C. Jan, R. Rajagopalan, D. P. Bartel. *Cell* **127**, 565 (2006).

[45] D. Klevebring, *et al. BMC Genomics* **10**, 620 (2009).

[46] H. Garcia-Ruiz, *et al. Plant Cell* **22**, 481 (2010).

[47] X. Qi, F. S. Bao, Z. Xie. *PLoS One* **4**, e4971 (2009).

[48] O. Voinnet. *Trends Plant Sci* **13**, 317 (2008).

[49] M. Alleman, *et al. Nature* **442**, 295 (2006).

[50] M. Wassenegger, G. Krczal. *Trends Plant Sci* **11**, 142 (2006).

[51] A. Peragine, *et al. Genes Dev* **18**, 2368 (2004).

[52] P. Dunoyer, *et al. EMBO J* **29**, 1699 (2010).

[53] O. Voinnet. *Cell* **136**, 669 (2009).

[54] B. Huettel, *et al. Biochim Biophys Acta* **1769**, 358 (2007).

[55] O. Borsani, *et al. Cell* **123**, 1279 (2005).

[56] S. Katiyar-Agarwal, *et al. Proc Natl Acad Sci U S A* **103**, 18002 (2006).

[57] U. Klahre, *et al. Proc Natl Acad Sci U S A* **99**, 11981 (2002).

[58] C. F. Chuang, E. M. Meyerowitz. *Proc Natl Acad Sci U S A* **97**, 4985 (2000).

[59] S. M. Elbashir, *et al. Nature* **411**, 494 (2001).

[60] E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon. *Nature* **409**, 363 (2001).

[61] B. Yu, *et al. Science* **307**, 932 (2005).

[62] J. Li, *et al. Curr Biol* **15**, 1501 (2005).

[63] Z. Yang, Y. W. Ebright, B. Yu, X. Chen. *Nucleic Acids Res* **34**, 667 (2006).

[64] X. Chen. *Annu Rev Cell Dev Biol* **25**, 21 (2009).

[65] F. Vazquez, S. Legrand, D. Windels. *Trends Plant Sci* **15**, 337 (2010).

[66] Y. Tomari, *et al. Science* **306**, 1377 (2004).

[67] A. Hamilton, O. Voinnet, L. Chappell, D. Baulcombe. *EMBO J* **21**, 4671 (2002).

[68] R. W. Carthew, E. J. Sontheimer. *Cell* **136**, 642 (2009).

[69] S. M. Elbashir, W. Lendeckel, T. Tuschl. *Genes Dev* **15**, 188 (2001).

[70] M. W. Jones-Rhoades, D. P. Bartel, B. Bartel. *Annu Rev Plant Biol* **57**, 19 (2006).

[71] E. Allen, Z. Xie, A. M. Gustafson, J. C. Carrington. *Cell* **121**, 207 (2005).

[72] D. C. Baulcombe. *Plant Mol Biol* **32**, 79 (1996).

[73] Z. Xie, E. Allen, A. Wilken, J. C. Carrington. *Proc Natl Acad Sci U S A* **102**, 12984 (2005).

[74] S. E. Schauer, S. E. Jacobsen, D. W. Meinke, A. Ray. *Trends Plant Sci* **7**, 487 (2002).

[75] L. H. Franzmann, E. S. Yoon, D. W. Meinke. *Plant J* **7**, 341 (1995).

[76] K. Robinson-Beers, R. E. Pruitt, C. S. Gasser. *Plant Cell* **4**, 1237 (1992).

[77] B. W. Schwartz, E. C. Yeung, D. W. Meinke. *Development* **120**, 3235 (1994).

[78] S. E. Jacobsen, M. P. Running, E. M. Meyerowitz. *Development* **126**, 5231 (1999).

[79] R. Margis, *et al. FEBS Lett* **580**, 2442 (2006).

[80] M. J. Axtell, J. A. Snyder, D. P. Bartel. *Plant Cell* **19**, 1750 (2007).

[81] F. E. Nicolas, *et al. Nucleic Acids Res* **38**, 5535 (2010).

[82] L. He, G. J. Hannon. *Nat Rev Genet* **5**, 522 (2004).

[83] R. F. Ketting, *et al. Genes Dev* **15**, 2654 (2001).

[84] S. W. Knight, B. L. Bass. *Science* **293**, 2269 (2001).

[85] S. M. Elbashir, *et al. EMBO J* **20**, 6877 (2001).

[86] G. Tang, B. J. Reinhart, D. P. Bartel, P. D. Zamore. *Genes Dev* **17**, 49 (2003).

[87] Z. Dong, M.-H. Han, N. Fedoroff. *Proc Natl Acad Sci U S A* **105**, 9970 (2008).

[88] Y. Qi, A. M. Denli, G. J. Hannon. *Mol Cell* **19**, 421 (2005).

[89] V. Gasciolli, A. C. Mallory, D. P. Bartel, H. Vaucheret. *Curr Biol* **15**, 1494 (2005).

[90] N. Bouche, D. Lauressergues, V. Gasciolli, H. Vaucheret. *EMBO J* **25**, 3347 (2006).

[91] I. J. Macrae, *et al. Science* **311**, 195 (2006).

[92] I. J. MacRae, K. Zhou, J. A. Doudna. *Nat Struct Mol Biol* **14**, 934 (2007).
[93] I. Papp, *et al. Plant Physiol* **132**, 1382 (2003).
[94] Z. Xie, *et al. PLoS Biol* **2**, E104 (2004).
[95] A. Hiraguri, *et al. Plant Mol Biol* **57**, 173 (2005).
[96] X. Tang, Y. Zhang, L. Tucker, B. Ramratnam. *Nucleic Acids Res* **38**, 6610 (2010).
[97] P. Dunoyer, C. Himber, O. Voinnet. *Nat Genet* **37**, 1356 (2005).
[98] G. Moissiard, E. A. Parizotto, C. Himber, O. Voinnet. *RNA* **13**, 1268 (2007).
[99] A. F. Fusaro, *et al. EMBO Rep* **7**, 1168 (2006).
[100] F. Vazquez. *Trends Plant Sci* **11**, 460 (2006).
[101] S. J. Curtin, *et al. FEBS Lett* **582**, 2753 (2008).
[102] J. P. Kastenmayer, P. J. Green. *Proc Natl Acad Sci U S A* **97**, 13985 (2000).
[103] S. Gazzani, *et al. Science* **306**, 1046 (2004).
[104] J. Curaba, X. Chen. *J Biol Chem* **283**, 3059 (2008).
[105] J. A. Diaz-Pendon, F. Li, W.-X. Li, S.-W. Ding. *Plant Cell* **19**, 2053 (2007).
[106] L. Donaire, *et al. J Virol* **82**, 5167 (2008).
[107] F. Qu, X. Ye, T. J. Morris. *Proc Natl Acad Sci U S A* **105**, 14732 (2008).
[108] Z. Xie, B. Fan, C. Chen, Z. Chen. *Proc Natl Acad Sci U S A* **98**, 6516 (2001).
[109] D. Yu, B. Fan, S. A. MacFarlane, Z. Chen. *Mol Plant Microbe Interact* **16**, 206 (2003).
[110] F. Vazquez, *et al. Mol Cell* **16**, 69 (2004).
[111] T. Dalmay, *et al. Cell* **101**, 543 (2000).
[112] P. Mourrain, *et al. Cell* **101**, 533 (2000).
[113] F. Qu, *et al. J Virol* **79**, 15209 (2005).
[114] N. Muangsan, C. Beclin, H. Vaucheret, D. Robertson. *Plant J* **38**, 1004 (2004).
[115] M. Yoshikawa, A. Peragine, M. Y. Park, R. S. Poethig. *Genes Dev* **19**, 2164 (2005).
[116] C. Himber, *et al. EMBO J* **22**, 4523 (2003).
[117] O. Voinnet, P. Vain, S. Angell, D. C. Baulcombe. *Cell* **95**, 177 (1998).
[118] F. E. Vaistij, L. Jones, D. C. Baulcombe. *Plant Cell* **14**, 857 (2002).
[119] A. Bleys, H. Van Houdt, A. Depicker. *RNA* **12**, 1633 (2006).
[120] R. Akbergenov, *et al. Nucleic Acids Res* **34**, 462 (2006).
[121] V. Ramachandran, X. Chen. *Trends Plant Sci* **13**, 368 (2008).
[122] G. Vilkaitis, A. Plotnikova, S. Klimasauskas. *RNA* **16**, 1935 (2010).
[123] J. Liu, *et al. Science* **305**, 1437 (2004).
[124] F. V. Rivas, *et al. Nat Struct Mol Biol* **12**, 340 (2005).
[125] N. Baumberger, D. C. Baulcombe. *Proc Natl Acad Sci U S A* **102**, 11928 (2005).
[126] H. Vaucheret. *Trends Plant Sci* **13**, 350 (2008).
[127] I. Gy, *et al. Plant Cell* **19**, 3451 (2007).
[128] F. F. Souret, J. P. Kastenmayer, P. J. Green. *Mol Cell* **15**, 173 (2004).
[129] B. Gregory, *et al. Dev Cell* **14**, 811 (2008).
[130] D. Gatfield, E. Izaurralde. *Nature* **429**, 575 (2004).
[131] S. Bail, *et al. RNA* **16**, 1032 (2010).
[132] E. Glazov, *et al. Plant J* **35**, 342 (2003).
[133] H.-W. Wang, *et al. Nat Struct Mol Biol* **16**, 1148 (2009).
[134] C. Matranga, *et al. Cell* **123**, 607 (2005).
[135] K. Miyoshi, *et al. Genes Dev* **19**, 2837 (2005).
[136] P. J. F. Leuschner, S. L. Ameres, S. Kueng, J. Martinez. *EMBO Rep* **7**, 314 (2006).
[137] T. Kawamata, H. Seitz, Y. Tomari. *Nat Struct Mol Biol* **16**, 953 (2009).

[138] S. Iwasaki, *et al. Mol Cell* **39**, 282 (2010).

[139] T. Iki, *et al. Mol Cell* **39**, 292 (2010).

[140] T. Kawamata, Y. Tomari. *Trends Biochem Sci* **35**, 368 (2010).

[141] J.-B. Ma, *et al. Nature* **434**, 666 (2005).

[142] M. Yoda, *et al. Nat Struct Mol Biol* **17**, 17 (2010).

[143] T. A. Rand, S. Petersen, F. Du, X. Wang. *Cell* **123**, 621 (2005).

[144] H. Addepalli, *et al. Nucleic Acids Res* **38**, 7320 (2010).

[145] B. Haley, P. D. Zamore. *Nat Struct Mol Biol* **11**, 599 (2004).

[146] K. Bohmert, *et al. EMBO J* **17**, 170 (1998).

[147] J.-B. Morel, *et al. Plant Cell* **14**, 629 (2002).

[148] K.-I. Nonomura, *et al. Plant Cell* **19**, 2583 (2007).

[149] R. E. Collins, X. Cheng. *FEBS Lett* **579**, 5841 (2005).

[150] J.-B. Ma, K. Ye, D. J. Patel. *Nature* **429**, 318 (2004).

[151] A. Boland, *et al. EMBO Rep* **11**, 522 (2010).

[152] D. S. Schwarz, *et al. Cell* **115**, 199 (2003).

[153] A. Khvorova, A. Reynolds, S. D. Jayasena. *Cell* **115**, 209 (2003).

[154] J.-J. Song, S. K. Smith, G. J. Hannon, L. Joshua-Tor. *Science* **305**, 1434 (2004).

[155] J. S. Parker, S. M. Roe, D. Barford. *EMBO J* **23**, 4727 (2004).

[156] Y. Wang, *et al. Nature* **461**, 754 (2009).

[157] Y. Qi, *et al. Nature* **443**, 1008 (2006).

[158] T. A. Montgomery, *et al. Cell* **133**, 128 (2008).

[159] D. S. Schwarz, G. Hutvagner, B. Haley, P. D. Zamore. *Mol Cell* **10**, 537 (2002).

[160] S. Mi, *et al. Cell* **133**, 116 (2008).

[161] A. Takeda, *et al. Plant Cell Physiol* **49**, 493 (2008).

[162] L. Wu, *et al. Mol Cell* **38**, 465 (2010).

[163] D. Chen, *et al. Bioinformatics* **26**, 1391 (2010).

[164] F. Frank, N. Sonenberg, B. Nagar. *Nature* **465**, 818 (2010).

[165] B. Moussian, *et al. EMBO J* **17**, 1799 (1998).

[166] K. Lynn, *et al. Development* **126**, 469 (1999).

[167] P. Brodersen, *et al. Science* **320**, 1185 (2008).

[168] E. R. Havecker, *et al. Plant Cell* **22**, 321 (2010).

[169] X. Zheng, J. Zhu, A. Kapoor, J.-K. Zhu. *EMBO J* **26**, 1691 (2007).

[170] V. Olmedo-Monfil, *et al. Nature* **464**, 628 (2010).

[171] D. Lobbes, *et al. EMBO Rep* **7**, 1052 (2006).

[172] S. Guang, *et al. Nature* **465**, 1097 (2010).

[173] V. Ambros. *Cell* **113**, 673 (2003).

[174] H. Siomi, M. C. Siomi. *Mol Cell* **38**, 323 (2010).

[175] V. N. Kim, J. Han, M. C. Siomi. *Nat Rev Mol Cell Biol* **10**, 126 (2009).

[176] V. K. Gangaraju, H. Lin. *Nat Rev Mol Cell Biol* **10**, 116 (2009).

[177] D. P. Bartel. *Cell* **136**, 215 (2009).

[178] E. J. Chapman, J. C. Carrington. *Nat Rev Genet* **8**, 884 (2007).

[179] N. Fahlgren, *et al. PLoS ONE* **2**, e219 (2007).

[180] M. Axtell. *Biochim Biophys Acta* **1779**, 725 (2008).

[181] M. J. Axtell, J. L. Bowman. *Trends Plant Sci* **13**, 343 (2008).

[182] A. Molnár, *et al. Nature* **447**, 1126 (2007).

[183] B. Zhang, *et al. Plant J* **46**, 243 (2006).

[184] R. Sunkar, G. Jagadeeswaran. *BMC Plant Biol* **8**, 37 (2008).

[185] T. Zhao, *et al. Genes Dev* **21**, 1190 (2007).

[186] A. Barakat, *et al. BMC Genomics* **8**, 481 (2007).

[187] A. Barakat, *et al. Plant J* **51**, 991 (2007).

[188] C.-Z. Zhao, *et al. BMC Plant Biol* **10**, 3 (2010).

[189] Y.-F. Li, *et al. Plant J* **62**, 742 (2010).

[190] C. Lelandais-Brière, *et al. Plant Cell* **21**, 2780 (2009).

[191] C. Arenas-Huertero, *et al. Plant Mol Biol* **70**, 385 (2009).

[192] V. Pantaleo, *et al. Plant J* **62**, 960 (2010).

[193] S. Griffiths-Jones, *et al. Nucleic Acids Res* **34**, D140 (2006).

[194] Z. Zhang, *et al. Nucleic Acids Res* **38**, D806 (2010).

[195] F. Vazquez, *et al. Nucleic Acids Res* **36**, 6429 (2008).

[196] N. Fahlgren, *et al. Plant Cell* **22**, 1074 (2010).

[197] J. Piriyapongsa, I. K. Jordan. *RNA* **14**, 814 (2008).

[198] F. F. d. Felippes, *et al. RNA* **14**, 2455 (2008).

[199] Z. Ma, C. Coruh, M. J. Axtell. *Plant Cell* **22**, 1090 (2010).

[200] Q.-H. Zhu, *et al. Genome Res* **18**, 1456 (2008).

[201] Y. Lee, *et al. EMBO J* **23**, 4051 (2004).

[202] X. Cai, C. H. Hagedorn, B. R. Cullen. *RNA* **10**, 1957 (2004).

[203] J. Bracht, *et al. RNA* **10**, 1586 (2004).

[204] X. Zhou, J. Ruan, G. Wang, W. Zhang. *PLoS Comput Biol* **3**, e37 (2007).

[205] Z. Xie, *et al. Plant Physiol* **138**, 2145 (2005).

[206] M. W. Jones-Rhoades, D. P. Bartel. *Mol Cell* **14**, 787 (2004).

[207] Y. Fang, D. L. Spector. *Curr Biol* **17**, 818 (2007).

[208] Y. Fujioka, M. Utsumi, Y. Ohba, Y. Watanabe. *Plant Cell Physiol* **48**, 1243 (2007).

[209] L. Song, M.-H. Han, J. Lesicka, N. Fedoroff. *Proc Natl Acad Sci U S A* **104**, 5437 (2007).

[210] B. Yu, *et al. Proc Natl Acad Sci U S A* **105**, 10073 (2008).

[211] S. Laubinger, *et al. Proc Natl Acad Sci U S A* **105**, 8795 (2008).

[212] M.-H. Han, S. Goud, L. Song, N. Fedoroff. *Proc Natl Acad Sci U S A* **101**, 1093 (2004).

[213] F. Vazquez, V. Gasciolli, P. Crete, H. Vaucheret. *Curr Biol* **14**, 346 (2004).

[214] Y. Kurihara, Y. Takashi, Y. Watanabe. *RNA* **12**, 206 (2006).

[215] L. Yang, *et al. Plant J* **47**, 841 (2006).

[216] B. Szarzynska, *et al. Nucleic Acids Res* **37**, 3083 (2009).

[217] A. L. Eamens, *et al. RNA* **15**, 2219 (2009).

[218] S. W. Yang, *et al. Structure* **18**, 594 (2010).

[219] B. C. Meyers, S. A. Simon, J. Zhai. *Curr Biol* **20**, R68 (2010).

[220] S. Werner, H. Wollmann, K. Schneeberger, D. Weigel. *Curr Biol* **20**, 42 (2010).

[221] L. Song, M. J. Axtell, N. V. Fedoroff. *Curr Biol* **20**, 37 (2010).

[222] J. L. Mateos, N. G. Bologna, U. Chorostecki, J. F. Palatnik. *Curr Biol* **20**, 49 (2010).

[223] J. T. Cuperus, *et al. Proc Natl Acad Sci U S A* **107**, 466 (2010).

[224] C. Addo-Quaye, *et al. RNA* **15**, 2112 (2009).

[225] N. G. Bologna, J. L. Mateos, E. G. Bresso, J. F. Palatnik. *EMBO J* **28**, 3646 (2009).

[226] W. Park, *et al. Curr Biol* **12**, 1484 (2002).

[227] H. A. Ebhardt, E. P. Thi, M.-B. Wang, P. J. Unrau. *Proc Natl Acad Sci U S A* **102**, 13398 (2005).

[228] S. Boutet, *et al. Curr Biol* **13**, 843 (2003).

[229] A. Telfer, R. S. Poethig. *Development* **125**, 1889 (1998).

[230] K. M. Bollman, *et al. Development* **130**, 1493 (2003).

[231] R. Yi, Y. Qin, I. G. Macara, B. R. Cullen. *Genes Dev* **17**, 3011 (2003).

[232] M. T. Bohnsack, K. Czaplinski, D. Gorlich. *RNA* **10**, 185 (2004).

[233] E. Lund, *et al. Science* **303**, 95 (2004).

[234] M. Y. Park, *et al. Proc Natl Acad Sci U S A* **102**, 3691 (2005).

[235] H. Vaucheret, F. Vazquez, P. Crete, D. P. Bartel. *Genes Dev* **18**, 1187 (2004).

[236] T. A. Montgomery, *et al. Proc Natl Acad Sci U S A* **105**, 20055 (2008).

[237] T. Z. Berardini, K. Bollman, H. Sun, R. S. Poethig. *Science* **291**, 2405 (2001).

[238] M. R. Smith, *et al. Proc Natl Acad Sci U S A* **106**, 5424 (2009).

[239] Y. Tomari, *et al. Cell* **116**, 831 (2004).

[240] K. Okamura, *et al. Nat Struct Mol Biol* **15**, 354 (2008).

[241] B. Czech, *et al. Mol Cell* **36**, 445 (2009).

[242] M. Megraw, *et al. RNA* **12**, 1612 (2006).

[243] M. Megraw, A. G. Hatzigeorgiou. *Methods Mol Biol* **592**, 149 (2010).

[244] J.-G. Joung, Z. Fei. *Bioinformatics* **25**, 387 (2009).

[245] W. Kim, *et al. Cell Res* **19**, 899 (2009).

[246] M.-N. Pouch-Pélissier, *et al. PLoS Genet* **4**, e1000096 (2008).

[247] B. Yu, *et al. Nucleic Acids Res* **38**, 5844 (2010).

[248] J. Tsang, J. Zhu, A. van Oudenaarden. *Mol Cell* **26**, 753 (2007).

[249] L. Gutierrez, *et al. Plant Cell* **21**, 3119 (2009).

[250] G. Wu, *et al. Cell* **138**, 750 (2009).

[251] E. Marin, *et al. Plant Cell* **22**, 1104 (2010).

[252] Z. Xie, K. D. Kasschau, J. C. Carrington. *Curr Biol* **13**, 784 (2003).

[253] H. Vaucheret, A. C. Mallory, D. P. Bartel. *Mol Cell* **22**, 129 (2006).

[254] A. C. Mallory, H. Vaucheret. *EMBO Rep* **10**, 521 (2009).

[255] A. C. Mallory, *et al. PLoS Genet* **5**, e1000646 (2009).

[256] J. Azevedo, *et al. Genes Dev* **24**, 904 (2010).

[257] A. Arvey, *et al. Mol Syst Biol* **6**, 363 (2010).

[258] J. M. Franco-Zorrilla, *et al. Nat Genet* **39**, 1033 (2007).

[259] T. Haraguchi, Y. Ozaki, H. Iba. *Nucleic Acids Res* **37**, e43 (2009).

[260] M. S. Ebert, J. R. Neilson, P. A. Sharp. *Nat Methods* **4**, 721 (2007).

[261] L. Poliseno, *et al. Nature* **465**, 1033 (2010).

[262] V. Ramachandran, X. Chen. *Science* **321**, 1490 (2008).

[263] S. Kennedy, D. Wang, G. Ruvkun. *Nature* **427**, 645 (2004).

[264] T. Iida, R. Kawaguchi, J.-i. Nakayama. *Curr Biol* **16**, 1459 (2006).

[265] S. Chatterjee, H. Grosshans. *Nature* **461**, 546 (2009).

[266] P. H. Olsen, V. Ambros. *Dev Biol* **216**, 671 (1999).

[267] D. V. Dugas, B. Bartel. *Plant Mol Biol* **67**, 403 (2008).

[268] M. W. Rhoades, *et al. Cell* **110**, 513 (2002).

[269] E. Lanet, *et al. Plant Cell* **21**, 1762 (2009).

[270] M. Gandikota, *et al. Plant J* **49**, 683 (2007).

[271] V. Stoppin-Mellet, J. Gaillard, M. Vantard. *Cell Biol Int* **27**, 279 (2003).

[272] J. Xu, J.-Y. Yang, Q.-W. Niu, N.-H. Chua. *Plant Cell* **18**, 3386 (2006).

[273] D. C. Goeres, *et al. Plant Cell* **19**, 1549 (2007).

[274] A. Eulalio, I. Behm-Ansmant, D. Schweizer, E. Izaurralde. *Mol Cell Biol* **27**, 3970 (2007).

[275] R. Bari, B. Datt Pant, M. Stitt, W.-R. Scheible. *Plant Physiol* **141**, 988 (2006).

[276] R. Schwab, *et al. Plant Cell* **18**, 1121 (2006).

[277] A. C. Mallory, *et al. EMBO J* **23**, 3356 (2004).

[278] P. Dunoyer, *et al. Plant Cell* **16**, 1235 (2004).

[279] N. Bao, K.-W. Lye, M. K. Barton. *Dev Cell* **7**, 653 (2004).

[280] X. Zhang, *et al. Cell* **126**, 1189 (2006).

[281] P. Chellappan, *et al. Nucleic Acids Res* **38**, 6883 (2010).

[282] L. Wu, *et al. Plant Cell* **21**, 3421 (2009).

[283] B. Khraiwesh, *et al. Cell* **140**, 111 (2010).

[284] E. Bonnet, J. Wuyts, P. Rouzé, Y. Van de Peer. *Proc Natl Acad Sci U S A* **101**, 11511 (2004).

[285] X.-J. Wang, J. L. Reyes, N.-H. Chua, T. Gaasterland. *Genome Biol* **5**, R65 (2004).

[286] B. Zhang, X. Pan, T. A. Anderson. *FEBS Lett* **580**, 3753 (2006).

[287] F. L. Xie, *et al. FEBS Lett* **581**, 1464 (2007).

[288] T. Dezulian, *et al. Bioinformatics* **22**, 359 (2006).

[289] X. Li, Y.-Z. Zhang. *Comput Biol Chem* **29**, 360 (2005).

[290] M. Lindow, A. Krogh. *BMC Genomics* **6**, 119 (2005).

[291] A. Adai, *et al. Genome Res* **15**, 78 (2005).

[292] M. Lindow, *et al. PLoS Comput Biol* **3**, e238 (2007).

[293] B. C. Meyers, M. Matzke, V. Sundaresan. *Trends Plant Sci* **13**, 311 (2008).

[294] K. D. Kasschau, *et al. Dev Cell* **4**, 205 (2003).

[295] R. Schwab, *et al. Dev Cell* **8**, 517 (2005).

[296] M. Zuker. *Nucleic Acids Res* **31**, 3406 (2003).

[297] I. L. Hofacker. *Nucleic Acids Res* **31**, 3429 (2003).

[298] L. Alves, Jr, *et al. Nucleic Acids Res* **37**, 4010 (2009).

[299] E. Bonnet, Y. He, K. Billiau, Y. Van de Peer. *Bioinformatics* **26**, 1566 (2010).

[300] B. H. Zhang, *et al. Cell Res* **15**, 336 (2005).

[301] W. Mhuantong, D. Wichadakul. *BMC Genomics* **10**, 366 (2009).

[302] M. A. German, *et al. Nat Biotechnol* **26**, 941 (2008).

[303] C. Addo-Quaye, T. W. Eshoo, D. P. Bartel, M. J. Axtell. *Curr Biol* **18**, 758 (2008).

[304] C. Addo-Quaye, W. Miller, M. J. Axtell. *Bioinformatics* **25**, 130 (2009).

[305] N. D. Mendes, A. T. Freitas, M.-F. Sagot. *Nucleic Acids Res* **37**, 2419 (2009).

[306] M. D. Howell, *et al. Plant Cell* **19**, 926 (2007).

[307] S. E. Heisel, *et al. PLoS One* **3**, e2871 (2008).

[308] S. Lu, Y.-H. Sun, H. Amerson, V. L. Chiang. *Plant J* **51**, 1077 (2007).

[309] M. Talmor-Neiman, *et al. Plant J* **48**, 511 (2006).

[310] T. Elmayan, *et al. FEBS J* **276**, 835 (2009).

[311] N. Kumakura, *et al. FEBS Lett* **583**, 1261 (2009).

[312] L. Vermeersch, N. De Winne, A. Depicker. *Plant J* **64**, 392 (2010).

[313] P. Hoffer, *et al. Proc Natl Acad Sci U S A* **108**, 409 (2011).

[314] Z. Luo, Z. Chen. *Plant Cell* **19**, 943 (2007).

[315] I. Hernandez-Pinzon, *et al. Plant J* **50**, 140 (2007).

[316] X. Adenot, *et al. Curr Biol* **16**, 927 (2006).

[317] Y. Nakazawa, A. Hiraguri, H. Moriyama, T. Fukuhara. *Plant Mol Biol* **63**, 777 (2007).

[318] G. Haas, *et al. EMBO J* **27**, 2102 (2008).

[319] H.-M. Chen, Y.-H. Li, S.-H. Wu. *Proc Natl Acad Sci U S A* **104**, 3318 (2007).

[320] K. Strässer, *et al. Nature* **417**, 304 (2002).

[321] A. G. Rondón, S. Jimeno, A. Aguilera. *Biochim Biophys Acta* **1799**, 533 (2010).

[322] N. E. Yelina, *et al.* *Proc Natl Acad Sci U S A* **107**, 13948 (2010).

[323] C. Furumizu, H. Tsukaya, Y. Komeda. *RNA* **16**, 1809 (2010).

[324] E. Allen, M. D. Howell. *Semin Cell Dev Biol* **21**, 798 (2010).

[325] F. F. Felippes, D. Weigel. *EMBO Rep* **10**, 264 (2009).

[326] L. Williams, C. C. Carles, K. S. Osmont, J. C. Fletcher. *Proc Natl Acad Sci U S A* **102**, 9703 (2005).

[327] C. Lurin, *et al.* *Plant Cell* **16**, 2089 (2004).

[328] N. O'Toole, *et al.* *Mol Biol Evol* **25**, 1120 (2008).

[329] M. S. Krasnikova, *et al.* *J Biomed Biotechnol* **2009**, 952304 (2009).

[330] M. Gu, C. D. Lima. *Curr Opin Struct Biol* **15**, 99 (2005).

[331] D. H. Chitwood, *et al.* *Genes Dev* **23**, 549 (2009).

[332] M. A. Matzke, J. A. Birchler. *Nat Rev Genet* **6**, 24 (2005).

[333] R. A. Mosher, F. Schwach, D. Studholme, D. C. Baulcombe. *Proc Natl Acad Sci U S A* **105**, 3145 (2008).

[334] M. Wassenegger, S. Heimes, L. Riedel, H. L. Sänger. *Cell* **76**, 567 (1994).

[335] C. S. Pikaard, J. R. Haag, T. Ream, A. T. Wierzbicki. *Trends Plant Sci* **13**, 390 (2008).

[336] M. Matzke, *et al.* *Curr Opin Cell Biol* **21**, 367 (2009).

[337] G. N. Rudenko, A. Ono, V. Walbot. *Plant J* **33**, 1013 (2003).

[338] B. J. Reinhart, D. P. Bartel. *Science* **297**, 1831 (2002).

[339] K. Nobuta, *et al.* *Proc Natl Acad Sci U S A* **105**, 14958 (2008).

[340] E. V. Dolgosheina, *et al.* *RNA* **14**, 1508 (2008).

[341] J. A. Casas-Mollano, *et al.* *Genetics* **179**, 69 (2008).

[342] M. Matzke, *et al.* *Curr Opin Plant Biol* **10**, 512 (2007).

[343] D. Zilberman, *et al.* *Nat Genet* **39**, 61 (2007).

[344] S. Jensen, M. P. Gassama, T. Heidmann. *Nat Genet* **21**, 209 (1999).

[345] D. Zilberman, X. Cao, S. E. Jacobsen. *Science* **299**, 716 (2003).

[346] B. Huettel, *et al.* *EMBO J* **25**, 2828 (2006).

[347] J. Liu, Y. He, R. Amasino, X. Chen. *Genes Dev* **18**, 2873 (2004).

[348] W. J. Soppe, *et al.* *Mol Cell* **6**, 791 (2000).

[349] S. W.-L. Chan, X. Zhang, Y. V. Bernatavichute, S. E. Jacobsen. *PLoS Biol* **4**, e363 (2006).

[350] I. R. Henderson, S. E. Jacobsen. *Genes Dev* **22**, 1597 (2008).

[351] A. M. Lindroth, *et al.* *Science* **292**, 2077 (2001).

[352] J. P. Jackson, A. M. Lindroth, X. Cao, S. E. Jacobsen. *Nature* **416**, 556 (2002).

[353] J. Vrbsky, *et al.* *PLoS Genet* **6**, e1000986 (2010).

[354] A. A. Aravin, G. J. Hannon, J. Brennecke. *Science* **318**, 761 (2007).

[355] M. A. Carmell, *et al.* *Dev Cell* **12**, 503 (2007).

[356] A. T. Wierzbicki, J. R. Haag, C. S. Pikaard. *Cell* **135**, 635 (2008).

[357] A. T. Wierzbicki, T. S. Ream, J. R. Haag, C. S. Pikaard. *Nat Genet* **41**, 630 (2009).

[358] L. Daxinger, T. Kanno, M. Matzke. *Cell* **135**, 592 (2008).

[359] Arabidopsis Genome Initiative. *Nature* **408**, 796 (2000).

[360] Y. Onodera, *et al.* *Cell* **120**, 613 (2005).

[361] N. A. Woychik, M. Hampsey. *Cell* **108**, 453 (2002).

[362] I. Grummt. *Genes Dev* **17**, 1691 (2003).

[363] L. Schramm, N. Hernandez. *Genes Dev* **16**, 2593 (2002).

[364] S. Lahmy, *et al.* *Proc Natl Acad Sci U S A* **106**, 941 (2009).

[365] L. Huang, *et al. Nat Struct Mol Biol* **16**, 91 (2009).

[366] T. S. Ream, *et al. Mol Cell* **33**, 192 (2009).

[367] M. Bellaoui, J. S. Keddie, W. Gruissem. *Plant Mol Biol* **53**, 531 (2003).

[368] S. Lahmy, *et al. Plant J* **39**, 809 (2004).

[369] C. F. Li, *et al. Cell* **126**, 93 (2006).

[370] M. El-Shami, *et al. Genes Dev* **21**, 2539 (2007).

[371] J. R. Haag, O. Pontes, C. S. Pikaard. *PLoS One* **4**, e4110 (2009).

[372] X.-J. He, *et al. Genes Dev* **23**, 318 (2009).

[373] T. Kanno, *et al. EMBO Rep* **11**, 65 (2010).

[374] O. Pontes, P. Costa-Nunes, P. Vithayathil, C. S. Pikaard. *Mol Plant* **2**, 700 (2009).

[375] J. A. Law, S. E. Jacobsen. *Nat Rev Genet* **11**, 204 (2010).

[376] A. J. Herr, M. B. Jensen, T. Dalmay, D. C. Baulcombe. *Science* **308**, 118 (2005).

[377] O. Pontes, *et al. Cell* **126**, 79 (2006).

[378] B. Zheng, *et al. Genes Dev* **23**, 2850 (2009).

[379] L. M. Smith, *et al. Plant Cell* **19**, 1507 (2007).

[380] T. Kanno, *et al. Nat Genet* **37**, 761 (2005).

[381] D. Pontier, *et al. Genes Dev* **19**, 2030 (2005).

[382] X. Zhang, *et al. Proc Natl Acad Sci U S A* **104**, 4536 (2007).

[383] O. Pontes, C. S. Pikaard. *Curr Opin Genet Dev* **18**, 197 (2008).

[384] C. F. Li, *et al. PLoS Genet* **4**, e27 (2008).

[385] J. A. Law, *et al. Curr Biol* **20**, 951 (2010).

[386] T. Kanno, *et al. Curr Biol* **14**, 801 (2004).

[387] T. Kanno, *et al. Nat Genet* **40**, 670 (2008).

[388] S. T. M. Allard, *et al. Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**, 647 (2005).

[389] N. Bies-Etheve, *et al. EMBO Rep* **10**, 649 (2009).

[390] X.-J. He, *et al. Cell* **137**, 498 (2009).

[391] I. Ausin, T. C. Mockler, J. Chory, S. E. Jacobsen. *Nat Struct Mol Biol* **16**, 1325 (2009).

[392] Z. Zheng, *et al. Plant J* **62**, 92 (2010).

[393] Z. Gao, *et al. Nature* **456**, 106 (2010).

[394] X.-J. He, *et al. Genes Dev* **23**, 2717 (2009).

[395] D. Zilberman, *et al. Curr Biol* **14**, 1214 (2004).

[396] Z. Lippman, *et al. PLoS Biol* **1**, E67 (2003).

[397] R. Lister, *et al. Cell* **133**, 523 (2008).

[398] L. Daxinger, *et al. EMBO J* **28**, 48 (2009).

[399] S. J. Cokus, *et al. Nature* **452**, 215 (2008).

[400] E. J. Richards. *Curr Biol* **12**, R694 (2002).

[401] K. Naumann, *et al. EMBO J* **24**, 1418 (2005).

[402] X. Cao, *et al. Curr Biol* **13**, 2212 (2003).

[403] X. Cao, S. E. Jacobsen. *Curr Biol* **12**, 1138 (2002).

[404] J. Bender. *Annu Rev Plant Biol* **55**, 41 (2004).

[405] S. W.-L. Chan, *et al. Science* **303**, 1336 (2004).

[406] R. K. Tran, *et al. Genome Biol* **6**, R90 (2005).

[407] Z. Gong, *et al. Cell* **111**, 803 (2002).

[408] F. Agius, A. Kapoor, J.-K. Zhu. *Proc Natl Acad Sci U S A* **103**, 11796 (2006).

[409] T. Kanno, *et al. EMBO Rep* **6**, 649 (2005).

[410] X. Zheng, *et al. Nature* **455**, 1259 (2008).

[411] J. Zhu, *et al. Curr Biol* **17**, 54 (2007).

[412] Z. Lippman, R. Martienssen. *Nature* **431**, 364 (2004).

[413] D. Jiang, W. Yang, Y. He, R. M. Amasino. *Plant Cell* **19**, 2975 (2007).

[414] S. B. Preuss, *et al. Mol Cell* **32**, 673 (2008).

[415] J. A. Jeddeloh, T. L. Stokes, E. J. Richards. *Nat Genet* **22**, 94 (1999).

[416] C. J. Hale, K. F. Erhard, Jr, D. Lisch, J. B. Hollick. *PLoS Genet* **5**, e1000598 (2009).

[417] F. Malagnac, L. Bartee, J. Bender. *EMBO J* **21**, 6842 (2002).

[418] A. M. Lindroth, *et al. EMBO J* **23**, 4286 (2004).

[419] M. L. Ebbs, L. Bartee, J. Bender. *Mol Cell Biol* **25**, 10507 (2005).

[420] M. L. Ebbs, J. Bender. *Plant Cell* **18**, 1166 (2006).

[421] Y. V. Bernatavichute, *et al. PLoS One* **3**, e3156 (2008).

[422] A.-V. Gendrel, *et al. Science* **297**, 1871 (2002).

[423] H. Saze, A. Shiraishi, A. Miura, T. Kakutani. *Science* **319**, 462 (2008).

[424] L. M. Johnson, *et al. PLoS Genet* **4**, e1000280 (2008).

[425] V. V. Sridhar, *et al. Nature* **447**, 735 (2007).

[426] Y. Jacob, *et al. Nat Struct Mol Biol* **16**, 763 (2009).

[427] H. Numa, *et al. EMBO J* **29**, 352 (2010).

[428] C. Yokthongwattana, *et al. EMBO J* **29**, 340 (2010).

[429] A. T. Wierzbicki. *EMBO J* **29**, 279 (2010).

[430] K. W. Earley, *et al. Genes Dev* **24**, 1119 (2010).

[431] R. A. Mosher. *New Phytol* **186**, 358 (2010).

[432] R. A. Mosher, C. W. Melnyk. *Trends Plant Sci* **15**, 204 (2010).

[433] T.-F. Hsieh, *et al. Science* **324**, 1451 (2009).

[434] M. Gehring, K. L. Bubb, S. Henikoff. *Science* **324**, 1447 (2009).

[435] R. A. Mosher, *et al. Nature* **460**, 283 (2009).

[436] R. K. Slotkin, *et al. Cell* **136**, 461 (2009).

[437] A. Werner. *RNA Biol* **2**, 53 (2005).

[438] X.-J. Wang, T. Gaasterland, N.-H. Chua. *Genome Biol* **6**, R30 (2005).

[439] C.-H. Jen, I. Michalopoulos, D. R. Westhead, P. Meyer. *Genome Biol* **6**, R51 (2005).

[440] S. H. Munroe, J. Zhu. *Cell Mol Life Sci* **63**, 2102 (2006).

[441] B. Lehner, G. Williams, R. D. Campbell, C. M. Sanderson. *Trends Genet* **18**, 63 (2002).

[442] N. Osato, *et al. Genome Biol* **5**, R5 (2003).

[443] R. Yelin, *et al. Nat Biotechnol* **21**, 379 (2003).

[444] B. C. Meyers, *et al. Nat Biotechnol* **22**, 1006 (2004).

[445] K. Numata, *et al. Gene* **392**, 134 (2007).

[446] S. Katiyar-Agarwal, S. Gao, A. Vivian-Smith, H. Jin. *Genes Dev* **21**, 3123 (2007).

[447] F. Liu, *et al. Science* **327**, 94 (2010).

[448] J. Georg, *et al. New Phytol* **186**, 615 (2010).

[449] H. Wang, N.-H. Chua, X.-J. Wang. *Genome Biol* **7**, R92 (2006).

[450] X. Zhou, *et al. Genome Res* **19**, 70 (2009).

[451] S. R. Henz, *et al. Plant Physiol* **144**, 1247 (2007).

[452] H. Jin, *et al. BMC Mol Biol* **9**, 6 (2008).

[453] M. Ron, *et al. Genes Dev* **24**, 1010 (2010).

[454] M. A. Matzke, M. Primig, J. Trnovsky, A. J. Matzke. *EMBO J* **8**, 643 (1989).

[455] F. Linn, I. Heidmann, H. Saedler, P. Meyer. *Mol Gen Genet* **222**, 329 (1990).

[456] C. Napoli, C. Lemieux, R. Jorgensen. *Plant Cell* **2**, 279 (1990).

[457] C. J. Smith, *et al. Mol Gen Genet* **224**, 477 (1990).

[458] A. R. van der Krol, *et al. Plant Mol Biol* **14**, 457 (1990).

[459] P. Meyer, H. Saedler. *Annu Rev Plant Physiol Plant Mol Biol* **47**, 23 (1996).

[460] J. A. Lindbo, L. Silva-Rosales, W. M. Proebsting, W. G. Dougherty. *Plant Cell* **5**, 1749 (1993).

[461] J. Goodwin, *et al. Plant Cell* **8**, 95 (1996).

[462] J. J. English, E. Mueller, D. C. Baulcombe. *Plant Cell* **8**, 179 (1996).

[463] D. C. Baulcombe. *Plant Cell* **8**, 1833 (1996).

[464] M. H. Kumagai, *et al. Proc Natl Acad Sci U S A* **92**, 1679 (1995).

[465] M. Ruiz, O. Voinnet, D. Baulcombe. *Plant Cell* **10**, 937 (1998).

[466] F. Ratcliff, B. D. Harrison, D. C. Baulcombe. *Science* **276**, 1558 (1997).

[467] S. N. Covey, N. S. Al-Kaff, L. Amagoia, D. S. Turner. *Nature* **385**, 781 (1997).

[468] J. C. Palauqui, *et al. Plant Physiol* **112**, 1447 (1996).

[469] J. C. Palauqui, T. Elmayan, J. M. Pollien, H. Vaucheret. *EMBO J* **16**, 4738 (1997).

[470] F. Li, S.-W. Ding. *Annu Rev Microbiol* **60**, 503 (2006).

[471] S.-W. Ding, O. Voinnet. *Cell* **130**, 413 (2007).

[472] X. Zhang, *et al. Genes Dev* **20**, 3255 (2006).

[473] N. Baumberger, *et al. Curr Biol* **17**, 1609 (2007).

[474] D. Bortolamiol, *et al. Curr Biol* **17**, 1615 (2007).

[475] L. Lakatos, *et al. EMBO J* **25**, 2768 (2006).

[476] P. Raja, B. C. Sanville, R. C. Buchmann, D. M. Bisaro. *J Virol* **82**, 8997 (2008).

[477] M.-B. Wang, M. Metzlaff. *Curr Opin Plant Biol* **8**, 216 (2005).

[478] G. Moissiard, O. Voinnet. *Proc Natl Acad Sci U S A* **103**, 19593 (2006).

[479] C. Beclin, S. Boutet, P. Waterhouse, H. Vaucheret. *Curr Biol* **12**, 684 (2002).

[480] F. Schwach, F. E. Vaistij, L. Jones, D. C. Baulcombe. *Plant Physiol* **138**, 1842 (2005).

[481] A. Depicker, M. V. Montagu. *Curr Opin Cell Biol* **9**, 373 (1997).

[482] Kooter, Matzke, Meyer. *Trends Plant Sci* **4**, 340 (1999).

[483] M. Fagard, H. Vaucheret. *Annu Rev Plant Physiol Plant Mol Biol* **51**, 167 (2000).

[484] H. Vaucheret, C. Béclin, M. Fagard. *J Cell Sci* **114**, 3083 (2001).

[485] T. Dalmay, R. Horsefield, T. H. Braunstein, D. C. Baulcombe. *EMBO J* **20**, 2069 (2001).

[486] P. Brodersen, O. Voinnet. *Trends Genet* **22**, 268 (2006).

[487] R. Fukunaga, J. A. Doudna. *EMBO J* **28**, 545 (2009).

[488] A. P. Wolffe, M. A. Matzke. *Science* **286**, 481 (1999).

[489] M. A. Matzke, A. J. Matzke, G. J. Pruss, V. B. Vance. *Curr Opin Genet Dev* **11**, 221 (2001).

[490] H. Vaucheret, M. Fagard. *Trends Genet* **17**, 29 (2001).

[491] J. B. Morel, P. Mourrain, C. Béclin, H. Vaucheret. *Curr Biol* **10**, 1591 (2000).

[492] A. Eamens, M.-B. Wang, N. A. Smith, P. M. Waterhouse. *Plant Physiol* **147**, 456 (2008).

[493] B. O. Petersen, M. Albrechtsen. *Plant Mol Biol* **58**, 575 (2005).

[494] H. Saze, T. Kakutani. *EMBO J* **26**, 3641 (2007).

[495] W. Aufsatz, *et al. Proc Natl Acad Sci U S A* **99 Suppl 4**, 16499 (2002).

[496] S. Mlotshwa, *et al. PLoS ONE* **3**, e1755 (2008).

[497] K. Okamura, *et al. Nature* **453**, 803 (2008).

[498] P. Dunoyer, *et al. Nat Genet* **39**, 848 (2007).

[499] P. Dunoyer, *et al. Science* **328**, 912 (2010).

[500] M. Kusaba. *Curr Opin Biotechnol* **15**, 139 (2004).

[501] R. A. Jorgensen, *et al. Cold Spring Harb Symp Quant Biol* **71**, 481 (2006).

[502] P. M. Waterhouse, C. A. Helliwell. *Nat Rev Genet* **4**, 29 (2003).

[503] D. Robertson. *Annu Rev Plant Biol* **55**, 495 (2004).

[504] Y. Tang, *et al. Plant Physiol* **153**, 632 (2010).

[505] F. E. Vaistij, L. Elias, G. L. George, L. Jones. *Plant Mol Biol* **73**, 391 (2010).

[506] A. L. Jackson, *et al. Nat Biotechnol* **21**, 635 (2003).

[507] P. Xu, *et al. Plant Physiol* **142**, 429 (2006).

[508] M. D. Schüssler, *et al. Plant J* **56**, 756 (2008).

[509] E. A. Parizotto, *et al. Genes Dev* **18**, 2237 (2004).

[510] J. P. Alvarez, *et al. Plant Cell* **18**, 1134 (2006).

[511] S. Ossowski, R. Schwab, D. Weigel. *Plant J* **53**, 674 (2008).

[512] N. Warthmann, *et al. PLoS ONE* **3**, e1829 (2008).

[513] A. I. Fernandez, *et al. Plant Physiol* **151**, 1729 (2009).

[514] B. Khraiwesh, *et al. Plant Physiol* **148**, 684 (2008).

[515] T. Zhao, W. Wang, X. Bai, Y. Qi. *Plant J* **58**, 157 (2008).

[516] A. Molnar, *et al. Plant J* **58**, 165 (2009).

[517] M. de la Luz Gutierrez-Nava, *et al. Plant Physiol* **147**, 543 (2008).

[518] O. Voinnet, D. C. Baulcombe. *Nature* **389**, 553 (1997).

[519] O. Voinnet. *FEBS Lett* **579**, 5858 (2005).

[520] H. Vogler, *et al. PLoS Pathog* **4**, e1000038 (2008).

[521] K. Kobayashi, P. Zambryski. *Plant J* **50**, 597 (2007).

[522] A. Molnar, *et al. Science* **328**, 872 (2010).

[523] B. D. Pant, A. Buhtz, J. Kehr, W.-R. Scheible. *Plant J* **53**, 731 (2008).

[524] E. M. Tretter, J. P. Alvarez, Y. Eshed, J. L. Bowman. *Plant Physiol* **147**, 58 (2008).

[525] A. Carlsbecker, *et al. Nature* **465**, 316 (2010).

[526] F. T. S. Nogueira, *et al. PLoS Genet* **5**, e1000320 (2009).

[527] M. T. Juarez, *et al. Nature* **428**, 84 (2004).

[528] A. Válóczi, *et al. Plant J* **47**, 140 (2006).

[529] E. Varkonyi-Gasic, *et al. Plant Methods* **3**, 12 (2007).

[530] S.-I. Lin, *et al. Plant Physiol* **147**, 732 (2008).

[531] A. Buhtz, *et al. Plant J* **53**, 739 (2008).

[532] A. Buhtz, J. Pieritz, F. Springer, J. Kehr. *BMC Plant Biol* **10**, 64 (2010).

[533] S.-I. Lin, T.-J. Chiou. *Plant Signal Behav* **3**, 730 (2008).

[534] L.-C. Hsieh, *et al. Plant Physiol* **151**, 2120 (2009).

[535] C. A. Brosnan, *et al. Proc Natl Acad Sci U S A* **104**, 14741 (2007).

[536] R. Schwab, *et al. PLoS One* **4**, e5980 (2009).

[537] C. A. Kidner, R. A. Martienssen. *Nature* **428**, 81 (2004).

[538] X. Chen, J. Liu, Y. Cheng, D. Jia. *Development* **129**, 1085 (2002).

[539] C. Lu, N. Fedoroff. *Plant Cell* **12**, 2351 (2000).

[540] I. Rubio-Somoza, J. T. Cuperus, D. Weigel, J. C. Carrington. *Curr Opin Plant Biol* **12**, 622 (2009).

[541] R. Sunkar, J.-K. Zhu. *Plant Cell* **16**, 2001 (2004).

[542] X. Zhou, G. Wang, W. Zhang. *Mol Syst Biol* **3**, 103 (2007).

[543] H.-H. Liu, *et al. RNA* **14**, 836 (2008).

[544] X. Zhou, *et al. Biochim Biophys Acta* **1779**, 780 (2008).

[545] D. Moldovan, *et al. J Exp Bot* **61**, 165 (2010).

[546] W. W. Kong, Z. M. Yang. *Plant Physiol Biochem* **48**, 153 (2010).

[547] B. Zhao, *et al. Biochem Biophys Res Commun* **354**, 585 (2007).

[548] S. Q. Huang, J. Peng, C. X. Qiu, Z. M. Yang. *J Inorg Biochem* **103**, 282 (2009).

[549] X. Jian, *et al. Genomics* **95**, 47 (2010).

[550] D. Ding, *et al. Ann Bot* **103**, 29 (2009).

[551] T.-J. Chiou. *Plant Cell Environ* **30**, 323 (2007).

[552] L. Shukla, V. Chinnusamy, R. Sunkar. *Biochim Biophys Acta* **1779**, 743 (2008).

[553] X.-Y. Lu, X.-L. Huang. *Biochem Biophys Res Commun* **368**, 458 (2008).

[554] A. A. Covarrubias, J. L. Reyes. *Plant Cell Environ* **33**, 481 (2010).

[555] R. C. Martin, P.-P. Liu, N. A. Goloviznina, H. Nonogaki. *J Exp Bot* **61**, 2229 (2010).

[556] G. Chuck, D. O'Connor. *Curr Opin Plant Biol* **13**, 40 (2010).

[557] D. Szakonyi, A. Moschopoulos, M. E. Byrne. *J Plant Res* **123**, 281 (2010).

[558] J. F. Emery, *et al. Curr Biol* **13**, 1768 (2003).

[559] J. R. McConnell, M. K. Barton. *Development* **125**, 2935 (1998).

[560] J. R. McConnell, *et al. Nature* **411**, 709 (2001).

[561] R. Zhong, Z.-H. Ye. *Plant Cell Physiol* **45**, 369 (2004).

[562] I. Ochando, *et al. Plant Physiol* **141**, 607 (2006).

[563] J. Kim, *et al. Plant J* **42**, 84 (2005).

[564] D. Otsuga, *et al. Plant J* **25**, 223 (2001).

[565] M. J. Prigge, *et al. Plant Cell* **17**, 61 (2005).

[566] L. Williams, *et al. Development* **132**, 3657 (2005).

[567] X. Yao, *et al. FEBS Lett* **583**, 3711 (2009).

[568] F. T. S. Nogueira, A. K. Sarkar, D. H. Chitwood, M. C. P. Timmermans. *Cold Spring Harb Symp Quant Biol* **71**, 157 (2006).

[569] J.-I. Itoh, K.-I. Hibara, Y. Sato, Y. Nagato. *Plant Physiol* **147**, 1960 (2008).

[570] L. Xu, *et al. Plant Cell Physiol* **47**, 853 (2006).

[571] D. Garcia, S. A. Collier, M. E. Byrne, R. A. Martienssen. *Curr Biol* **16**, 933 (2006).

[572] E. Levine, P. McHale, H. Levine. *PLoS Comput Biol* **3**, e233 (2007).

[573] F. T. S. Nogueira, *et al. Genes Dev* **21**, 750 (2007).

[574] J. Wang, *et al. J Exp Bot* **61**, 1885 (2010).

[575] F. T. Nogueira, M. C. Timmermans. *Plant Signal Behav* **2**, 519 (2007).

[576] E. K. Yoon, *et al. Nucleic Acids Res* **38**, 1382 (2010).

[577] M. Aida, *et al. Plant Cell* **9**, 841 (1997).

[578] A. C. Mallory, D. V. Dugas, D. P. Bartel, B. Bartel. *Curr Biol* **14**, 1035 (2004).

[579] P. Sieber, *et al. Development* **134**, 1051 (2007).

[580] C. Baker, P. Sieber, F. Wellmer, E. Meyerowitz. *Current Biology* **15**, 303 (2005).

[581] K. Hibara, S. Takada, M. Tasaka. *Plant Journal* **36**, 687 (2003).

[582] P. Laufs, A. Peaucelle, H. Morin, J. Traas. *Development* **131**, 4311 (2004).

[583] C. T. Larue, J. Wen, J. C. Walker. *Plant J* **58**, 450 (2009).

[584] K. Nikovics, *et al. Plant Cell* **18**, 2929 (2006).

[585] Y. Berger, *et al. Development* **136**, 823 (2009).

[586] A. Brand, N. Shirding, S. Shleizer, N. Ori. *Planta* **226**, 941 (2007).

[587] A. Peaucelle, H. Morin, J. Traas, P. Laufs. *Development* **134**, 1045 (2007).

[588] S. Raman, *et al. Plant J* **55**, 65 (2008).

[589] H.-S. Guo, Q. Xie, J.-F. Fei, N.-H. Chua. *Plant Cell* **17**, 1376 (2005).

[590] J. H. Kim, *et al. Science* **323**, 1053 (2009).

[591] S. A. Oh, *et al. Plant J* **12**, 527 (1997).

[592] J. F. Palatnik, *et al. Nature* **425**, 257 (2003).

[593] C. Schommer, *et al. PLoS Biol* **6**, e230 (2008).

[594] Y. He, H. Fukushige, D. F. Hildebrand, S. Gan. *Plant Physiol* **128**, 876 (2002).

[595] A. Swiatek, *et al. FEBS Lett* **572**, 118 (2004).

[596] L. Pauwels, *et al. Proc Natl Acad Sci U S A* **105**, 1380 (2008).

[597] T. Koyama, M. Furutani, M. Tasaka, M. Ohme-Takagi. *Plant Cell* **19**, 473 (2007).

[598] N. Ori, *et al. Nat Genet* **39**, 787 (2007).

[599] J. H. Kim, D. Choi, H. Kende. *Plant J* **36**, 94 (2003).

[600] D. Liu, Y. Song, Z. Chen, D. Yu. *Physiol Plant* **136**, 223 (2009).

[601] P. M. Donnelly, *et al. Dev Biol* **215**, 407 (1999).

[602] R. E. Rodriguez, *et al. Development* **137**, 103 (2010).

[603] G. Horiguchi, G.-T. Kim, H. Tsukaya. *Plant J* **43**, 68 (2005).

[604] M. M. Alonso-Peral, *et al. Plant Physiol* **154**, 757 (2010).

[605] R. S. Poethig. *Science* **301**, 334 (2003).

[606] A. Telfer, K. M. Bollman, R. S. Poethig. *Development* **124**, 645 (1997).

[607] H. Tsukaya, K. Shoda, G. T. Kim, H. Uchimiya. *Planta* **210**, 536 (2000).

[608] R. S. Poethig. *Curr Opin Genet Dev* **19**, 374 (2009).

[609] M. R. Willmann, R. S. Poethig. *Curr Opin Plant Biol* **8**, 548 (2005).

[610] G. Chuck, *et al. Nat Genet* **39**, 1517 (2007).

[611] Y. Jiao, *et al. Nat Genet* **42**, 541 (2010).

[612] J. H. Clarke, *et al. Plant J* **20**, 493 (1999).

[613] M. J. Prigge, D. R. Wagner. *Plant Cell* **13**, 1263 (2001).

[614] G. Wu, R. S. Poethig. *Development* **133**, 3539 (2006).

[615] S. Schwarz, *et al. Plant Mol Biol* **67**, 183 (2008).

[616] J.-W. Wang, *et al. Plant Cell* **20**, 1231 (2008).

[617] M. J. Aukerman, H. Sakai. *Plant Cell* **15**, 2730 (2003).

[618] X. Chen. *Science* **303**, 2022 (2004).

[619] J.-H. Jung, *et al. Plant Cell* **19**, 2736 (2007).

[620] N. Lauter, *et al. Proc Natl Acad Sci U S A* **102**, 9412 (2005).

[621] C. Hunter, H. Sun, R. S. Poethig. *Curr Biol* **13**, 1734 (2003).

[622] C. Hunter, *et al. Development* **133**, 2973 (2006).

[623] T. Usami, G. Horiguchi, S. Yano, H. Tsukaya. *Development* **136**, 955 (2009).

[624] N. Fahlgren, *et al. Curr Biol* **16**, 939 (2006).

[625] J. L. Reyes, N.-H. Chua. *Plant J* **49**, 592 (2007).

[626] T. J. Guilfoyle, G. Hagen. *Curr Opin Plant Biol* **10**, 453 (2007).

[627] A. C. Mallory, D. P. Bartel, B. Bartel. *Plant Cell* **17**, 1360 (2005).

[628] J.-W. Wang, *et al. Plant Cell* **17**, 2204 (2005).

[629] P.-P. Liu, *et al. Plant J* **52**, 133 (2007).

[630] M.-F. Wu, Q. Tian, J. W. Reed. *Development* **133**, 4211 (2006).

[631] P. Ru, L. Xu, H. Ma, H. Huang. *Cell Res* **16**, 457 (2006).

[632] Q. Xie, *et al. Nature* **419**, 167 (2002).

[633] I. Ochando, *et al. Int J Dev Biol* **52**, 953 (2008).

[634] L. Navarro, *et al. Science* **312**, 436 (2006).

[635] H. Jin. *FEBS Lett* **582**, 2679 (2008).

[636] Y. Li, *et al. Plant Physiol* **152**, 2222 (2010).

[637] D. MacLean, *et al. PLoS One* **5**, e9901 (2010).

5

Genome-wide small RNA profiling during leaf development of *Arabidopsis*

Frederik Coppens[1,2], Kristin D. Kasschau[3,4], Eric Bonnet[1,2], Ramakrishna K. Chodavarapu[5], Suhua Feng[5,6], Matteo Pellegrini[5,7], Steven E. Jacobsen[5,6,7], Dirk Inzé[1,2], James C. Carrington[3,4], Gerrit T.S. Beemster[1,2,8]

[1]Department of Plant Systems Biology, Flanders Institute for Biotechnology (VIB), 9052 Ghent, Belgium
[2]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Ghent, Belgium
[3]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA
[4]Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331, USA
[5]Department of Molecular, Cell, and Developmental Biology, University of California Los Angeles, Los Angeles, CA 90095, USA
[6]Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, CA 90095, USA
[7]Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA 90095, USA
[8]Department of Biology, University of Antwerp, 2020 Antwerp, Belgium

Summary

The transition of cells from proliferation to expansion and subsequently to maturity determines the growth of multicellular organs. These transitions are closely associated with large-scale transcriptional reprogramming. During the last decade, the importance of small RNA (sRNA) molecules in regulation of developmental processes became clear and has been extensively studied, in plants particularly at the level of entire seedlings and inflorescence tissues. To unravel their role in organ growth, we profiled sRNAs by deep sequencing during the three developmental stages of leaf development in *Arabidopsis thaliana*: proliferation, expansion and maturity, which we used before for mRNA profiling.

We found that microRNAs (miRNAs) involved in nutrient stress response, such as phosphate and sulfate limitation, are specifically upregulated during proliferation, suggesting that during early phases of development these nutrients are limiting for growth. Next to known mature miRNAs, we also used the sequencing data to study the extent of small RNA biogenesis from miRNA hairpins and to discover new, differentially expressed miRNAs. This analysis revealed that a diverse small RNA population is generated from the region around the mature miRNA. In some cases such an alternative sRNA was expressed much higher compared to the mature miRNA.

We also studied differential sRNA expression at known, annotated features (based on TAIR8). Unlike what we expected, we found a positive correlation between mRNA and sRNA expression for protein-coding genes. Only for about 5% of the protein-coding genes, this positive correlation did not occur: while the mRNA is low expressed, sRNA expression is relatively high. We could show that a subset of these 'uncorrelated' genes is targeted by *trans*-acting siRNAs (ta-siRNAs) or miRNAs, triggering ta-siRNA production from these regions. The remaining genes were specifically associated with 24-nt siRNAs. This class of small RNAs is involved in *cis*-regulatory DNA methylation, especially in CHH sequence context. A genome-wide survey of DNA methylation during the three developmental stages confirmed that the 24-nt associated, uncorrelated genes are indeed methylated to a higher degree in all sequence contexts. The level of methylation increased during development in CHH context, while it decreased for CG and CHG.

Small RNAs matching transposons are generally 24-nt long and their expression increases during development. We could also confirm that transposon methylation in CHH context increases as the leaf ages, while CG and CHG methylation decreases.

In conclusion, our data confirm an important role for small RNA molecules in leaf development. Especially 24-nt siRNAs associated with DNA methylation appear to constitute an important regulatory layer, which has not been addressed previously in organ development.

5.1 Introduction

5.1.1 Regulation of shoot development

Regulation of developmental processes is crucial in multicellular organisms requiring the simultaneous onset or inactivation of a complex network of genes. Both transcription factors (TF) and microRNAs (miRNAs) are able to regulate the expression of different genes or gene families and are therefore ideally suited to be at the core of developmental regulatory processes. In plants it has become clear that a disproportionate number of targets of miRNAs is involved in development (1). Moreover, feedback mechanisms have been reported between miRNAs and their targets, thereby forming a layer of regulation that enables finetuning of transcriptional control and provides robustness to the regulatory system (2, 3, 4).

Leaves produced by the shoot can be distinguished based on size, shape and trichome distribution (5, 6, 7, 8). The first leaf pair is small, flat and round, has a smooth margin and only trichomes on the adaxial surface. Subsequent leaves become progressively larger, are oval, curl downwards, show serrations, have relatively short petioles, a higher number of hydathodes and can produce both abaxial and adaxial trichomes. These morphological changes in subsequent leaves are referred to as heteroblasty. This phase transition from juvenile traits in the first leaf pair to adult characteristics in subsequent leaves, is regulated by two miRNAs (miR156 and miR172) that act antagonistically. While miR156 is necessary and sufficient for juvenile characteristics (2, 8), miR172 stimulates adult traits (9, 10, 11). A second regulatory mechanism of this phase change involves *trans*-acting short interfering RNAs (ta-siRNAs) generated by the *TAS3* locus (12, 13, 14, 15, 16). The transcript of the *TAS3* locus is cleaved by miR390 and this induces the generation of phased siRNAs: the RNaseIII-like enzyme DCL4 cleaves iteratively 21-nt ta-siRNAs, starting from the miRNA cleavage site (for more information about this pathway I refer to Chapter 4.3.2). These ta-siRNAs target *AUXIN RESPONSE FACTOR 3* (*ARF3*)/*ETTIN* and *ARF4* that regulate heteroblasty in *Arabidopsis* (17, 18).

5.1.2 Regulation of transitions in leaf development

Our previous research has made it clear that developmental changes during leaf development are associated with genome-wide changes in gene expression patterns (19). Leaf development can be subdivided in three phases. After a leaf primordium is initiated, cells proliferate, cell division accompanied by cell expansion to maintain cell size homeostasis. Next, cell division ceases in a tip-to-base gradient (20) and cells expand rapidly. When cell size is stabilizing, the leaf reaches its mature phase (19, 21). Despite tissue heterogeneity, expression profiles in entire leaf material from these three developmental stages closely reflects the developmental phases. The enrichment of core cell cycle genes in the proliferation phase, validated this ap-

proach and allowed us to identify unknown cell cycle related genes (19, Chapter 3). However, the microarray analysis using Affymetrix ATH1 only allows profiling of known, annotated genes.

Small RNAs are also involved in the regulation of maturation of the leaf. Upon overexpression of miR319a (*jaw-D* mutant) the proliferation phase is prolonged, due to downregulation of *TEOSINTE BRANCHED/CYCLOIDEA/PCF* (*TCP*) target genes that are involved in the regulation of the transition from proliferation phase to expansion phase (22). *GROWTH-REGULATING FACTOR* (*GRF*) TFs are know regulators of leaf growth (23, 24, 25). The majority of the *GRF*s are targeted by miR396 and upon overexpression of this miRNA the levels of mitotic genes decrease and cell proliferation is reduced (25, 26). It has been recently shown that cell number is quantitatively regulated by the balance between miR396 and *GRF*s (26). These two regulatory modules, miR319a/*TCP*s and miR396/*GRF*s, are linked as TCP4, a miR319a target, promotes miR396 expression (26).

5.1.3   Temporal changes in silencing in plants

Temporal variations in gene silencing have been observed for transposons in maize (27, 28, 29). An inactive *Suppressor*-mutator (*Spm*) transposable element can be re-activated during development and the likelihood of reactivation increases with the age of the plant (28). It has been suggested that this is similarly controlled as the developmental switch from juvenile-to-adult phase (27). For transposable elements from the Robertson's *Mutator* (*Mu*), DNA methylation and concomitant suppression of the mutant phenotype were shown to be clonally inherited. The proportion of wild type tissue (*Mu* element is methylated) increased in sectored plants in each successive leaf, with total loss of mutant tissue typically occurring in the seventh leaf. Also here, a change in methylation characteristics occurs during the development of maize (29, 30). A gradual nuclear (instead of cytoplasmatic) accumulation of transposon-derived RNA during development of maize, was found to correlate with the initiation of silencing (31).

Elmayan and Vaucheret (32) observed mitotically heritable but meiotically reversible silencing of a transgene. Every generation, the transgene was progressively silenced throughout development and during meiosis this was reset. Transgenic plants could be classified into two classes. The first class shows a rapid decline with complete silencing after one month of growth, while the other class has a more gradual decline over 4 months in homozygotes and more than one year in heterozygotes. Progressive silencing during development and resetting in the progeny has been found often with transgenics (33, 34, 35, 36, 37, 38) but also onset later in development (39, 40) or in specific tissue that have high homologous endogenous gene expression (41, 42, 43) has been described.

The inactivation of *Spm* in maize is correlated with DNA methylation in the $5'$ region (27). This is also the case for other transposable elements, such as *Mu*, and

Activator (*Ac*) (29). In general terms, gene silencing gets stronger as development progresses, which has been associated with an increase in DNA methylation in maize (28, 29, 30, 32, 33, 44, 45).

A link between DNA methylation and transgene silencing was also found in *Arabidopsis* (46, 47, 48). Silencing through methylation of transgenes and homologous sequences was correlated to the expression of small RNA molecules of circa 24-nt in length (49). This heterochromatic siRNA biogenesis pathway has since been studied thoroughly and siRNAs have been shown to induce DNA methylation. This silencing mechanism is important for genome integrity as it represses the expression of transposable elements but is also involved in development, plant defense and possibly stress (50, 51, 52, 53, 54, 55, 56, 57). For more details about this biogenesis pathway, I refer to Chapter 4.3.3.

During organ development, cells mature in a largely irreversible way, suggesting a potential role for methylation in the developmental control. To date however, methylation has not been extensively studied in this context.

5.1.4 Research objectives

To examine the potential roles of small RNAs in organ development, we profiled small RNAs for the three stages of leaf development: proliferation, expansion and maturity. We focused on small RNAs that are differentially expressed during leaf development. Next to the analysis of known microRNAs, we also predicted a new miRNA candidate gene. We studied the expression of ta-siRNA generating loci throughout development as well as potential new sites of phased sRNA production. Finally, we investigated the potential role of small RNAs on gene silencing through guidance of DNA methylation, we analyzed the small RNA expression profiles for protein-coding genes and transposons and compared those to DNA methylation patterns.

5.2 Results

5.2.1 Sequence analysis of small RNAs during leaf development

Small RNA (sRNA) libraries from the three developmental stages of leaf development were sequenced to determine to what extent biogenesis of sRNAs changes during development of this organ. To this end the first true leaf pair of *Arabidopsis thaliana* ecotype Columbia was used as a model system. The three distinct developmental stages of leaf development were profiled - proliferation, expansion and mature - each in two replicate libraries from independent pools of plants. Deep sequencing using Illumina technology yielded in total more than 60 million reads (Table 5.1). From these, only sequences of 18 to 30 bp in length for which the se-

quence contained the first 6 bases of the $3'$ adapter were retained. On average 52.6% of the sequences fulfilled both criteria. The obtained sRNA sequences were then mapped to the *Arabidopsis* genome based on perfect matching. Depending on the sample, 25.3 to 34.2% of the sequences could be mapped to the genome (Table 5.1). A total of 1,236,369 unique reads were sequenced, of which 71% was represented by a single read. Assuming stable levels of total sRNAs, samples were normalized for population size to 3 million mapped reads per sample, corresponding roughly to the average amount of mapped reads across all samples. Then the small RNA data was repeat-normalized by dividing the number of reads for a sequence by the number of perfect matches of that sequence to the genome. This addresses the bias towards highly repetitive sequences. All further analyses were performed using the repeat-normalized data.

Table 5.1: Number of sequences obtained with Illumina sequencing for the three developmental stages of leaf development (two biological repeats): total number of reads, number of sequences that were retained after parsing (based on the identification of the first 6 bases of the $3'$ adaptor) and after mapping to the *Arabidopsis* genome requiring perfect matches. For the parsed and mapped reads, the retained fraction of the total reads is indicated in brackets.

Sample	Total	Parsed	Mapped
Proliferation 1	8,308,435	4,532,173 (54.5)	2,749,065 (33.1)
Proliferation 2	8,506,930	3,605,964 (42.4)	2,151,983 (25.3)
Expansion 1	8,856,541	4,998,515 (56.4)	2,877,833 (32.5)
Expansion 2	11,435,714	6,668,228 (58.3)	3,808,746 (33.3)
Mature 1	9,156,449	5,501,019 (60.1)	3,127,533 (34.2)
Mature 2	14,694,596	6,778,211 (46.1)	3,140,091 (21.4)
Total	60,958,665	32,084,110 (52.6)	17,855,251 (29.3)

5.2.1.1   Genome distribution

To assess the source of small RNAs on the genome, a sliding window approach was used, applying a window of 50,000 bp wide and a stepsize of 10,000 bp using all mappable reads. First, all sRNA generating loci were visualized along the chromosome. These loci are most abundant at the centromere and pericentromeric regions for all chromosomes (illustrated for chromosome 1 for mature leaves in Figure 5.1A). The same approach was then applied to the repeat-normalized data (Figure 5.1B). Here, the pericentromeric regions are most abundant. The analysis was repeated using only 21-nt (Figure 5.1C and D) and 24-nt sequences (Figure 5.1E and F), as these represent the two major classes of sRNAs in *Arabidopsis* (49). For both size classes the sRNA generating loci are still biased towards the centromeric

region, although more outspoken for 24-nt. A region rich in pentatricopeptide repeat (PPR) genes also contains numerous 21-nt sRNA generating loci. These *PPR* genes are targeted by *TAS1-* and *TAS2*-derived ta-siRNAs and are also a known source of phased sRNAs (58). The repeat-normalized counts are overall low for the 21-nt size class, with the exception of solitary peaks. The five most prominent loci on chromosome 1 correspond to *MIR159a, MIR159b, MIR161*, the mentioned region rich in *PPR* genes and *TAS1b*, known as a ta-siRNA generating locus (13, 59). The profiles for 20- and 22-nt sRNAs are similar to these observed for the 21-nt class (data not shown). The 24-nt profile of sRNA generating loci shows a more even spread along the chromosome, indicating that this class has a higher diversity compared to the 21-nt class. The repeat-normalized counts indicate that the peak at the centromere is mainly due to highly repetitive sequences, as this is largely lost. The pericentromeric regions still show an enrichment, confirming previously reported patterns of highly diverse and abundant sRNA generating loci at these sites (60, 61, 62). The solitary peak in the 24-nt repeat-normalized data corresponds to *MIR163*, known to have a mature sequence of 24-nt (www.mirbase.org). The peak in the number of 24-nt generating loci (marked by a question mark) for 24-nt sRNAs, could not be linked to a known feature. The 23-nt sRNAs have a similar profile as observed for the 24-nt class (data not shown). This general profile is similar for all chromosomes (graphs for the mature stage of the other chromosomes in Supplemental Figure 5.1, 5.2, 5.3, 5.4). Throughout development, the described profiles for the number of loci generating sRNAs are very similar, both in the pattern as the amplitude. The repeat-normalized data show an increased amplitude for the 24-nt sequences at maturity compared to earlier phases of development (Supplemental Figure 5.5, 5.6). This is also the case for 23-nt sequences (data not shown).

5.2.1.2   Size distribution

The small RNA population in *Arabidopsis* can be divided based upon their length in two major classes: 21-nt and 24-nt long. The 21-nt class contains mostly miRNA and ta-siRNAs while heterochromatic siRNAs dominate the 24-nt class. Analysis of the distribution of sRNA generating loci according to the their size shows a bias towards 24-nt sRNAs in all developmental stages (Figure 5.2A), consistent with previous observations (60, 61). Weighing the sRNAs according to the repeat-normalized reads (Figure 5.2B) reveals that the 21-nt sRNAs are most abundant, which is due to the high expression of a limited number of miRNAs. The two highest expressed miRNAs, miR159a and miR159b, represent circa 600K and 180K reads per sample or 20% and 6% of all reads, respectively. The profile of 24-nt sRNAs during development hints at an increase in expression, although this is not statistically significant.
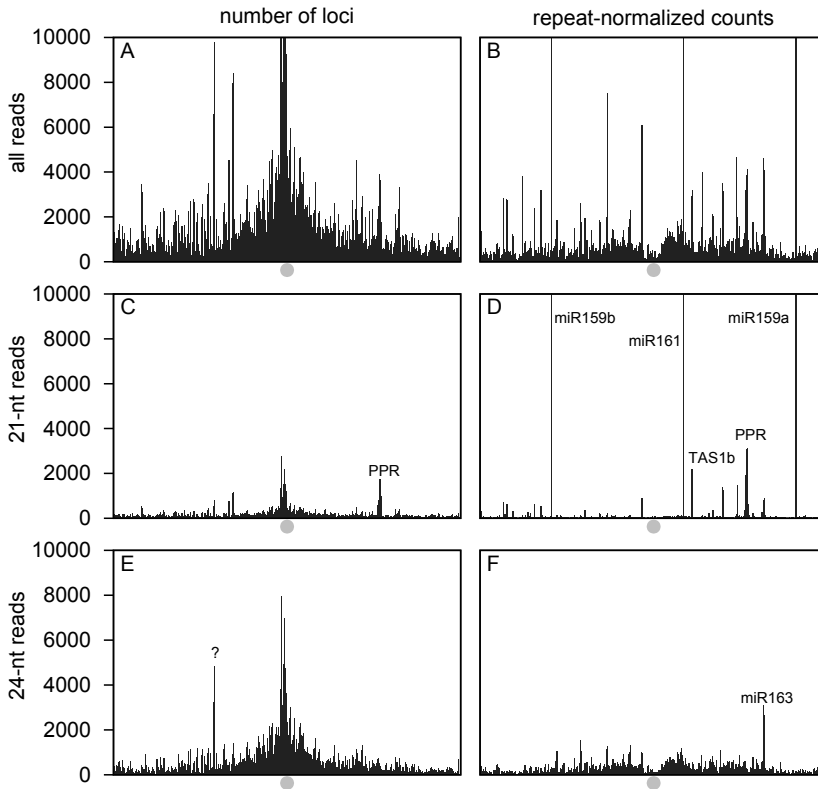
Figure 5.1: Genome-wide distribution of small RNA data of mature leaves for Chromosome 1 using a sliding window approach (50,000 bp window, 10,000 bp shift). The number of sRNA generating loci (A,C,E) and repeat-normalized counts (B,D,F) are represented for all sequences (A,B), the 21-nt (C,D) and 24-nt (E,F) size class. Due to the high abundance of miR159a (669,417), miR159b (169,337) and miR161 (17,253), these expression levels are cut off in D. The position of the centromere is marked by a grey circle.

5.3   Differentially expressed small RNAs

To identify differentially expressed sRNAs, I used only sequences that have at least a total of 10 reads in our dataset. This 'background correction' retains 90% of the reads but only 5% of the number of unique sequences. Sequences that are differentially expressed during leaf development were identified using *limma* (63) with a False Discovery Rate (FDR) according to Benjamini and Hochberg of maximally 0.05. Also a fold-change cutoff of two-fold was applied. This approach identified 3595 sRNAs that are differentially expressed between developmental stages. The
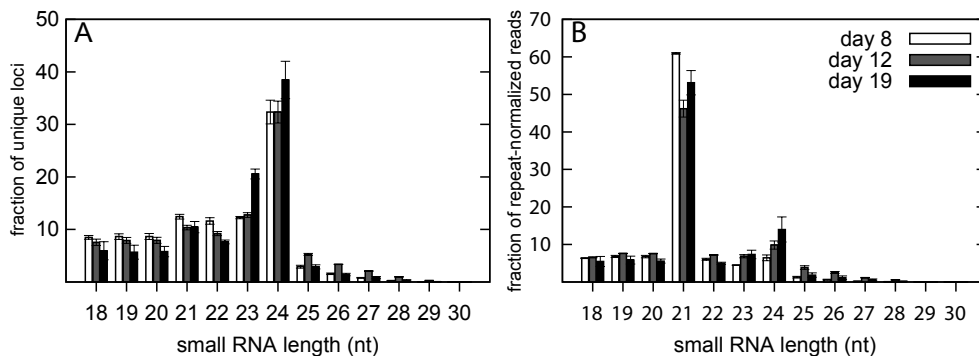
Figure 5.2: Distribution of sequences according to the length of small RNAs. (A) For unique sequences and (B) for repeat-normalized reads for all data (mean ± SEM, n=2)
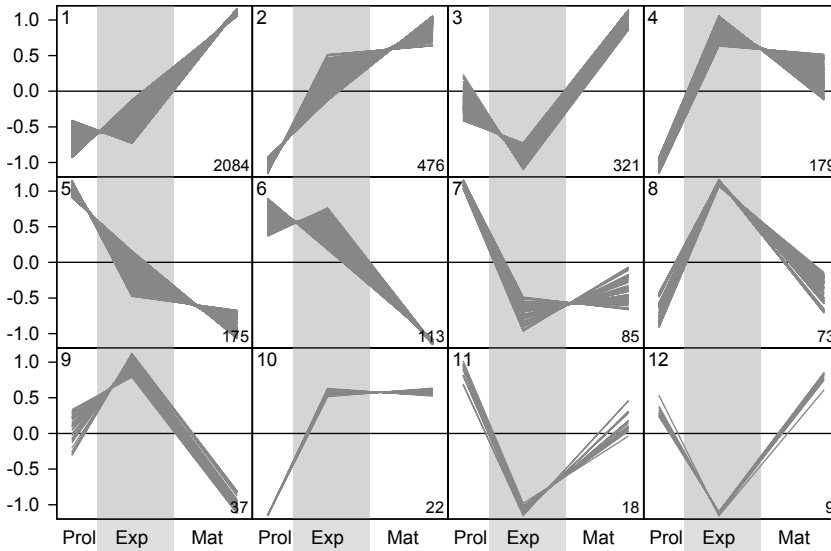
expression patterns of the differential sRNAs were normalized and clustered with the QT clustering method using the R-package *flexclust* (64, 65). A diameter of 0.4 resulted in 12 clusters (Figure 5.3) with three sRNAs not assigned to any cluster.

The majority of the differential sequences (85%) shows an increase in expression during development. These are separated into five clusters depending on the expression in expansion: cluster 1 (2084 sRNAs) shows a gradual increase over development, cluster 2 (476 sRNAs) has a steeper increase towards expansion and cluster 3 (321 sRNAs) shows a decrease from proliferation to expansion followed by a larger increase to mature stage. The final two clusters that increase during development slightly peak at expansion (cluster 4 – 179 sRNAs) or stay at the same expression level in expansion and maturity (cluster 10 – 22 sRNAs).

For a second group of considerably smaller clusters the expression decreases during leaf development. This general pattern represents only 10% of the sequences, comprising clusters 5, 6 and 7 with 175, 113 and 85 sRNAs, respectively. Again the expression in expansion differentiates these clusters with a gradual decrease for cluster 5 and a high (cluster 6) or low (cluster 7) level of expression during expansion. The final 5% of the differential sRNAs are specifically high (clusters 8 and 9) or low (clusters 11 and 12) expressed at expansion.

To link these clusters to a functional categorization of the corresponding sRNAs, I looked into the distribution of the lengths of the sRNAs per cluster (Table 5.2) and to which genomic features these sequences are matching (Table 5.3).

Curiously, there is a striking correlation between the predominant classes of small RNAs and their expression profiles. There are three distinctly different distributions of small RNA lengths: (1) Distributions where the majority of sequences have a length of 20- to 22-nt. Clusters 5, 6, 7 and 11 fit this profile and all have

Figure 5.3: Expression patterns of differentially expressed small RNAs during the three stages of leaf development: proliferation (Prol), expansion (Exp) and mature (Mat). Data were normalized and clustered using *qtclust* from the R package *flexclust*. The number of sRNAs in each cluster is denoted at the right bottom.

a profile with a strongly decreasing character throughout development. (2) Distributions with predominantly 23- and 24-nt sRNAs. This is the case for clusters 1, 2 and 3, notably these are the largest clusters, which show a strong increase in their expression levels as leaf development progresses. (3) Finally, clusters 4, 8, 9 and 10 contain mostly sRNAs of 25- to 28-nt and have a profile characterized by a maximum expression during the cell expansion phase.

Next to the length of the sequences, the region where the sRNA matches the genome is an important clue to the function of these RNA sequences. Again there are clear differences related to the expression pattern. The increasing expression clusters 1, 2 and 3 contain mostly sRNAs that match transposable elements and transposon fragments (> 40%) and are low in exon-matching sRNAs (9.5–13%). This in contrast to the other clusters that have between 24 and 38% of exon-matching sRNAs. Remarkably, clusters 4 and 10, two cluster that also show high expression in the mature stage, show no enrichment in transposable elements. These however, are enriched in tRNA matching sequences (19 and 11%, respectively). Cluster 7 is also enriched in sRNAs matching transposable elements and transposon fragments (20%) but it also contains sRNAs that match exons (30%). For the last two clusters, the number of sRNAs was too small to get meaningful results with this type of analysis.

Table 5.2: Total number of sequences and length distribution (%) for the clusters of differentially expressed sRNAs in Figure 5.3. The different shades of grey indicate three distinct distributions: predominantly 20- to 22-nt (lightest), 23- to 24-nt (darker) and 25- to 28-nt (darkest).

Cluster		1	2	3	4	5	6	7	8	9	10	11	12
Total		2084	476	321	179	175	113	85	73	37	22	18	9
	19	0.48	1.89	0.93	0	4	7.08	3.53	0	0	0	16.67	0
Sequence length	20	1.3	1.68	1.56	1.68	5.14	12.39	2.35	0	5.41	0	11.11	11.11
	21	5.9	4.83	12.77	1.68	22.86	17.7	36.47	2.74	2.7	4.55	38.89	11.11
	22	1.58	2.94	5.92	3.91	28	13.27	35.29	1.37	21.62	0	27.78	22.22
	23	41.36	12.39	41.12	6.7	4	7.96	4.71	10.96	16.22	9.09	0	22.22
	24	49.23	73.74	37.69	20.67	34.29	35.4	17.65	13.7	13.51	31.82	5.56	33.33
	25	0.14	0.84	0	21.79	1.71	4.42	0	17.81	16.22	18.18	0	0
	26	0	0.84	0	20.67	0	1.77	0	20.55	10.81	22.73	0	0
	27	0	0.84	0	15.64	0	0	0	19.18	2.7	v9.09	0	0
	28	0	0	0	7.26	0	0	0	9.59	8.11	4.55	0	0
	29	0	0	0	0	0	0	0	4.11	2.7	0	0	0

Table 5.3: Features matching small RNAs (in %) for clusters of differentially expressed sRNAs in Figure 5.3. The shading highlights the most important feature categories per cluster

Cluster	1	2	3	4	5	6	7	8	9	10
exon	10.81	13.01	9.45	38.11	29.99	24.63	30.00	24.99	26.08	29.42
five prime UTR	0.18	0.86	0.41	0	0	1.45	0	0	0	0
intron	6.30	4.62	6.99	0	2.86	4.35	0	0	0	0
mRNA	14.17	12.60	14.00	9.52	11.43	14.50	14.01	15.62	17.39	11.76
miRNA	0.96	3.77	0.83	4.76	10	2.90	7.99	0	0	0
ncRNA	0.69	0.41	0.83	0	8.56	4.35	2.00	3.13	0	0
other RNA	0.69	0.41	0.83	0	8.56	4.35	2.00	3.13	0	0
protein	9.15	7.97	9.45	9.52	4.29	10.15	7.99	12.49	17.39	5.88
protein coding gene	10.81	9.66	10.69	9.52	7.14	11.60	7.99	12.49	17.39	5.88
pseudogene	1.30	1.30	0.83	0	0	2.90	0	3.13	4.35	5.88
pseudogenic exon	1.14	0.86	0.83	0	0	2.90	0	3.13	4.35	5.88
pseudogenic transcript	1.30	1.27	0.41	4.76	0	2.90	0	3.13	4.35	5.88
rRNA	0.18	0.41	0.41	4.76	1.44	2.90	2.00	3.13	4.35	5.88
snRNA	0	0	0	0	0	0	4.00	0	0	0
snoRNA	0	0	0	0	1.44	0	0	0	0	0
tRNA	0.45	1.27	0.41	19.04	1.44	4.35	0	3.13	0	11.76
three prime UTR	1.21	0.86	0.83	0	2.86	1.45	2.00	3.13	0	0
transposable element	36.70	37.83	38.28	0	4.29	0	14.01	6.25	4.35	5.88
transposable element gene	3.99	2.94	4.10	0	5.70	4.35	6.00	3.13	0	5.88

5.4 microRNAs

The first microRNAs (miRNAs) were found to target proteins that function in development (66, 67). Therefore, I analyzed the expression profiles of known miRNAs and tried to discover new miRNAs in our developmental time series.

5.4.1 Known miRNAs

All known miRNAs were obtained from the microRNA database (www.mirbase.org, release 15). I identified perfectly matching reads for 157 out of 218 (72%) known miRNAs in *Arabidopsis*. As multiple miRNAs can have the same mature sequence, these 157 miRNAs correspond to 112 unique sequences, for which 24 had less than 10 reads in our total dataset. The 88 remaining sequences (133 miRNAs) were tested for differential expression during development according to the methodology described earlier. 27 miRNA sequences were differentially expressed, corresponding to 34 miRNAs (Table 5.4 and 5.5). For this, the FDR cutoff was slightly relaxed to 0.1. Full details of the miRNA expression and the significance of the differential expression in our dataset can be found in Supplemental Table 5.1. Within the differentially expressed miRNAs, two expression patterns can be distinguished: a first group has a decreasing trend (Table 5.4), while the second set increases during development (Table 5.5).

miR396 regulates cell proliferation through its targets, *GROWTH REGULATING FACTOR* (*GRF*) transcription factors (26). *Arabidopsis* encodes nine *GRF*s of which seven have a miR396 target site, only *GRF5* and *GRF6* lack this motif. miR396 is low expressed in leaf primordia and steadily increases during development (26), which is confirmed in our dataset (Table 5.5). With the exception of *GRF9*, the other eight *Arabidopsis GRF*s are present in the microarray data. Both *GRF2* and *GRF3* decrease circa 4-fold during development (Table 5.7), in the case of *GRF2* this was shown previously to be caused by miR396 (26). These two *GRF*s are the only ones that are differentially downregulated during leaf development in the microarray data (19). Only *GRF1-4* are expressed above the background of which GRF4 is differentially increasing over time and *GRF1* is not differential. These results are in contrast with the qPCR analysis performed in Rodriguez *et al.* (26), where all studied *GRF*s (*GRF1-7,9*) decrease in expression during the development of the fifth leaf.

During shoot development, miR156 and miR172 are important regulators of the phase transition from juvenile to adult leaves. In our data, during the development of juvenile leaves, miR172e is differentially expressed and increases over time. Also other members of the miR172 family show an increasing trend, but these are not statistically significant (Supplemental Table 5.1). Two of the targets of miR172, *AP2* and *TOE2*, indeed decrease during development. Moreover, SPL10 increases during development (FDR 0.03, fold change 1.4). *SPL10* is a positive regulator of miR172 (2).

The previous analysis was based on the mature sequences defined in miR-Base. However, also other sequences perfectly match the hairpin that generates known miRNAs. I identified all sequences that map to the genomic loci of all known miRNA generating hairpins, also described by miRBase. First, I analyzed the distribution of these sequences along the hairpin. For nearly all hairpins, the sequences were grouped in either one or two regions, corresponding to the miRNA and miRNA* site. The starting base of the different sequences matching to these regions lay, for the vast majority, in a 5-base window. The sequence length varied from 19 to 24 nucleotides. The sequences are highly strand-specifically generated as only 2% of the matches corresponded to the opposite strand of the miRNA hairpin for all but two miRNA loci. The majority (> 90%) of these sRNAs matching the opposite strand were represented by less then 5 reads. For two miRNA precursors this strand-specificity was not observed: miR781 and miR783. The hairpin of miR781 is such that the inverse complement is identical, causing the mature miRNA to match perfectly on both strands. The locus of miR783 generates a large variety of small RNAs from both strands and thus has an atypical profile for a miRNA. Two miRNAs, miR779.2 and miR841, generate a small RNA corresponding to the 21- or 22-nt downstream of the mature miRNA, suggesting a phased biogenesis could take place for these two miRNAs.

For most hairpins, the mature miRNA, as described in miRBase, is the highest expressed sRNA matching the locus. There are however exceptions in our dataset. The most prominent one is miR824, for which a 21-nt sequence 593 bp upstream of the start of the mature miRNA is more than eleven times higher expressed, yielding one of the highest expressed sequences in our dataset. This sequence corresponds to the miRNA* of the mature miRNA for this long hairpin. For further analysis, we focused on sequences that are differentially expressed during leaf development (Table 5.6).

For miR393a and b, the mature 21-nt miRNA sequence was represented by only 4 reads in our dataset. Two sRNAs that start at the same position, 19- and 20-nt in length have 69 and 265 reads, respectively, and are both significantly upregulated during development. Target prediction, using TAPIR (bioinformatics.psb.ugent.be/webtools/tapir/) which was recently developed in our lab (68), showed that they have the same specificity as the described mature miR393, regulating *TRANSPORT INHIBITOR RESPONSE 1* (*TIR1*) and its homologous sequences *AUXIN-SIGNALING F-BOX PROTEIN 1* (*AFB1*), *AFB2* and *AFB3* (67). Three of these targets (*AFB1*, *AFB2* and *TIR1*) decrease during development and thus have an antagonistic expression pattern to these *MIR393*-derived sRNAs (Table 5.6 and Table 5.7).

Table 5.4: Differentially expressed miRNAs decreasing during development (mature sequence according to miRBase) with FDR adjusted p-value, average expression per time point (Prol: Proliferation, Exp: Expansion, Mat: Mature) and validated and/or predicted targets. Targets were determined using TAPIR (68), experimentally validated targets are indicated in bold, predicted targets that have been found using TAPIR in this study and were not described before to our knowledge, are indicated in italics.

miRNA	FDR	Prol	Exp	Mat	Target	Ref
miR158a	0.08	94.08	19.61	67.95	PPR proteins (AT1G64100, AT3G03580), *WD-40 repeat family protein (AT1G49910)*	(69)
miR171a	0.06	422.9	122.82	63.67	**SCL6 (II, III, IV)**	(70, 71)
miR173	0.07	14.91	1.97	2.39	**TAS1a/b/c, TAS2**	(59)
miR394a-b	0.05	833.98	184.12	144.54	**F-box family protein (AT1G27340)**, *loricrin-related (AT5G09670)*	(67)
miR395a,d,e	0.04	12.15	1.18	6.7	**APS1/3/4, AST68**, *GUN5*	(59, 67)
miR395b,c,f	0.04	17.56	1.44	3.67	**APS4, SULTR2;1**	
miR399a	0.06	8.46	1.04	0	-	
miR399b-c	0.03	11.58	2.22	0.24	**PHO2**	(59, 72, 73)
miR399f	0.04	84.54	20.31	7.18	**PHO2**	
miR775	0.05	960.49	221.4	165.61	**Galactosyl transferase (AT1G53290)**, *DCL1*	(74)
miR779.2	0.09	19.03	5.63	3.35	*2-Cysteine peroxiredoxin B (AT5G06290), AT1G15600*	
miR827	0.05	43.29	2.35	11.97	**SPX (AT1G63010), NLA**	(74)
miR839	0.07	25.34	3.52	4.78	-	
miR841	0.07	80.32	13.46	21.56	-	
miR862-5p	0.09	21.37	5.24	4.78	-	
miR2111a-b	0.04	8.55	0	1.2	**F-box (AT3G27150)**, *calcineurin-like phosphoesterase (AT1G07010)*	(75)

Table 5.5: Differentially expressed miRNAs increasing during leaf development (mature sequence according to miR-Base) with FDR adjusted p-value, average expression per time point (Prol: Proliferation, Exp: Expansion, Mat: Mature) and validated and/or predicted targets. Targets were determined using TAPIR (68), experimentally validated targets are indicated in bold, predicted targets that have been found using TAPIR in this study and were not described before to our knowledge, are indicated in italics.

miRNA	FDR	Prol	Exp	Mat	Target	Ref
miR163	0.09	469.38	966	2343.2	**S-adenosyl methyltransferases (AT1G66690, AT1G66700, AT1G66720, AT3G44860, AT3G44870)**	(69, 76, 77)
miR167d	0.03	153.73	1192.53	4587.84	ARF6, ARF8	(78)
miR172e	0.05	2.64	5.11	27.28	**AP2, TOE1/2/3**, SNZ,SMZ	(9, 10, 76, 79, 80, 81)
miR391	0.04	4.43	31.92	82.39	*STT3a*	
miR396a	0.03	54.98	276.39	2371.4	**GRF1/2/3/4/7/8/9**	(26, 67, 82)
miR396b	0.08	1158.26	1541.44	5439.77	**GRF1/2/3/4/7/8/9**, *AT2G34530, ATTPS10*	(26, 67, 82)
miR398a	0.07	10.88	7.19	39.25	**Cytochrome C oxidase (AT3G15640), CSD1, CSD2, CCS1**	(67, 83, 84, 85)
miR842	0.07	0.7	2.75	10.05	**Jacalin lectins (AT1G52070, AT1G52100, AT1G52120, AT1G57570, AT1G60110, AT1G60130, AT1G61230, AT2G25980, AT5G28520, AT5G38550, AT5G49850, AT5G49870)**	(60, 74)
miR843	0.03	46.98	248.25	1046.8	ACS4, F-box (AT1G11810, AT3G13830)	(74)
miR846	0.04	190.13	149.98	1293.12	Jacalin lectin (AT1G33790, AT1G52050, AT1G52060, AT1G52070, AT1G52110, AT1G52130, AT1G57570, AT1G60095), *uridylyltransferase-related (AT1G16880)*, DUR3	(74)
miR861-3p	0.07	4.58	4.45	25.36	-	

Table 5.6: Expression of sRNAs from miRNA hairpins that differ from the mature sequence (based on miRBase). Here only the highest expressed are listed, a full list can be found in Supplemental Table 5.2.

miRNA	FDR	Predicted targets	Prol	Exp	Mat	Comment
miR156c	0.06	-	635.80	1036.83	3642.96	miRNA*; 9 fold higher than miRNA
miR157a,b	0.05	-	212.88	33.54	39.00	miRNA*; 2.4 fold higher than miRNA
miR169d	0.04	-	10.21	27.47	108.35	miRNA*; 22 fold higher than miRNA
mir391	0.03	STT3A	1.09	11.39	144.59	miRNA*; 1.3 fold higher than miRNA
miR393a,b	0.04	AFB1,AFB2, AFB3,TIR1,bHLH (AT3G23690)	5.20	6.02	52.90	miRNA has only 4 reads; 19- (data not shown) and 20-nt species at same position
miR396a	0.03	-	54.98	276.39	2371.40	15 sRNAs that are differentially expressed with an increasing trend during development
miR405a,b,d	0.04	Transposable elements (AT3G28005, AT4G37570, AT5G36010, AT5G36090), LRR protein (AT3G17640)	0.23	0.48	5.74	miRNA not found, 165 different sequences matching these three hairpins, 65% 23- and 24-nt, mostly at the miRNA region; most have only a few reads, those that are differential all increase during development
miR773	0.04	DMT2, putative methyltransferase	1742.04	626.20	163.20	miRNA had only 9 reads; 22-nt species at same position 500-fold higher reads;
miR783	0.04	F-box (AT1G66290, AT1G66300, AT1G66320)	787.76	540.23	71.84	a lot of species, nearly half differential (FDR 0.1), all decreasing over development

Table 5.7: Differentially expressed predicted targets of miRNAs in Table 5.4, 5.5 and 5.6 in leaf development (19). Sequences corresponding to the miRNA/miRNA* region are indicated by an asterisk "*", the remaining sequences originate from the same region of, but are distinct from, the mature miRNA (based on miRBase).

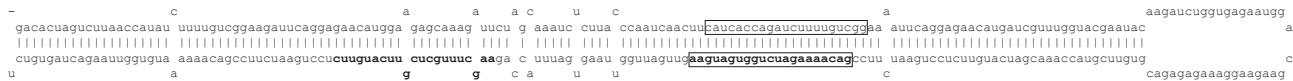| AGI	miRNA	name target	Prol	Exp	Mat	FDR adj. P	Fold
AT4G00150	miR171a	SCL6(IV)	115.5	230.8	217.7	2.5E-02	2.0
AT4G36920	miR172e	AP2	34.6	25.0	22.5	2.5E-02	1.5
AT5G60120	miR172e	TOE2	29.5	25.4	19.6	4.6E-02	1.5
AT5G19690	miR391*	STT3A	294.8	197.9	175.8	6.1E-04	1.7
AT3G23690	miR393a-b	bHLH	157.9	131.1	85.8	7.8E-03	1.8
AT3G26810	miR393a-b	AFB2	162.7	120.8	99.3	1.1E-02	1.6
AT3G62980	miR393a-b	TIR1	359.4	265.7	205.0	4.4E-02	1.8
AT4G03190	miR393a-b	AFB1/GRH1	282.5	209.5	103.2	4.7E-03	2.7
AT1G27340	miR394a-b	F-box	142.4	193.4	270.6	1.1E-03	1.9
AT3G22890	miR395a,d,e	APS1	1104.2	1919.3	2277.9	5.4E-03	2.1
AT5G10180	miR395b,c,f	SULTR2;1	101.4	254.7	258.9	3.3E-03	2.6
AT5G43780	miR395b,c,f	APS4	77.4	268.5	394.3	9.6E-04	5.1
AT2G36400	miR396a	GRF3	122.9	30.1	30.2	1.1E-02	4.1
AT4G37740	miR396a	GRF2	52.5	14.1	13.7	2.8E-05	3.8
AT1G60140	miR396b	ATTPS10	47.8	122.6	94.1	6.1E-02	2.6
AT2G33770	miR399f	PHO2	101.7	188.6	265.1	1.1E-02	2.6
AT3G17640	miR405a	leucine rich repeat	59.5	123.0	83.6	1.7E-02	2.1
AT5G06290	miR779.1,miR779.2	2-Cysteine peroxiredoxin B	1384.8	941.8	657.0	1.7E-03	2.1
AT1G02860	miR827	NLA	31.3	109.4	113.8	2.8E-04	3.6
AT1G52100	miR842	Jacalin lectin	351.6	39.3	37.5	4.7E-05	9.4
AT5G45380	miR846	DUR3	8.4	9.7	16.7	3.2E-02	2.0
AT1G07010	miR2111a-b	calcineurin-like phosphoesterase	228.2	460.5	776.5	7.6E-04	3.4

5.4.2 miRNA discovery

Based on our dataset, we tried to discover new microRNAs that are differentially expressed during leaf development. Therefore I selected all small RNAs of 20- to 22-nt and filtered out the sequences that have at least a total of 10 reads in our dataset. Sequences that are differentially expressed during leaf development were identified using *limma* (63) with a FDR cutoff of 0.05. This resulted in 761 candidate sequences. The genomic locus of these candidates was investigated to identify possible stem-loops and only loci that yielded both a miRNA and miRNA* region where retained.

This resulted in a new candidate miRNA at chromosome 5 (Candidate 1 in Table 5.8) with a mature miRNA sequence of TTGATGATTCGACAAAGTGAA and miRNA* CACTTTGTCGAGTCACCAAGG. Both the predicted miRNA and miRNA* increase during development (Table 5.9). Target prediction using TAPIR, resulted in four candidate targets. These genes are all annotated as transposable element genes: three from the non-LTR retrotransposon family (LINE) (*AT1G10160*, *AT5G01335*, *AT5G32616*) and one gypsy-like retrotransposon (*AT5G32358*). None of these genes are included in the microarray, but in a similar time series experiment spanning proliferation and expansion for leaf 3 of *Arabidopsis* using the AGRO-NOMICS1 Tiling Array, they were not differentially expressed (personal communication, M. Andriankaja).

Additionally, a candidate miRNA/miRNA* pair was found on the *MIR783* hairpin, different from the mature sequence described in miRBase. Remarkably, due to the extensive perfect pairing in the stem region, this miRNA/miRNA* duplex can be formed from both the Crick and Watson strand (Candidate 2a and 2b in Table 5.8, Figure 5.4). Both sequences are highest expressed during proliferation (Table 5.9). TAPIR identified two genes encoding unknown proteins (*AT2G31110* and *AT5G07790*) as possible targets for the miRNA, and 5 genes could be targeted by the miRNA*: coding for an unknown protein (*AT5G58000*), an F-box domain and F-box-domain like protein (*AT1G66290*, *AT1G66640*), ARGININE/SERINE-RICH SPLIC-ING FACTOR 35 (*AT4G25500*) and a copia-like retrotransposon (*AT1G46120*). Only *AT2G31110*, *AT5G58000* and *AT4G25500* are represented in the microarray data, but both are not differentially expressed during leaf development. In the tiling array data of all these targets only *AT2G31110* was differentially expressed, increasing from proliferation to expansion. Why the ATH1 and tiling array data differ for this gene, is unclear.

Table 5.8: Candidate miRNA stemloops. Position on the genome: chromosome (C), start position (start) and end position (stop), strand and sequence of the mature miRNA and miRNA*. The ID is the number used as a reference for the stemloop.

ID	C	Start	Stop	Strand	miRNA	miRNA*
1	5	22339940	22340135	-	TTGATGATTCGACAAAGTGAA	CACTTTGTCGAGTCACCAAGG
2a	1	24724861	24725036	+	GACAAAGATCTGGTGATGAA	CATCACCAGATCTTTTGTCGG
2b	1	24724861	24725036	-	GACAAAGATCTGGTGATGAA	CATCACCAGATCTTTTGTCGG


```
-                    c                                    a        a    a c     u     c                              a                          aagaucuggugagaaugg
  gacacuagucuuaaccauau uuuugucggaagauucaggagaacaugga gagcaaag uucu g aaauc cuua ccaaucaacuucaucaccagaucuuuugucggaa auucaggagaacaugaucguuugguacgaauac                    a
  ||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||| |||| | ||||| |||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
  cugugaucagaauuggugua aaaacagccuucuaaguccucuuguacuu cucguuuc aaga c uuuag gaau gguuaguugaaguaguggucuagaaaacagccuu uaaguccucuuguacuagcaaaccaugcuugug                    c
u                    a                                g        g    c a     u     u                              c                          cagagagaaaggaagaag
```

Figure 5.4: miR783 hairpin according to miRBase. In bold the mature miRNA (based on miRBase) is indicated, the boxed sequences form a possible alternative miRNA/miRNA* duplex, with the likely miRNA indicated in bold. Due to the extensive perfect pairing of the stem of the hairpin, this miRNA duplex can be formed from both the Crick as the Watson strand.

Table 5.9: Expression profile of predicted candidate miRNA and miRNA* in proliferation (Prol), expansion (Exp) and mature (Mat) phase of development for both the miRNA and miRNA* for each of the stemloops in Table 5.8 (ID between parentheses). The number of hits on the genome for each sequence is indicated.

sRNA	hits	Prol	Exp	Mat
miRNA (1)	1	89.99	134.39	1159.30
miRNA* (1)	1	1.09	4.32	22.03
miRNA (2)	4	53.65	16.34	0.96
miRNA* (2)	4	7.15	0.52	0.00

5.4.3 *Trans*-acting siRNAs

The biogenesis of *trans*-acting short interfering sRNAs (ta-siRNA) starts with cleavage of a *TAS* locus guided by a miRNA. This triggers the production of sRNAs from this cleaved end through repeated DCL activity, yielding consecutive 21-nt sequences from the *TAS* locus. Due to the nature of this biogenesis mechanism the produced ta-siRNAs have a predominant 21-nt phasing. This allows the production of a predictable and limited number of ta-siRNA species. In *Arabidopsis* 8 *TAS* loci, grouped in 4 families, have been identified (13, 58, 59, 60). In our dataset *TAS1a*, *TAS1b*, *TAS1c*, *TAS2* and *TAS3a* are abundantly expressed with over 400 sequences matching each of these loci. *TAS3b*, *TAS3c* and *TAS4* generate much less ta-siRNAs during leaf development, with 14, 18 and 22 sRNAs, respectively.

The phased production at a locus can be visualized in several ways. Howell *et al.* (58) have developed an algorithm that yields a phase score for each position in a locus, based upon the abundance of ta-siRNAs in a 21-nt register up to 8 cycles downstream of that position. A phase score is only calculated if at least three of these 8 positions have sRNA expression associated with it. The score increases when more and/or higher expressed sRNAs are formed that are spaced (a multiple of) 21 bases to that position. A correction needs to be applied for sequences on the opposite strand, due to the 2-nt $3'$ overhang of DCL products. For *TAS* loci this yields a peak-pattern in which the amplitude is a measure of the extent of phased sRNA production. I applied this algorithm to our dataset for the *TAS* loci, using the repeat-normalized data for each sample and averaging the two biological replicates using the geometric average. Figure 5.5 illustrates this for *TAS2*, the *TAS* locus with highest sRNA expression in leaf development. The major peaks coincide with the grid and are thus in register with the miRNA cleavage site (Figure 5.5A), as expected. Deviations from this phase occur, which was also reported by Howell *et al.* (58). The here obtained profiles have a similar amplitude and overall resemble those published in Howell *et al.* (58), with the exception of higher phase scores

for positions between the predominant phase, likely due to the greater sequencing depth in this study.

Another visualization method for phasing was presented by Axtell (86) and shows the fraction of sRNAs for the whole locus that start in the same 21-nt register. Here, all sRNAs that match the locus are divided over 21 bins based on their starting position. This can be done by dividing the start position relative to the miRNA cleavage site by 21 and assigning the sRNA to the bin designated by the remainder of this division, thus using modular arithmetic. Again a correction needs to be applied for sequences on the opposite strand. The result of this algorithm is visualized by a radial graph and loci can be ranked according to the highest fraction, but also the number of distinct sequences and the total abundance of sRNAs matching the locus should be taken into consideration. Figure 5.5B depicts the radial graphs for the three developmental stages for the *TAS2* locus. The repeat-normalized data for each sample was used as input and the results for the two biological replicates were averaged (geometric average). The highest value for the ratio, the fraction of sRNAs that are in the predominant phase, confirms the highly phased sRNA production at this locus, moreover this is in phase with the miRNA cleavage site (phase 1). The secondary phase is shifted by one nucleotide.

For *TAS1c*, Howell *et al.* (58) observed a shift in the phasing, increasing with the number of cycles, which is also present in our dataset and can be clearly seen from the lack of a predominant phase in the radial graphs (Supplemental Figure 5.9). The profile for *TAS3a* shows that an important fraction of the ta-siRNAs are out-of-phase, shifted by 11 bp, which was hypothesized by Howell *et al.* (58) to be due to secondary cleavage by a ta-siRNA (Supplemental Figure 5.10).

The results for *TAS1a* and *TAS1b* also indicate a clear phased pattern of sRNA production. Remarkably, the phase seems to shift during development. The expected phase, in register with the miR173 cleavage site, is replaced by one shifted 5 and 13 bp for *TAS1a* and *TAS1b*, respectively (Supplemental Figure 5.7 and 5.8).

For *TAS3b*, *TAS3c* and *TAS4*, phased ta-siRNA production could not be verified due to the low abundance of sRNAs at this locus in our dataset. The sequences that match *TAS3c* and *TAS4* are mainly proliferation specific and the highest expressed sRNAs were indeed separated by 21 nt (data not shown). These ta-siRNA generating loci could therefore play a role early in leaf development.

Next, I determined which sites of the genome also showed a phased sRNA production. Therefore, I assembled the sRNAs into 'islands' such that sRNAs which are separated less than 200 bp are assigned to the same island (61). Our dataset resulted in 95233 islands for the *Arabidopsis* genome. The advantage of this approach, rather than starting from e.g. the genome annotation, lies in the data-centered generation of the islands, resulting in a grouping based upon the sRNA data rather than upon *a priori* knowledge.

Both algorithms, using phase score and binning into 21 phases, were applied
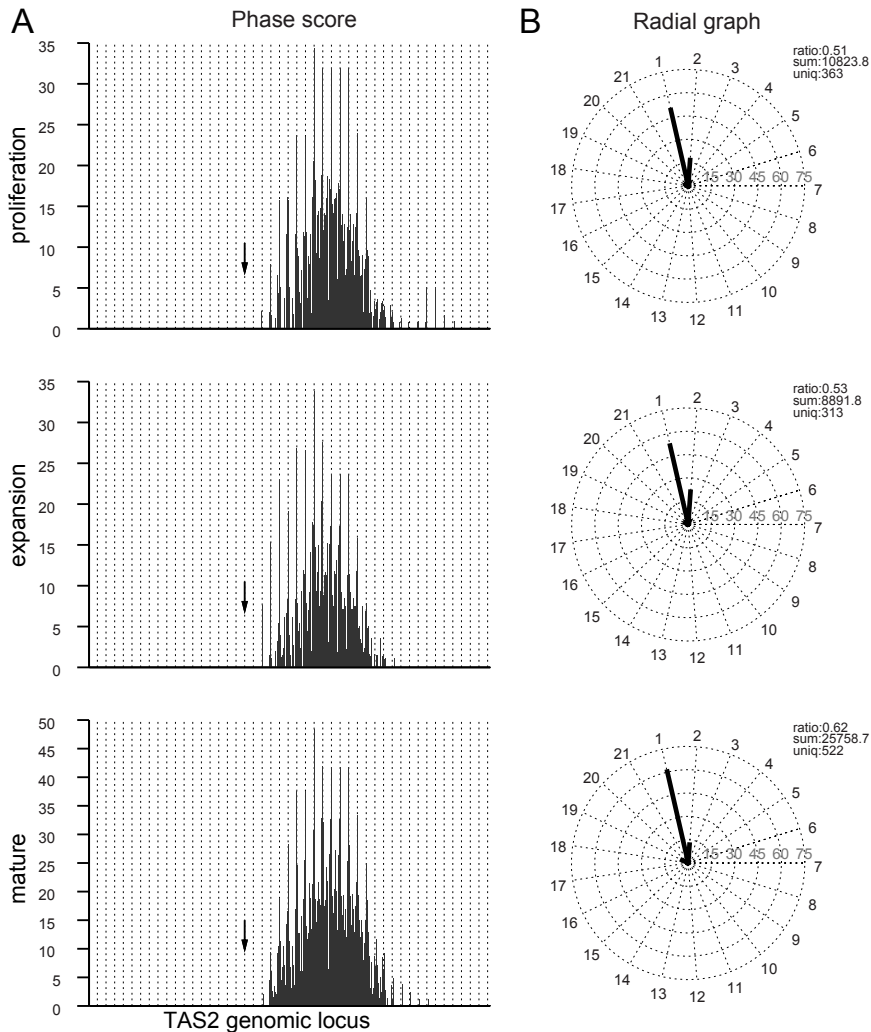
Figure 5.5: Visualization of phased siRNA production for the *TAS2* locus. (A) The phase score plot (58) is shown for the three developmental stages. The x-axis represents the entire *TAS2* genomic locus (TAIR8) and the arrow indicates where the miR173-mediated cleavage occurs that initiates the ta-siRNA cascade. The grid is spaced 21-nt and is in register with this cleavage site. (B) The radial graphs (86) for the three developmental stages show the percentage of sRNAs in each of the 21 possible phase registers according to their start position. The highest fraction is indicated (ratio) as well as the total number of repeat-normalized reads (sum) and the number of sRNAs (uniq) for each graph.

to these islands. Only regions that for at least one of the samples had an amplitude of 5 using the phase score algorithm and a ratio of at least 25% for the highest phase bin, were selected for further analysis (58, 86). Additionally, I selected for regions with more then 10 distinct sRNAs that have a total of at least 200 repeat-normalized reads. I focused on regions that are differential during leaf development, quantified by the fold change of the sum of all sRNAs matching the island, using a cutoff of 2-fold. This identified 58 regions where the predominant sequences are phased in a 21-nt register and their expression has the same differential trend during leaf development. The phasing diagrams are available online (www.psb.ugent.be/~frcop/PhaseDiagrams.html).

This approach successfully identified *TAS1a*, *TAS1b*, *TAS1c*, *TAS2* and *TAS3a* as loci of phased sRNA production, but *TAS1c* did not meet the ratio requirement (22%), while the maximum fold change for *TAS3a* was only 1.7. *TAS1c* had the highest fold change of all *TAS* loci (3.6), from the *TAS* loci that passed all criteria *TAS2* was most differential with a fold change of 2.9. Respectively, 35 and 28% of the sequences that match *TAS1c* and *TAS2* had an FDR adjusted p-value of less than 0.1. These sRNAs represented more than 40% of the reads matching both loci. The expression predominantly increases during leaf development with 90% of the significant sRNAs showing this behavior. Also sequences that are not significant generally had an upwards trend. For *TAS1a* and *TAS1b*, 18% of the sRNAs were differentially expressed, representing circa 30% of the reads and these similarly increase during development. For *TAS3a* only 6.5% of the reads are associated with differentially expressed sRNAs. There is a consistent downwards trend during leaf development for these, as well as also for non-significant sequences. This decrease during development can also be observed in all significant sequences matching *TAS3c* and *TAS4* (data not shown).

Next to the known *TAS* loci, also predicted and known targets of ta-siRNAs were found using this approach. For *TAS1* and *TAS2*, pentatricopeptide (PPR) genes have been shown as targets and their cleavage results in a phased sRNA production from these loci. The PPR genes *AT1G63130* and *AT1G63070* were identified as loci of phased sRNA production and both are differentially expressed, increasing during development.

The hairpin of *MIR841* has proliferation specific phased siRNA production, initiated upstream of the miRNA region and extending over the loop into the stem region of the miRNA*. Also the mature miRNA is differentially expressed, decreasing during development (Table 5.4).

In the intron of *AT5G48860*, coding for an unknown protein that is not expressed above the background during leaf development, phased sRNAs are found in the mature stage of leaf development. The second exon of *AT1G53880* and *AT1G53900*, two identical genes encoding for GTP binding proteins, also generates phased sRNAs at the mature stage of development. Also downstream of the F-box protein-coding

gene *AT1G31163*, mature specific phased sRNAs are produced. These last three genes are not present in the microarray data.

A remarkably high number of regions with phased siRNA production, contain transposable elements. For *AT2TE82000*, *AT3TE70410*, *AT4TE07570*, *AT4TE69050* and *AT5TE05480*, sRNAs are generated in a 21-nt register from the body of the transposable element. For others, the phased production takes place at the 3′ end or downstream of the element, this is the case for *AT1TE57570*, *AT1TE66780*, *AT1TE81210*, *AT2TE29150*, *AT2TE42355*, *AT3TE41755*, *AT4TE27090* and *AT5TE80340*. For all but one, phased sRNA biogenesis increases towards maturity or is only present at the mature stage of development. The only exception is *AT1TE81210*, which has a decreasing trend during development.

5.4.4 Loci differentially expressing sRNAs

To calculate differentially expressed sRNAs, sequences that are low expressed (less than 10 reads in our dataset, across the six samples) were removed. As a consequence, loci generating a large variety of low expressed sRNAs, are not picked up using this method. I wanted to assess if annotated features, described by TAIR8, differentially express sRNAs during development. These features include all described gene models as well as transposons. The sum of all sRNAs per feature per sample was used to calculate differentially expressed loci, using the same methodology as for individual sRNAs, selecting for loci with at least 100 repeat-normalized reads in our dataset. Using an FDR adjusted p-value of 0.05 and two-fold as a cutoff, 825 out of a total of 29391 features were found to be differentially expressed (Table 5.10). A full list is available online (www.psb.ugent.be/~frcop/signLoci.xlsx). This analysis was done separately for 'transposable element genes' and 'transposable elements' (referred to as transposons in the rest of the text), yielding 1297 differential transposons out of a total of 35089. A full list is available online (www.psb.ugent.be/~frcop/signTransposons.xlsx).

This methodology identified 68 miRNA loci as differential (Supplemental Table 5.3), these encompass most of the differential miRNAs described earlier. The very low expressed miRNAs are not present, as they do not meet the criterion of at least 100 reads matching the locus. Here, 20 additional *MIR* genes were identified (miR156e/f, miR162a/b, miR165a/b, miR167c, miR170, miR172a/b/d, miR319a/b, miR397b, miR398b/c, miR400, miR403, miR472a, miR823) that are differential when all sRNAs matching the locus are assessed as a group, but for which no individual sequences were differential.

The feature type 'other RNA' encompasses *TAS* loci and all three *TAS1* loci and *TAS2* were confirmed as differential but not *TAS3a* due to the presence of a large number of non-differential sRNAs. *TAS3b* was also identified as differentially decreasing through development while no individual sRNAs were found to be differentially expressed.

Table 5.10: Overview of TAIR8 features and matching sRNAs. The total number of features in TAIR8, the number of loci with at least 100 sRNAs matching and the number of loci for which the sum of all repeat-normalized reads is differentially expressed during leaf development (FDR < 0.05, fold change > 2)

Feature type	Total	≥ 100 sRNAs	significant
miRNA	176	164	68
other RNA	326	197	13
protein-coding gene	27235	18760	624
pseudogenic transcript	866	512	30
rRNA	15	4	0
snoRNA	71	67	8
snRNA	13	13	1
tRNA	689	630	81
transposable element gene	3900	145	111
transposable element	31189	1366	1186

5.4.4.1   sRNAs matching protein-coding genes

Of all loci annotated as protein-coding genes, 624 have differential sRNA expression when all matching sRNAs are regarded. Gene Ontology (GO) enrichment analysis (87) of this gene set indicates an overrepresentation of ribosome-related and photosystem II-related genes. As these genes are typically highly expressed, I investigated if the expression level is positively correlated with sRNA expression. To mine the data, I applied a scrolling window approach and developed a custom-made genome browser (see 5.7.6.1). The most prominent observation upon visual inspection of the small RNA data using this tool, led to the same conclusion: high expressed genes often have high sRNA expression matching the gene body. This suggests a link between gene expression and sRNA production for protein-coding genes. Therefore, I examined the correlation between the expression level of protein-coding genes and the expression of small RNAs matching a coding region for each developmental stage.

Correlation small RNA expression and mRNA expression
Figure 5.6A shows a plot for the mature stage where the mRNA expression of a locus is plotted against the sum of the repeat-normalized expression data for all sRNAs that match the coding region of that gene, divided by the length of the gene to correct for the increased chance of matching of sRNAs with increased gene length. The plot contains 15119 genes that have at least one sRNA matching the coding sequence and that are present on ATH1. This confirms that for the majority of genes there is a positive correlation between mRNA expression and small RNAs matching the gene. The same pattern is observed for the other stages of develop-

ment (data not shown). As expected, a large number of the low-expressed genes are not expressed above the background (MAS5 analysis, Figure 5.6B). The black dots represent genes that are not expressed above the background, but also at the lower end of the gene expression, grey dots are still present, signifying genes expressed above the background during leaf development. The genes that are expressed below a level of 30 contain 90.7% of all genes that are below the background and 18.2% of the genes that are expressed above the background.

While for the majority of the protein-coding genes the sRNA expression is correlated with the gene expression, there is a small proportion (about 5%) that has high small RNA expression and a relatively low mRNA expression. This set of genes could therefore be negatively regulated through small RNA silencing. To be able to select this set of genes easily, a transformation of the data was done by dividing the repeat-normalized and gene-length-normalized sRNA data (Figure 5.6A, B) by the gene expression (Figure 5.6C). This results in a measure for sRNA expression relative to the total amount of mRNA (expression level x gene length) produced from a locus. As a results a cutoff of this measure can be applied (0.0002) to select for the genes that do not show a positive correlation of sRNA and gene expression, for convenience the thus selected gene set will be referred to as the '*uncorrelated gene set*' while the gene set below the cutoff is named '*correlated gene set*'. We also required that there were at least 10 matching sRNA reads in our dataset for a gene to be selected. This yielded a selection of 489 genes that fulfill these criteria in at least one development stage of which 155 had gene expression above the background (Figure 5.6D). This subset of 155 genes of the uncorrelated gene set will be referred to as the '*selected gene set*'. All gene lists are available online (www.psb.ugent.be/~frcop/GeneSets.zip).

Bias towards functional sRNA sizes
To assess if there is a functional difference between the correlated and uncorrelated small RNA populations I compared the size distribution of the sRNAs associated with the two gene sets. Mainly two classes of sRNA sizes, 21- and 24-nt, are known to be functional in plants. sRNAs of 21-nt function through cleavage or translational inhibition of RNA targets while 24-nt sRNAs cause DNA methylation of complementary DNA, mostly by *cis*-regulation. For each protein-coding gene the distribution of the lengths of matching sRNAs was determined for each developmental stage. This was averaged over all genes in a gene set. First I assessed the length distribution for all protein-coding genes for which sRNAs matching the coding sequence were present in our dataset (Figure 5.7A). This population of sRNAs is broadly spread over the length range with about 15% for 19 to 24-nt sequences and gradually declining for longer sequences. We observed a correlation between the mRNA expression and the sRNA production for most protein-coding genes. Selecting for genes that show a correlation between mRNA and sRNA levels (correlated gene set, grey dots in Figure 5.6C), gives the same pattern as for all genes
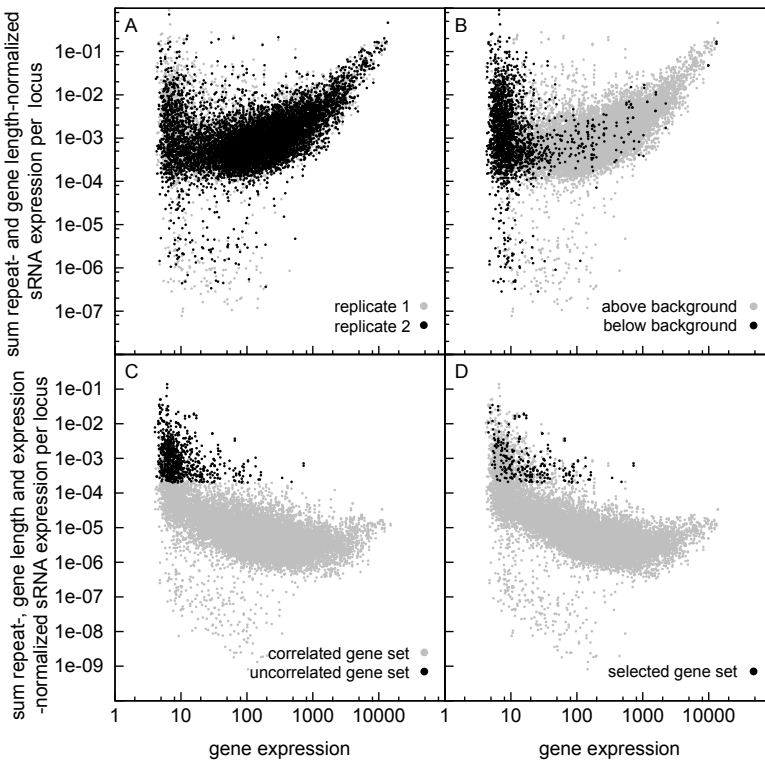
Figure 5.6: Correlation gene expression and sRNA expression. (A) Gene expression versus repeat- and gene length-normalized sRNA expression of sequences matching the locus for mature stage of leaf development. The black and grey dots represent the two biological replicates of our sRNA dataset. (B) The same data points as in A, now the black dots represent the genes whose expression during leaf development is below the background, the grey dots are detected above the background. (C) The data in A is transformed by dividing the repeat-normalized and gene-length-normalized sRNA data (Figure 5.6A, B) by the gene expression, this allows for an easier determination of a cutoff to isolate the genes for which gene expression and sRNA expression are not positively correlated: the black dots are above 0.0002 and are based upon at least 10 sRNAs. The black dots constitute the '*uncorrelated gene set*', while the grey dots make up the '*correlated gene set*'. (D) The same data points as in C, the black dots highlight the genes that are part of the uncorrelated gene set and for which the gene is expressed above the background during leaf development ('*selected gene set*').

(Figure 5.7B). This is expected as this makes up the bulk of the analyzed genes. The uncorrelated gene set (black dots in Figure 5.6) shows a bias towards 24-nt sequences, but also 21- to 23-nt sRNAs are present at higher frequencies compared to the shorter sequences of 19- and 20-nt (Figure 5.7C).
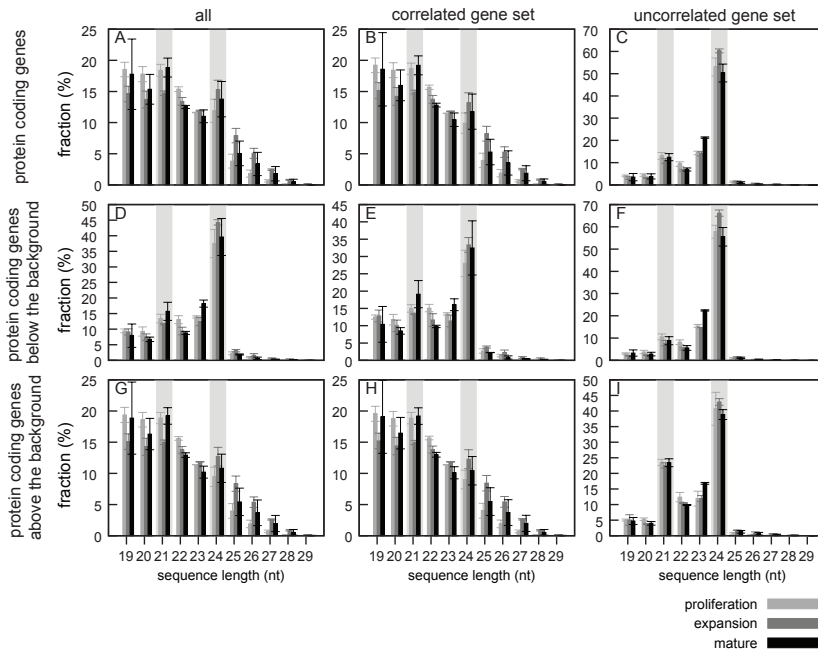
Figure 5.7: Length distribution for different sets of protein-coding genes. (A) All protein-coding genes for which sRNAs matching their coding sequence were present in our dataset. The other subsets were defined based upon two criteria: expression below (D, E, F) or above (G, H, I) the background and the correlation of sRNA and mRNA levels as defined by the 0.002 cutoff in Figure 5.6C: genes that follow the general trend of correlation between sRNA and mRNA levels (B, E, H) and those that deviate from this (C, F, I). The grey bars emphasize the two major classes of sRNAs involved in regulation: 21- and 24-nt sRNAs.

Next, I asked if there is difference in sRNA population between genes that are expressed below and above the background. Genes that are above the background show a similar pattern as observed for all genes (Figure 5.7G) while genes expressed below the background have a bias towards 24-nt sequences (Figure 5.7D), but this is not as outspoken as for the genes that show no correlation between sRNA and mRNA levels (Figure 5.7C). When splitting the genes that are expressed below the background according to the correlation criterion, the subset that is correlated is similar to the whole set below the background (Figure 5.7E), while the uncorrelated

subset has a stronger bias towards 24-nt sequences (Figure 5.7F), resulting in a distribution quite similar to all uncorrelated genes (Figure 5.7C). Also the genes that are expressed above the background show similar patterns for all genes and the subset that exhibits correlation between sRNA and mRNA levels (Figure 5.7G, H). The selected gene set (black dots in Figure 5.6D) has a bias towards both 21- and 24-nt sequences (Figure 5.7I). Therefore, I hypothesize that the selection in the uncorrelated gene set is enriched in genes that are regulated by 24-nt heterochromatic siRNAs, while further selection based upon the expression of the gene above the background specifically retains genes that generate 21-nt sRNAs.

The length distribution of the sRNAs that match protein-coding genes expressed above the background and for which there is a lower correlation with the mRNA levels (selected gene set) suggests that these genes could be involved in an sRNA regulatory node or regulated by sRNAs. Therefore, I investigated the selected genes in more detail. The set of genes contains 11 pentatricopeptide (PPR) repeat-containing protein of which 7 are listed as known ta-siRNA targets (*AT1G12620*, *AT1G12775*, *AT1G62910*, *AT1G62930*, *AT1G63080*, *AT1G63150*, *AT1G63400*) (58, 88, 89) and 8 have miR161.1/miR161.2 target regions that could initiate ta-siRNA production (*AT1G06580*, *AT1G62910*, *AT1G62930*, *AT1G63080*, *AT1G63150*, *AT1G63400*, *AT1G64580*, *AT5G16440*) (90). Seven additional genes are also targeted by ta-siRNAs: *AUXIN SIGNALING F-BOX 2* (*AFB2*) and a bHLH (*AT3G23690*) which are also targeted by miR393a/b, two disease resistance proteins of the CC-NBS-LRR class (*AT5G43730* and *AT5G43740*) that are targeted by miR472a and the TIR-NBS-LLR class *AT5G38850*, as well as two unknown proteins *AT5G49440* and *AT1G11700*. This yields in total 14 ta-siRNA targeted genes out of 46 listed (58, 88, 89) and annotated as protein-coding genes in TAIR8. ta-siRNAs as well as the secondary sRNAs produced by ta-siRNA targets are predominantly 21-nt in length. This is confirmed by the length distribution of the ta-siRNA and miRNA targeted genes in our selection as more than 60% of the sRNAs (weighed according to the repeat-normalized expression) are 21-nt in length and about 15% are 22-nt long (Figure 5.8A). The non-targeted genes in the selection are more biased towards 23- and 24-nt sRNAs (circa 16% and 47%, respectively) (Figure 5.8B). These data show that genes that have disproportionally high sRNA expression relative to their length and gene expression are likely to be targeted by miRNAs and/or ta-siRNAs, or have a higher likelihood to be regulated through chromatin modulation.

Profile of sRNA expression

In the previous analyses the sRNA expression at a region of interest was reduced to the sum of their expression during development. Another way to visualize the sRNAs generated, is based upon the position of the sRNAs on the locus. The relative position was used, ranging from 0% to 100% from the start until the final position of the locus, which allows for easy comparison of loci with different lengths. Also 1000 bp up- and downstream were included in the profile (also shown in per-
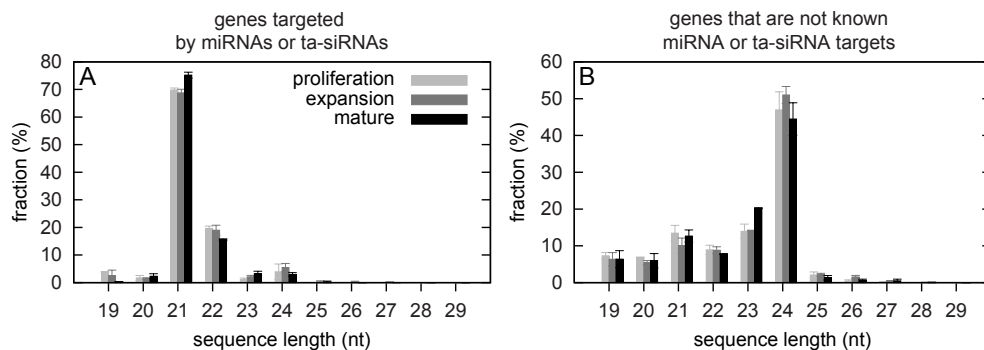
Figure 5.8: Length distribution of the selected gene set. Subdivided in (A) genes that are known as ta-siRNA or miRNA targets and (B) genes that are not known as such. Data is represented as mean ± SEM.

centage). The repeat-normalized data were used to calculate the total expression for each position and a sliding window of 30 (%) was applied to smooth the data. The use of the repeat-normalized data as a weighing factor can introduce a bias in the data when one or a few sequences are much higher expressed compared to the bulk of the sRNAs. To address this, a boxplot of the data was determined per replicate and all data points above the outer whisker (defined as 1.5 times the interquartile region, from the first quartile to the third quartile) where reduced to the average expression. In all analyses it was verified that this modification does not change the general trend of the profile. This was done for each size class separately as we expect differences in the profile depending on the length of the sRNAs.

To introduce this concept, I analyzed the sRNA profile of all miRNAs (as annotated in TAIR8) (Figure 5.9). A miRNA precursor forms a stem-loop, therefore the 5′ and 3′ region of a *MIR* gene code for the stem and are separated by the loop region. As the mature miRNA sequence can be located in both the 5′ and 3′ region, depending on the miRNA, the profile is expected to contain two peaks with low expression around the middle of the locus, where the loop region is situated. This is indeed the case, as can be observed in Figure 5.9A for sRNAs of 21-nt. The peaks are broad because the position of the mature miRNA in the stem is variable and extend 15% into the up- and downstream region due to the sliding window approach. While for longer sRNAs the same pattern is maintained, the amplitude is much lower (Figure 5.9B-D), this is due to a subset of *MIR* genes that generate a mature miRNA longer than 21-nt and through the production of sRNAs that deviate from the mature sequence, which I found to be mostly generated around the region of the miRNA duplex. I analyzed also the newly identified miRNA and confirmed that this locus has the same pattern (Figure 5.9E-H).
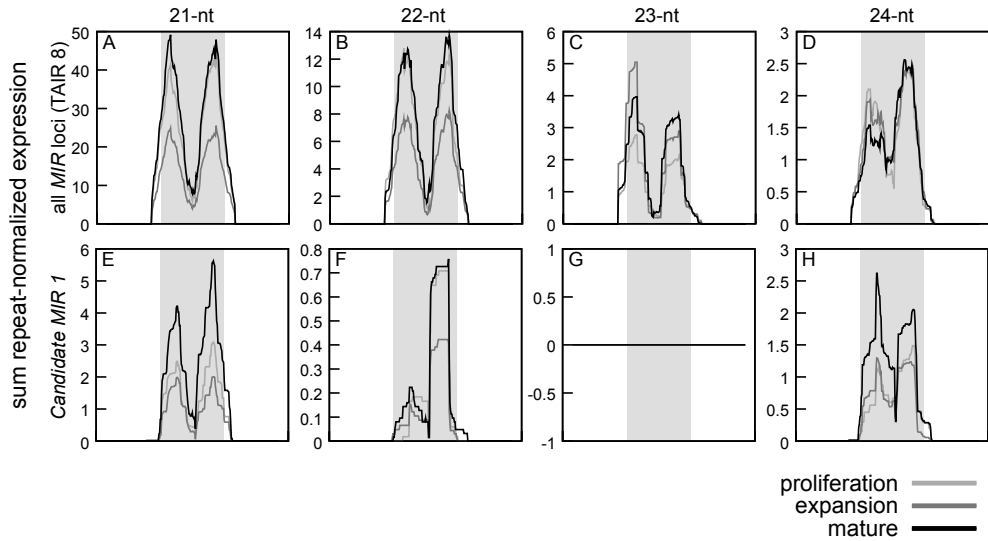
Figure 5.9: Profile of localization of sRNAs matching miRNA-generating loci.
(A-D) for all *MIR* loci annotated in TAIR8 for 20- to 24-nt sRNAs and (E-H) for the
in this study identified candidate *MIR1* (Table 5.8). No sRNAs of 23-nt in length
matched *MIR1*, therefore no profile could be determined. The grey area represents
the coding sequence (CDS) and it is flanked by 1000 bp up- and downstream (white
area).

I applied this methodology to the gene sets defined in the previous paragraph
(Figure 5.10). The criteria I used, selected for protein-coding genes that have high
sRNA expression associated with their coding regions, resulting in an 'uncorrelated
gene set'. Therefore, when the sRNAs are plotted according to the position where
they match to the gene, for all size classes mainly localization in the coding region
is obtained (Figure 5.10A-D). While 24-nt sRNAs are spread over the whole coding
sequence (CDS), 21-nt sequences show a peak in the first half, which could be due
to mis-annotated miRNAs in TAIR8. In all size classes for this uncorrelated gene
set, the expression in the mature stage of leaf development is increased compared
to the proliferation and expansion phase (Figure 5.10A-D). The correlated gene set
has a different profile (Figure 5.10E-H). While the 21- and 22-nt sequences mainly
originate from the CDS, these sRNAs are lowest expressed during the mature stage
of development. For the longer sRNAs of 23- and 24-nt there is a shift in the pattern
between proliferation/expansion and mature stage of development. While earlier
in development the CDS localization holds true, at late stages the majority of the
23- and 24-nt sRNAs are generated from the flanking regions. For all size classes
a drop in sRNA expression is present at the beginning and end of the CDS. When

only genes that are expressed above the background are analyzed (Figure 5.10I-L), a similar profile as for the correlated set is obtained. This similarity is expected as the correlated gene set and the genes that are expressed above the background contain the majority of the protein-coding genes included in this analysis. The genes that are expressed below the background level (Figure 5.10M-P), have a less outspoken peak at the CDS for 21- and 22-nt sRNAs, but the decreased expression at the edges of the CDS is present. 23- and 24-nt sRNAs are equally spread over the whole analyzed region with exception of a decrease around the start site. The selected gene set (Figure 5.10Q-T) has a similar behavior as the uncorrelated gene set, of which it is a subset.
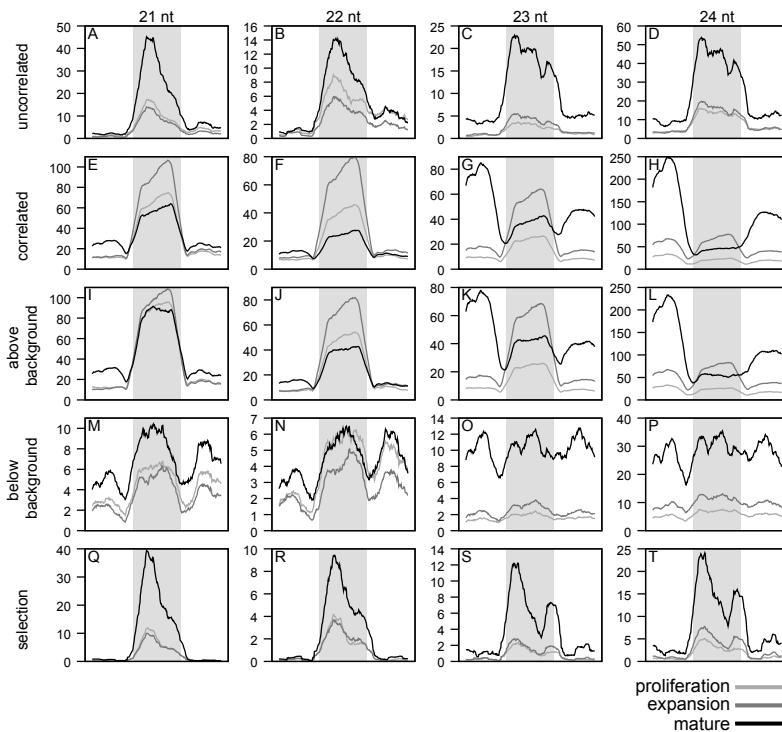
Figure 5.10: Profile of localization for defined sets of genes for 21- to 24-nt sRNA sequences: genes for which sRNA and mRNA expression are uncorrelated (A-D) or correlated (E-H), genes that are expressed above (I-L) or below (M-P) the background and the selected gene set (Q-T): genes for which sRNA and mRNA expression are uncorrelated and that are expressed above the background. The grey area represents the coding sequence (CDS) and it is flanked by 1000 bp up- and downstream (white area).

If from the selected gene set only genes that are known miRNA and ta-siRNAs targets are regarded, profiles can only be determined for 21- and 22-nt sRNAs (Figure 5.11A-D) and these are similar to the whole set (Figure 5.10Q-T). The remaining genes, not known as targets, have a low expression of 21- and 22-nt sRNAs while the profiles for 23- and 24-nt match those of the whole set (Figure 5.11E-H). This corroborates that the selection criteria identified two distinct types of sRNA generating protein-coding genes: miRNA and ta-siRNA targeted genes that also generate sRNAs of mainly 21-nt and genes that have 23- and 24-nt sRNA expression likely to be generated through the heterochromatic siRNA pathway.
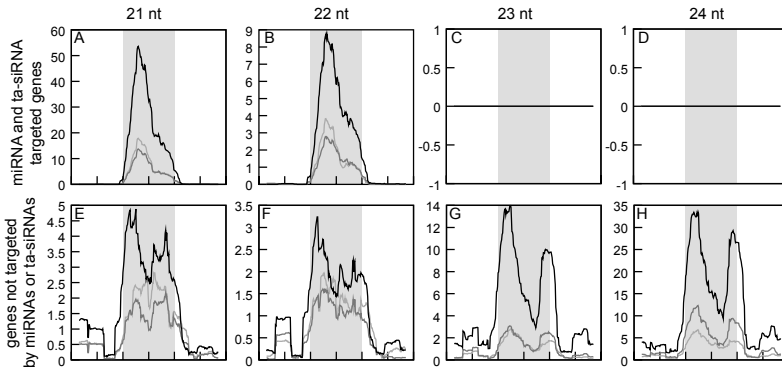
Figure 5.11: Profiles for subsets of the selected gene set for 21- to 24-nt sRNA sequences: genes for which sRNA and mRNA expression are uncorrelated and expressed above the background, subdivided for miRNA and ta-siRNA targets (A-D) and genes that are not known targets of small RNAs (E-H). The grey area represents the coding sequence (CDS) and it is flanked by 1000 bp up- and downstream (white area).

5.4.4.2   sRNAs matching transposons

Highly repetitive sequences such as transposons are known to generate small RNAs through the heterochromatic biogenesis pathway. These small RNAs are circa 24-nt long. Here, I studied the distribution of small RNAs matching transposons and their behavior throughout development.

First I verified if the 24-nt sRNAs are indeed overrepresented for transposons. Therefore, I used all 'transposable fragments' annotated in TAIR8 and determined the sRNA distribution according to the size classes, as described earlier. Circa 60% of the sRNAs, weighed according to the repeat-normalized reads, is 24-nt in length and 15% (proliferation and expansion) up to 25% (mature) are 23-nt while both 21- and 22-nt sRNAs are below 10% (Figure 5.12).
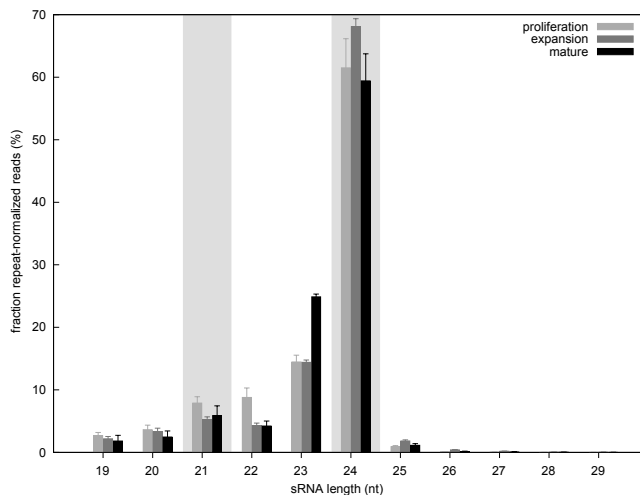
Figure 5.12: Length distribution of sRNAs matching transposable elements. The grey bars emphasize the two major regulatory classes for sRNAs: 21- and 24-nt.

The heterochromatic biogenesis pathway has as major effector protein ARG-ONAUTE4 (AGO4). The loading of sRNAs into AGO proteins is affected by the $5'$ nucleotide and for AGO4 it has been shown that it has a preference for adenosyl as a first base (91). Therefore, I determined the frequency of the 4 nucleotides as a starting base for sRNAs matching transposable elements, grouped according to the length of the sRNA. The GC-content of the *Arabidopsis* genome is 36%, leading to expected frequencies of 18% for G and C, and 32% for A and T (Table 5.11). For 21- to 23-nt sRNAs no statistically significant difference between the observed and expected frequencies was detected while for 24-nt the frequency of A as a starting base was 46%, significantly higher than expected. This result, together with the preference for 24-nt sRNAs, confirms that the sequenced sRNAs matching transposable elements are indeed predominantly heterochromatic siRNAs.

Table 5.11: Frequency of the start base for sRNAs of 21- to 24-nt matching transposable elements. The p-value was calculated with the Chi-square test using R.

length	A	C	G	T	p-value
21	34.97	12.50	17.09	35.45	0.50
22	34.98	14.33	19.56	31.13	0.76
23	29.89	14.90	17.78	37.43	0.66
24	46.26	10.25	14.43	29.05	0.01

The profile of sRNAs on transposable elements was determined using the approach described earlier. This yields for all size classes a similar pattern with a peak in sRNA generation at the 5′ and 3′ ends of the transposable element for the mature stage of development (Figure 5.13A-D). The amplitude does differ among the different sizes, with 23- and 24-nt an order of magnitude higher compared to 21- and 22-nt sRNAs. These profiles corroborate the mature specific upregulation of 23- and 24-nt sRNAs and moreover show that for transposon-associated sRNAs this upregulation holds true for all size classes. The high density of sRNAs at both ends of the transposon suggests that some mechanism is at play that generates sRNAs specifically at these sites.
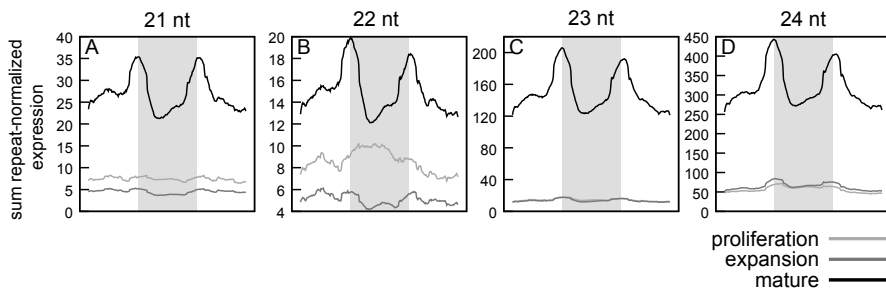
Figure 5.13: Profile of localization for sRNAs matching 'transposable elements' and 'transposable element genes': (A-D) 21- to 24-nt sRNAs. The grey area represents the transposon and it is flanked by 1000 bp up- and downstream (white area).

5.4.5   Methylation patterns

The role of 24-nt siRNAs in guiding DNA methylation has been well established in plants (49, 51, 52, 54, 55, 57). To asses this, we profiled DNA methylation at the same three stages of leaf development. The methylation was analyzed in three sequence contexts: CG, CHG and CHH (where H is A, C or T). For each stage one sample was prepared through bisulfite conversion (92) and the resulting libraries were sequenced with a coverage of circa 1x (circa 15% of CHG and CHH sites and 10% for CHH sites had datapoints). This does not allow to analyze DNA methylation at the single base level, but is sufficient to assess the DNA methylation for a set of loci, e.g. all protein-coding genes, the subsets I have defined earlier (correlated, uncorrelated) or all transposable elements.

First, a genome-wide analysis was performed using a sliding window of 100 kb (illustrated for chromosome 1 in Figure 5.14). All types of DNA methylation show a clear bias towards the centromeric region. Overall CG and CHG methylation

decrease during development while CHH methylation increases. Consistent with previous reports (92, 93), CG methylation is present at higher levels compared to CHG methylation and CHH sites are methylated only at low levels. During development, the level of CG and CHG methylation decreases while CHH increases steadily.
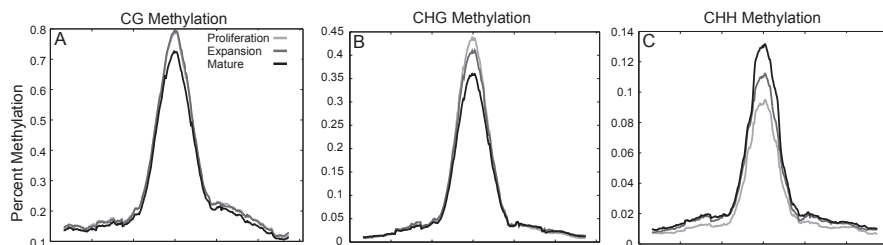

Figure 5.14: DNA methylation of chromosome 1. (A) CG, (B) CHG and (C) CHH sequence context (where H is A, C or T) for the three stages in development: proliferation, expansion and mature. The x-axis represents the full length of chromosome 1.

Next, we analyzed the methylation pattern for the same gene sets we used for the sRNA profiles. All protein-coding genes have CG methylation mainly at the CDS with a decrease at the start and end of the gene (Figure 5.15A). CHG and CHH show an inverse pattern, with lower methylation in the genic region (Figure 5.15B,C). These patterns are similar as observed previously (92). When focusing on the uncorrelated gene set, the overall methylation is increased compared to the correlated gene set. In the case of CG methylation, the methylation level increases with about 50% but the pattern is similar. The increase is much more outspoken for CHG and CHH methylation and, more importantly, the profile differs dramatically. For the genes that are in the uncorrelated gene set, the methylation at CHG and CHH sites is higher at the CDS compared to the flanking region. Both start and end of the gene show peaks for CHG methylation while only the start position is markedly higher for CHH methylation.

Transposable elements are highly methylated regions in the genome of *Arabidopsis* and a link between sRNAs and DNA methylation has been established (49, 51, 52, 54, 57). The profiles show in all sequence contexts a preference for DNA methylation at the transposon region compared to the flanking regions (Figure 5.16). Both CG and CHG decrease gradually, while CHH methylation increases steadily during leaf development, similar to what we observed in the genome-wide analysis (Figure 5.14). Remarkably, while the patterns for CG and CHG exhibit a sharp decline from the transposon region towards both flanking regions, this is not the case for CHH methylation where the flanking regions show elevated levels of DNA methylation in this sequence context.
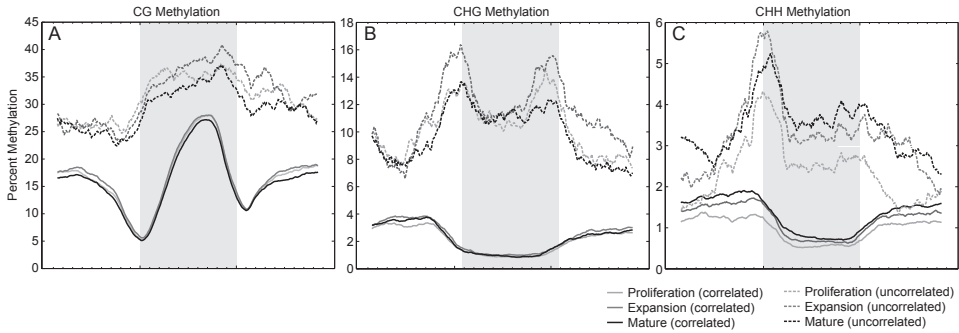
Figure 5.15: DNA methylation at protein-coding genes. (A) CG, (B) CHG and (C) CHH sequence context (where H is A, C or T) for the coding sequence (CDS, grey) and flanking regions (1000 bp). The protein-coding genes are subdivided in a 'correlated gene set' for which mRNA expression and sRNA expression at the CDS are positively correlated and an 'uncorrelated gene set' containing genes that have high sRNA expression and a relatively low gene expression. The data for the three developmental phases (proliferation, expansion and mature) is depicted.
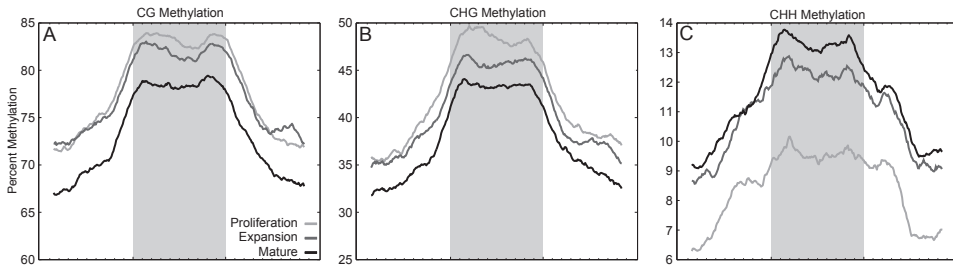
Figure 5.16: DNA methylation at transposable elements. (A) CG, (B) CHG and (C) CHH sequence context (where H is A, C or T) for the transposon region (grey) and flanking regions (1000 bp).

5.5   Discussion

We profiled the small RNA population in the first leaf pair of *Arabidopsis* during its development from proliferation phase, characterized by both cell division and cell expansion, over expansion phase when cells mainly expand, until maturity. While previous research focused mainly on inflorescence tissue, mature tissue or whole seedlings (60, 61, 89, 94), we specifically harvested the first leaf pair at these three developmental stages. To our knowledge, small RNA profiling of developmentally staged samples has only been performed in *Physcometrilla patens*, but this research was not specifically focused on the patterns during development (95).

5.5.1 Genome-wide increase of 24-nt sRNAs

The genome-wide distribution of the small RNAs (sRNAs) is similar to what has been reported previously (60, 61). Especially the pericentromeric regions generate a diverse and abundant small RNA population. When the data for 21- and 24-nt sRNAs, the two most important regulatory classes (49), are considered separately, the patterns are strikingly different (Figure 5.1). The 21-nt sRNAs are limited to highly abundant, solitary peaks matching miRNAs and ta-siRNA generating loci. The length distribution in our dataset confirms that a small number of loci is very high expressed. In contrary, 24-nt sRNAs show a more even spread along the chromosome. This class of sRNAs makes up the majority of the sequences, but their individual expression is relatively low.

While specific miRNAs and ta-siRNAs are differentially expressed during development, there is an overall increasing trend for 23- and 24-nt sRNAs. Both the genome-wide analysis as the length distribution indicate this (Figures 5.1 and 5.2). Additionally, the clustering of differentially expressed sRNAs showed that the clusters enriched for 23- and 24-nt sRNAs all show an increasing trend during development. The 24-nt siRNAs are associated with heterochromatic regions, arising mainly from pericentromeric regions, heterochromatic DNA, repeat elements, retroelements and some methylated DNA regions, silencing the expression of these regions (61, 96). This was confirmed by the feature enrichment of the 23- and 24-nt enriched clusters, that contained mainly transposons. Our data indicate that 24-nt sRNA generation from transposons, and thus by association the silencing of these loci, increases gradually during leaf development.

5.5.2 Nutrient stress-related miRNAs are upregulated at proliferation

Leaf growth is regulated by GROWTH-REGULATING FACTOR (GRF) transcription factors (TFs) (23, 82). Rodriguez *et al.* (26) showed that miR396 increases steadily during leaf development, which is confirmed in our dataset for both miR396a and b. Two of its targets, *GRF2* and *GRF3*, show indeed an inverse expression pattern, suggesting these genes are regulated by miR396 during leaf development. Through the repression of GRF activity, validated for *GRF2* (26), it regulates cell proliferation and thus early phases of leaf development.

The differential expression for miR396 was expected from previous work. However, a remarkably high number of miRNAs that are differentially expressed during leaf development, have been implicated in nutrient stress respons. Sulfate deprivation induces miR395, of which all six loci are decreasing as development progresses, regulating sulfur metabolism through repression of *ATP SULFURYLASES* (*APS*) and *SULFATE TRANSPORTER 2;1* (*SULTR2;1*)/*ARABIDOPSIS SULFATE TRANSPORTER 68* (*AST68*) (59, 67, 97, 98, 99, 100, 101). APS proteins regulate accumula-

tion of sulfate in the shoot, while translocation between leaves is under SULTR2;1 control. Overexpression of miR395 leads to increased sulfate levels in leaves and impaired distribution from older to younger leaves (101). The higher expression of miR395 during proliferation could enhance the availability of sulfate. The reduction of miR395 expression when the leaf ages results in increased levels of its targets *APS1*, *APS4* and *SULTR2;1* (Table 5.7), which could enhance distribution of sulfate towards younger leaves ensuring proper sulfate homeostasis.

Upon phosphate (P_i) limitation miR399 is induced which leads to repression of its target *PHOSPHATE2* (*PHO2*)/*E2-UBIQUITIN-CONJUGATING ENZYME24* (*UBC24*) (72, 102, 103, 104). This increases P_i uptake and relocation from roots to shoots, while remobilization within leaves is blocked (73, 105). The high expression of miR399 at proliferation could ensure that sufficient phosphate is available while reducing the expression of miR399 during leaf development could help in keeping the available phosphate mobile. The expression of its target *PHO2* is indeed inverse to that of miR399, with low levels at proliferation and a 2.6 fold increase towards maturity (Table 5.7).

Also miR827 was found to be upregulated upon P_i limitation and decreases during development (104) and upon P_i starvation a decrease of *NITROGEN LIM-ITATION ADAPTATION* (*NLA*) was observed (106). *NLA*, an E3 ligase, is a target of miR827 (74) and is specifically downregulated in proliferation. *NLA* is, as its name indicates, involved in the response upon nitrogen (N) limitation. An *nla* mutant showed reduced anthocyanin production and early onset of senescence when grown under low N conditions (107). However, upon simultaneous N and P_i limitation both anthocyanin production as timing of senescence returned to wild type situation, indicating that these pathways interact.

miR2111 decreases during development (albeit already low expressed during proliferation) and has also been found to be induced when P_i is limiting. Also miR2111 targets an E3 ligase, *AT3G27150*, but this gene is specifically expressed in the root (104). In the phloem sap, miR2111 was highly abundant upon P_i limitation and it could thus be part of a systemic regulatory pathway (104). miR398a is repressed upon phosphate, nitrate and carbon starvation (104) and its expression is low at proliferation and increases during development.

However, not all miRNAs that are reported as differential upon P_i starvation are differential during leaf development, e.g. miR778 was previously found to be strongly induced (75, 104) but was not represented by any small RNAs in our dataset. Also members of the miR169 family have been found to be reduced upon P_i and N limitation, but these were not differentially expressed in our dataset. This could be due to tissue specific differences as these studies used whole shoot or seedlings while we specifically profiled the first leaf pair.

The expression profiles of the miRNAs involved in nutrient stress, suggest that early in development the macro elements phosphate, sulfate and nitrogen are limit-

ing. For sulfur and phosphate this is also reflected in the expression of the miRNA target genes involved in their homeostasis. Early in development, the leaf is a sink tissue, requiring high amounts of nutrients to sustain the rapid growth through cell division and cell expansion. This could cause a local depletion of macro elements nutrients, despite its abundance in the growth medium. It has been shown that cell cycle activity dictates the requirement for P_i and treatments that enhance growth by proliferation increase the demand for P_i (108). The growth medium used contains 1% of sucrose, which stimulates growth and could thus cause the increased demand for P_i and the onset of the P_i starvation response.

5.5.3   Differential phased siRNA production during development

In *Arabidopsis*, four families of loci are known that generate ta-siRNAs: *TAS1-4*. Only for *TAS3* a biological function has been unraveled. Through regulation of its targets *AUXIN RESPONSE FACTOR 3* (*ARF3*) and *ARF4* it is controlling abaxial-adaxial polarity and vegetative phase change (16, 18, 59, 95, 109, 110). This takes place early on in the development of the leaf. Indeed, the biogenesis of ta-siRNAs from the *TAS3a* locus decreases during leaf development and correspondingly both *ARF3* and *ARF4* increase significantly (fold change 1.4 and 1.7, respectively, 19). The phasing patterns (Supplemental Figure 5.10) illustrate that the phasing becomes progressively more pronounced at one phase, shifted 11 bp from the miR390 cleavage site, which could be due to cleavage of the *TAS3a* transcript by a *TAS3*-produced ta-siRNA that is in phase with the cleavage site, as hypothesized before by Howell *et al.* (58).

ta-siRNAs produced by the *TAS1* and *TAS2* loci are present at high levels in our dataset and their abundance increases during development. It is known that these ta-siRNAs target PPR genes and trigger phased siRNA production from these target loci. We have indeed found two PPR genes that also produce phased siRNAs and both have an increased sRNA expression over time. Also for *TAS1a* and *TAS1b* ta-siRNAs, a shift in phasing is observed during development. Early in development ta-siRNAs are predominantly in phase with the miR173 cleavage site for both loci, but an additional phase becomes progressively apparent when the leaf matures. Remarkably, for all *TAS* loci the phasing of ta-siRNAs is less precise when the expression of small RNAs from the locus increases. For the *TAS3a* locus, the shift corresponds to the position at which a ta-siRNA would cleave the *TAS* transcript. For the other loci, this is not the case. Therefore, other mechanisms could be at play e.g. cleavage by other small RNAs *in trans*.

The identification of islands with a phased sRNA production, yielded a remarkably high number of regions containing transposable elements (14 out of 58). The phased sRNAs were either generated from the body of the transposable element or biased towards the 3′ end. For the majority (13 out of 14), the phased sRNA expression was highest at maturity. For the 24-nt siRNAs, I also observed an increase

during development. This similar pattern suggests that this could be an additional strategy to silence transposable elements and thus protect genome integrity.

5.5.4 Selection for sRNA generating loci

The analysis of differential expression for individual small RNA molecules has its limitations. Low expressed sRNAs need to be removed prior to the analysis to avoid false positives and to prevent a bias towards these sRNAs in the results. This type of analysis is geared towards miRNAs and ta-siRNAs, which are very specific sequences that are often highly expressed. Heterochromatic siRNAs are produced from a broad region and generally exhibit a broad sequence diversity but low expression for each siRNA. This is apparent from the length distribution of small RNAs in our dataset. The fraction of unique 24-nt sequences is very high (Figure 5.2) while when the length distribution is weighed according to the repeat-normalized data 21-nt sRNAs are predominant due to the high expression level of a limited number of miRNAs. Therefore, I deployed a different strategy to analyze heterochromatic siRNAs. For all annotated features of the *Arabidopsis* genome, I calculated the sum of all matching sRNAs and determined loci with differential expression of sRNAs. As miRNA target loci generally do not match perfectly with the regulating miRNA, these are not detected this way, but only regions that potentially generate sRNAs are selected for. I focused on transposons and protein-coding genes.

5.5.4.1 Transposons are progressively silenced during development

In general, transposons were associated with 24-nt sRNAs that have an increased expression level at maturity. These sequences predominantly have an adenosyl at the 5′ end, suggesting that these sequences are indeed heterochromatic siRNAs as this is a hallmark for loading into AGO4 (91), the main effector of this pathway. This type of small RNAs has been associated with DNA methylation, more specifically in CHH sequence context (49, 51, 111). The level of CHH methylation at transposons indeed increased during development, while those of the symmetric CG and CHG methylation decreased. This kind of compensation behavior has been observed before (92, 112). The increase of CHH methylation is however more gradual during development while the siRNAs are relatively low expressed during proliferation and expansion and increase strongly at maturity. This suggests that the siRNA production is downstream of DNA methylation and probably is part of a mechanism to maintain it, rather than that the siRNAs are the initial trigger causing DNA methylation. However, this hypothesis is based upon only three time points and low resolution DNA methylation data and will therefore require further validation. While the profile of methylation has predominance for the body of transposons (Figure 5.16), sRNAs are mainly associated with the start and end of this

region (Figure 5.13). Transcription of a target locus is required for the silencing of that locus (for more details on the biogenesis pathway of heterochromatic siRNAs, see Chapter 4.3.3). At the edges of the transposons, the chromatin structure could be more open and therefore allow a higher level of transcription, resulting in increased siRNA levels specifically there. The reduction in CG and CHG methylation could be the cause for the increase in CHH methylation and siRNA expression. In conclusion, transposons seem to be progressively targeted by heterochromatic siRNAs during development, thus protecting the genome integrity by silencing these elements. The question remains however, what the biological relevance is of the progressive increase of this safeguard mechanism. While an increased protection of genome integrity in germ cells can be linked to the necessity to protect the next generation, its role during vegetative development is unclear. Recently, the importance of monomethylation of histone 3 lysine 27 (H3K27me) in prevention of re-replication of heterochromatic regions was shown (113). Both the H3K27me mark as the production of 24-nt small RNAs are required for silencing of heterochromatic regions. Therefore, the increased activity of the heterochromatic siRNA pathway could play a role in protection against overreplication as differentiation progresses.

5.5.4.2   A subset of protein-coding genes is regulated by heterochromatic siRNAs and DNA methylation

Small RNA distributions for protein-coding genes show two different patterns: sRNAs of 19- to 22-nt are mainly located in the body of this feature while 23- and 24-nt sequences are mainly generated in the gene body early in development. This latter class shows preference for the flanking sequences in the mature stage of development. For protein-coding genes, a depletion of siRNAs at the gene body was observed previously by Cokus *et al.* (92). In this study, five-week old plants were used and this profile is indeed confirmed in our data for mature leaves for 23- and 24-nt sRNAs (Figure 5.10H).

   We observed a positive correlation between the level of sRNA expression associated with the coding sequence (CDS) of protein-coding genes and their mRNA expression level. The distribution of the lengths of sRNAs matching the CDS (Figure 5.7A) shows an equal presence of 19- to 24-nt sequences. The absence of a bias in length distribution suggests that protein-coding genes are generally not regulated by these sRNAs but that these could be degradation products. When we select for genes that do not comply to the general positive correlation between sRNA and mRNA expression, a clear bias towards 24-nt sRNAs becomes apparent (Figure 5.7C). This indicates that the uncorrelated genes are likely downregulated by heterochromatic siRNAs. Also genes that are not expressed above the background in our microarray experiment during leaf development (19), have predominantly 24-nt sRNAs associated with their CDS (Figure 5.7D). This suggests

that (a subset of) these genes could be silenced through the heterochromatic siRNA pathway by RNA-directed DNA methylation (RdDM). The bias towards 24-nt for the uncorrelated genes can only partially be explained by the selection for mainly genes below the background, as the genes in this set that are expressed above the background also show a clear bias towards 24-nt sRNAs and an additional peak for 21-nt is present (Figure 5.7I). Further analysis of this selected gene set (uncorrelated and mRNA expressed above the background) showed that the 21-nt sequences are produced by genes that are targeted by miRNAs or ta-siRNAs which triggers a phased siRNAs production from these loci. The remaining genes are mainly a source of 24-nt sRNAs suggesting these genes are under control of RdDM.

When analyzing all protein-coding genes, CG methylation is mostly present in the gene body with a marked absence around start and stop sites (92, our data). CHG is lower in the gene body compared to the flanking regions. These patterns are stable throughout development. CHH methylation has a similar pattern as CHG methylation but increases over development both in the flanking regions as in the gene body. The uncorrelated genes however, deviate strongly from this general pattern. CG methylation at the CDS is less outspoken but this region is specifically methylated at CHG and CHH. This in contrast to the correlated genes that show a depletion of CHG and CHH methylation at the CDS (Figure 5.15). This confirms our hypothesis that the uncorrelated genes are likely regulated by RdDM through 24-nt heterochromatic siRNAs. Also here, siRNAs only increase strongly at mature phase while methylation increases more gradually during development. Moreover, CHG methylation is specifically high at both the 5' and 3' end of the CDS while CHH methylation only peaks at the 5' end. This could be related to specific sequence characteristics at these sites.

5.6   Conclusion and perspectives

We have profiled the small RNA population in *Arabidopsis thaliana* during leaf development. This revealed differential expression of small RNAs. Our data clearly indicate an integrated role of small RNAs in the regulation of organ development. Early in development, miRNAs involved in nutrient stress response are abundant, suggesting that macro elements are limiting at this stage. During maturation of the leaf, 24-nt siRNAs accumulate and likely silence both transposons as well as specific protein-coding genes. To assess this further, a higher resolution of DNA methylation during development is required. Analysis of DNA methylation at the single gene level would allow to evaluate the correlation with siRNA production at the gene level. Additionally, histone modifications could be assessed to deepen our understanding of the relation between siRNAs, DNA methylation and chromatin modifications. Unraveling this regulatory layer in leaf development could help us understand how plants control growth and development.

5.7 Experimental procedures

5.7.1 Plant material

The first pair of true leaves from *Arabidopsis thaliana* ecotype Columbia was harvested during proliferation, expansion and mature stage, at 8, 12 and 19 days after stratification, respectively. The first leaf pair of 1024, 512 and 256 plants was harvested directly in liquid nitrogen for each of two biological replicates at day 8, 12 and 19, respectively. Plants were grown in petri dishes (Integrid Petri Dish, Beckon Dickinson Labware, Le Pont de Claix, France) in long day conditions (16h light) at 22 °C. The growth medium contained 0.5x Murashige and Skoog (MS) basal salts (Duchefa Biochemie, Haarlem, The Netherlands) (114), 0.5 g/l MES (Duchefa Biochemie, Haarlem, The Netherlands), 1% sucrose (Acros Organics, Geel, Belgium) and 0.8% Plant Tissue Culture Agar (Lab M Limited, Bury, Lancaster, UK). Sterilisation was performed by autoclaving at 1 bar and 120 °C for 20 minutes.

5.7.2 RNA extraction

RNA was extracted using TriReagent (Molecular Research Center, Cincinnati, Ohio, USA) and chloroform (Merck, Darmstadt, Germany). The aqueous layer was purified by repeated extractions with chloroform. RNA was precipitated with isopropanol (Merck, Darmstadt, Germany) and subsequently dissolved in 0.1 TE buffer with 50 mM Tris-HCl (Biosolve B.V., Valkenswaard, The Netherlands) and 50 mM NaCl (Merck, Darmstadt, Germany). Further purification was performed using repeated additions of phenol:chloroform:isoamyl alcohol (25:24:1) pH4.3 (MP Biomedicals, Inc., Illkirch, France) until no material was visible at the interface of organic and aqueous phase. Finally RNA was extracted with chloroform, precipitated using ammonium acetate (BDH Laboratory Supplies, Poole, Dorset, UK) and ethanol (Merck, Darmstadt, Germany) at -80 °C overnight and dissolved in 0.1 TE buffer. Integrity of total RNA was checked on a 2% agarose gel.

5.7.3 Construction of small RNA library

Small RNA molecules of 17 to 30 nucleotides were purified using 17% polyacrylamide gel electrophoresis (PAGE). Radioactive end-labeled oligonucleotide markers of 18 and 24 bp were used to select the appropriate size class. Total RNA samples were spiked with radioactively labeled RNA transcript markers using the Maxiscript kit (Ambion, Austin, Texas, USA). Next, the small RNA library was constructed: First the 3′ adaptor (Illumina, Inc., San Diego, California, USA) was added using T4 RNA ligase (Ambion, Austin, Texas, USA). After PAGE purification of the ligated product, the 5′ adaptor was ligated using T4 RNA ligase (Ambion, Austin, Texas, USA) in an ATP dependent reaction. cDNA was prepared from purified

small RNAs with both adaptors using the SuperScript III First Strand Synthesis System (Invitrogen, Carlsbad, California, USA). cDNA was amplified with Phusion High Fidelity DNA polymerase (Bioké, Leiden, The Netherlands) and PAGE purified.

5.7.4   Sequencing & parsing of the small data

Sequencing of the obtained small RNA libraries was done, in two biological repeats, using the Illumina 1G Genome Analyzer (Illumina, Inc., San Diego, California, USA). Base calling was done using the Illumina software. Sequences of 18 to 30 bp were parsed based on the identification of the first 6 bases of the $3'$ adaptor. The obtained sequences were mapped to the *Arabidopsis* genome based on perfect matching. All samples were normalized for population size to 3 million mapped reads per sample.

5.7.5   Profiling of methylation status

Genomic DNA was extracted from leaf 1 and 2 at 8, 12 and 19 days after stratification using the DNeasy Plant Mini kit (Qiagen, Venlo, The Netherlands) according to the manufacturers recommendations. Bisulfite conversion was done as described previously (92). Library generation and ultra-high-throughput sequencing were carried out according to manufacturer instructions (Illumina, Inc., San Diego, California, USA).

5.7.6   Data visualization

5.7.6.1   Genome browser

The small RNA data was visualized for analysis using a custom-made genome browser written in Processing (processing.org), a Java-based programming language. This custom browser was made to overcome some of the shortcomings that the currently available genome browsers have. We have implemented (1) interactive scrolling, (2) easy visual comparison of the different datasets, in this case developmental stages, (3) the possibility to rapidly switch between different subsets of the data.

The data was smoothed using a sliding window approach using a 1000 bp interval and a step size of 200 bp. This was done for all developmental stages and for every length class of small RNAs separately. The data can be plotted with or without this sliding window approach. The source code is available upon request from the corresponding author.

5.7.6.2   Small RNA distribution plots

The distribution of small RNAs for specific features of the genome (protein-coding genes, microRNAs, transposable elements) was calculated by dividing every feature in 100 bins (each containing 1/100th of the total number of basepairs of the feature). Also 1000 bp upstream and downstream of the feature were incorporated and divided in 100 bins. The small RNA reads were assigned to the appropriate bin according to their starting position. A sliding window approach (30 bins) was used to smooth the data.

5.7.7   Microarray data

To asses gene expression at similar time points of leaf development, microarrays published by (19) were re-analyzed using the *affy* package of Bioconductor (115) applying a custom-made CDF (116).

5.7.8   miRNA candidate and target prediction

Prediction of miRNA candidate sequences was done starting from differentially expressed 20 to 22-nt sequences in our dataset. The potential to form a stem-loop from the genomic region surrounding these sequences match was analyzed using the methodology in Bonnet *et al.* (117).

TAPIR (bioinformatics.psb.ugent.be/webtools/tapir/) was used to identify targets for the candidate miRNAs (68). The 'fast' mode was used to determine possible target genes starting from all annotated genes in TAIR8. This was done using non-stringent parameters (score $\leq$ 8, free energy ration $\geq$ 0.5). The obtained targets were then used as input for the 'precise' mode. This method uses the RNAhybrid algorithm that is conceived to determine miRNA:mRNA duplexes with a high accuracy (118). Here, the default parameters were used (score $\leq$ 4, free energy ration $\geq$ 0.7). Further details on the TAPIR program are available in the 'manual' section on bioinformatics.psb.ugent.be/webtools/tapir/.

Contributions

Frederik Coppens made the small RNA libraries with the help of Kristin Kasschau. The conversion of the raw sequencing data was performed by Carrington lab. All further analyses of the small RNA data were performed by Frederik Coppens except the determination of putative new microRNAs based on the small RNA data, which was done by Eric Bonnet. Suhua Feng made the libraries for genome-wide profiling of DNA methylation status. Ramakrishna K. Chodavarapu analyzed the DNA methylation sequencing data.

References

[1] M. W. Jones-Rhoades, D. P. Bartel, B. Bartel. *Annu Rev Plant Biol* **57**, 19 (2006).

[2] G. Wu, *et al. Cell* **138**, 750 (2009).

[3] L. Gutierrez, *et al. Plant Cell* **21**, 3119 (2009).

[4] E. Marin, *et al. Plant Cell* **22**, 1104 (2010).

[5] A. Telfer, K. M. Bollman, R. S. Poethig. *Development* **124**, 645 (1997).

[6] H. Tsukaya, K. Shoda, G. T. Kim, H. Uchimiya. *Planta* **210**, 536 (2000).

[7] R. S. Poethig. *Science* **301**, 334 (2003).

[8] R. S. Poethig. *Curr Opin Genet Dev* **19**, 374 (2009).

[9] M. J. Aukerman, H. Sakai. *Plant Cell* **15**, 2730 (2003).

[10] X. Chen. *Science* **303**, 2022 (2004).

[11] J.-H. Jung, *et al. Plant Cell* **19**, 2736 (2007).

[12] C. Hunter, H. Sun, R. S. Poethig. *Curr Biol* **13**, 1734 (2003).

[13] A. Peragine, *et al. Genes Dev* **18**, 2368 (2004).

[14] Z. Xie, E. Allen, A. Wilken, J. C. Carrington. *Proc Natl Acad Sci U S A* **102**, 12984 (2005).

[15] M. Yoshikawa, A. Peragine, M. Y. Park, R. S. Poethig. *Genes Dev* **19**, 2164 (2005).

[16] X. Adenot, *et al. Curr Biol* **16**, 927 (2006).

[17] N. Fahlgren, *et al. Curr Biol* **16**, 939 (2006).

[18] C. Hunter, *et al. Development* **133**, 2973 (2006).

[19] G. T. S. Beemster, *et al. Plant Physiol* **138**, 734 (2005).

[20] K. A. Pyke, J. L. Marrison, A. M. Leech. *Journal of Experimental Botany* **42**, 1407 (1991).

[21] P. M. Donnelly, *et al. Dev Biol* **215**, 407 (1999).

[22] J. F. Palatnik, *et al. Nature* **425**, 257 (2003).

[23] J. H. Kim, D. Choi, H. Kende. *Plant J* **36**, 94 (2003).

[24] G. Horiguchi, G.-T. Kim, H. Tsukaya. *Plant J* **43**, 68 (2005).

[25] B. H. Lee, *et al. Plant Physiol* **151**, 655 (2009).

[26] R. E. Rodriguez, *et al. Development* **137**, 103 (2010).

[27] J. A. Banks, P. Masson, N. Fedoroff. *Genes Dev* **2**, 1364 (1988).

[28] N. V. Fedoroff, J. A. Banks. *Genetics* **120**, 559 (1988).

[29] R. Martienssen, A. Barkan, W. C. Taylor, M. Freeling. *Genes Dev* **4**, 331 (1990).

[30] R. Martienssen, A. Baron. *Genetics* **136**, 1157 (1994).

[31] G. N. Rudenko, A. Ono, V. Walbot. *Plant J* **33**, 1013 (2003).

[32] T. Elmayan, H. Vaucheret. *The Plant Journal* **9**, 787 (1996).

[33] F. de Carvalho, *et al. EMBO J* **11**, 2595 (1992).

[34] C. Dehio, J. Schell. *Proc Natl Acad Sci U S A* **91**, 5538 (1994).

[35] F. Dorlhac de Borne, M. Vincentz, Y. Chupeau, H. Vaucheret. *Mol Gen Genet* **243**, 613 (1994).

[36] H. Vaucheret, J. C. Palauqui, T. Elmayan, B. Moffatt. *Mol Gen Genet* **248**, 311 (1995).

[37] F. Meins. *Plant Mol Biol* **43**, 261 (2000).

[38] E. Glazov, *et al. Plant J* **35**, 342 (2003).

[39] W. Boerjan, G. Bauw, M. Van Montagu, D. Inzé. *Plant Cell* **6**, 1401 (1994).

[40] J. C. Palauqui, *et al. Plant Physiol* **112**, 1447 (1996).

[41] C. J. Smith, *et al. Mol Gen Genet* **224**, 477 (1990).

[42] A. R. van der Krol, *et al. Plant Cell* **2**, 291 (1990).

[43] A. R. van der Krol, *et al. Plant Mol Biol* **14**, 457 (1990).

[44] S. M. Cocciolone, K. C. Cone. *Genetics* **135**, 575 (1993).

[45] O. A. Hoekenga, M. G. Muszynski, K. C. Cone. *Genetics* **155**, 1889 (2000).

[46] T. Elmayan, *et al. Plant Cell* **10**, 1747 (1998).

[47] C. Beclin, S. Boutet, P. Waterhouse, H. Vaucheret. *Curr Biol* **12**, 684 (2002).

[48] S. Boutet, *et al. Curr Biol* **13**, 843 (2003).

[49] A. Hamilton, O. Voinnet, L. Chappell, D. Baulcombe. *EMBO J* **21**, 4671 (2002).

[50] S. Jensen, M. P. Gassama, T. Heidmann. *Nat Genet* **21**, 209 (1999).

[51] D. Zilberman, X. Cao, S. E. Jacobsen. *Science* **299**, 716 (2003).

[52] Z. Xie, *et al. PLoS Biol* **2**, E104 (2004).

[53] E. J. Chapman, J. C. Carrington. *Nat Rev Genet* **8**, 884 (2007).

[54] M. Matzke, *et al. Curr Opin Plant Biol* **10**, 512 (2007).

[55] D. Zilberman, *et al. Nat Genet* **39**, 61 (2007).

[56] C. S. Pikaard, J. R. Haag, T. Ream, A. T. Wierzbicki. *Trends Plant Sci* **13**, 390 (2008).

[57] M. Matzke, *et al. Curr Opin Cell Biol* **21**, 367 (2009).

[58] M. D. Howell, *et al. Plant Cell* **19**, 926 (2007).

[59] E. Allen, Z. Xie, A. M. Gustafson, J. C. Carrington. *Cell* **121**, 207 (2005).

[60] R. Rajagopalan, H. Vaucheret, J. Trejo, D. P. Bartel. *Genes Dev* **20**, 3407 (2006).

[61] K. D. Kasschau, *et al. PLoS Biol* **5**, e57 (2007).

[62] N. Fahlgren, *et al. Plant Cell* **22**, 1074 (2010).

[63] G. K. Smyth. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).

[64] F. Leisch. *Computational Statistics and Data Analysis* **51**, 526 (2006).

[65] T. Scharl, F. Leisch. In A. Rizzi, M. Vichi, eds., *Compstat 2006—Proceedings in Computational Statistics*, 1015–1022 (Physica Verlag, Heidelberg, Germany, 2006).

[66] J. C. Carrington, V. Ambros. *Science* **301**, 336 (2003).

[67] M. W. Jones-Rhoades, D. P. Bartel. *Mol Cell* **14**, 787 (2004).

[68] E. Bonnet, Y. He, K. Billiau, Y. Van de Peer. *Bioinformatics* **26**, 1566 (2010).

[69] M. W. Rhoades, *et al. Cell* **110**, 513 (2002).

[70] C. Llave, Z. Xie, K. D. Kasschau, J. C. Carrington. *Science* **297**, 2053 (2002).

[71] K. D. Kasschau, *et al. Dev Cell* **4**, 205 (2003).

[72] H. Fujii, *et al. Curr Biol* **15**, 2038 (2005).

[73] K. Aung, *et al. Plant Physiol* **141**, 1000 (2006).

[74] N. Fahlgren, *et al. PLoS ONE* **2**, e219 (2007).

[75] L.-C. Hsieh, *et al. Plant Physiol* **151**, 2120 (2009).

[76] W. Park, *et al. Curr Biol* **12**, 1484 (2002).

[77] E. Allen, *et al. Nat Genet* **36**, 1282 (2004).

[78] M.-F. Wu, Q. Tian, J. W. Reed. *Development* **133**, 4211 (2006).

[79] M. Schmid, *et al. Development* **130**, 6001 (2003).

[80] N. Lauter, *et al. Proc Natl Acad Sci U S A* **102**, 9412 (2005).

[81] R. Schwab, *et al. Dev Cell* **8**, 517 (2005).

[82] D. Liu, Y. Song, Z. Chen, D. Yu. *Physiol Plant* **136**, 223 (2009).

[83] H. Yamasaki, M. Pilon, T. Shikanai. *Plant Signal Behav* **3**, 231 (2008).

[84] L. Beauclair, A. Yu, N. Bouché. *Plant J* **62**, 454 (2010).

[85] N. Bouché. *Plant Signal Behav* **5**, 684 (2010).

[86] M. J. Axtell. *Methods Mol Biol* **592**, 59 (2010).

[87] S. Maere, K. Heymans, M. Kuiper. *Bioinformatics* **21**, 3448 (2005).

[88] H.-M. Chen, Y.-H. Li, S.-H. Wu. *Proc Natl Acad Sci U S A* **104**, 3318 (2007).

[89] C. Lu, *et al. Genome Res* **16**, 1276 (2006).

[90] A. M. Gustafson, *et al. Nucleic Acids Res* **33**, D637 (2005).

[91] S. Mi, *et al. Cell* **133**, 116 (2008).

[92] S. J. Cokus, *et al. Nature* **452**, 215 (2008).

[93] R. Lister, *et al. Cell* **133**, 523 (2008).

[94] C. Lu, *et al. Science* **309**, 1567 (2005).

[95] M. J. Axtell, C. Jan, R. Rajagopalan, D. P. Bartel. *Cell* **127**, 565 (2006).

[96] M. A. Matzke, J. A. Birchler. *Nat Rev Genet* **6**, 24 (2005).

[97] N. Yoshimoto, *et al. Plant J* **29**, 465 (2002).

[98] Leustek, Saito. *Plant Physiol* **120**, 637 (1999).

[99] K. Saito. *Plant Physiol* **136**, 2443 (2004).

[100] C. G. Kawashima, *et al. Plant J* **57**, 313 (2009).

[101] G. Liang, F. Yang, D.-Q. Yu. *Plant J* (2010).

[102] R. Bari, B. Datt Pant, M. Stitt, W.-R. Scheible. *Plant Physiol* **141**, 988 (2006).

[103] T.-J. Chiou, *et al. Plant Cell* **18**, 412 (2006).

[104] B. D. Pant, *et al. Plant Physiol* **150**, 1541 (2009).

[105] O. Valdés-López, *et al. Plant Cell Environ* **31**, 1834 (2008).

[106] R. Morcuende, *et al. Plant Cell Environ* **30**, 85 (2007).

[107] M. Peng, Y.-M. Bi, T. Zhu, S. J. Rothstein. *Plant Mol Biol* **65**, 775 (2007).

[108] F. Lai, J. Thacker, Y. Li, P. Doerner. *Plant J* **50**, 545 (2007).

[109] L. Williams, C. C. Carles, K. S. Osmont, J. C. Fletcher. *Proc Natl Acad Sci U S A* **102**, 9703 (2005).

[110] D. H. Chitwood, *et al. Genes Dev* **23**, 549 (2009).

[111] I. R. Henderson, S. E. Jacobsen. *Nature* **447**, 418 (2007).

[112] T.-F. Hsieh, *et al. Science* **324**, 1451 (2009).

[113] Y. Jacob, *et al. Nature* (2010).

[114] T. Murashige, F. Skoog. *Physiologia Plantarum* **15**, 473 (1962).

[115] L. Gautier, L. Cope, B. M. Bolstad, R. A. Irizarry. *Bioinformatics* **20**, 307 (2004).

[116] T. Casneuf, Y. Van de Peer, W. Huber. *BMC Bioinformatics* **8**, 461 (2007).

[117] E. Bonnet, J. Wuyts, P. Rouzé, Y. Van de Peer. *Proc Natl Acad Sci U S A* **101**, 11511 (2004).

[118] J. Krüger, M. Rehmsmeier. *Nucleic Acids Res* **34**, W451 (2006).

Supplemental data

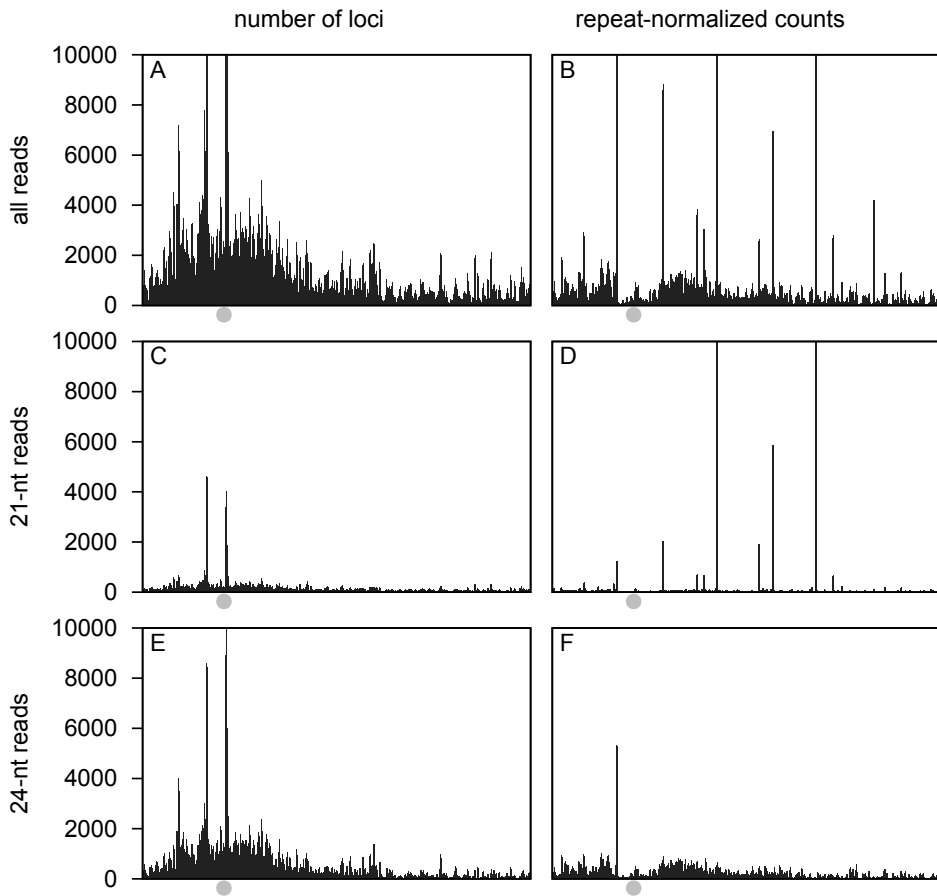number of loci   repeat-normalized counts
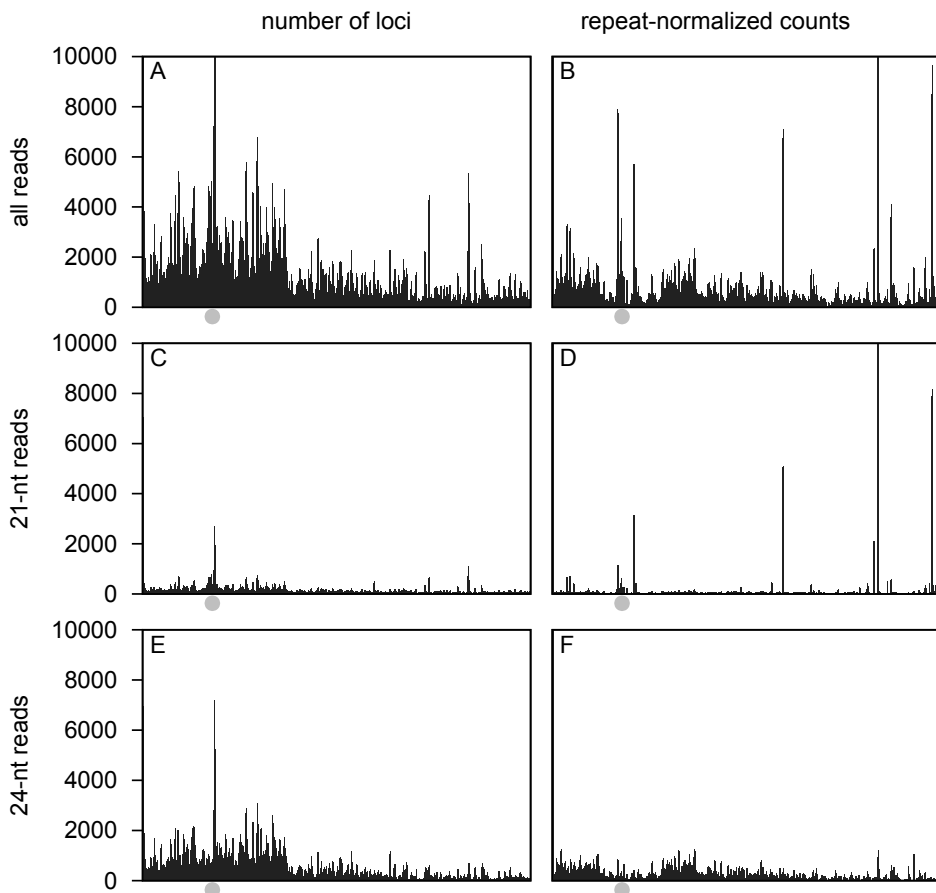
Supplemental Figure 5.1: Genome-wide distribution of small RNA data of mature leaves for Chromosome 2 using a sliding window approach (50,000 bp window, 10,000 bp shift). The number of sRNA generating loci (A,C,E) and repeat-normalized counts (B,D,F) are represented for all sequences (A,B), the 21-nt (C,D) and 24-nt (E,F) size class. The position of the centromere is marked by a grey circle.
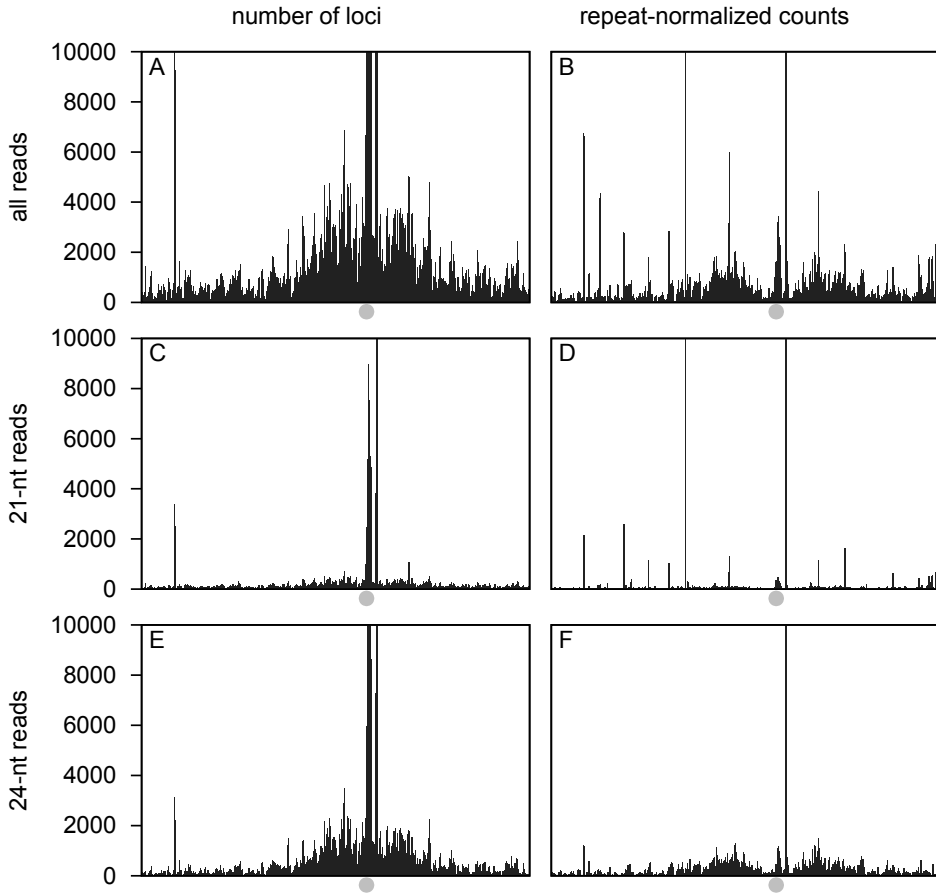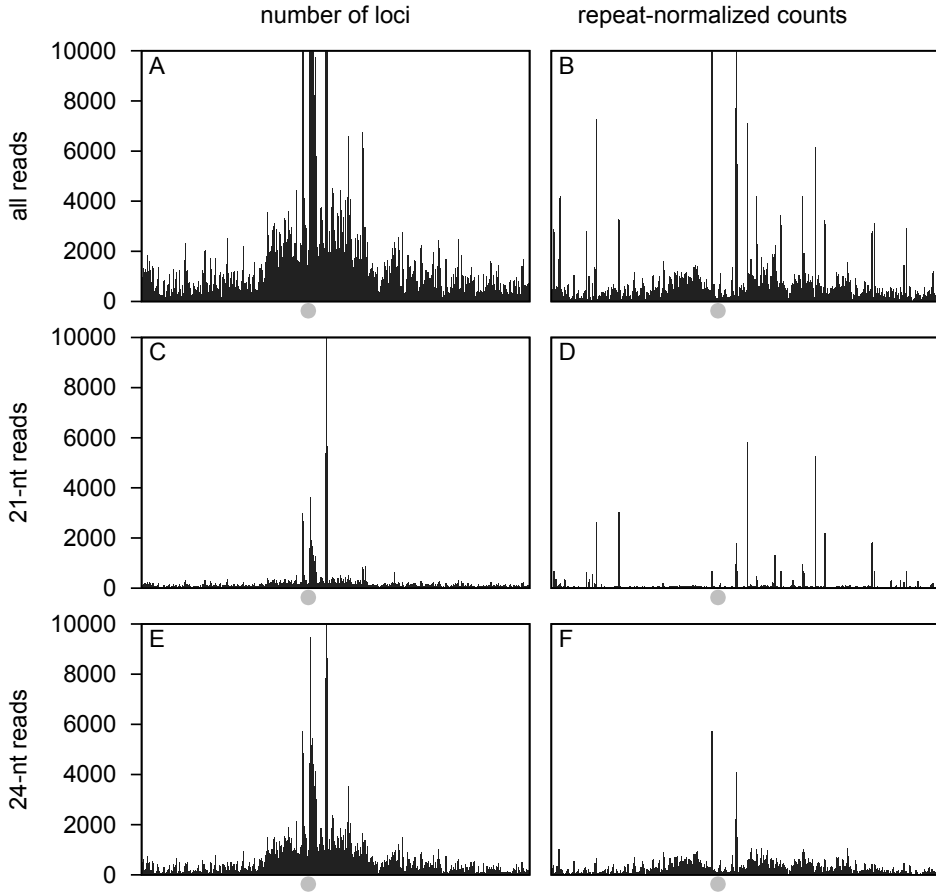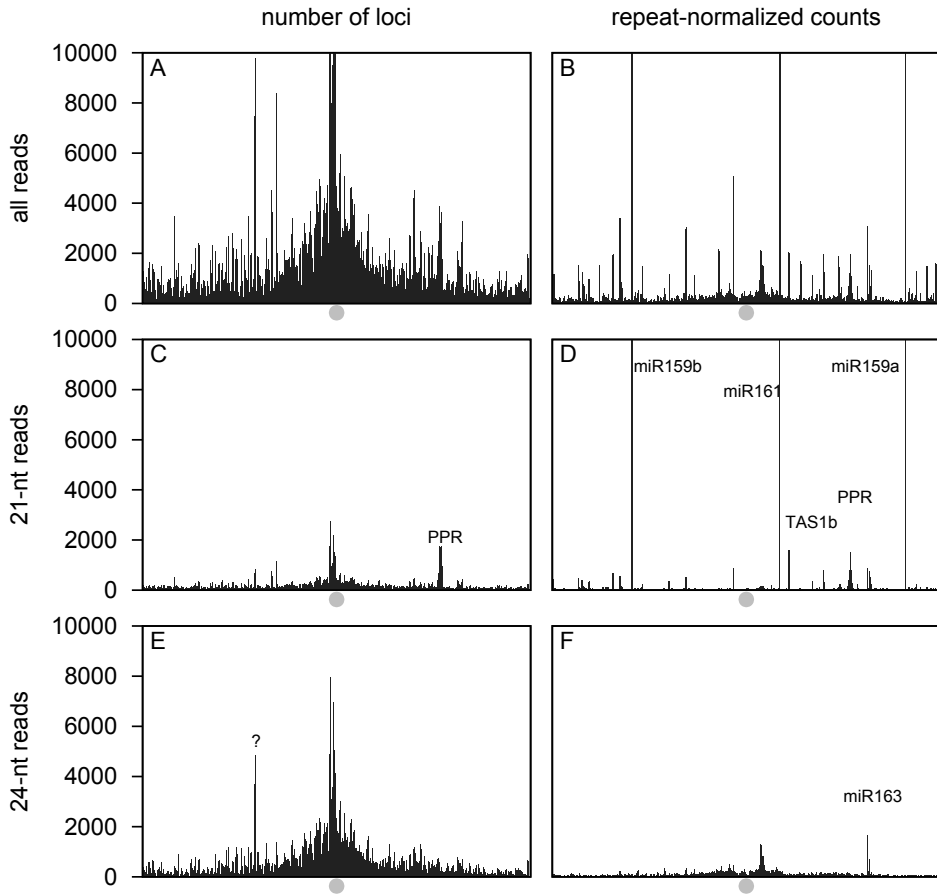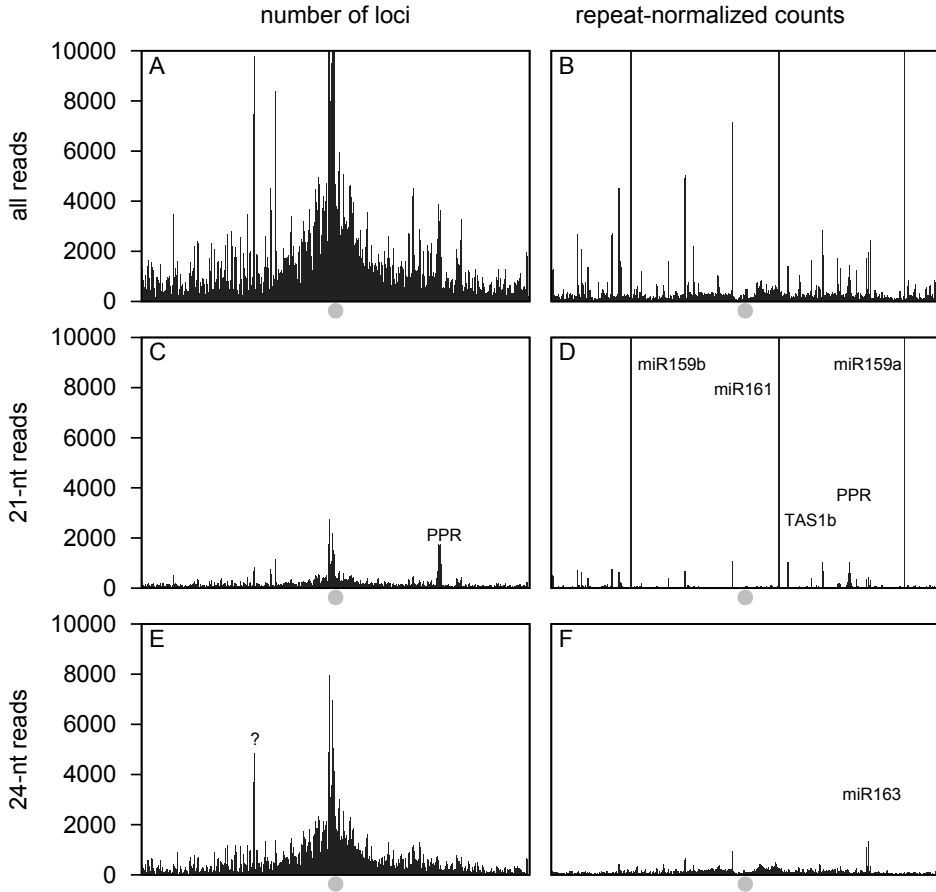
Supplemental Figure 5.2: Genome-wide distribution of small RNA data of mature leaves for Chromosome 3 using a sliding window approach (50,000 bp window, 10,000 bp shift). The number of sRNA generating loci (A,C,E) and repeat-normalized counts (B,D,F) are represented for all sequences (A,B), the 21-nt (C,D) and 24-nt (E,F) size class. The position of the centromere is marked by a grey circle.

Supplemental Figure 5.3: Genome-wide distribution of small RNA data of mature leaves for Chromosome 4 using a sliding window approach (50,000 bp window, 10,000 bp shift). The number of sRNA generating loci (A, C, E) and repeat-normalized counts (B, D, F) are represented for all sequences (A, B), the 21-nt (C, D) and 24-nt (E, F) size class. The position of the centromere is marked by a grey circle.

Supplemental Figure 5.4: Genome-wide distribution of small RNA data of mature leaves for Chromosome 5 using a sliding window approach (50,000 bp window, 10,000 bp shift). The number of sRNA generating loci (A, C, E) and repeat-normalized counts (B, D, F) are represented for all sequences (A, B), the 21-nt (C, D) and 24-nt (E, F) size class. The position of the centromere is marked by a grey circle.

Supplemental Figure 5.5: Genome-wide distribution of small RNA data of proliferating leaves for Chromosome 1 using a sliding window approach (50,000 bp window, 10,000 bp shift). The number of sRNA generating loci (A, C, E) and repeat-normalized counts (B, D, F) are represented for all sequences (A, B), the 21-nt (C, D) and 24-nt (E, F) size class. The position of the centromere is marked by a grey circle.

Supplemental Figure 5.6: Genome-wide distribution of small RNA data of expanding leaves for Chromosome 1 using a sliding window approach (50,000 bp window, 10,000 bp shift). The number of sRNA generating loci (A, C, E) and repeat-normalized counts (B, D, F) are represented for all sequences (A, B), the 21-nt (C, D) and 24-nt (E, F) size class. The position of the centromere is marked by a grey circle.

Supplemental Table 5.1: Overview of miRNA expression during development. The name of the miRNA, the number of perfect hits of the mature sequence on the *Arabidopsis* genome, the repeat-normalized reads (average of two biological repeats) for proliferation (Prol), expansion (Exp) and mature (Mat) phase and the False Discovery Rate (FDR) are listed.

miRNA	hits	Prol	Exp	Mat	FDR
miR156a-f	6	31.53	12.42	56.59	0.12
miR157a-c	3	34.18	11.63	26.99	0.30
miR157d	4	0.28	0.00	1.20	0.07
miR158a	1	94.09	19.61	67.96	0.08
miR158b	1	0.55	0.00	0.00	
miR159a	1	731338.59	455018.60	669344.10	0.64
miR159b	1	247395.64	159710.78	168894.17	0.67
miR159c	1	235.56	123.40	155.91	0.56
miR160a-c	3	1117.87	1005.94	1846.04	0.49
miR161.1	1	20083.28	10146.81	7339.01	0.23
miR161.2	1	699.26	358.16	988.44	0.26
miR162a-b	2	183.79	51.31	140.82	0.15
miR163a	1	469.38	966.01	2343.20	0.09
miR164a-b	2	201.22	273.73	327.46	0.65
miR164c	1	1.79	1.58	5.75	0.24
miR165a-b	2	166.78	301.28	108.45	0.26
miR166a-g	7	128.47	290.40	255.21	0.42
miR167a-b	2	27220.84	17695.84	24941.52	0.70
miR167c	1	1.24	0.40	5.27	0.09
miR167d	1	153.74	1192.53	4587.85	0.03
miR168a-b	2	8300.01	9644.62	5148.08	0.52
miR169a	1	1516.91	2652.59	2550.16	0.51
miR169b-c	2	5.12	2.75	3.35	0.64
miR169d-g	4	0.49	1.01	1.92	0.25
miR169h-n	7	1.60	1.37	3.70	0.25
miR170	1	57.84	13.57	34.03	0.19
miR171a	4	422.90	122.82	63.67	0.06
miR171b-c	2	16.36	9.42	14.60	0.57
miR172a-b	2	485.86	396.22	1114.79	0.21
miR172c-d	2	0.63	0.20	1.67	0.57
miR172e	1	2.64	5.11	27.28	0.05
miR173a	1	14.92	1.97	2.40	0.07
miR319a-b	2	151.37	142.46	65.80	0.29
miR319c	1	191.83	112.35	386.71	0.15
miR390a-b	2	204.47	105.03	220.67	0.36
miR391	1	4.43	31.93	82.39	0.04

Supplemental Table 5.1: Overview of miRNA expression during development.
(Continued)

miRNA	hits	Prol	Exp	Mat	FDR
miR393a-b	2	0.55	0.20	0.24	
miR394a-b	2	833.98	184.12	144.54	0.05
miR395a,d,e	3	12.16	1.18	6.70	0.04
miR395b,c,f	3	17.56	1.44	3.67	0.04
miR396a	1	54.98	276.40	2371.40	0.03
miR396b	1	1158.27	1541.44	5439.77	0.08
miR397a	1	20.25	11.12	18.18	0.72
miR397b	1	16733.65	7054.33	7252.91	0.28
miR398a	1	10.88	7.20	39.25	0.07
miR398b-c	2	831.10	540.21	1379.23	0.29
miR399a	1	8.46	1.04	0.00	0.06
miR399b-c	2	11.58	2.22	0.24	0.03
miR399d	1	5.43	1.04	0.00	0.17
miR399f	1	84.54	20.31	7.18	0.04
miR400	2	4.28	1.11	7.91	0.11
miR402	1	0.00	0.79	0.00	
miR403	1	25.83	10.45	37.83	0.18
miR408	1	5765.25	3232.42	7784.61	0.34
miR447a-b	2	20.08	18.25	15.32	0.86
miR472	1	18.10	9.02	17.24	0.44
miR771	1	1.09	0.92	0.48	
miR773	1	2.49	1.57	0.96	
miR775	1	960.49	221.40	165.61	0.05
miR777	1	15.70	10.20	16.28	0.65
miR779.2	1	19.04	5.63	3.36	0.09
miR780.1	1	0.55	0.00	0.00	
miR781	2	5.40	4.39	15.56	0.16
miR783	1	1.64	0.92	1.92	
miR822	1	62.08	65.55	202.00	0.13
miR823	1	230.54	66.68	61.77	0.10
miR824	1	1091.93	743.82	1900.51	0.30
miR825	1	487.46	336.26	382.49	0.79
miR826	1	0.70	0.00	0.00	
miR827	1	43.29	2.36	11.98	0.05
miR828	1	2.73	0.40	0.96	
miR829.1	1	0.55	0.92	0.00	
miR830	1	1.24	0.00	2.88	
miR833-5p	1	1.09	0.40	4.79	0.14
miR834	1	0.55	0.00	0.00	

Supplemental Table 5.1: Overview of miRNA expression during development. (Continued)

miRNA	hits	Prol	Exp	Mat	FDR
miR835-5p	1	0.00	0.92	0.96	
miR837-5p	2	0.00	0.46	2.16	0.11
miR838	1	16.40	10.86	27.76	0.31
miR839	1	25.34	3.52	4.79	0.07
miR840	1	2.19	1.70	5.75	0.35
miR841	1	80.33	13.46	21.56	0.07
miR842	1	0.70	2.75	10.05	0.07
miR843	1	46.98	248.25	1046.81	0.03
miR844	1	3.18	1.96	4.79	0.59
miR845a	1	5.52	1.04	6.22	0.14
miR845b	1	0.00	0.40	0.00	
miR846	1	190.14	149.98	1293.12	0.04
miR847	1	44.10	53.45	97.68	0.39
miR848	1	92.63	46.26	51.70	0.42
miR850	1	2.73	0.79	0.48	
miR851-5p	1	0.55	1.44	0.48	
miR853	1	4.03	0.92	0.96	0.37
miR857	1	85.72	26.56	73.22	0.16
miR858	1	17.70	10.59	9.58	0.56
miR859	1	2.33	1.97	1.92	0.78
miR860	1	4.43	1.70	3.36	0.41
miR861-3p	1	4.58	4.45	25.36	0.07
miR861-5p	1	1.09	0.52	2.88	
miR862-3p	1	2.88	0.00	0.00	
miR862-5p	1	21.37	5.24	4.79	0.09
miR863-5p	1	3.82	5.10	16.28	0.16
miR864-5p	1	26.28	19.75	47.42	0.41
miR865-5p	1	0.70	0.40	1.92	
miR866-3p	1	0.55	0.40	1.92	
miR869.2	1	1.94	2.36	1.92	0.99
miR870	1	0.00	0.40	0.96	
miR1886.2	1	91.53	96.03	212.57	0.29
miR1886.3	1	0.00	0.00	0.48	
miR1888	1	0.55	0.92	5.76	0.34
miR2111a-b	2	8.55	0.00	1.20	0.04
miR2112-3p	1	1.40	1.44	6.71	0.21
miR2934	1	0.70	0.00	0.48	

Supplemental Table 5.2: All differential sRNAs that match a miRNA hairpin (based on miRBase). sRNAs corresponding to the mature miRNA are indicated in bold (based on miRBase). The name of the miRNA, the length of the mature sequence in basepairs (bp) and the repeat-normalized reads (average of two biological repeats) for proliferation (Prol), expansion (Exp) and mature (Mat) phase are listed.

miRNA	length (bp)	Prol	Exp	Mat
miR156c	21	6.36	17.40	92.84
miR156d	21	16.94	13.61	108.26
miR157a	22	32.25	4.77	1.91
miR157a	21	212.88	33.54	39.00
miR157b	21	212.88	33.54	39.00
miR157b	22	32.25	4.77	1.91
miR159b	22	7.61	2.10	0.00
miR160a	21	21.55	53.69	197.15
miR160c	21	21.55	53.69	197.15
miR161.1	21	6.06	0.00	0.48
miR161.2	21	6.06	0.00	0.48
miR164c	21	1.24	6.01	24.41
miR167d	22	153.73	1192.53	4587.84
miR169d	21	10.21	27.47	108.35
miR169d	20	0.41	1.18	5.90
miR169e	20	0.55	3.27	15.31
miR169f	20	0.41	1.18	5.90
miR169g	21	10.21	27.47	108.35
miR169g	20	0.41	1.18	5.90
miR169g*	21	10.21	27.47	108.35
miR169g*	20	0.41	1.18	5.90
miR169i	22	0.00	9.02	17.73
miR169j	20	0.70	19.08	11.49
miR169l	20	0.97	11.11	13.40
miR169n	20	0.97	11.11	13.40
miR171a	23	34.25	9.80	0.96
miR2111a	21	8.55	0.00	1.20
miR2111b	21	8.55	0.00	1.20
miR319c	20	4.82	0.00	3.83
miR390b	21	8.00	5.88	0.48
miR390b	19	14.91	7.59	0.00
miR390b	21	35.40	13.19	0.48
miR390b	20	83.14	37.49	1.44
miR390b	19	46.92	26.13	0.48
miR390b	20	6.46	1.83	0.00
miR391	21	4.43	31.92	82.39
miR391	21	1.09	11.39	144.59
miR393a	19	1.09	1.89	13.64
miR393a	20	5.20	6.02	52.90

Supplemental Table 5.2: All differential sRNAs that match a miRNA hairpin (based on miRBase). sRNAs corresponding to the mature miRNA are indicated in bold (based on miRBase) (Continued)

miRNA	length (bp)	Prol	Exp	Mat
miR393a	21	0.00	0.39	22.01
miR393a	20	0.00	0.00	12.44
miR393b	19	1.09	1.89	13.64
miR393b	20	5.20	6.02	52.90
miR395a	21	12.15	1.18	6.70
miR395b	21	17.56	1.44	3.67
miR395c	21	17.56	1.44	3.67
miR395d	21	12.15	1.18	6.70
miR395e	21	12.15	1.18	6.70
miR395f	21	17.56	1.44	3.67
miR396a	21	0.00	0.39	6.70
miR396a	22	0.00	0.00	5.74
miR396a	21	0.55	0.00	5.75
miR396a	23	0.00	1.04	12.45
miR396a	21	1.79	4.05	91.92
miR396a	22	28.64	157.23	1462.48
miR396a	20	0.55	2.10	47.37
miR396a	19	1.39	0.00	12.93
miR396a	21	1.24	2.75	42.13
miR396a	21	54.98	276.39	2371.40
miR396a	20	37.08	56.56	283.32
miR396b	20	37.08	56.56	283.32
miR399b	21	11.58	2.22	0.24
miR399c	21	11.58	2.22	0.24
miR399f	21	84.54	20.31	7.18
miR405a	23	0.23	0.48	5.74
miR405b	23	0.23	0.48	5.74
miR405d	23	0.23	0.48	5.74
miR773	22	1742.04	626.20	163.20
miR779.1	22	9.85	5.75	0.00
miR779.2	22	9.85	5.75	0.00
miR783	24	13.58	7.20	0.48
miR783	24	60.43	23.69	4.31
miR783	23	1.83	0.23	0.00
miR783	21	13.41	4.09	0.24
miR783	24	2.93	1.14	0.00
miR783	22	2.52	0.26	0.00
miR783	21	1.79	0.13	0.00
miR783	22	7.22	4.14	0.32
miR783	21	2.95	1.48	0.00
miR783	24	2.62	1.22	0.00
miR783	24	2.62	1.22	0.00

Supplemental Table 5.2: All differential sRNAs that match a miRNA hairpin (based on miRBase). sRNAs corresponding to the mature miRNA are indicated in bold (based on miRBase) (Continued)

miRNA	length (bp)	Prol	Exp	Mat
miR783	22	7.22	4.14	0.32
miR783	21	2.95	1.48	0.00
miR783	21	15.70	1.31	1.68
miR783	24	56.10	21.60	1.92
miR783	24	20.11	7.40	1.44
miR783	21	16.35	2.22	2.63
miR783	24	2.93	1.14	0.00
miR783	21	1.79	0.13	0.00
miR783	22	2.52	0.26	0.00
miR783	23	1.83	0.23	0.00
miR783	21	13.41	4.09	0.24
miR783	24	60.43	23.69	4.31
miR783	21	11.73	1.83	0.00
miR783	24	25.49	8.36	1.44
miR783	24	8.00	11.62	0.00
miR783	24	787.76	540.23	71.84
miR783	24	9.94	4.71	0.48
miR783	22	41.59	14.63	3.83
miR783	24	6.06	1.83	0.00
miR783	21	16.35	2.22	2.63
miR783	24	20.11	7.40	1.44
miR827	21	43.29	2.35	11.97
miR835-3p	23	8.15	1.70	0.00
miR835-5p	23	8.15	1.70	0.00
miR839	22	7.15	0.39	0.48
miR841	21	15.46	1.11	2.16
miR841	24	6.61	0.39	0.00
miR843	20	0.00	1.44	10.05
miR843	21	46.98	248.25	1046.80
miR843	20	0.00	3.93	10.05
miR843	22	2.33	12.82	66.04
miR843	21	29.52	117.29	477.80
miR843	21	1.09	7.33	33.04
miR844	21	1.24	0.00	5.75
miR844*	21	1.24	0.00	5.75
miR846	21	0.62	1.51	19.86
miR846	21	190.13	149.98	1293.12
miR863-3p	21	9.55	30.13	98.63
miR863-5p	21	9.55	30.13	98.63

Supplemental Table 5.3: List of all miRNA loci that are found to be differentially expressed based upon the sum of all sRNA reads.

miR156A (AT2G25095)	miR393A (AT2G39885)
miR156B (AT4G30972)	miR393B (AT3G55734)
miR156C (AT4G31877)	miR394A (AT1G20375)
miR156D (AT5G10945)	miR394B (AT1G76135)
miR156E (AT5G11977)	miR396A (AT2G10606)
miR156F (AT5G26147)	miR396B (AT5G35407)
miR158A (AT3G10745)	miR397B (AT4G13555)
miR161 (AT1G48267)	miR398A (AT2G03445)
miR162A (AT5G08185)	miR398B (AT5G14545)
miR162B (AT5G23065)	miR398C (AT5G14565)
miR163 (AT1G66725)	miR399C (AT5G62162)
miR164/miR164C (AT5G27807)	miR399F (AT2G34208)
miR165/miR165A (AT1G01183)	miR400 (AT1G32582)
miR165/miR165B (AT4G00885)	miR403 (AT2G47275)
miR167C (AT3G04765)	miR405A (AT2G22668)
miR167D (AT1G31173)	miR472a (AT1G12294)
miR169D (AT1G53683)	miR773a (AT1G35501)
miR169E (AT1G53687)	miR775a (AT1G78206)
miR169F (AT3G14385)	miR779a (AT2G22496)
miR169G (AT4G21595)	miR823a (AT3G13724)
miR169I (AT3G26812)	miR827a (AT3G59884)
miR169J (AT3G26813)	miR835a (AT1G76062)
miR169L (AT3G26816)	miR836a (AT2G25011)
miR169N (AT3G26819)	miR838a (AT1G01046)
miR170 (AT5G66045)	miR839a (AT1G67481)
miR171A (AT3G51375)	miR840a (AT2G02741)
miR172/miR172A (AT2G28056)	miR841a (AT4G13564)
miR172/miR172B (AT5G04275)	miR843a (AT3G48057)
miR172D (AT3G55512)	miR846a (AT1G61226)
miR172E (AT5G59505)	miR850a (AT4G13493)
miR173 (AT3G23125)	miR852a (AT4G14504)
miR319/miR319a (AT4G23713)	miR857a (AT4G13554)
miR319/miR319B (AT5G41663)	miR861a (AT3G48201)
miR319C (AT2G40805)	miR862a (AT2G25171)
miR391 (AT5G60408)	
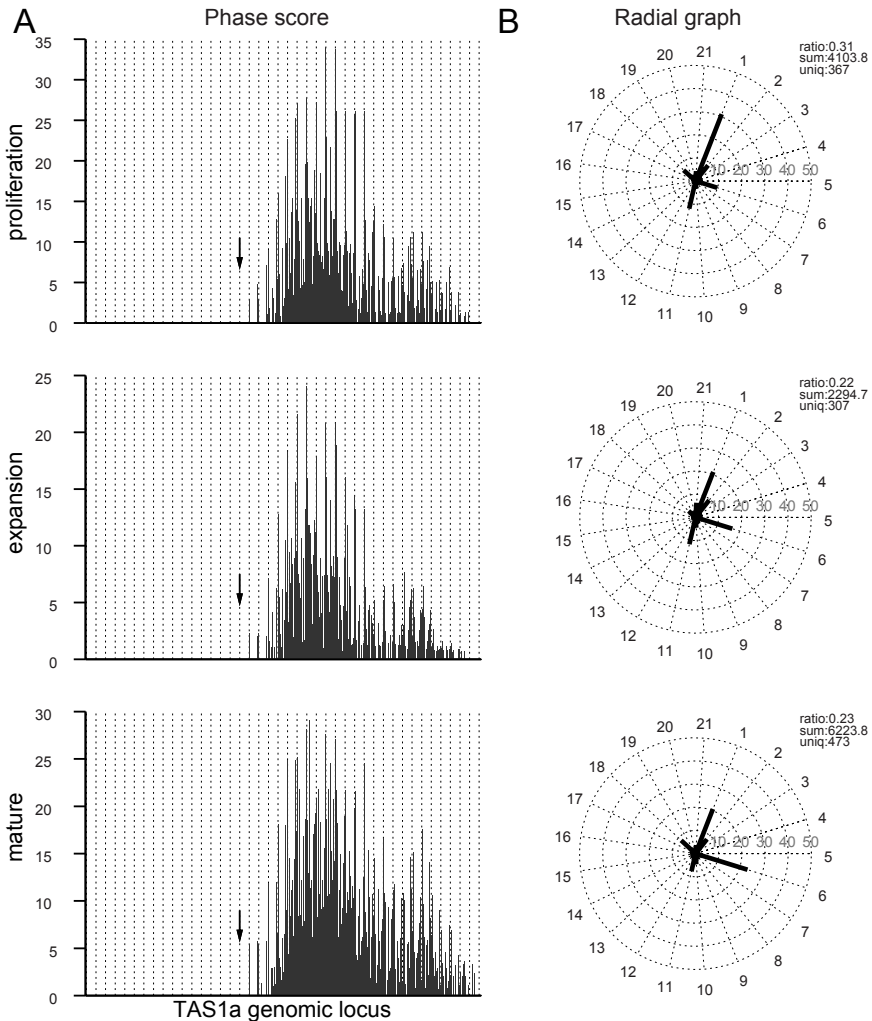
Supplemental Figure 5.7: Visualization of phased siRNA production for the
TAS1a **locus.** (A) The phase score plot (58) is shown for the three developmental stages. The x-axis represents the entire *TAS1a* genomic locus (TAIR8) and the arrow indicates where the miR173-mediated cleavage occurs that initiates the ta-siRNA cascade. The grid is spaced 21-nt and is in register with this cleavage site. (B) The radial graphs (86) for the three developmental stages show the percentage of sRNAs in each of the 21 possible phase registers according to their start position. The highest fraction is indicated (ratio) as well as the total number of reads (sum) and the number of sRNAs (uniq) for each graph.
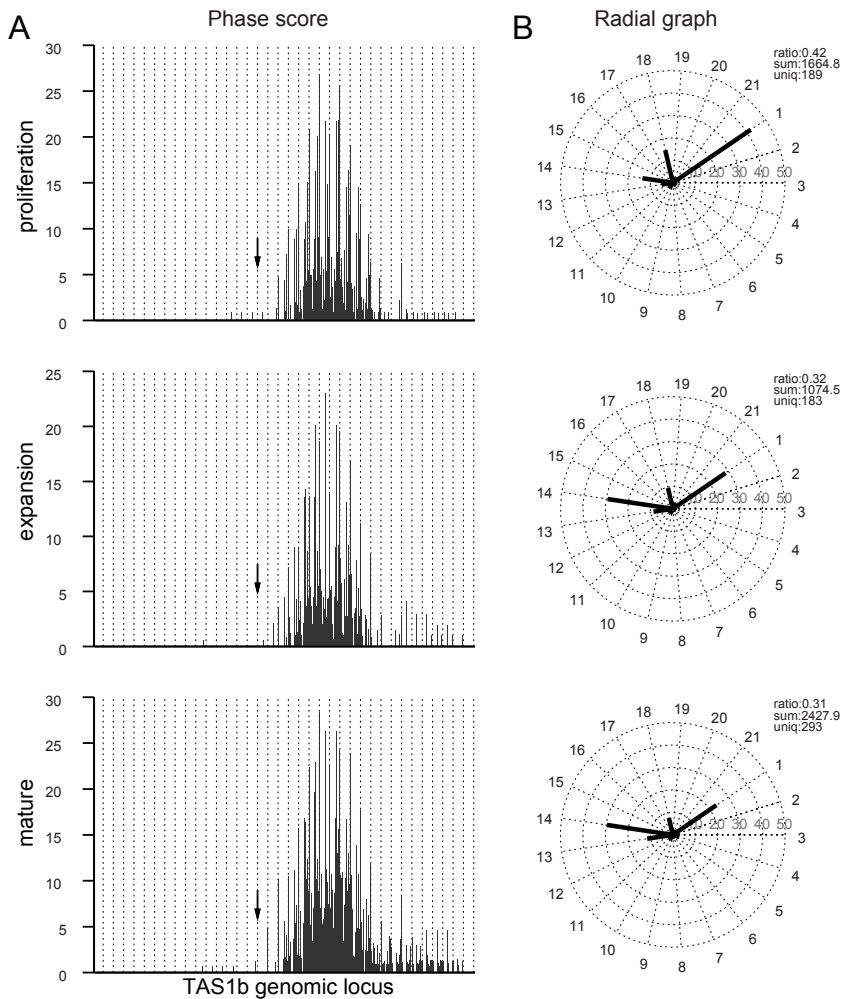
Supplemental Figure 5.8: Visualization of phased siRNA production for the *TAS1b* locus. (A) The phase score plot (58) is shown for the three developmental stages. The x-axis represents the entire *TAS1b* genomic locus (TAIR8) and the arrow indicates where the miR173-mediated cleavage occurs that initiates the tasiRNA cascade. The grid is spaced 21-nt and is in register with this cleavage site. (B) The radial graphs (86) for the three developmental stages show the percentage of sRNAs in each of the 21 possible phase registers according to their start position. The highest fraction is indicated (ratio) as well as the total number of reads (sum) and the number of sRNAs (uniq) for each graph.
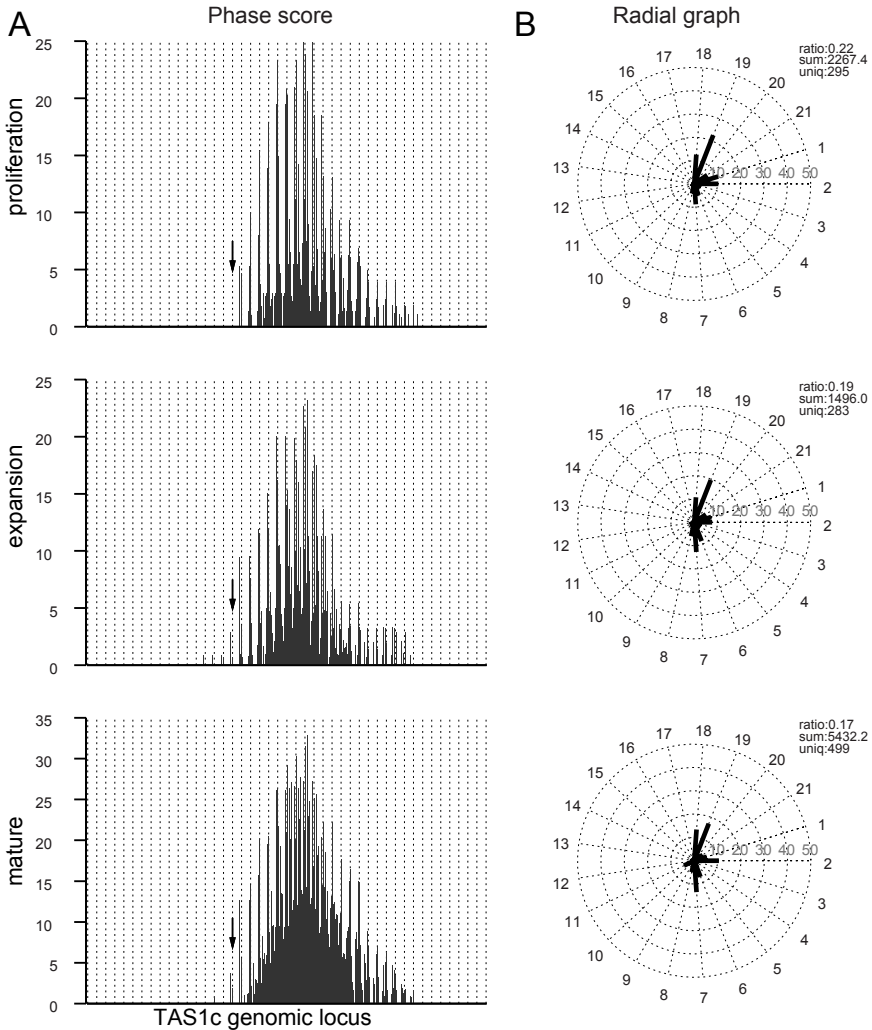
Supplemental Figure 5.9: Visualization of phased siRNA production for the
***TAS1c* locus.** (A) The phase score plot (58) is shown for the three developmental stages. The x-axis represents the entire *TAS1c* genomic locus (TAIR8) and the arrow indicates where the miR173-mediated cleavage occurs that initiates the ta-siRNA cascade. The grid is spaced 21-nt and is in register with this cleavage site. (B) The radial graphs (86) for the three developmental stages show the percentage of sRNAs in each of the 21 possible phase registers according to their start position. The highest fraction is indicated (ratio) as well as the total number of reads (sum) and the number of sRNAs (uniq) for each graph.
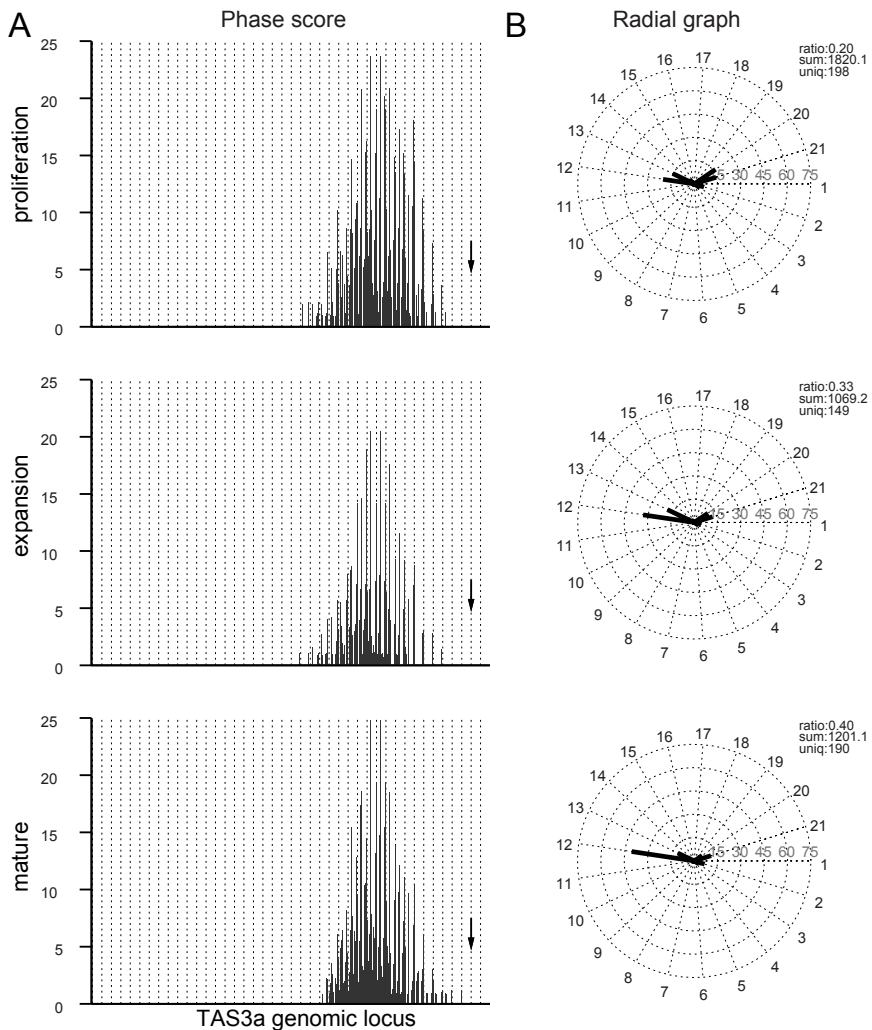
Supplemental Figure 5.10: Visualization of phased siRNA production for the *TAS3a* **locus.** (A) The phase score plot (58) is shown for the three developmental stages. The x-axis represents the entire *TAS3a* genomic locus (TAIR8) and the arrow indicates where the miR390-mediated cleavage occurs that initiates the ta-siRNA cascade. The grid is spaced 21-nt and is in register with this cleavage site. (B) The radial graphs (86) for the three developmental stages show the percentage of sRNAs in each of the 21 possible phase registers according to their start position. The highest fraction is indicated (ratio) as well as the total number of reads (sum) and the number of sRNAs (uniq) for each graph.

Part III

Protocol standardization and automation

6

Kinematic analysis of cell division and expansion

This chapter describes a detailed protocol to perform kinematic analysis of leaves and roots in both monocots and dicots. Thsi was published as a book chapter in the series 'Methods in Molecular Biology' in the issue 'Plant Developmental Biology'.

Reference

Rymen, B., Coppens, F., Dhondt, S., Fiorani, F. and Beemster, G. T. S. (2010). Kinematic analysis of cell division and expansion, *Methods Mol Biol*, 655, 203–27

Contributions

Frederik Coppens and Stijn Dhondt contributed to the description of the kinematic analysis of *Arabidopsis* leaves, Bart Rymen contributed to the part about maize leaves.

Kinematic Analysis of Cell Division and Expansion

Bart Rymen, Frederik Coppens, Stijn Dhondt, Fabio Fiorani, and Gerrit T.S. Beemster

Abstract

Plant growth is readily analysed at the macroscopic level by measuring size and/or mass. Although it is commonly known that the rate of growth is determined by cell division and subsequent cell expansion, relatively few studies describing growth phenotypes include studies of the dynamics of these processes. Kinematic analyses provide a powerful and rigorous framework to perform such studies and have been adapted to the specific characteristics of various plant organs. Here we describe in detail how to perform these analyses in root tips and leaves of the model species *Arabidopsis thaliana* and in the leaves of the monocotyledonous crop species, *Zea mays*. These methods can be readily used and adapted to suit other species in most laboratories.

Key words: Cell division, cell expansion, kinematic analysis, image analysis, epidermal cells, Arabidopsis, leaf growth, root growth, maize.

1. Introduction

In plant sciences, growth is a key characteristic that is widely used to evaluate genotypes and responses to environmental conditions. Towards commercial application, growth rate under optimal or limiting conditions is the primary trait that ultimately determines crop yield. Measurements of growth at the whole plant or organ level are relatively straightforward: they involve measuring size or weight at multiple times and calculation of the rate of increase. However, in many cases the purpose of growth experiments is also to learn about the mechanisms that drive the observed differences in growth rate. A number of strategies have been developed

to this end, each focussing on different aspects of growth regulation. Firstly, classical growth analysis focuses on the evolution of plant (dry) weight and its utilisation in roots, stems and leaves. Here, the main emphasis is on understanding relative growth rates (RGR), which expresses on how efficiently biomass is used to generate more biomass through photosynthesis (1).

A second strategy focuses on how the initiation and growth of individual leaves contribute to the development of the shoot as a whole. This approach is described in detail in **Chapter 7** of this volume.

The aim of the kinematic approaches that are the subject of this chapter is to understand how processes that operate at the cellular level, division and expansion, contribute to differences in rates of growth at the whole organ level. These methods have been pioneered halfway last century (2, 3) and a rigorous mathematical framework was developed based on laws of fluid dynamics a few decades later (4). Since then there has been a gradual increase in the experimental use of these methods, largely supported by the increasing power of computers and the availability of powerful, easy-to-use public-domain image-analysis software like ImageJ (http://rsbweb.nih.gov/ij/).

In this contribution, we describe the methods for analysis of cell division and expansion parameters in three different experimental systems: the first leaf pair of *Arabidopsis thaliana*, the primary root of the same species and the leaf of the monocotyledonous species maize. We focus here in detail on the practical implementation of such measurements. More extended reviews about its conceptual basis and the derivation of the formulae used have been published before (4, 5). Together these three experimental systems provide a comprehensive overview of the possibilities that this approach currently offers and the protocols can easily be adapted to suit other species.

2. Materials

2.1. Analysis of Arabidopsis Leaves

1. Round Petri dishes (12 cm) preferably with a grid on the bottom (*see* **Note 1**) and porous tape for sealing.

2. Murashige and Skoog (MS) medium: Mix $0.5 \times$ MS salts, 10 g/L sucrose, 0.5 g/L 2-(*N*-morpholino)ethanesulfonic acid (MES) and 0.8 g/L plant tissue culture agar in nanopure water. Adjust pH to 5.8 before adding agar. Autoclave medium at 1 bar over-pressure for 20 min. In a flow bench, pour 50 mL of medium into each plate when the agar is still

hot. Allow the agar to set with the lid opened and close only when the medium is at room temperature to prevent excess condensation. Plates can be stored in a plastic bag at 4°C for about 2 weeks.

3. 3.5% bleach solution.

4. Sterile water.

5. 70% ethanol.

6. Lactic acid.

7. Hoyer medium: 80 g Chloral hydrate, 20 mL glycerol and 10 mL water (6).

8. Mounting material: Object slides and cover slips.

9. Binocular microscope equipped with a camera.

10. Microscope with 20 × and 40 × Plan Differential Interference Contrast (DIC) lenses and a drawing tube (*see* **Note 2** on the principle of a drawing tube).

11. Fladbed scanner.

12. Computer running image-analysis software (e.g. ImageJ; Public domain image-analysis software, freely available from http://rsbweb.nih.gov/ij/).

13. Spreadsheet (e.g. MS Excel or OpenOffice (freely available from http://www.openoffice.org/)).

2.2. Analysis of **Arabidopsis** *Primary Roots*

1. For preparation of MS medium and plates, *see* **Section 2.1**.

2. Square Petri dishes (12 cm) and porous tape for sealing.

3. 3.5% bleach solution.

4. Sterile water.

5. Mounting material: Object slides, cover slips and sticky tape (clear).

6. Tooth-pick with short hair (ca. 2 cm, e.g. eyelash) glued on it.

7. Toner powder in Petri dish.

8. Microscopy: For velocity measurements, a simple microscope with a Plan 5 × and 10 × long working distance lens. Ideally this microscope is capable of working in horizontal orientation, so that stage is vertical allowing the roots to grow unperturbed with regards to the gravistimulus. For cell-length measurements, a microscope with 20 × and 40 × plan differential interference contrast (DIC) lenses (*see* **Note 3** on importance of plan lenses). Digital camera for acquiring images on both microscopes.

9. Flatbed scanner.

10. Computer running image-analysis software (e.g. ImageJ; Public domain image-analysis software, freely available from http://rsbweb.nih.gov/ij/).

11. Spreadsheet (e.g. MS Excel or OpenOffice (freely available from http://www.openoffice.org/)).

12. R (Public domain statistical software, freely available from http://www.r-project.org/).

2.3. Analysis of Maize Leaves

1. Ruler.

2. Lactic acid.

3. 3:1 (v/v) absolute ethanol:acetic acid.

4. Mounting material: Object slides and coverslips.

5. 4′,6-Diamidino-2-phenylindole (DAPI) solution.

6. Buffer solution: 50 mM NaCl, 5 mM EDTA and 10 mM TRIS-HCl, pH 7.0.

7. Microscopy: For cell-length measurements: Microscope with $20 \times$ and $40 \times$ plan differential interference contrast (DIC) lenses. For meristem length measurements: An epifluorescence microscope with an excitation filter at wavelength 350 nm and emission filter at wavelength 420 nm, equipped with $20 \times$ and $40 \times$ lenses. Digital camera connected to a personal computer for acquiring images on both microscopes.

8. Computer running image-analysis software (e.g. ImageJ; Public domain image-analysis software, freely available from http://rsbweb.nih.gov/ij/).

9. Spreadsheet (e.g. MS Excel or OpenOffice; freely available from http://www.openoffice.org/).

10. R (Public domain statistical software, freely available from http://www.r-project.org/).

3. Methods

3.1. Analysis of *Arabidopsis* Leaves

After initiation at the shoot apex, leaves of dicotyledonous plants go through subsequent stages of cell division and expansion before they reach maturity. Although the transitions between these stages occur in a tip to base gradient (7), the approach outlined here calculates average rates of division and expansion across the entire leaf. The analysis is based on two sets of primary data: leaf blade area and cell area. For the first leaf pair of *A. thaliana* ecotype Columbia 0, we measure these variables from 2 days after germination (DAG) until 22 DAG. This period

spans from leaf emergence until maturity under our environmental conditions (growth chamber: 21°C, fluorescent (cool white) light 80 μE/m^2/s, 16/8 h day/night and shelves cooled at 19°C to prevent condensation against the lids of the Petri dishes).

3.1.1. Preparation of Plant Material

1. Sterilise seeds for 15 min in 3.5% bleach solution in 1.5-mL microcentrifuge tubes and wash three times with sterile water. After the last wash leave the water in the tubes.

2. Sow seeds using a 20-μl pipette with tips of which the top is cut 2 mm with a razor blade to increase the opening. Aspirate about a dozen seeds with enough water to dispense them. Plate the seeds individually on the agar. In this step, it is important to gently touch the medium with the tip in order to break the surface of the agar layer, generating a small crack that allows the root to penetrate the medium (*see* **Note 4**).

3. Leave the plates open in the flow for a few minutes to let excess water evaporate.

4. Close the Petri dishes using porous tape that allows gas exchange.

3.1.2. Mounting on Slides

1. Harvest whole plants for the earliest stages of development. Dissect leaves once the first true leaves have formed a petiole (ca. 5 DAG).

2. Place the material in 70% ethanol to remove chlorophyll overnight.

3. Transfer material to lactic acid and incubate overnight for clearing.

4. Mount material on a microscopic slide using the lactic acid as mounting medium (*see* **Note 5**). For cell analysis, the abaxial epidermis is used; therefore, it is important to place the leaf on the microscopic slide with the adaxial-trichome-containing side down. For the earliest stages, expose the first leaf pair by gently pulling apart the cotyledons before placing the coverslip.

3.1.3. Leaf Area Measurements

1. Acquire images for about 10 leaves per genotype/treatment and time point using a binocular microscope equipped with a camera. Adjust the magnification to the size of the leaves under examination (×1.25–6.3) and make sure to photograph a ruler at the same magnification for image calibration (*see* **Note 6**).

2. Measure the leaf area using an image-analysis program, e.g. ImageJ. Calibrate the program using the ruler image, which allows converting pixels to corresponding distances in mm.

3. Use the 'polygon selection' tool to outline the leaf blade and measure the area. Because the leaves of Arabidopsis often have a curved surface, the leaf edge is sometimes folded double. To accommodate for this, it is necessary to measure the area of the folded parts of the leaf that can be easily recognised in addition to the outline of the whole leaf area.

4. Copy the measurements into a spreadsheet program and calculate the total leaf blade area. For the kinematic analysis, select at least 5 median-sized leaves and calculate average area and standard errors.

3.1.4. Measurements of Cell Area and Stomata Number

1. Visualise the cells in the cleared samples with a DIC microscope. Depending on the size of the cells, use a 20, 40 or 63 × magnification. Use preferably the abaxial epidermis of the first leaf pair for the analysis (*see* **Note 7** for a discussion on cell types). As there is a gradient in cell development from leaf tip to bottom, analyse the cells at two positions of the leaf, at about a quarter from the tip and bottom of the leaf and halfway between the leaf margin and the mid-vein. Avoid regions directly above the vasculature, because the epidermal cells can be more elongated and harder to visualise due to the optical disturbance of the underlying dense vascular strands.

2. Outline the leaf epidermal cells using a drawing tube (for an alternative *see* **Note 8**). To gather sufficient data, draw about 50 cells for each epidermal area examined. To minimise a bias in cell sizes, avoid the edge of the paper and avoid drawing a disproportionately high number of small cells. This problem may arise because small cells more easily fit on the paper compared to bigger ones. Draw a calibration grid for each magnification to allow the correct scaling.

3. The number of stomata per drawing is counted manually and recorded.

3.1.5. Processing of the Images

1. When the drawings are finished, they need to be digitised. Correct potential artefacts by closing cell walls that show gaps and erasing cell walls at the edge of the drawing, which do not delineate full cells. Also, avoid cell walls that touch the edges of the paper. Record the origin of the drawing by annotating next to the drawn area: genotype/treatment, time point, position in the leaf (tip or bottom) and magnification used for visualisation (for a typical example of the epidermis and a drawing, *see* **Fig. 14.1**).

2. Scan the drawings. Depending on the available scanner, adjusting the settings might take some trial and error. Scanning at 300 dpi and saving the scanned images as jpeg file are generally sufficient.
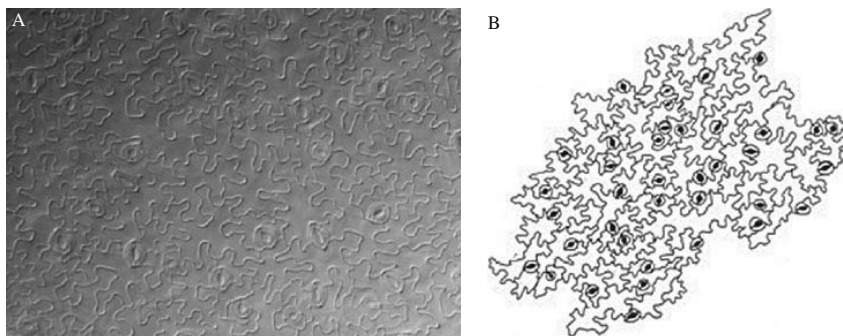
Fig. 14.1. **Image processing of the abaxialf epidermis of *Arabidopsis thaliana* leaves.** A. DIC image of the abaxial epidermis of cleared leaves. B. Typical drawing of cell outlines resulting from a similar specimen.

3. Process the images using image-analysis software (here described for ImageJ). Set the calibration by opening the appropriate calibration image and use the straight line tool to select a length of the grid. Use the 'Analysis/Set Scale' command and fill in the known distance and the units. Apply these settings for all images that are subsequently opened. Select the option 'global' under 'Analysis/Set Measurements' to verify that 'Area' and 'Display label' are selected.

4. Open an image and convert it to 8-bit greyscale under 'Image/Type.'

5. Apply a threshold using 'Image/Adjust/Threshold' that allows the cell walls to be best visualised.

6. Apply a closing step 'Process/Binary/Close' to close small gaps in the drawing possibly created by the thresholding.

7. Use the 'Magic Wand' tool and click to the left of the drawn area. This selects the outline of the entire area of the cell drawing. If the selection includes the interior of some cells not all cell walls are closed. Close gaps by selecting a line across the opening and clicking 'Edit/Draw' or 'Ctrl-D.' As this is tedious, check drawings beforehand for these artefacts.

8. Measure the drawn area by clicking 'Analyze/Measure' or 'Ctrl-M.'

9. Count all cells in the drawing. To easily keep track of which cells were counted, select every cell with the 'Magic Wand' tool and click 'Edit/Fill' or 'Ctrl-F' to fill it.

10. Copy the obtained data from the Result Window to a spreadsheet program (e.g. MS Excel).

3.1.6. Calculations

Following the above procedures, the primary data for each leaf consist of the leaf blade area and for two drawings (tip and base

of leaf): number of cells, total area of the drawn cells and number of stomata. From these derived parameters can be calculated as follows:

1. Determine average leaf area as the average (and standard error) of all leaf areas per genotype/treatment at each time point (Units: mm^2). Due to the exponential nature of the growth process, present this parameter on a log-scale.

2. Divide the total area of the drawn cells by the number of cells in it (pavement cells + guard cells (2 per stoma)); this yields the average cell area in that drawing. Calculate an average (and standard error) of all the cell areas for all the leaves per genotype and time point.

3. Calculate for each drawing the Stomatal Index (*SI*) by dividing the number of guard cells (number of stomata (*S*) multiplied by two to correct for the presence of two guard cells) by the total number of cells (pavement cells (*PC*) + guard cells):

$$SI = \frac{2S}{(PC + 2S)}.$$

3. Divide the leaf area by the average cell area from the same leaf to obtain the number of cells per leaf. Calculate averages (and standard error) of the number of cells per leaf for all the leaves per genotype/treatment and time point. Due to the exponential nature of the division process, represent leaf numbers in log-scale with base 2.

4. The leaf expansion rate (LER) is the derivative of leaf area over time (on a $\log_n$ scale). Calculate this derivative by using the LocPoly algorithm or use a spreadsheet such as MS Excel. An R-script for the LocPoly or an example of an Excel sheet with those calculations (*see* **Note 9**) can be obtained from the corresponding author of this chapter.

5. Average cell division rate is the derivative of the cell number data with respect to time. The calculation is similar to the LER, using the $\log_2$ of the number of cells (cell/cell/h). Calculate cell-cycle duration as the inverse of cell-division rate. For additional information on effects on cell-cycle phase duration, these data can be combined with flow-cytometry measurements (*see* **Note 10**).

3.2. Analysis of Arabidopsis Primary Roots

The analysis consists of two parts, measurement of the velocity profile and cell-length distribution, and these need to be combined in order to calculate cell-division rates. In order for this to work, it is imperative to work relative to a common reference point. In the root typically the quiescent centre (QC) is a

convenient point as it forms the origin of all cell files. Unfortunately, the QC is not recognisable in surface view of the root, which is needed for the velocity-profile measurements. Therefore, these measurements are made relative to the tip of the root and later corrected with the distance between the QC and tip of the root as measured on a median view using DIC optics.

3.2.1. Plant Growth and Measurements of Whole Root Growth Rates

To allow for roots to grow and be accessible for observations, grow plants in Petri dishes with agar-solidified growth medium. Although different nutrient mixes can be used, according to our experience full-strength MS salts are a convenient choice. First, this mix is available as ready-made salts so that no mixing of stock solutions is required. Second, the characteristics of the plants are good. The shoots look vigorous and green and although the roots grow slower than on, for example, Hoagland media, most ecotypes we tested grow at a steady rate from early after germination rather than accelerating (8), which simplifies the analysis significantly.

1. Sterilise seeds for 15 min in 3.5% bleach solution in 1.5-mL microcentrifuge tubes and wash three times with sterile water. After the last wash, leave the water in the tubes.

2. Sow seeds using a 20-µl pipette with tips from which the top 2 mm is cut with a razor blade to increase the opening. Aspirate about a dozen seeds with enough water to dispense them. Place the seeds individually on the agar. Pick up around 10 seeds and sufficient water and distribute 8–10 seeds at equal distance at about 1 cm from the top of the plate. Make sure not to touch the agar with the tip as breaking the agar surface will result in roots entering the agar rather than growing along the surface. When accidentally multiple seeds are sown at the same position, remove the additional ones with the pipette tip. Adding a bit of additional water often simplifies this task. Use about 10 roots for each line or treatment and grow another 20 or so in parallel plates for independent measurements of root growth rates.

3. Seal plates with porous tape to allow air-exchange and place under an angle of 80–85° in a growth chamber. As steady-state conditions are assumed, it is best to use continuous light (about 60–80 $\mu E/m^2/s$ photosynthetic active radiation) and temperature (ca 21°C) conditions.

4. Start measurements of root growth rates as soon as most seeds have germinated. Make each day a small scratch at the back of the plate perpendicular to the growth direction of the root marking the position of the tip. Keep track of the time when these marks are placed. Continue these measurements for at least 2 weeks.

5. Scan the plates using a flatbed scanner (*see* **Note 11**). Make sure always to use the same scanner resolution and also scan in a ruler as a scale for calibration.

6. Open the scanned images of plates with image-analysis software such as ImageJ. Calibrate the scale of all images using the image of the ruler.

7. Use the 'Freehand' tool to measure the distance between the marks.

8. Transfer these data for each root into a spreadsheet program. Calculate average root-elongation rates for each root at a particular day by dividing the measured distance between marks by the time difference between when the marks were made using a time scale of hours.

9. Calculate average velocity and standard error for each time interval across all roots. Use the growth rates calculated this way as a first screen to decide on which lines or treatments to concentrate on in more detail. Typically, preliminary experiments are done for this purpose in which also mature cortical cell length is measured (about 20 cells per root).

10. Divide root-elongation rate by mature cell length to obtain cell-production rates. As a general rule, differences for root-elongation rate, mature cell length or cell-production rate need to exceed 10% for kinematic analysis to be feasible, because too many replicates need to be measured to obtain significant differences for individual kinematic parameters. For kinematic analysis, measure ca 20 undisturbed roots growing on separate plates to control for effects of the experimental manipulations.

3.2.2. Velocity Measurements

The aim of these measurements is to determine the velocity at which cells are moving in function of position along the root. As a basis for this, it is necessary to make several images of the root over a specified time interval. Although the analysis can be done semi-automatically (*see* **Note 12**), we will describe here how to do this manually. This approach requires least setting up, is probably still the most reliable and increases understanding of the principles involved. The automation could be a valuable next step for those implementing the analysis for more routine usage.

1. Sprinkle the surface of the root with contrasting and recognisable marks. A convenient substance for this is toner powder that is used in copiers and laser printers. Pour some of this powder from a fresh cartridge into a small Petri dish. Open the plate(s) with roots to be analysed under a binocular and dip an eyelash or other fine hair glued to a toothpick in this powder. Then hold the hair above the root and

gently tap it to release the powder evenly, but thinly on the surface from the root tip to the region where mature root hairs appear. It is no problem to spill on the agar, but do not apply too much on the root as this will interfere with the cell-length measurements in the second stage. A good amount is shown in **Fig. 14.2**. After all the roots on a plate have been marked, quickly reseal the plate and bring it back to the growth room, leaving it to recuperate for an hour.
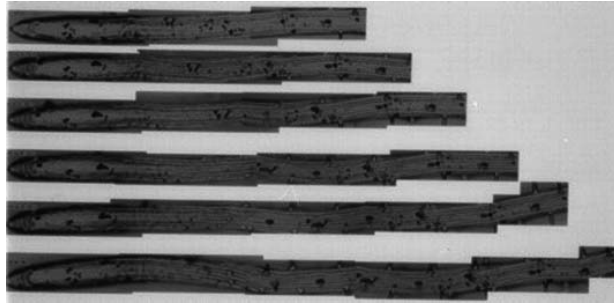
Fig. 14.2. **Time-lapse observations for a single Arabidopsis root tip.** Successive images were taken at intervals of 1 h and composed of a set of overlapping frames. The dark spots are larger aggregates of toner powder, showing the appropriate density at which they need to be applied. They also illustrate the rate of displacement in different parts of the root.

2. Mount the whole plate, preferentially on a horizontally oriented microscope fitted with a camera so as to keep the orientation of the plate vertical (*see* **Note 13**). Using a 5 or 10 × magnification make a series of overlapping images starting from the tip of the root till well into the region of fully developed root hairs, focussing on the root surface and toner particles. It is convenient to orient the root horizontally in the resulting images, so adjust the camera angle to achieve this.

3. Repeat this for all roots on the plate and make at least three sets of observations at hourly intervals. Also make sure that an image is made of a calibration grid using the same magnification.

4. Use the obtained images to make a composite image by stitching the individual images together (*see* **Note 14**). The individual particles form small aggregates with recognisable shapes that can be found back between two images of the same root taken at different times (*see* **Fig. 14.2**).

5. Draw a series of vertical lines through specific particles in a pair of images that are open at the same time. The distance between these lines should ideally be less than half of the root diameter. At some point along the root, cell walls and

root hairs become convenient natural landmarks that can (also) be used.

6. Open the resultant composite images in the image-analysis software, which is calibrated using the image of the calibration scale. Measure the distance from the tip of the root to the first mark along the mid-line of the root, followed by the distance between subsequent marks (*see* **Note 15**).

7. Repeat this procedure for each of the three images of the same root taken at hourly intervals so that the displacement of the reference marks can be determined.

8. Copy the distances from the tip to each particle into the spreadsheet and calculate the cumulative distances from the tip to define the position of each particle in every image.

9. Calculate the average position, $X_{\text{avg}} = \frac{X_{t_1} + X_{t_2}}{2}$, with X_t the position at a particular time t as well as velocity $V(x) = \frac{X_{t_2} - X_{t_2}}{t_2 - t_1}$, for all particles for each pair of images. Combine the x, $V(x)$ data series for the subsequent time intervals to increase the number of observations.

10. To obtain X values relative to the QC, subtract the distance between QC and root tip. These data will be needed in combination with cell-length data to calculate division rates.

11. Determine the velocity at specific, equally spaced positions along the root to allow averaging between roots and further calculations by combining these measurements with the cell-length data. A local polynomial smoothing procedure has been developed (9) and implemented as an R script that is available upon request from the corresponding author. This script gives a series of data with increasing smoothing. Smoothing should just remove the local noise, but not affect the overall curve.

12. Use the script to calculate the local derivative of the velocity curve, which equals the relative rate of cell expansion RLER or $R(x)$. Average the data for $V(x)$ and $R(x)$ (at x) between roots and calculate standard errors.

3.2.3. Cell-Length Measurements

While the velocity measurements give a detailed insight into the spatial distribution of cell expansion along the root tip as a whole, it lacks a direct link to the underlying cellular components. The same results would be obtained if a single giant cell was analysed instead of a root of same proportions consisting of thousands of cells. In order to get this cellular perspective, we need to measure the length of cells in function of position along the axis. Although typically this is done for a single cell type, like we will describe here, it is possible to compare multiple cell types (9).

The Arabidopsis root tip has the advantage that it has a small diameter (ca. 140 μm) because only a single layer of each tissue (10) is present. This allows for high-quality microscopy images using a whole-mount procedure that saves time and precludes potential problems such as shrinkage during an embedding procedure.

Although the mounting procedure is relatively simple, care needs to be taken to avoid the collapse of cells, particularly those in the meristem, which can be caused by both physical and osmotic pressures. Therefore, we recommend the use of spacers (a strip of sticky tape on both sides of the object slide to rest the coverslip on) and a solution that is isotonic to the growth medium of the root. The easiest for the latter is to make the same solution as used for the agar medium without adding the agar.

1. Cut the roots that were used for the velocity observations described above from the shoot with a scalpel or razor blade (make sure to keep track of individual plants to be able to combine velocity and cell-length data for each plant). Cut at least a centimetre of root and handle close to the cut to avoid damaging the tip where the measurements will be made. Place a droplet of the mounting solution on the surface of the object slide in the middle between the two spacers. Place the root in the droplet and lower the cover slide gently onto it.

2. Place the specimen under a microscope fitted with DIC optics. Find the tip of the root at low magnification (5 or 10 ×) and switch to 40 × for DIC imaging. Orient the root horizontally in the image; we will assume the tip to be pointing to the left. Adjust the microscope settings to optimise the image focussing on a median section in the region of the tip where the quiescent centre (QC) cells can be found. From the QC, files of cells can be seen radiating out.

3. Take a photo focused on the QC and make sure that the tip of the root cap is in the same frame. This image is required for determining the offset between the tip of the root and the QC in the velocity data (as described above). Capture and save the image for analysis at a later stage.

4. Typically, the cortical cells are easiest to analyse because they are the biggest in diameter; therefore, it is advisable to focus on those although other types can also be used. For the next frame, move the microscope so that the QC is close to the left edge of the frame. Focus on the cortical files so that a clear row of cells can be seen on at least one side of the root. Capture and save this image. In addition to this medium plane, it is possible to image additional cells by zooming out (without changing X, Y position on the stage) to the tangential plane through the cortical cell layer.

5. Make a series of images by moving along the root from the tip to the region where the root hairs are fully mature, capturing both the median and tangential plane whenever possible. Make sure that these images overlap by 10–20% to allow them to be aligned and stitched for further analysis.

6. Before making cell-length measurements, combine/stitch the images from each plane of view into a single image so that each file can be followed over the length of the root. Use the same approach as outlined for the images of the marks on the surface. When measuring cell sizes, it is important to bear in mind that they need to be positioned relative to the QC. For the median section, this is straightforward: position the line tool to select the length of the first cell adjacent to the QC and measure it. Then select the next cell in the file (in ImageJ by moving only one side of the line tool, leaving the side that marks the wall between the first and second cell in place) and measure its length. Continue doing this until the end of the file is reached (*see* **Note 16**).

7. Because this QC is only visual on the median section, use this image to measure the distance from the base of the QC to the left border of the image. On the tangential images, measure the distance from the first unambiguous cell to the left border of the image and subsequently measure all cells in this file.

8. Copy all cell lengths and distances that do not refer to a cell length for each file separately into a spreadsheet program.

9. Calculate the distance from the QC to the midpoint of each cell. For median files, add the total length of all preceding cells and sections that are not cells + 0.5 × the length of the cell itself.

10. Change the calculations of these positions from formulae into values (In MS Excel by 'Edit/Copy,' 'Edit/Paste Special: Values') and then remove the data for the sections that do not refer to cell lengths.

11. Combine the data from all files into a single pair of columns and sort for ascending position (x).

12. Use these data for interpolation and generation of equally spaced data using the same Locpoly routine in R that was used for the Velocity data (*see* **Section 3.1.6** and step 4). Calculate average cell length and standard errors between replicate roots from the interpolated data.

3.2.4. Calculation of Cell-Division Rates

1. For the calculation of cell-division rates, combine the equally spaced data for velocity and cell length and calculate flux ($F(x)$), the number of cells that are passing at a particular

position, by dividing local velocity by cell length: $F(x) = \frac{V(x)}{l(x)}$. In steady-state conditions (for non-steady state *see* **Note 17**), the derivative of the flux function $\left(\frac{dF(x)}{dx}\right)$ denotes the local rate of cell production per unit of length $(P(x))$. Calculate this derivative using 5-point equations (11).

2. Calculate local cell-division rates $(D(x))$ from the cell-production rates by dividing by local density, which equates to multiplying them with the local cell length: $D(x) = l(x) \times P(x)$.

3.2.5. Calculation of Kinematic Parameters

Based on the calculation of spatial data for cell size, velocity, cell expansion and division rates, it is relatively straightforward to calculate organ scale data for each replicate root.

1. Determine the size of the growth zone (L_{gz}) as the position where $V(x)$ becomes maximal and thus its derivative, $R(x)$, goes to 0.

2. Determine the size of the meristem (L_{mer}) as the position where $F(x)$ becomes constant, hence where $D(x)$ becomes 0.

3. Determine the size of the elongation zone (L_{el}) as the difference between L_{gz} and L_{mer}.

4. To determine the number of cells in each zone, calculate the number of cells for each interval of the interpolated cell-length data by dividing the size of the interval by the average size of the cells in it (by averaging the size of cells at the beginning and endpoint).

5. The number of cells in the meristem (N_{mer}) equals the cumulative number of cells in all the intervals located within the meristem.

6. The number of cells in the elongation zone (N_{el}) equals the cumulative number of cells in all the intervals located within the elongation zone.

7. The number of cells in the whole of the growth zone (N_{gz}) equals $N_{mer} + N_{el}$.

8. Determine the root-elongation rate (E) as the velocity in the mature part of the root by averaging the values obtained in this region.

9. Calculate mature cell size l_{mat} as the average cell length in the mature region.

10. Calculate cell production in the whole of the meristem as E/l_{mat}.

11. The average cell-division rate (D_{avg}) equals cell-production rate divided by N_{mer} and the average cell-cycle duration $T_c = \frac{\ln(2)}{D_{avg}}$.

12. Calculate all these parameters for each individual root and calculate averages and perform statistics.

3.3. Analysis of Maize Leaves

Similar to roots, the cells in the epidermis of monocotyledonous leaves are arranged in linear files, allowing the same basic approach as in the root system to analyse cell division (*see* above). The main difference between roots and monocotyledonous leaves, however, is that it is not possible to determine the velocity profile by direct observation of the epidermal cells in the growth zone, because the older leaves encapsulate the growing younger ones. Therefore, an indirect method based on cell-length profiling only has to be used. This method assumes that during steady-state growth, the cell-length profiles are constant. It entails leaf-elongation rates measurements, measurements of the cell-length profile along the axis of the leaf and estimation of the size of the leaf basal meristem. Unfortunately, it is not possible to perform all these measurements on the same individual leaves. Therefore, it is unavoidable to combine measurements of several distinct plants grown under the same conditions.

The method will be explained and discussed based on our experience with maize, but can be adapted to other monocotyledonous species.

3.3.1. Plant Growth and Measurements of Leaf-Elongation Rates

1. To perform a kinematic analysis of leaf growth in maize, a batch of at least 15 plants per condition is necessary. Record the leaf-elongation rate (LER) in function of time for a first subset of these plants by measuring with a ruler the length of the leaf (soil surface to leaf tip) under study at regular time intervals, preferably daily (*see* **Note 18**) for automation and a higher resolution approach. Straighten the leaf by hand and take caution not to break or damage the leaf as touching may influence growth rates. Start leaf measurement from leaf appearance (emergence from the sheath at the base of the shoot) until its complete extension.

2. Calculate LER from the recorded data as the difference in leaf length on two successive time points divided by the time interval between them (in mm/h). For a typical monocotyledonous leaf, leaf elongation is generally linear during the first days after appearance, followed by a period of progressive decline depending on leaf position and treatment (environmental conditions). The first days of linear increase can be considered as a situation of steady-state growth. At the cell level, it is assumed that also cell production and cell-length profiles in the growth zone are constant during the same time period (12, *see* **Note 19**).

3.3.2. Cell-Length Measurements

1. For profiling the cell length, harvest the whole growth zone during steady-state growth. The size of the growth zone depends on the environmental conditions, the species, the genotype and the developmental stage examined. Therefore, it is important to make an estimate of this size to make sure that the samples encompass the full extent of the growth zone (*see* **Note 20**).

2. To sample the growth zone, remove the older leaves that surround the growing leaf. Take special care not to damage the basal part of the growing leaf because this is where the leaf basal meristem is located. Some practise beforehand to optimise the dissection technique is required.

3. Segment the growth zone into smaller pieces (for example, segments of 10 mm) if the growth zone is too large to mount as a whole on microscope slides. Place the samples in absolute ethanol for 48 h for chlorophyll removal and fixation. To obtain better and faster clearing, renew the ethanol after 6 h. For further clearing and storage, place the samples in lactic acid.

4. For the cell-length measurements, mount the leaf (segments) on microscope slides. Unroll the samples on the object slide and remove the mid-vein with a scalpel. Mount the samples in lactic acid, so that the abaxial epidermis, which generally contains fewer stomata compared to the adaxial side, is placed face up.

5. Analyse the specimens under a microscope fitted with differential interference contrast (DIC) optics. Find the orientation of the samples and start from the most basal part. In maize, use the trichomes at the edge of the leaf, which point towards the leaf tip, as a reference to distinguish base from tip directions. Adjust the microscope to optimise the image focussing on the epidermal cell walls. Moving in distal direction from the base of the leaf sample, measure all the cells belonging to one epidermal cell file. Choose a cell file adjacent to stomatal rows on the abaxial side because these cell files are convenient to recognise and consist of a single cell type (*see* **Note 21**).

6. To take into account the variation in cell size at different positions across the leaf, repeat these measurements for a few equivalent cell files (e.g. cell files adjacent to stomatal rows) for each leaf.

7. Calculate the distance from the base to the mid-point of each cell in a spreadsheet. This is done by determining the cumulative lengths of all cells in more basal positions in the same file + 0.5 × the length of the cell itself.

8. Combine all data for a leaf by copying the data for the different cell files into the same position and size columns and sorting them for ascending position.

9. Use these data for interpolation and generation of equally spaced data using the same Locpoly routine in R that was used for the root data (*see* **Section 3.1.6** and step 4).

10. Calculate averages and standard errors between replicate leaves using the interpolated data.

3.3.3. Estimation of the Linear Extent of the Leaf Basal Meristem

1. Estimate the size of the meristematic zone of the leaves by locating the most distal mitosis in the cell files of interest. Sample the basal part of the growth zone in a fashion similar to the approach used for cell-length measurements. Again, sampling should occur during the steady-state growth and without damaging the most basal part. Place the samples in 3:1 (v/v) absolute ethanol:acetic acid for fixation of cell walls and clearing of chlorophyll. Keep the samples in this solution at 4°C from 24 h up to several weeks.

2. Rinse the samples with a buffer solution containing 50 mM NaCl, 5 mM EDTA and 10 mM TRIS-HCl, pH7.

3. Visualise the nuclei by incubating the samples in the dark for 1–20 min in the same buffer solution containing DAPI at a concentration of 1 μg/mL (for an alternative stain, *see* **Note 22**). Avoid too intense staining, since only staining of the epidermal cells and not of underlying cell layers is desired. Therefore, check the samples for fluorescent signal emission after a short incubation.

4. According to this first microscopic assessment, stop the reaction when a satisfactory signal is achieved. To stop the reaction, remove the DAPI by rinsing in the buffer solution. Sample re-incubation is always possible if it is necessary to achieve a higher signal level.

5. Mount the samples in a drop of the buffer solution on a microscope slide and cover with a coverslip. For image analysis, the same setup as for the cell-length measurements is required; a digital camera connected to a personal computer with image-analysis software enabling measurements if possible directly on the live acquired image.

6. Starting from the leaf base, score recognisable mitotic cells (metaphase, anaphase and telophase) in cell files adjacent to stomatal rows. In these stomatal rows, asymmetrical divisions take place at the more distal end of the division zone. They represent a convenient landmark for locating the region where the last divisions in the adjacent files can be found. Since mitosis is a relative rare event, it is necessary to examine at least 10 cell files per leaf to find the most distal mitotic event.

7. Determine the distance between the most distal mitotic cell, considered as a proxy for distal boundary of the meristem, and the base of the leaf. Draw and measure a straight line on the images with image-analysis software, between the most distal mitotic cell and the base of the leaf. When the meristem does not fit into a single microscope frame, add up the measurements between landmarks visible in successive overlapping microscopic frames.

3.3.4. Calculation of Overall Kinematic Parameters

1. Based on the measured LER, cell-length profile and meristem size, calculate the kinematic parameters for leaf elongation. The calculations are similar to the ones described above for root growth. Calculate the number of cells in each zone (meristem N_{mer} and elongation zone N_{el}), the mature cell length (l_{mat}), cell production of the meristem (P), the average cell-division rate (D_{avg}) and the average cell-cycle duration (T_c) exactly with the formulae, as described (*see* **Section 3.2.5**).

2. The size of the different zones is calculated in a slightly different way. Estimate the size of the growth zone (L_{gz}) as the distance from the leaf base to the position where the cells reach 95% of their mature length on the smoothed cell-length profile.

3. Estimate the meristem size (L_{mer}) in leaves as the distance at which mitotic cells occur with respect to the base of the leaf (*see* above).

4. Calculate the size of the elongation zone (L_{el}) as the difference between L_{gz} and L_{mer}.

5. In addition, the local cell-elongation rate $R(x)$ can be estimated in monocotyledonous leaves from the cell-length profile. Calculate this parameter for all positions as the position derivative of cell length multiplied by the cell production: $R(x) = P \times \frac{\partial l}{\partial x}$. Because in the meristem the flux, which is estimated by P, is not constant, this formula cannot be used in the meristem.

3.3.5. Flow Cytometry

Similar to the approach used for the Arabidopsis leaves, a deeper insight into the role of cell-cycle progression in the leaf elongation can be obtained by complementing the kinematics with flow cytometry. This technique allows the relative quantification of DNA nuclear content in the different regions of the growth zone and thereby it enables inferences of cell-cycle progression status in the meristem and endoreduplication in the elongation zone (13). However, bear in mind that this approach uses the entire leaf rather than the specific cell type analysed using the microscopy, so that some caution needs to be taken into account with the interpretation.

4. Notes

1. To facilitate regular spacing of seeds, plates with a grid on the bottom are convenient. These provide 32 full squares on a 12-cm dish. For analysis of young leaf material up to 7 DAG, the seeds can be sown densely with four seeds per square. For the time points 8–11 DAG, two seeds are sown per square while at later time points only one seed per square can be used. This way the plants have enough space to develop without overlapping with each other.

2. A drawing tube is a mirror system mounted on the microscope in the light path between the objective and the oculars that allows to view the specimen mounted on the microscope stage and at the same time a sheet of paper on the desk next to the microscope. This setup is used for outlining the epidermal cells in cleared leaf specimen and has much higher accuracy, particularly drawing the small pavement cells around the stomata, compared to using digital images, because it allows for focussing while drawing the cell outlines.

3. As the calculations of expansion and division rates involve local derivatives of velocity and cell-length data, small systematic errors may influence the data. One common source of error is the use of lenses with barrel or pincushion distortion. This type of lenses is not suitable for the measurements required for these kinematic analyses. PLAN lenses are designed to avoid this problem and need to be used, particularly for the analyses of cell lengths and velocity along an axis.

4. Plating for the analysis of leaf growth on horizontal agar plates and root growth on vertical plates requires an opposite strategy with regards to placing the seeds on the agar: While it is desirable for the root analysis to have the root growing on the surface of the agar and not penetrate it, for optimal and homogenous leaf growth the roots need to be able to penetrate the medium. To facilitate entry into the agar on horizontal plates, slightly touch the agar when placing the seeds, making small cracks in the surface that serve as an entry point for the roots. Avoid this when preparing vertical plates so the roots do not find an entry point and instead remain growing at the agar surface.

5. Sometimes the leaf material does not clear completely after transfer to lactic acid. This is typically due to the presence of relatively high amounts of starch, which makes the imaging of the epidermal cell walls very difficult. To improve the

clearing of the leaves, they can be transferred to Hoyer medium. The duration of clearing in Hoyer medium (6) depends on the stage of the leaf material and amount of starch present. For young leaves, 15–30 min are usually sufficient, while older leaves may require up to 2 h. It takes some trial and error to determine the desired clearing time. If the leaf material is kept too long in Hoyer medium, the clearing is excessive and the cell walls are no longer visible using DIC. Finally, the material needs to be transferred back to lactic acid after Hoyer treatment and mounted on object slides using lactic acid as mounting medium. The cell analysis needs to be done shortly (hours) after the Hoyer clearing, as the effect of this clearing dissipates and clearing needs to be redone.

6. For the earliest time points, it is not possible to determine the leaf area using a binocular. In these cases the outline of the primordium is drawn using a DIC microscope and drawing tube (*see* **Note 2**).

7. For practical reasons, we have opted to work on the abaxial epidermis of the first leaf pair as it does not have trichomes that complicate the cell measurements, and cell divisions are in transverse direction only. However, conceptually it is also possible to perform the calculations for the adaxial epidermis by either ignoring or taking into account the trichomes and accompanying cell complex. The palisade parenchyma could be analysed either by ignoring the divisions perpendicular to the surface that increase the number of mesophyl layers or by including measurements of the number of palisade cell layers or by including measurements of the number of palisade cell layers. In the spongy mesophyll, individual cell layers are not easily defined, complicating the analysis probably beyond what is feasible to work on. For addressing specific research questions it may be interesting to compare calculations on multiple cell layers.

8. In absence of a drawing tube, images of cells can also be made using a digital camera mounted to the microscope and tracing the walls on a computer using the mouse or a drawing pad. The disadvantage of this method is that some areas of the image may not be focussed optimally and/or the resolution of the camera is poor at low enough magnification to view sufficient numbers of cells, resulting in loss of accuracy.

9. Smoothing and interpolation of the primary cell-length and velocity data is an important step in the calculations. By using a local polynomial approach instead of fitting

a predefined function, no assumptions as to the overall shape of the curve needs to be made. The basis of the methods is a local quadratic fitting of a polynomial to a small section of the data, which allows smoothing the data and calculating the local differential. In Excel, this requires setting up three columns: 1. time (in days after germination/sowing), 2. the quadrate of time and 3. the natural log of average leaf area. Typically a five-point quadratic fitting is used to calculate adjusted leaf area and its derivative (*RLER*) for the mid-point. Because this is not possible for the first and last two points of the series, these are calculated from the same fit as the third from the beginning and end of the series, respectively. Calculations start from the third time point: In columns 4–7, the polynomial coefficients will be calculated using the function 'Linest'; the first argument for this function is the *y*-values for 5 points (C2:C6) and the second argument are the *X* and X^2-values for 5 points (A2:B6). The third and fourth arguments are 'TRUE' for the use of a constant and 'FALSE' for the output of the fitting statistics (these co-ordinates for cells are assuming that the data columns contain a row of headers). This results in the following formula: =Linest(C2:C6,A2:A6,TRUE,FALSE). To invoke the array calculation after entering, instead of pushing 'Enter', the combination 'Ctrl-Shift-Enter' needs to be used. This will write in the three selected cells the values for the coefficients *a*, *b* and *c* of the quadratic fitting ax^2+bx+c. This array calculation can then be copied down up to the third last point. Now it is possible to calculate the smoothed *y*-values $aX^2 + bX + c$ and RLER, the derivative $2aX + b$, in two extra columns (*see* **Note 9** for the rationale behind this smoothing approach). The first and last 2 points are calculated from their own *X* values and the coefficients of the fitting for the third points. When the *X* values are given in days, the obtained *RLER* needs to be divided by 24 to obtain the rates as $mm^2/mm^2/h$.

10. Flow cytometry can be used to analyse the (nuclear) DNA content of the cells of the leaf using DAPI as a fluorochrome. This allows estimating the relative cell-cycle phase duration during cell proliferation. Because the standard preparation involves chopping the entire leaf blade, this analysis is not restricted to a single tissue as for the case of cell measurements. Nevertheless, we found that the transitions from proliferation to expansion and from expansion to maturity as defined by the kinematic analysis are closely reflected in the DNA profile (13), so that with some caution these data can effectively be combined.

11. Optimise the scanning conditions to get sufficient contrast between the root and the background. Often opening the lid or putting a dark piece of paper between the plates and the lid improves the scannings. Modern scanners using LED light typically do not give good results.

12. Although the time-lapse observations can be done manually as described here, a number of automated image-analysis programs have been developed that can be used to analyse image sequences generated at much shorter intervals (14, 15).

13. It is well known that roots quickly respond to gravistimulus associated with their altered orientation. As this may impact the growth rate of the root, the plate has to be kept in a vertical orientation as much as possible. One way to minimise the potential effect of this perturbation is to place the microscope used for the time-lapse observations in the growth chamber and place it horizontally, so that the stage is in vertical orientation. Many microscopes have a flat back that allows this; otherwise the microscope needs to be fitted with an extra set of supports. It may also be required to make some adjustments to the stage to easily fit the plates and increase the resistance of the stage control to avoid gravity move the stage down during the observations. Although this setup is ideal, it is possible to make the observations at a horizontal stage. This requires keeping the observation time short enough to preclude effects on the growth. This can be tested by comparing the elongation rates calculated from the time lapse with those obtained from the control set of roots by marking the rate of displacement of the tip of the root. As also other experimental perturbations during the kinematic observations can influence the growth rate, it is good practice to include this quality control.

14. Several commercial photo-stitching programs are available, recent versions of Photoshop (Adobe) have a routine to do this (file/automate/photomerge) and there are also public domain ImageJ plug-ins available that can do this. When any of these routines are used make sure that there is no deformation of the images to make them fit better as this will lead to errors in the measured cell sizes.

15. ImageJ has a nice feature for this as it allows to leave one of the ends of the line selection in place while moving the other allowing to use the software like a digital calliper, which avoids systematic overlap or gaps between subsequent measurements.

16. Several problems are often encountered that make it impossible to measure all cells in a single file. Firstly, the root is often not entirely straight so that the file that is being measured rotates out of focus and another one replaces it. Secondly, there can be some problem in part of the specimen that prevents cells to be unambiguously measured. In such cases, it is a useful trick to measure the distance from the last unambiguous cell wall to the first one that is again unmistakable. This distance is then measured, but a note has to be made that this measurement does not correspond to the length of a cell. This trick also allows 'jumping' from one cortical file to the next.

17. For roots that are not growing steady-state (either accelerating or decelerating growth), additional cell-length measurements on a set of parallel grown roots need to be done at an interval of 1 or 2 days before and after the velocity measurements. Add the local rate of cell-density change into the calculation of cell-production rate: $P(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} + \frac{\mathrm{d}\rho(x)}{\mathrm{d}t}$. Calculate this rate as the difference in cell density $\rho(x) = \frac{1}{l(x)}$ obtained from the extra cell-length measurement after the time of the velocity measurements minus the cell density of the extra measurement prior to the time of velocity measurements divided by the time difference between these two observations. Note that as the cell-length measurements are destructive these adjustments can only be done at a population level rather than for each individual root.

18. A more advanced alternative to the daily manual (ruler) measurements of leaf length is using electronic systems, based on linear velocity displacement transducers (LVDTs). Using LVDTs, the measurements of leaf extension are acquired at much higher time resolution (minutes or even seconds) (14, 15).

19. To ensure that the plants grow at steady-state, they should preferentially be grown under controlled conditions to avoid variations of leaf-elongation rate. Especially fluctuations in temperature, light intensity or humidity should be minimised, because they strongly influence growth rates (12, 13).

20. For maize, the growth zone of leaf 4 of maize inbred line B73 grown at 25°C spans about 100 mm (13), which can be used as a reference for the analysis of maize leaf growth.

21. The easiest and fastest way to handle these measurements is to directly measure the cells on the images captured 'live' by a digital camera connected to a PC running image-

analysis software. For ImageJ, plug-ins for a range of cameras are available from the program Web site. Alternatively, the same method as for the root can be used; capturing a series of overlapping picture, which are then merged together before measuring the cell lengths off-line.

22. An alternative staining protocol for nuclei visualisation is using Feulgen staining, which allows doing the analysis without the need of an epifluorescence microscopy (15).

References

1. Poorter, H. and Garnier, E. (1996) Plant growth analysis: An evaluation of experimental design and computational methods. J Exp Bot **47**, 1343–1351.

2. Goodwin, R. H. and Stepka, W. (1945) Growth and differentiation in the root tip of *Phleum pratense*. Amer J Bot **32**, 36–46.

3. Erickson, R. O. and Sax, K. B. (1956) Rates of cell division and cell elongation in the growth of the primary root of *Zea mays*. Proc Am Phylos Soc **100**, 499–514.

4. Silk, W. K. and Erickson, R. O. (1979) Kinematics of plant growth. J Theor Biol **76**, 481–501.

5. Fiorani, F. and Beemster, G. T. S. (2006) Quantitative analyses of cell division in plants. Plant Mol Biol **60**, 963–979.

6. Candela, H., Martínez-Laborda, A., and Micol, J. L. (1999) Venation pattern formation in *Arabidopsis thaliana* vegetative leaves. Dev Biol **205**, 205–216.

7. Donnelly, P. M., Bonetta, D., Tsukaya, H., Dengler, R. E., and Dengler, N. G. (1999) Cell cycling and cell enlargement in developing leaves of Arabidopsis. Dev Biol **215**, 407–419.

8. Beemster, G. T. S., De Vusser, K., De Tavernier, E., De Bock, K., and Inzé, D. (2002) Variation in growth rate between *Arabidopsis thaliana* ecotypes is correlated with cell division and A-type cyclin dependent kinase activity. Plant Physiol **129**, 854–864.

9. Beemster, G. T. S. and Baskin, T. I. (1998) Analysis of cell division and elongation underlying the developmental acceleration of root growth in *Arabidopsis thaliana*. Plant Physiol **116**, 515–526.

10. Dolan, L., Janmaat, K., Willemsen, V., Linstead, P., Poethig, S., Roberts, K., and Scheres, B. (1993) Cellular organisation of the *Arabidopsis thaliana* root. Development **119**, 71–84.

11. Erickson, R. O. (1976) Modeling of plant growth. Ann Rev Plant Physiol **27**, 407–434.

12. Ben-Haj-Salah, H. and Tardieu, F. (1995) Temperature affects expansion rate of maize leaves without change in spatial distribution of cell length. Analysis of the coordination between cell division and cell expansion. Plant Physiol **109**, 861–870.

13. Beemster, G. T. S., De Veylder, L., Vercruysse, S., West, G., Rombaut, D., Van Hummelen, P., Galichet, A., Gruissem, W., Inzé, D., and Vuylsteke, M. (2005) Genome-wide analysis of gene expression profiles associated with cell cycle transitions in growing organs of Arabidopsis. Plant Physiol **138**, 734–743.

14. Basu, P., Pal, A., Lynch, J. P., and Brown, K. M. (2007) A novel image-analysis technique for kinematic study of growth and curvature. Plant Physiol **145**, 305–316.

15. van der Weele, C. M., Jiang, H. S., Palaniappan, K. K., Ivanov, V. B., Palaniappan, K., and Baskin, T. I. (2003) A new algorithm for computational image analysis of deformable motion at high spatial and temporal resolution applied to root growth. Roughly uniform elongation in the meristem and also, after an abrupt acceleration, in the elongation zone. Plant Physiol **132**, 1138–1148.

7

Automation of kinematic analysis

7.1 Introduction

Growth of plants can be largely attributed to two processes: cell division and cell expansion. The final shape and size of an organ is the result of a developmental program regulating these processes. The shoot apical meristem (SAM) initiates lateral organs forming a leaf primordium starting from a small number of founder cells (1). At first the cells making up the leaf primordium continue to divide, accompanied by cell expansion to maintain cell size homeostasis (2). In this early stage of leaf development, the increasing mass of cells is the driving force for leaf growth. Next, cell division ceases in a tip-to-base gradient and vacuole-driven cell expansion becomes the predominant growth process (3). Recent research in our group has revealed that this transition from proliferation to expansion is rapidly established with roughly the top half of the leaf expanding and the bottom half still actively dividing. The fraction of dividing cells decreases over time (Andriankaja, Dhondt and Inzé, unpublished results). During expansion phase, cells of the stomatal lineage continue to divide to produce additional stomata (4). The mature stage of development is reached when cell expansion ceases and the leaf has obtained its final size.

Plant growth is a key characteristic for both commercial applications as academic research. This requires quantification of growth related parameters in different genetic backgrounds and in contrasting environmental conditions. Ultimately the goal is to enhance growth, leading to an improved crop yield. Growth is an ambiguous

term covering biomass, size, number of tillers, seed size and number. We use the *Arabidopsis* leaf as a model system and try to understand what determines the mature size of a leaf. The first step to evaluate the effect of transgenes or treatments to determine leaf size is therefore to measure mature leaf size either on a single leaf of the rosette or by analyzing all the leaves to determine at which position leaf size is affected (leaf series analysis, 5). This however does not allow to discern what are the mechanisms leading to the observed organ size differences. Measuring the average cell size of mature epidermal cells and estimating the number of cells in a mature leaf (from the ratio of leaf blade area over mature cell area), can give some more insight if cell division or cell expansion contribute most to the observed phenotypic changes. Recently, we have shown that increased leaf size in known *Arabidopsis* mutants and transgenic lines with increased leaf size is predominantly associated with an increase in cell number (5). However, quantification of the endpoint situation, the mature leaf, does not include the dynamics of the underlying mechanism. Different mechanisms can lead to an increase in cell number e.g. delay of the transition from proliferation to expansion phase, increase in cell division rate, increased divisions of the meristemoids of the stomatal lineage or a larger number of cells in the initial leaf primordium. To be able to distinguish these different mechanisms, the evolution of parameters such as leaf size, cell number and size need to be followed throughout the development of the leaf. This type of analysis is known as 'kinematic analysis'. Based on the primary measurements (leaf size, cell size and cell number) throughout development it is possible to calculate cell division rate (CDR) and relative leaf expansion rate (RLER). The mathematical background of kinematic analyses has been reviewed by Fiorani and Beemster (6). We recently published an overview of methodologies to perform kinematic analysis in different types of organs such as *Arabidopsis* root tips and leaves and monocotyledonous leaves (see Chapter 6, 7). I contributed the part about *Arabidopsis* leaves. The method is based on harvesting the first leaf pair of *Arabidopsis* on a daily basis from its emergence until maturity (in our conditions from 5 to 25 days after stratification). Leaf size is measured using image analysis software (e.g. ImageJ, rsbweb.nih.gov/ij/) based on digital images. The cells of the abaxial epidermis are drawn using a microscope equipped with a drawing tube and these drawings are also analyzed using ImageJ. This experimental approach allows calculating the CDR and RLER using a spreadsheet program such as MS Excel (Microsoft, Redmond, Washington, USA).

While a kinematic analysis gives detailed insight in the underlying mechanism to understand the effect of a mutation, transgene or treatment, it is not commonly used as it is rather specialistic and labor intensive. Especially the drawing of the epidermal cells is very time consuming, but also the manual analysis of the images can take several weeks. Therefore, we set out to automate the data analysis. This is done in a two-step process. First the drawings are analyzed using image analysis algorithms that threshold the cell walls, recognize guard cells and pavement cells

and measure cell area for individual cells (developed by Stijn Dhondt). I set out to develop software that takes these data and automatically calculates all relevant parameters.

7.2   Automation

Automation can be implemented at several steps in the protocol. The three most time consuming steps, and thus most desirable for optimization, are (a) drawing of the abaxial leaf epidermal cells, (b) measurement of the cell and leaf areas, and (c) data analysis. I will briefly go over the steps that have been taken for the first two parts and give some more detail about the data analysis algorithms which I have implemented.

7.2.1   Segmentation

To be able to determine cell cycle parameters, cell size needs to be measured and number of cells per leaf need to be calculated (for details see Chapter 6). This is preferentially done on the abaxial epidermis of leaf 1 and 2, because this side does not contain trichomes. During development there is a tip-to-base gradient in cell division (3, 8). To accommodate for this, two drawings are made per leaf, one at about one quarter from the tip of the leaf and one at one quarter from the base (9). A trained person takes about 15 minutes per drawing. Per genotype and per time point, five replicates are drawn. For a full kinematic analysis around 20 time points are needed. This adds up to 400 drawings for a wild type versus mutant/treatment comparison or about three weeks full time work.

The automation of this requires taking photos of the epidermis and analyzing these images to filter out the cell walls. There are several technical hurdles that need to be overcome: (a) One digital image does not capture enough cells, thus stitching of several tiled images is needed. (b) When mounted on slide, the leaf is not entirely flat which requires photos to be taken at different focal planes. The curvature of the leaf is strongly dependent on the genotype and growth conditions. Moreover, an autofocus feature will always focus on the palisade parenchyma, as this layer typically contains more contrast than the epidermis. (c) Cell walls and cytoplasm have limited contrasts, requiring boosting of the contrast in order to detect cell walls reliably.

The first two points can be addressed by using automated microscopes with high-precision stage control. The third step however, is the most challenging as this involves changing the perception of the image. There are several possible solutions to enhance contrast. The current manual method takes advantage of Differential Interference Contrast (DIC). Alternatively, fluorescent labeling of the cell wall is a possibility, but this involves additional work in making transgenic marker lines

that need to be crossed with mutant lines of interest. Staining of cell walls, works well for early time points, but for older leaves it is difficult to get a homogeneous staining. However, improved protocols have recently been developed (10).

We decided to try to start from DIC images to extract the cell walls. This does not require additional bench work but is quite challenging from the informatics point of view. Therefore, we started a collaboration with the Department of Telecommunications and Information Processing (TELIN) at the University of Ghent. The first attempts look promising, but some additional hurdles need to be taken: (a) folds in the leaf are also recognized as cell walls, and (b) the light-shadow effect typical of DIC images makes the detection of cell walls dependent on the orientation relative to the direction of DIC. To address this second issue, we are trying to implement the use of two DIC images with the DIC directions perpendicular to each other. This work is ongoing and coordinated by Stijn Dhondt from our group.

7.2.2   Image analysis

The segmentation yields a binary image of the epidermal cell walls, independent of the way it is obtained. Next, the total area of the registered cells needs to be determined as well as the number of cells. Also the number of stomata needs to counted. Additionally, individual cell areas of all cells can yield interesting information. While total area, cell number and stomata were counted manually, it is not feasible to measure individual cell areas as this would be too time consuming.

The analysis was automated by Stijn Dhondt. I will only briefly address the major principles, the detailed analysis will be published elsewhere. Some small adaptations in the segmentation are needed in order to facilitate automatic analysis of the images. The pores of stomata need to be completely filled, this avoids recognition of the pore as a cell and allows for counting of the stomata as well as identifying the adjacent guard cells. The quality of the input image determines the accuracy of the analysis, therefore images need to be improved manually prior to analysis. Closing of all cell walls and ensuring that stomates are drawn in a cartoon-like way, is most important. This is efficiently done using a Cintiq tablet (Wacom, Krefeld, Germany).

The image analysis was implemented using the *mmorph* toolbox (SDC Information Systems, Naperville, Illinois, www.mmorph.com). This determines the registered epidermal area, the number of cells, the number of stomata and all individual cell sizes subdivided in pavement and guard cells. An outline of the analysis methodology is shown in Figure 7.1. The measured parameters are saved in tab-delimited text files and used as input for the data analysis algorithms.
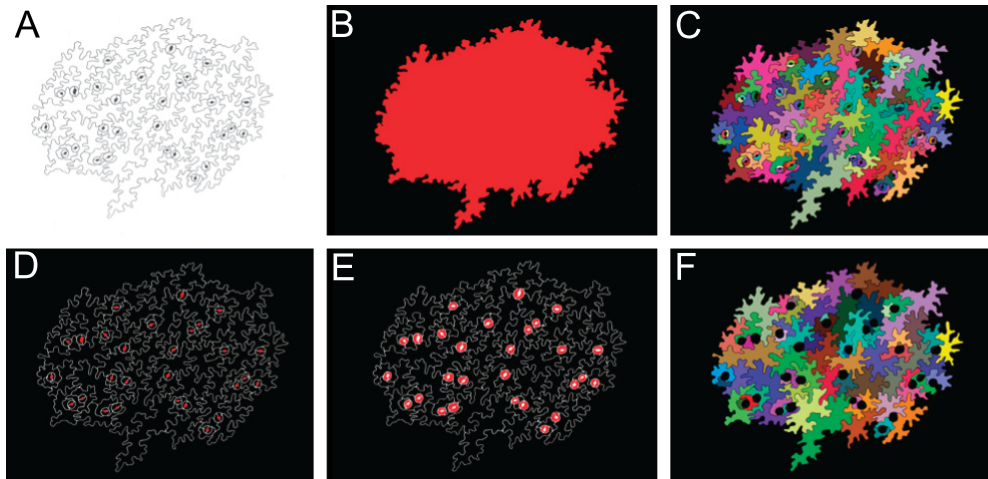
Figure 7.1: Image analysis of microscopic drawings. (A) Original drawing. (B) The area of the drawing is determined automatically. (C) Total cell number is measured after automatic correction and optimization of the drawing. (D) The number of stomata is identified by automatic recognition of the stomatal pores. (E) Guard cells are identified as cells adjacent to a stomatal pore, allowing the measurement of individual guard cell sizes. (F) Non-guard cells are identified as pavement cells, allowing the measurement of individual pavement cell sizes.

7.2.3 Data analysis

The data analysis involves a highly hierarchical step-wise analysis, which was originally performed using a spreadsheet such as MS Excel (Microsoft, Redmond, Washington, USA). This is laborious, prone to errors and no proper statistical analysis using error propagation was used. Therefore, we set out to automate and improve this analysis. The first implementation was done using the scripting language Perl (www.perl.org). However, given the large amount of data produced by a kinematic analysis and the additional features we implemented, the organization of the data became an important issue. Also, to be able to use these algorithms and make validation of correct processing easier, a graphical user interface (GUI) would be helpful. Creating a database and GUI using Perl is a time-consuming task. Therefore, a different programming environment was used. I chose to use the application-programming interface (API) Cocoa which defines a framework called CoreData that handles data storage and retrieval. Integrated with Cocoa is Interface Builder, which provides an intuitive approach to make a GUI. Using these frameworks allows to concentrate on programming the data analysis algorithms. The downside of this API is its limitation to computers running Mac OS X (Apple, Cupertino, CA, USA).

To be able to perform the data analysis, the input data needs to be organized in a structured way. Therefore, I implemented a hierarchical Data Analysis Model with '*Genotype*' as top level (Figure 7.2). This represents a specific genetic background such as wild type, transformants or mutants as well as treatments. A *Genotype* contains '*Plants*' that are characterized by their age (e.g. 12 days after stratification). Each *Plant* is linked to a specific *Genotype*, but *Genotypes* have multiple *Plants* that make up the developmental time series. The next level is a '*Leaf*'. All the leaves that are analyzed in replicate for a certain genetic background or treatment at a specific time point are linked to the same *Plant*. Each *Leaf* has an associated leaf area and is linked to one or more '*Drawings*'. Each *Drawing* represents the image analysis data extracted from a segmented image. A *Drawing* has several measured parameters: drawn area, number of cells and number of stomata. By dividing the drawn area by the number of cells, the average cell area for that *Drawing* is calculated. This is done for all *Drawings* associated with a *Leaf* and these values are again averaged to obtain an average cell area for the *Leaf*. Division of the leaf area by this value gives the estimated cell number per leaf. The number of stomata per *Drawing* is used to calculate the stomatal index (SI), the fraction of guard cells in the epidermal layer. Combining the SI data for different *Drawings* results in the SI for a *Leaf*. All the data points for the different *Leaves* associated with a *Plant* are again averaged to obtain the average leaf area, average cell area, estimated cell number and SI for a *Plant*. The derivative of the leaf area over time (thus combining different *Plants* of the same genotype but with different age) gives the relative leaf expansion rate (RLER), while the derivative of the cell number results in the cell division rate (CDR). These derivatives are calculated by a local second degree polynomial fit to the log transformed data using the least squares method which is implemented in a generic way using matrix calculations. For all the described calculations the standard error is calculated according to the rules of error propagation (11). The result of the data analysis is represented in a set of graphs within the program to facilitate interpreting the data. Two screenshots of the program are included in Figure 7.3 as illustration of the GUI.

The automated image analysis also allows us to introduce some new features based on individual cells. Therefore, each *Drawing* is associated with a number of '*Cells*'. For all the cells the size is measured individually and the contour and its circularity are determined. Cells that are adjacent to a stomatal pore (recognizable on the drawings as these are rather large filled-in areas) and that comply to empirically determined size thresholds (based upon visual inspection) are considered as *Guard Cells*. This allows to distinguish between pavement cells and guard cells. The individually measured cell sizes are not used to calculate the average cell areas described earlier. The size that is measured is the area contained within the drawn cell walls. Thus, the line representing the cell wall is not included in the individual cell sizes, while this is the case when the drawn area is divided by the number of
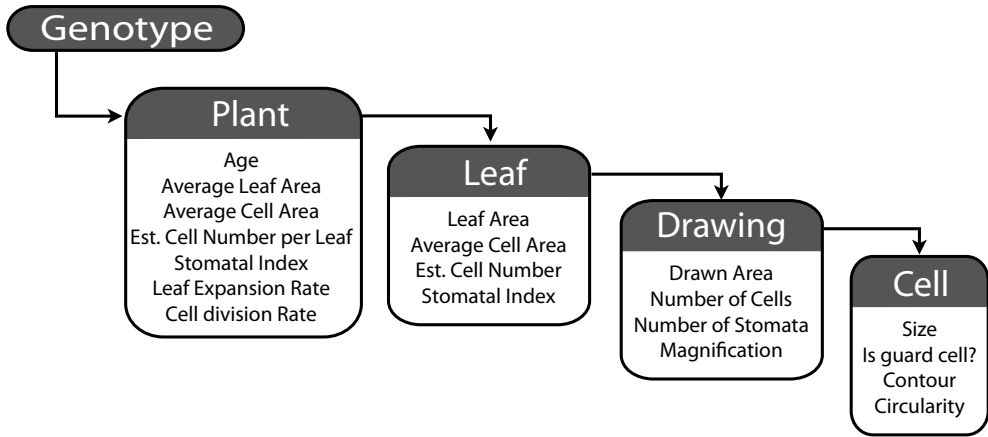
Figure 7.2: Data analysis model used to organize the data generated by a kinematic analysis. The rectangles represent the different levels of the hierarchical organization. Within the rectangles, the most important parameters are listed. The arrows indicate the hierarchical structure of the data. The top level is '*Genotype*' (the genetic background such as wild type, transformants or mutants as well as treatments). Each *Genotype* contains several '*Plants*', with as main characteristic their age, making up the developmental time series. Every '*Leaf*' represents one of the replicate leaves that was analyzed for a specific age and genotype (through its link with a Plant and Genotype, respectively). The '*Drawings*' for a *Leaf* contain the results of the image analysis of the drawings of the abaxial epidermis. Each individual cell of the drawing has a '*Cell*' entry, linked to its *Drawing*.

cells. The difference between the average cell area calculated with the two different methods can be up to 10%. Therefore, we opted not to use the individual cell sizes for the averaged cell areas, but incorporate also the cell wall by starting from the drawn area, so that the thickness of the lines does not influence the results. The individual cell areas were used to determine cell size distributions and calculate the average pavement cell area and guard cell area. These data also allow assessing shape characteristics of the cells such as the circularity and makes it possible to calculate a relative cell expansion rate. The cell distributions are a useful addition to the data analysis. While the average cell area already contains valuable information, the individual cell data allow us to assess the spread of cell areas.

7.3   Publications

The described methodology has been applied to a quadruple-*DELLA* and *sly1-10* mutant (12) as well as for the *SHORT-ROOT* (*SHR*) and *SCARECROW* (*SCR*) mu-
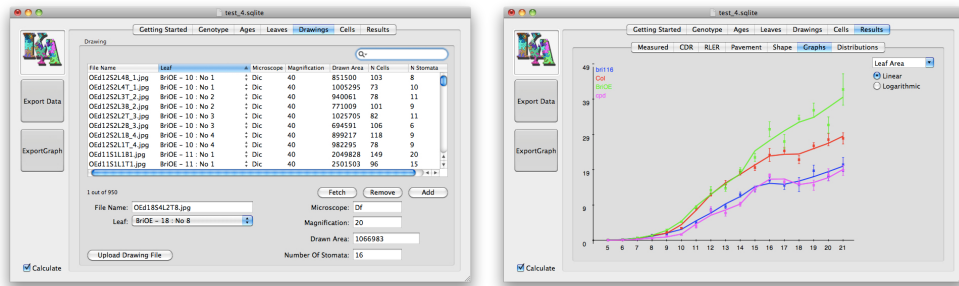
Figure 7.3: Screenshot of the Kinematic Analysis program: input of the data for *Drawings* (left) output of the leaf area in a graph (right). The tabs at the top are similar to the hierarchical structure of the data (Figure 7.2). Data is typically inserted by going through these tabs from left to right. Data input can be done manually or through copy-paste. For *Leaves*, *Drawings* and *Cells* a tab-delimited text file can be uploaded. Selecting the 'Results' tab automatically performs the data analysis (this can be avoided by deselecting the tick box at the left bottom). The 'Results' tab (right) contains sub-tabs to organize the data: 'Measured' contains the averages of the measured data (leaf area, cell area, cell number per leaf, SI), 'CDR' the cell division rates, 'RLER' the relative leaf expansion rates, 'Pavement' contains the average pavement cell area and number of pavement cells per leaf as well as a pavement cell division rate and pavement cell expansion rate (calculated based upon the derivative of the pavement cell number and pavement cell area). 'Shape' contains average circularity and complexity of the cells. 'Graphs' contains a graphical representation for the data in the first three tabs (Measured, CDR, RLER) and gives the possibility to see the data linear or in log scale. The last tab, 'Distributions', shows a bar graph with the fraction of cells in each size bin. The default size bins can be adjusted and a drop-down box allows to select for all cells or only pavement cells or guard cells.

tants (13). These analyses are both published and the first page of the article is included on the following pages. Also for the *cyca2;234* triple mutant a kinematic analysis was performed, this publication is submitted and included in this dissertation (Chapter 2). Using the cell distributions for the *cyca2;234* mutant, we could determine that the increased average cell size is due to two phenomena: the increased size of aberrant guard cells and the presence of extremely large cells. For the *shr* mutant, an increase in the size specifically of the guard cells was observed (13). Finally, a manuscript is in preparation describing the growth parameters of a gain-of-function of *BRASSINOSTEROID INSENSITIVE1* (*BRI1*), a *bri1* loss-of-function mutant and a loss-of-function mutant of *CONSTITUTIVE PHOTOMORPHOGENIC DWARF* (*CPD*), involved in brassinosteroid biosynthesis (Zhiponova M., unpublished results).

Report

Gibberellin Signaling Controls Cell Proliferation Rate in *Arabidopsis*

Patrick Achard,[1,5] Andi Gusti,[1,5] Soizic Cheminant,[1]
Malek Alioua,[1] Stijn Dhondt,[2,3] Frederik Coppens,[2,3]
Gerrit T.S. Beemster,[2,4] and Pascal Genschik[1,*]
[1]Institut de Biologie Moléculaire des Plantes, Centre National
de la Recherche Scientifique, Unité Propre de Recherche
2357, Conventionné avec l'Université de Strasbourg,
67084 Strasbourg, France
[2]Department of Plant Systems Biology, Flanders Institute of
Biotechnology (VIB), B-9052 Gent, Belgium
[3]Department of Plant Biotechnology and Genetics,
Gent University, B-9052 Gent, Belgium
[4]Department of Biology, University of Antwerp,
B-2020 Antwerp, Belgium

Summary

Plant growth involves the integration of many environmental and endogenous signals that together with the intrinsic genetic program determine plant size. At the cellular level, growth rate is regulated by the combined activity of two processes: cell proliferation and expansion. Gibberellins (GA) are plant-specific hormones that play a central role in the regulation of growth and development with respect to environmental variability [1–3]. It is well established that GA promote growth through cell expansion by stimulating the destruction of growth-repressing DELLA proteins (DELLAs) [1, 4–6]; however, their effects on cell proliferation remain unknown. Kinematic analysis of leaf and root meristem growth revealed a novel function of DELLAs in restraining cell production. Moreover, by visualizing the cell cycle marker cyclinB1::β-glucuronidase in GA-signaling mutants, we show that GA modulate cell cycle activity in the root meristem via a DELLA-dependent mechanism. Accordingly, expressing gai (a nondegradable DELLA protein [4]) solely in root meristem reduced substantially the number of dividing cells. We also show that DELLAs restrain cell production by enhancing the levels of the cell cycle inhibitors Kip-related protein 2 (KRP2) and SIAMESE (SIM). Therefore, DELLAs exert a general plant growth inhibitory activity by reducing both cell proliferation and expansion rates, enabling phenotypic plasticity.

Results and Discussion

GA promote important processes of plant growth through cell elongation by promoting the disappearance of nuclear DELLAs [1, 6]. Binding of GA to the GA receptors GID1 promotes interaction of GID1 with the DELLAs [7–9], subsequent polyubiquitination of the DELLAs via the E3 ubiquitin-ligase SCF[SLY1], and eventual destruction of DELLAs by the 26S proteasome [10–13]. Thus, mutants that stabilize DELLAs, such as the GA-deficient ga1-3 or the F box mutant sly1-10, are dwarf, a phenotype that is reverted by the lack of DELLA

*Correspondence: pascal.genschik@ibmp-ulp.u-strasbg.fr
[5]These authors contributed equally to this work

function [1, 12, 13]. Although it is clear that DELLAs are important regulators of plant growth by restraining cell expansion, it remains unknown whether they also act on cell proliferation. To investigate this issue, we performed a kinematic analysis of leaf growth [14–16]. For this purpose, we determined the leaf blade area, the average cell area, and total cell number of abaxial epidermal cells of the first true leaf pair of sly1-10, gai-t6 rga-t2 rgl1-1 rgl2-1 (also called quadruple-DELLA mutant [2]) and wild-type plants grown side by side under the same growth conditions (Figure 1; Figure S1 available online). Under these experimental conditions, leaf growth can be subdivided in three developmental phases (Figure S1). First, until 10 days after sowing, leaf growth is associated mainly with proliferation, where cell division and expansion is balanced, resulting in stable cell size [17]. In the second phase, between days 10 and ~17, the division rate decreases faster than expansion rates, causing average cell size to increase. Finally, after day 18, both cell division and expansion have stopped. Thus, final leaf size depends on rates and duration of cell proliferation and subsequent expansion [16, 17].

We found at 10 days after sowing, sly1-10 plants exhibited a reduction of 41% of the leaf blade area compared to wild-type plants, whereas in the quadruple-DELLA mutant, we observed an increase of 35% (Figure 1). The mutations did not affect the cell size during proliferation; therefore, the difference in leaf area was due to a difference in cell number (Figure 1B). Indeed, in comparison to wild-type leaves, average leaf cell number was decreased by 42% in sly1-10 and increased by 38% in the quadruple-DELLA mutant, respectively. Supporting this observation, the cell division rate representing the total number of cells produced per unit of time and per meristematic cell [15] was, respectively, lower in sly1-10 leaves and higher in quadruple-DELLA mutant leaves than in wild-type leaves at earliest time points (Figure S1E). Thus, DELLAs slow down early leaf growth by restraining the cell division rate (Figures S1B and S1E). Interestingly, we also found that the cell division rate decreased slower in sly1-10 leaves and faster in the quadruple-DELLA mutant compared to wild-type leaves (Figure S1E). The prolonged cell division activity observed in sly1-10 and shortened activity in quadruple-DELLA mutant leaves could be explained by a compensatory mechanism for the respective mutant's deficit or excess cell production [16].

During the second phase of leaf development, growth is mainly driven by cell expansion [17]. During this phase, DELLAs repress leaf growth through their effects on cell elongation (Figure 1). Hence, mature cells of sly1-10 were significantly smaller than those of wild-type. However, because of the prolonged division, mature sly1-10 leaves contained 30% more cells than wild-type leaves, indicating that DELLAs modify the balance between cell division and expansion and that reduced epidermal cell size is partially compensated by an increase in the cell number. It is noteworthy that we found similar kinetics of the appearance of stomatal complexes in wild-type, sly1-10, and quadruple-DELLA mutant leaves, indicating that guard cell differentiation proceeded normally (Figure S1F). Taken together, these data show that accumulation of DELLAs restrains leaf growth by a dual mechanism, first

SHORT-ROOT and SCARECROW Regulate Leaf Growth in Arabidopsis by Stimulating S-Phase Progression of the Cell Cycle[1][W][OA]

Stijn Dhondt, Frederik Coppens, Freya De Winter, Kamal Swarup, Roeland M.H. Merks, Dirk Inzé*, Malcolm J. Bennett, and Gerrit T.S. Beemster

Department of Plant Systems Biology, Flanders Institute for Biotechnology, 9052 Ghent, Belgium (S.D., F.C., F.D.W., R.M.H.M., D.I., G.T.S.B.); Department of Plant Biotechnology and Genetics, Ghent University, 9052 Ghent, Belgium (S.D., F.C., F.D.W., R.M.H.M., D.I., G.T.S.B.); School of Biosciences and Centre for Plant Integrative Biology, University of Nottingham, Sutton Bonington LE12 5RD, United Kingdom (K.S., M.J.B.); Netherlands Consortium for Systems Biology, 1090 6B Amsterdam, The Netherlands (R.M.H.M.); CWI, 1090 6B Amsterdam, The Netherlands (R.M.H.M.); and Department of Biology, University of Antwerp, 2020 Antwerp, Belgium (G.T.S.B.).

SHORT-ROOT (SHR) and SCARECROW (SCR) are required for stem cell maintenance in the Arabidopsis (*Arabidopsis thaliana*) root meristem, ensuring its indeterminate growth. Mutation of *SHR* and *SCR* genes results in disorganization of the quiescent center and loss of stem cell activity, resulting in the cessation of root growth. This paper reports on the role of SHR and SCR in the development of leaves, which, in contrast to the root, have a determinate growth pattern and lack a persistent stem cell niche. Our results demonstrate that inhibition of leaf growth in *shr* and *scr* mutants is not a secondary effect of the compromised root development but is caused by an effect on cell division in the leaves: a reduced cell division rate and early exit of the proliferation phase. Consistent with the observed cell division phenotype, the expression of *SHR* and *SCR* genes in leaves is closely associated with cell division activity in most cell types. The increased cell cycle duration is due to a prolonged S-phase duration, which is mediated by up-regulation of cell cycle inhibitors known to restrain the activity of the transcription factor, E2Fa. Therefore, we conclude that, in contrast to their specific roles in cortex/endodermis differentiation and stem cell maintenance in the root, *SHR* and *SCR* primarily function as general regulators of cell proliferation in leaves.

Stem cells are undifferentiated, totipotent cells that are able to duplicate themselves and to form offspring that differentiates into multiple cell types. They are situated in a microenvironment, the stem cell niche,

where extracellular signals maintain stem cell division at low rates and prevent differentiation (Ohlstein et al., 2004; Li and Xie, 2005). In plants, the best studied stem cell niches are within the root and shoot apical meristems. There, stem cells produce somatic daughter cells that go on dividing and expanding, thereby forming the postembryonic tissues and organs that make up the body of the plant. It is the balance between stem cell maintenance within the meristem and differentiation of cells that exit the niche that facilitates indeterminate root and shoot growth.

SHORT-ROOT (SHR) and SCARECROW (SCR) are members of the GRAS family of transcription factors (Pysh et al., 1999; Lee et al., 2008), required for stem cell maintenance in the root apical meristem. Mutation of *SHR* and *SCR* genes causes a disorganization of the quiescent center (QC) and loss of stem cell activity, resulting in the depletion of proliferating cells in the root meristem and, consequently, cessation of root growth. Essentially, loss of SHR/SCR function renders root growth determinate. Furthermore, *shr* and *scr* mutants lack longitudinal cell divisions that separate the cortex/endodermis initial daughter cells, resulting in only one ground tissue cell layer (Benfey et al., 1993; Scheres et al., 1995; Di Laurenzio et al., 1996; Helariutta et al., 2000; Sabatini et al., 2003; Heidstra et al., 2004). In

References

[1] T. A. Steeves, I. M. Sussex. *Patterns in Plant Development* (Cambridge University Press, Cambridge, UK, 1989).

[2] P. Green. *Botanical Gazette* **137**, 187 (1976).

[3] P. M. Donnelly, *et al. Dev Biol* **215**, 407 (1999).

[4] D. C. Bergmann, F. D. Sack. *Annu Rev Plant Biol* **58**, 163 (2007).

[5] N. Gonzalez, *et al. Plant Physiol* **153**, 1261 (2010).

[6] F. Fiorani, G. T. S. Beemster. *Plant Mol Biol* **60**, 963 (2006).

[7] B. Rymen, *et al. Methods Mol Biol* **655**, 203 (2010).

[8] K. A. Pyke, J. L. Marrison, A. M. Leech. *Journal of Experimental Botany* **42**, 1407 (1991).

[9] L. De Veylder, *et al. Plant Cell* **13**, 1653 (2001).

[10] N. Wuyts, *et al. Plant Methods* **6**, 17 (2010).

[11] P. R. Bevington, D. K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences, 3rd edition* (McGraw-Hill Inc., New York, NY, USA, 2003).

[12] P. Achard, *et al. Curr Biol* **19**, 1188 (2009).

[13] S. Dhondt, *et al. Plant Physiol* **154**, 1183 (2010).

8

Quantitative PCR

8.1 Introduction

Quantifying the number of target sequences in an RNA pool can be done in different ways. Classical methods are RNA gel blot analysis, *in situ* hybridization and microarrays, and more recently sequencing. Combining reverse transcriptase and PCR (RT-PCR) is a powerful technique to quantify e.g. mRNA. Because of the amplification this technique has a high sensitivity and requires only small amounts of starting material. The first applications measured the amount of amplicon at the end of the reaction using agarose gelelectroforesis and ethidiumbromide, poly-acrylamidegelelectroforesis and fluorescent labels or DNA gel blot analysis using radioactivity or chemiluminescence. This is laborious as it requires post-PCR processing and moreover uses poisonous chemicals (1).

The introduction of fluorescence changed the landscape. Fluorescent dyes that bind specifically to double-stranded DNA (dsDNA) can be used to measure the amount of PCR product in a reaction while the PCR is ongoing, hence the name real-time RT-PCR. This technique can be performed starting from only picograms of total RNA. Post-PCR manipulations can be eliminated completely which enables medium- to high-throughput experiments. Nowadays this technique is referred to as quantitative PCR (qPCR) and is regarded as the gold standard for expression quantification and often used for validating microarray results (2). Although qPCR is a very useful technique, it only measures the expression level of the mRNA which can be different from the protein level while the activity can be regulated by

translation efficiency, post-translational modifications, complex formation, protein localization,...Nevertheless, mRNA levels have proven their use in analyzing the function of a gene.

Although qPCR has become a routine technique to measure the expression level of genes, the experimental setup and data analysis are often not up to standard. The high sensitivity of qPCR makes it especially prone to the detection of contaminants such as genomic DNA or primer-dimers. Therefore, it is necessary to include appropriate controls, which will be addressed here. Recently, the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines have been established to increase the quality of published qPCR experiments (3).

Data analysis was typically performed using a spreadsheet. This requires thorough knowledge of the analysis algorithms and is laborious. Moreover, due to the repetitive nature of this manual analysis it is easy to make mistakes that are difficult to track. Therefore, I have developed and implemented an automated algorithm allowing fast and accurate data analysis for qPCR experiments.

8.2   Workflow considerations

At each step of the qPCR workflow it is critical to adopt good laboratory practices. In this paragraph I want to point out some of the issues that need to be considered to obtain reliable results.

8.2.1   Sample preparation

8.2.1.1   RNA extraction

A large part of the workflow deals with RNA molecules. Therefore, all due precautions need to be taken to avoid degradation of RNA. The material that is harvested needs to be frozen in liquid nitrogen as fast as possible. All further processing needs to be done using RNase free material according to the Good Laboratory Practices involving RNA.

The total RNA that is obtained should be of high quality to avoid inhibitory affects of contaminating constituents on the PCR reaction. Several protocols can be used. For *Arabidopsis* plant material, a combination of TriReagent (Applied Biosystems, Lennik, Belgium) and Qiagen RNeasy Cleanup (Qiagen,Venlo,The Netherlands) or extraction using the Qiagen Mini kit (Qiagen, Venlo,The Netherlands) give good quality RNA. It is also recommendable to include a DNase treatment to avoid contamination with genomic DNA. If this step is not performed, this has to be taken in consideration during the primer design (see 8.2.2.3).

8.2.1.2 RNA quality control

As the RNA quality is critical to obtain reliable results, this should be controlled rigorously. Both RNA quantity and quality can be assessed using a spectrophotometer (e.g. NanoDrop-1000, Thermo Scientific, Erembodegem, Belgium). Absorption at 260 nm is correlated to the amount of nucleotides in the solution, while the ratio of 260/280 reflects the purity of the sample. This ratio should be optimally around 2 for RNA. Values lower than 1.7 indicate the presence of proteins, phenolics or other contaminants that strongly absorb at 280 nm. This is however a 'rule of thumb' as it also depends on the constitution of the RNA molecules (e.g. uracil nucleotides have a 260/280 ratio of 4 and thymine of 1.47, www.nanodrop.com). A second measurement for purity is the 260/230 ratio, which should ideally be between 2 and 2.2. Carbohydrates and phenol absorb at 230 nm and the presence of these contaminants can thus lead to low 260/230 ratios. To ensure RNA samples are not degraded, RNA can be loaded on a gel and visually checked or RNA integrity can be verified using e.g. a 2100 Bioanalyzer (Agilent Technologies, Diegem, Belgium).

8.2.1.3 cDNA formation

After extraction of RNA it has to be converted to cDNA. This is done using a viral enzyme, reverse transcriptase. The First Strand Synthesis System III for RT-PCR (Invitrogen, Carlsbad, California, USA) or iScript (BioRad, Nazareth, Belgium) are commonly used. Deprez *et al.* (4) compared the efficiency of different primers for cDNA synthesis. They found that the efficiency of gene specific and oligo(dT)$_n$ primers is high, while random hexamers have a lower efficiency. Gene specific primers are the best option if you only look at efficiency, however they have the major drawback that they only generate one specific product, requiring a separate reaction for each target under study. This is a source of variability because reverse transcriptase enzymes are sensitive to environmental conditions. When oligo(dT)$_n$ primers are used, you can use the resulting product to determine the quantity of different genes, increasing the reproducibility of the experiment. Another possibility is to perform the cDNA formation and qPCR in one step instead of separating the two processes. However, the two-step process generates less primer-dimers and is therefore preferred over the one-step process (5).

8.2.2 Experimental setup

8.2.2.1 Controls

Next to high quality samples, a good experimental setup is required to obtain reliable results. Most important is the use of a negative control to ensure that the right amplicon is produced. This can be done by adding extra samples containing water instead of cDNA (no template control, NTC). Amplification in this negative control

indicates the presence of contaminating cDNA or the production of primer-dimers. In both cases, the expression levels for the target sequence in the samples cannot be reliable determined as it is not possible to establish what the contribution of the contaminant or primer-dimers is to the fluorescent signal. Then optimization of the primers and/or PCR conditions is required. Another type of negative control is a 'no amplification control' (NAC) using as a sample the product of a reverse transcription reaction without addition of reverse transcriptase. This controls for the presence of genomic DNA (gDNA) contamination. However, this requires a large amount of RNA, which is not always easily obtained and increases the workload. By performing a DNase treatment, this becomes less of a problem and therefore we usually do not perform this type of negative control. A positive control is also recommendable as this allows assessing the primers used to amplify a target.

8.2.2.2   Normalization using housekeeping genes

The goal of normalization is to eliminate differences between samples caused by the total amount and quality of RNA you started with. This in order to detect the real differences in expression level (6). qPCR is a technique with high sensitivity, reproducibility and dynamic range, which makes a good normalization indispensable. Usually cDNA is produced starting from 1 μg of total RNA. This is the first step to ensure that the initial amount of input RNA is the same for all the samples. However, only a small fraction of the total RNA is mRNA (3-5%) and also the quality of the RNA is not assessed (6). Therefore, the quantification of total RNA is not sufficient to ensure equal input for all samples, but serves as a starting point to ensure the input is in the same order of magnitude. In the case of microarrays, samples are normalized to the total amount of hybridization signal. This is only possible because of the large number of probes on the arrays. This is not the case for qPCR experiments, therefore targets need to be added to the experiment that amplify 'housekeeping genes'. These are genes that are stably expressed in all experimental conditions and are thus not influenced by e.g. the introduction of a transgene, a treatment or during a time series. This requires prior knowledge of such housekeeping genes based on microarray data, literature,...The theoretical concept of a gene that is not influenced by any treatment or mutation is however not usable in practice. Vandesompele *et al.* (6) developed a method to use multiple genes to normalize the data. A measure for the stability is calculated based on the non-normalized expression levels, this allows to assess the quality of the selected housekeeping genes.

8.2.2.3   Primer design

The design of primers to amplify a part of the target sequence is a crucial step in quantification of the expression. The primer pair should amplify specifically

and efficiently the target sequence of interest. To achieve this, a set of guidelines for the design of primers need to be followed. These include: 17-28 bases in length, GC content of 50-60%, a G or C at the $3'$ end increases priming efficiency, self-complementarity should be avoided, runs of three or more identical nucleotides should be avoided, melting temperature of 59 ± 2 °C and the product length should be 80 to 250 bp. To facilitate primer design several software packages are available such as the freely available Primer3 (7), but also commercial solutions as Beacon Designer (version 4.0, Premier Biosoft, Palo Alto, USA) can be used. The specificity can be assessed using BLAST (8). To combine both primer design and validation of the specificity I have developed a Perl (www.perl.org) script combining the wrapper for Primer3 in BioPerl (7, 9) and NCBI Standalone Blast (8). Recently a similar tool, Primer-BLAST, has become available through the NCBI website (www.ncbi.nlm.nih.gov/tools/primer-blast/). Regardless, these *in silico* analyses have to be confirmed using melting curves and gelelectroforesis (10).

A melting curve analysis is performed after the PCR reaction is finished. After a full denaturation, the temperature is decreased to 65 °C forming dsRNA and yielding a high fluorescence signal. Next, the temperature is steadily increased up to 97 °C. At the melting temperature (T_m) of the amplified product(s), the fluorescent signal decreases sharply. This melting curve is typically analyzed through its derivative, which gives a peak at the T_m of the product(s). If multiple products are formed (due to aspecific amplification or formation of primer-dimers) this leads to multiple peaks or very broad peaks. Thus, a single, sharp peak is required to quantify the target of interest reliably, otherwise the primers need to be re-designed or the PCR protocol adapted.

8.2.2.4 Protocol

The cycling protocol will depend upon the polymerase and the equipment used to perform the qPCR reactions. In our lab the LightCycler 480 Sybr Green I Master (Roche Applied Science, Vilvoorde, Belgium) is used on the LightCycler 480 (Roche Applied Science, Vilvoorde, Belgium) according to the manual instructions. The detailed protocol can be found in Table 8.1.

8.2.3 Data analysis

8.2.3.1 Theoretical modeling of a PCR reaction

Theoretically, the product of a PCR reaction increases exponentially over time, doubling after every cycle. This can be described with formula 8.1 (11):

$$N_n = N_0.2^n \tag{8.1}$$

Table 8.1: Protocol for the LightCycler 480 Sybr Green I Master kit.

Step		Cycles	Temperature (°C)	Time (mm:ss)
Activation		1	95	10:00
Amplification	denaturation	45	95	00:10
	annealing		60	00:10
	extension		72	00:15
Melting curve		1	95	00:05
			65	01:00
			97	00:01
Cooling		1	40	∞

where n: number of cycles, N_0: starting amount of amplicon, N_n: quantity of the amplicon after n cycles.

However, the efficiency (E) of a PCR reaction, the factor with which the amount of PCR product increases every cycle, has to be taken into account. When the yield of the amplification is 100%, the PCR reaction follows the theoretical curve. Such a high efficiency is usually not the case, a value of 80-90% is more realistic. This requires an adaptation of formula 8.1:

$$N_n = N_0.(1+E)^n \tag{8.2}$$

Formula 8.2 allows to determine the starting amount of an amplicon as well as its efficiency from the amplification curve. Therefore, the data is linearized by a log-transformation (formula 8.3), which makes it possible to easily calculate the starting amount (formula 8.4) and efficiency (formula 8.5).

$$\log(N_n) = \log(N_0) + \log(1+E).n \tag{8.3}$$
$$N_0 = 10^{\text{intercept}} \tag{8.4}$$
$$E = 10^{\text{slope}} - 1 \tag{8.5}$$

When the progress of a PCR reaction is followed in real time, using a marker that fluoresces when bound to double-stranded DNA, such as Sybr Green, the exponential increase only holds true for a part of the observed fluorescence. Until the exponential amplification of the amplicon surpasses the background level, the fluorescent output fluctuates around that background level (Figure 8.1A). When the background level is surpassed, the exponential accumulation of the amplicon is reflected in the fluorescence signal (Figure 8.1B). In this phase of the reaction, the initial amount of target sequence can be estimated based upon the theoretical PCR curves described (formula 8.2). The exponential increase attenuates due to the

accumulation of double-stranded product, the decrease in dNTPs and primers, the increasing phosphate concentration in the buffer and the progressive inactivation of the polymerase (12). Once this plateau phase is reached (Figure 8.1C), quantification is not possible as the relation between the initial amount of amplicon and the measured fluorescence is lost (13).
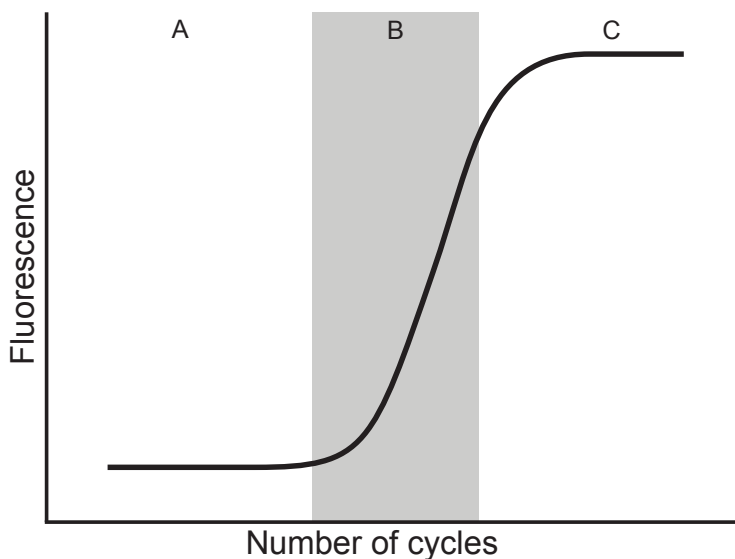
Figure 8.1: Illustration of the progression of the fluorescence signal during a PCR reaction. (A) signal is below the background, (B) exponential increase of the amplicon, (C) plateau phase.

8.2.3.2 Efficiency

Quantification is very sensitive to variations in the PCR efficiency. This efficiency is mostly dependent on the primers but also sample-to-sample variations occur. The classical approach to calculate the efficiency is to make a standard curve and determine the efficiency based on the slope. That efficiency is then always used for the specific primer pair involved. This requires the quantification of a dilution series of a plasmid containing the target sequence or (a mixture of) cDNA samples. Cloning all target sequences is laborious and expensive, especially in a research lab where a large number of different targets are quantified. Also, this approach does not take into account possible inhibitory components in the samples and rules out mispriming as only the desired target is present. Making a mixture of cDNA samples is therefore better, however this is not always possible due to the limited availability of cDNA. Additionally, the quantification of a dilution series requires

a large number of reactions and can thus become very expensive when a large number of primer pairs need to be evaluated.

We have implemented a different algorithm to estimate the efficiency of a primer pair, based on the PCR curves. This approach is based on the loglinear transformation of the fluorescence output (formula 8.3), which allows the calculation of the efficiency (formula 8.5). For all (successful) reactions in the experiment, the average efficiency is calculated for each primer pair. This averaging is required to minimize variability. This methodology was implemented as an applet in Excel (Microsoft, Redmond, Washington, USA) by Ramakers *et al.* (14). The algorithm uses 4 to 6 points in the region that best fits exponential amplification and the efficiency is determined based upon a linear regression on the loglinear transformed data.

8.2.3.3   Expression levels

Calculation of the initial amount of mRNA, the expression level of the target gene, is based upon the correlation of fluorescence and amount of amplicon. A certain fluorescence signal is chosen, above but close to the background. The x-value at which the PCR curve reaches this fluorescence threshold is determined through interpolation of the loglinear regression using four points around the intersection (Figure 8.2). This x-value is referred to as the threshold cycle (C_T), but also quantification cycle (C_Q) or crossing point (C_P) are used in literature.

To calculate expression levels from the C_T value, the $\Delta\Delta C_T$ method is used (10). This methodology determines the ratio of the expression of a target gene (formula 8.6) compared to a control sample (formula 8.7). This implies that the expression level is always a relative value and does not allow determining the absolute expression of a target gene. Determination of the absolute expression requires the quantification of a dilution series of samples for which the number of target molecules is known.

$$\text{expression} = (1 + E^{\text{target}})^{-C_T} \tag{8.6}$$

$$\text{ratio} = (1 + E^{\text{target}})^{C_T(\text{control-sample})} \tag{8.7}$$

8.3   Software implementation

I have developed a Perl (www.perl.org) script with Common Gateway Interface (CGI) technology, called *qPCRAnalysis.pl*. The methodology is based upon to the normalization methodology of GeNorm (6) and the data analysis in qBase (15). This script was developed because the first versions of qBase were implemented as a macro in Excel (Microsoft, Redmond, Washington, USA), which suffered from instability of the program. More recently, a standalone version, qBasePlus (Biogazelle, Ghent, Belgium), was released, which is more stable but requires a license fee. Ad-
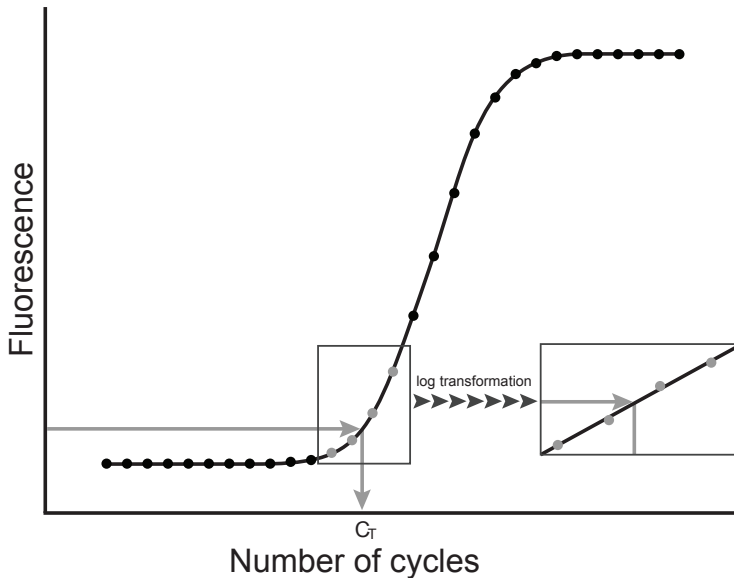
Figure 8.2: Calculation of the threshold cycle (C_T) at which a certain fluorescence level is reached. A fluorescence threshold is chosen at which the PCR curve is in the exponential phase. The C_T-value is determined through interpolation based upon the loglinear regression of four points around the intersection (grey points).

ditionally, the script I developed is fully integrated with the qPCR equipment in the lab, combining the output from the Janus Mini robot (Perkin Elmer, Zaventem, Belgium) that performs the pipetting for the qPCR reactions in 384 well plates and the raw output from the LightCycler 480 (Roche Applied Science, Vilvoorde, Belgium). The output file from the robot contains information on the reactions (samples, targets) and their location in the multi-well plate, while the data from the qPCR apparatus describes the measured PCR curves.

8.3.1 Installation

To run qPCRAnalysis.pl, both Perl (www.activestate.com/activeperl) and Apache Webserver (www.apache.org) need to be installed. Two additional Perl modules are also required: CGI and Statistics::Regression (www.cpan.org). Next, the scripts need to be placed in the correct directory, usually 'cgi-bin', on the web server. The web-based user interface of the analysis program can be accessed through this URL: localhost/cgi-bin/qPCR/qPCRAnalysis.pl. The scripts are available upon request.

8.3.2 Methodology

The analysis methodology that was implemented is based upon the algorithm of qBase (15). Therefore, I will not give a fully detailed mathematical explanation here, but only touch upon the principles of the algorithms.

For each reaction, the C_T value and efficiency are calculated from the PCR curve, as explained previously. Next, the average C_T value is calculated for all replicates of a sample-target combination (typically three technical replicates are performed). The efficiency for all reactions with the same target (over all different samples) is averaged and this value is used to determine the expression value (based on formula 8.6). The same methodology is also applied to the reference genes and for each sample a normalization factor is calculated using the geometric average of the expression values of all reference genes for that sample. All samples are normalized through division with this normalization factor. Finally, the results are rescaled by dividing through the maximum expression level per target.

8.3.3 Walkthrough

The first page of the program displays a list of previous experiments (Figure 8.3A). By filling in the name of an existing experiment, this experiment is loaded, while typing a new name while start a new experiment. Clicking on 'Continue' takes you to the next page where 'runs' can be added to the experiment (Figure 8.3B). When an experiment encompasses several multi-well plates (each considered as a run), all data will be combined and analyzed together. By selecting the xml-file containing the raw data from the LightCycler 480 and the text-file from the robot and clicking 'Add', a run is added to the experiment. When all the runs are added, the analysis is started by clicking on 'Results'. The first step in the analysis is the calculation of the C_T value and efficiency for all reactions. Reactions that have the same sample name and target are considered as technical replicates. The reactions are ordered such that the replicates are grouped together, which is highlighted by the background color. When the difference between the minimum and the maximum C_T value is larger then 1 (this parameter is configurable in the previous page), the background is red to highlight this (Figure 8.3C). Now the user can deselect reactions that should not be included in the analysis based upon a outlying C_T values or low efficiency. When the selection is made, the next step is reached by clicking 'Calculate' at the bottom of the page. This last step encompasses the selection of the reference genes (Figure 8.3D). Multiple genes can be selected by holding 'Ctrl' (Windows) or 'Cmd' (Mac OS X). Upon clicking on 'Continue', the results are displayed: the expression values with standard error (Figure 8.3E), the normalization factor used (Figure 8.3F), the average efficiency (Figure 8.3G) and a quality assessment for the selected reference genes (Figure 8.3H-I). These quality controls are based on the calculations in GeNorm, more details can be found in Vandesompele *et al.*

(6). When the quality requirements, defined in this publication, are not reached, the results are displayed in red. Additionally, graphs of the expression results are made using gnuplot (www.gnuplot.info), which are accessed through the 'Graphs' button at the bottom of the results page. All results are saved as a tab-delimited text file or can be copied into a spreadsheet from the browser.

Figure 8.3: Screenshots from the different pages in the qPCR analysis program.
(A) Selection of the experiment. If the experiment exists the data is loaded, otherwise a new experiment is generated. (B) Input of the data. Each run requires two files: an xml-file from containing the qPCR curves (exported from the LightCycler 480) and an annotation file with the sample and target for each well. (C) Selection of the data. In this step outlying data points can be excluded from the analysis. (D) Selection of the reference genes for normalization. (E-I) Results of the analysis. (E) Expression values for each sample and target (average and standard error), (F) the applied normalization factors, (G) the efficiency of the primers for each target (1+E), (H,I) quality controls for the normalization based on the calculations of GeNorm (6).

8.3.4   Future improvements

The current implementation is installed on several computers in the lab as the way it saves the data (in a folder on the hard drive) is not suited for deployment on the server. To optimize this, data should be saved in a database and preferable be linked to the login of the user. Excluding all reactions of a certain sample or target is at this point only possible by deselecting all these reactions manually. Providing an easier way to do this would enhance the user experience greatly. Functionality involving biological replicates would also be interesting to add.

8.4   Publications

Next to the data analysis script, I was involved in setting up the whole qPCR workflow, including the programming of the pipetting robots: first the Xiril100 (Xiril AG, Hombrechtikon, Switzerland), which was recently replaced by a Janus Mini (Perkin Elmer, Zaventem, Belgium). This qPCR pipeline was used for expression analysis in collaboration with Paulo Ferreira (16) and Nathalie Verbruggen (17, 18). The first page of the publications in which I am co-author are included.

Report

The Arabidopsis Anaphase Promoting Complex (APC)
Regulation Through Subunit Availability in Plant Tissues

Nubia Barbosa Eloy[1]

Frederik Coppens[3]

Gerrit T.S. Beemster[3]

Adriana Silva Hemerly[1, 2]

Paulo Cavalcanti Gomes Ferreira[1,2,*]

[1]Instituto de Bioquímica Médica, CCS; Universidade Federal do Rio de Janeiro; Rio de Janeiro, Brazil

[2]Laboratório de Biologia Molecular de Plantas; Instituto de Pesquisas do Jardim Botânico do Rio de Janeiro; Rio de Janeiro, Brazil

[3]Department of Plant Systems Biology; Flanders Interuniversity Institute for Biotechnology; Ghent University; Gent, Belgium

*Correspondence to: P.C.G. Ferreira; Laaboratorio de Biolgia Molecular de Plantas; Instituto de Pesquisas do Jardim Botanico do Rio de Janeiro; Rua Pacheco Leao 915; 22460-030 Rio de Janeiro, RJ; Tel.: +55.21.32042086; Fax: +55.21.22948696; Email: paulof@bioqmed.ufrj.br

KEY WORDS

ABBREVIATIONS

| APC | anaphase promoting complex |
| CDC | cell division cycle |

ACKNOWLEDGEMENTS

ABSTRACT

Sister-chromatid separation and exit from mitosis require ubiquitin-mediated proteolysis of cell cycle regulators such as cyclin B and securin. The specificity of the reaction is controlled by an ubiquitin-ligase multiprotein complex known as APC (Anaphase Promoting Complex). Comparison of the coding sequences of *Arabidopsis* genes with the Genbank database reveals extensive homology of the predicted ORFs with the corresponding proteins of other eukaryotes, indicating that the APC is well conserved in plants. However, different from other eukaryotes, the *Arabidopsis* genes have some particular characteristics, such as the presence of two copies of the *CDC27* gene. Furthermore, expression analyses of the *AtAPC* genes disclose complex profiles that differ, depending on the tissue examined. In actively dividing cell suspensions there is a direct correspondence between the rates of proliferation and mRNA levels from the *AtAPC* components. On the other hand, in plant organs, dark-grown seedlings and during leaf growth, this correlation is lost and the *AtAPC* genes are highly expressed in tissues with low overall cell division. Moreover, expression patterns diverge between the subunit genes, raising the possibility that there could be more than one form of the APC, which would execute distinct functions during plant development. The results suggest that an important layer of regulation of APC/C in plants could operate through subunit availability in specific tissues and/or cellular compartments.

INTRODUCTION

Ubiquitination-mediated proteolysis is a primary mechanism by which the levels of regulatory proteins are controlled. The process of ubiquitination of a substrate involves the activity of a cascade of three enzymes, the ubiquitination-activating enzyme (E1), the ubiquitin-conjugating enzyme (E2), and the ubiquitin-protein ligase (E3). Normally the substrate specificity and regulation of ubiquitination is conferred by the E3 ubiquitin-protein ligase, which binds directly to the target protein and is the rate-limiting step in the ubiquitination cascade (reviewed in refs. 1 and 2). In the eukaryotic cell cycle, the ubiquitination reactions are mediated by two proteolytic pathways: the Skp1-Cullin-F-box-protein (SCF) and the Anaphase Promoting Complex or Cyclosome (APC/C).[3-7] Although the APC/C and SCF are evolutionarily related multiprotein complexes, they play different regulatory tasks during cell cycle progression. The SCF controls the G_1 to S transition by driving the degradation of G_1 cyclins and CDK inhibitors (reviewed in ref. 8). The activity of the APC/C is essential for the regulation of metaphase to anaphase transition and exit from mitosis by ordered destruction of mitotic regulators, including securin—an inhibitor of chromosome separation—cyclin A, cyclin B and many of the mitotic regulatory kinases (reviewed in refs. 2 and 9). The SCF ubiquitin-ligase is a macromolecular complex formed of a cullin subunit, Cul1; a RING-H2 protein, Hrt1/Rbx1; an F-box subunit and a linker subunit, Skp1.[10] The APC/C, like the SCF, contains a catalytic core formed of a cullin subunit, APC2, and a RING-H2 protein, APC11. The two proteins form an ubiquitin ligase core and this complex has the ability to bind E2s and stimulate non-specific ubiquitination activity.[11] The APC/C was first purified from *Xenopus* extracts and clam oocytes[6-7] and initially eight and nine subunits, respectively, were identified. Subsequently, the subunit composition of the complex has then been identified in different organisms[11-14] (reviewed in ref. 15), ranging from 8 to 13 subunits.[16-17] This complexity is unexpected because many other ubiquitin ligases are only composed of one or a few subunits, meaning that ubiquitin ligase activity does not inevitably depend on multiple subunits. Therefore, it remains puzzling why the APC is composed of so many protein components and what their individual functions are.

Early transcriptomic changes induced by magnesium deficiency in *Arabidopsis thaliana* reveal the alteration of circadian clock gene expression in roots and the triggering of abscisic acid-responsive genes

Christian Hermans[1], Marnik Vuylsteke[2,3], Frederik Coppens[2,3], Adrian Craciun[1], Dirk Inzé[2,3] and Nathalie Verbruggen[1]

[1]Laboratory of Plant Physiology and Molecular Genetics, Université Libre de Bruxelles, Bd du Triomphe, B-1050 Brussels, Belgium; [2]Department of Plant Systems Biology, VIB, B-9052 Ghent, Belgium; [3]Department of Plant Biotechnology and Genetics, Ghent University, B-9052 Ghent, Belgium

Author for correspondence:
Christian Hermans
Tel: +32 2 650 54 17
Email: chermans@ulb.ac.be

Summary

• Plant growth and development ultimately depend on environmental variables such as the availability of essential minerals. Unravelling how nutrients affect gene expression will help to understand how they regulate plant growth.

• This study reports the early transcriptomic response to magnesium (Mg) deprivation in Arabidopsis. Whole-genome transcriptome was studied in the roots and young mature leaves 4, 8 and 28 h after the removal of Mg from the nutrient solution.

• The highest number of regulated genes was first observed in the roots. Contrary to other mineral deficiencies, Mg depletion did not induce a higher expression of annotated genes in Mg uptake. Remarkable responses include the perturbation of the central oscillator of the circadian clock in roots and the triggering of abscisic acid (ABA) signalling, with half of the up-regulated Mg genes in leaves being ABA-responsive. However, no change in ABA content was observed.

• The specificity of the response of some Mg-regulated genes was challenged by studying their expression after other mineral deficiencies and environmental stresses. The possibility to develop markers for Mg incipient deficiency is discussed here.

Introduction

Magnesium (Mg) is one of the nine essential macro-nutrients that plants utilize in relatively large amounts for their growth and development (Williams & Salt, 2009). The best recognized physiological functions of Mg in plants are the harvesting of solar energy (Beale, 1999), the cation-mediated grana stacking of thylakoid membranes (Kaftan *et al.*, 2002), the activation of enzymes in metabolic biochemistry (Cowan, 2002), including those in the Calvin cycle (Pakrasi *et al.*, 2001), the chelation to nucleotidyl phosphate forms in the cell energy budget (Igamberdiev & Kleczkowski, 2001) and the involvement in nucleic acid folding and the chemical catalysis of RNA splicing (Pyle, 2002). The earliest symptoms observed within days to

weeks of Mg deficiency consist of an impairment in sugar partitioning, leading to starch accumulation (Fisher & Bremer, 1993; Cakmak *et al.*, 1994a,b; Mehne-Jakobs, 1995; Fisher *et al.*, 1998; Hermans *et al.*, 2004, 2005; Hermans & Verbruggen, 2005, 2008; Ding *et al.*, 2006) and the enhancement of antioxidative mechanisms (Cakmak & Marschner, 1992; Tewari *et al.*, 2004, 2006; reviewed in Cakmak & Kirkby, 2008). According to the plant species, root growth is not affected in the first weeks of Mg shortage, in contrast to the shoot, resulting in an increase of the root : shoot biomass ratio (Hermans *et al.*, 2004, 2005, 2006; Hermans & Verbruggen, 2005). As knowledge of the impact on other processes is scarce, a transcriptome analysis can provide nonbiased hints about early targets of Mg starvation. The rapid transcriptional changes

Research

Systems analysis of the responses to long-term magnesium deficiency and restoration in *Arabidopsis thaliana*

Christian Hermans[1], Marnik Vuylsteke[2,3], Frederik Coppens[2,3], Simona M. Cristescu[4], Frans J. M. Harren[4], Dirk Inzé[2,3] and Nathalie Verbruggen[1]

[1]Laboratory of Plant Physiology and Molecular Genetics, Université Libre de Bruxelles, Bd du Triomphe, B-1050 Brussels, Belgium; [2]Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium; [3]Department of Plant Biotechnology and Genetics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium; [4]Department of Molecular and Laser Physics, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands

Author for correspondence:
Christian Hermans
Tel: +32 2 650 5417
Email: chermans@ulb.ac.be

Summary

• Unravelling mechanisms that control plant growth as a function of nutrient availability presents a major challenge in plant biology. This study reports the first transcriptome response to long-term (1 wk) magnesium (Mg) depletion and restoration in *Arabidopsis thaliana*.
• Before the outbreak of visual symptoms, genes responding to Mg starvation and restoration were monitored in the roots and young mature leaves and compared with the Mg fully supplied as control.
• After 1 wk Mg starvation in roots and leaves, 114 and 2991 genes were identified to be differentially regulated, respectively, which confirmed the later observation that the shoot development was more affected than the root in Arabidopsis. After 24 h of Mg resupply, restoration was effective for the expression of half of the genes altered. We emphasized differences in the expression amplitude of genes associated with the circadian clock predominantly in leaves, a higher expression of genes in the ethylene biosynthetic pathway, in the reactive oxygen species detoxification and in the photoprotection of the photosynthetic apparatus. Some of these observations at the molecular level were verified by metabolite analysis.
• The results obtained here will help us to better understand how changes in Mg availability are translated into adaptive responses in the plant.

Introduction

Plants require magnesium (Mg) to harvest solar energy and to drive photochemistry. This is probably one the most important physiological functions of this metal as the central atom of chlorophyll (Wilkinson *et al.*, 1990; Hörtensteiner, 2009). Signs of Mg deficiency in most plants usually manifest belatedly as a chlorophyll breakdown between the veins and make their appearance first in mature leaves, systematically progressing from these towards the youngest ones (Bennett, 1997; Hermans & Verbruggen, 2008). The knowledge about Mg^{2+} uptake by roots, transport to shoots and recycling between organs is relatively limited (Gardner, 2003; Karley & White, 2009). The

few physiological reports essentially describe an early impairment in sugar partitioning (in Arabidopsis, Hermans & Verbruggen, 2005; bean plants, Fisher & Bremer, 1993; Cakmak *et al.*, 1994a,b; rice, Ding *et al.*, 2006; spinach, Fisher *et al.*, 1998; spruce, Mehne-Jakobs, 1995 and sugar beet, Hermans *et al.*, 2004, 2005). One dramatic effect of Mg starvation is sugar accumulation in source leaves, before any noticeable effect on photosynthetic activity. Sugar accumulation in source leaf tissues, rather than a reduction in the amount of Mg available for chlorophyll biosynthesis, could be at the root of the decrease in chlorophyll content (Hermans *et al.*, 2004; Hermans & Verbruggen, 2005). A later effect of Mg deficiency is a reduction of plant growth and a modification of

References

[1] A. Giulietti, *et al. Methods* **25**, 386 (2001).

[2] S. N. Peirson, J. N. Butler, R. G. Foster. *Nucleic Acids Res* **31**, e73 (2003).

[3] S. A. Bustin, *et al. Clin Chem* **55**, 611 (2009).

[4] R. H. L. Deprez, A. C. Fijnvandraat, J. M. Ruijter, A. F. M. Moorman. *Analytical Biochemistry* **307**, 63 (2002).

[5] J. Vandesompele, A. D. Paepe, F. Speleman. *Anal Biochem* **303**, 95 (2002).

[6] J. Vandesompele, *et al. Genome Biol* **3**, RESEARCH0034 (2002).

[7] S. Rozen, H. Skaletsky. *Methods Mol Biol* **132**, 365 (2000).

[8] S. F. Altschul, *et al. J Mol Biol* **215**, 403 (1990).

[9] J. E. Stajich, *et al. Genome Res* **12**, 1611 (2002).

[10] M. W. Pfaffl. *Nucleic Acids Res* **29**, e45 (2001).

[11] J. Peccoud, C. Jacob. *Gene quantification*, chap. Statistical estimation of PCR amplification rates, 111–128 (New York Birkhäuser, 1998).

[12] P. Kainz. *Biochim Biophys Acta* **1494**, 23 (2000).

[13] W. M. Freeman, S. J. Walker, K. E. Vrana. *Biotechniques* **26**, 112 (1999).

[14] C. Ramakers, J. M. Ruijter, R. H. L. Deprez, A. F. M. Moorman. *Neurosci Lett* **339**, 62 (2003).

[15] J. Hellemans, *et al. Genome Biol* **8**, R19 (2007).

[16] N. B. Eloy, *et al. Cell Cycle* **5**, 1957 (2006).

[17] C. Hermans, *et al. New Phytol* **187**, 119 (2010).

[18] C. Hermans, *et al. New Phytol* **187**, 132 (2010).

General Conclusion

In this thesis, I set out to deepen our understanding of the regulation of leaf development in plants. To this end I studied known cell regulators as well as unknown proteins with a proliferation-specific expression profile. For the unknown proteins my results confirmed that these genes indeed play a role in the M-phase of the cell cycle. While I did not unravel the function of these genes in detail yet, my results constitute a start with the functional analysis that will allow further investigation of their role in development. The selection procedure was thus successful in identifying genes that are involved in cell cycle regulation. We focused on two gene families but the remaining genes in our list of proliferation genes will certainly yield more interesting leads to better understand how plants regulate proliferation and development.

On a different regulatory level, I profiled small RNAs during leaf development. The results showed that small RNAs are differentially regulated during leaf development and are likely to play an important role in regulation of organ development of plants. Especially regulation by DNA methylation appears to be an important mechanism to silence genes in the course of development. Profiling of methylation at higher resolution than used in the experiment I conducted would allow assessing this at the gene level.

While there are still many outstanding questions in the regulation of cell cycle and organ development, my work contributed to the international effort to steadily gather the pieces of the puzzle. Ultimately, we hope that through this research of basic developmental processes, we will be able to improve plant growth in specific conditions and contribute to solve the upcoming challenges in our society concerning food, feed and bioenergy.

Curriculum Vitae

Personalia

Name	Frederik Coppens
Address	Gravenstraat 32A, 9968 Bassevelde, Belgium
Email	frederik.coppens@gmail.com
Mobile	+32 (0) 477/13 63 20
Date of birth	July 26, 1982
Place of birth	Eeklo, Belgium
Nationality	Belgian
Marital status	married to Kathleen De Graeve
Languages	Dutch: native, English: fluent, French: good, German: basic

Current position

Position	Pre-doctoral fellow
Institute	Department of Plant Systems Biology
	Flanders Institute of Biotechnology, Gent, Belgium
Address	Technologiepark 927, 9052 Gent, Belgium
Email	frcop@psb.vib-ugent.be
Tel	+32 (0) 9/33 13 953

Education

2005 – 2010 Ph.D. fellowship

Department Plant Biotechnology and Genetics, Ghent University, Gent, Belgium; Department of Plant Systems Biology, Flanders Institute of Biotechnology (VIB), Gent, Belgium

2005 – 2009: Ph.D. fellowship of the Research Foundation Flanders (FWO)

2009 – 2010: Ph.D. fellowship of the Special Research Fund of Ghent University (BOF fund)

Promoter: Prof. Dr. Dirk Inzé

Co-promoters: Prof. Dr. Gerrit T.S. Beemster, Prof. Dr. James C. Carrington

2000 – 2005 Master in Bioscience Engineering

>   Faculty of Bioscience Engineering, Ghent University, Gent, Belgium
>   Undergraduate dissertation: "Functional analysis of cell cycle genes in *Arabidopsis thaliana* leaf development"
>   Promoter: Prof. Dr. Dirk Inzé, Prof. Dr. Godelieve Gheysen
>   Great distinction

Experience

2005 – 2010   PhD fellow Ghent University

Sep 2007   Short visit: small RNA sequenering experiment (Solexa), Carrington Lab, Oregon State University, OR, USA

Sep – Dec 2009   Microarray analysis for DevGen

2005 – 2006   Supervising Stijn Dhondt, Master Bioscience Engineering, "Functionele analyse van proliferatiegenen uit de HMGB en ENODL familie in *Arabidopsis thaliana*"

2005 – 2006   Supervising Yonca Surgun, undergraduate student, "Functional analysis of 3 *HMGB* genes with a potential role in cell cycle regulation in *Arabidopsis thaliana*"

2006 – 2007   Supervising Liesbeth De Geest, Bachelor biomedische laboratoriumtechnologie, "Fenotypische analyse van potentiële, ongekende celcyclusregulatoren en de cycline A2 familie in *Arabidopsis thaliana*"

2006 – 2007   Supervising Joanna Rosołowska, 1st year Master Food Sciences and Biotechnology, "Functional analysis of new cell cycle regulators in *Arabidopsis thaliana*"

2007 – 2008   Supervising Joanna Rosołowska, Master Food Sciences and Biotechnology, "Functional analysis of new cell cycle regulators in *Arabidopsis thaliana*"

2006 – 2009   Assistance practical course Plant Physiology by Prof. Dr. Dirk Inzé

2009 – 2010   Responsable practical course Plant Physiology by Prof. Dr. Dirk Inzé

Publications

Peer-reviewed publications

Eloy, N. B., Coppens, F., Beemster, G. T. S., Hemerly, A. S. and Ferreira, P. C. G. (2006). The *Arabidopsis* anaphase promoting complex (APC): regulation through subunit availability in plant tissues, *Cell Cycle*, 5, 1957–1965

Achard, P., Gusti, A., Cheminant, S., Alioua, M., Dhondt, S., Coppens, F., Beemster, G. T. S. and Genschik, P. (2009). Gibberellin signaling controls cell proliferation rate in *Arabidopsis*, *Curr Biol*, 19, 1188–1193

Hermans, C., Vuylsteke, M., Coppens, F., Craciun, A., Inzé, D. and Verbruggen, N. (2010). Early transcriptomic changes induced by magnesium deficiency in *Arabidopsis thaliana* reveal the alteration of circadian clock gene expression in roots and the triggering of abscisic acid-responsive genes, *New Phytol*, 187, 119–131

Hermans, C., Vuylsteke, M., Coppens, F., Cristescu, S. M., Harren, F. J. M., Inzé, D. and Verbruggen, N. (2010). Systems analysis of the responses to long-term magnesium deficiency and restoration in *Arabidopsis thaliana*, *New Phytol*, 187, 132–144

Dhondt, S., Coppens, F., De Winter, F., Swarup, K., Merks, R. M. H., Inzé, D., Bennett, M. J. and Beemster, G. T. S. (2010). SHORT-ROOT and SCARE-CROW regulate leaf growth in *Arabidopsis* by stimulating S-phase progression of the cell cycle, *Plant Physiol*, 154, 1183–1195

Manuscripts in preparation

Vanneste, S., Coppens, F., Lee, E., Donner, T., Xie, Z., Van Isterdael, G., Dhondt, S., De Winter, F., De Rybel, B., De Veylder, L., Friml, J., Inzé, D., Grotewold, E., Scarpella, E., Sack, F., Beemster, G. T. S. and Beeckman, T. (submitted). Plant CYCA2s are G2/M regulators that are transcriptionally repressed during differentiation
S.V. and F.C. contributed equally to this work

Skirycz, A., Claeys, H., De Bodt, S., Yusuke, J., Oikawa, A., Andriankaja, M., Maleux, K., Eloy, N. B., Coppens, F., Kamiya, Y. and Inzé, D. (submitted). Pause-and-stop – the effects of osmotic stress on cell proliferation and a role for ethylene signalling

Andriankaja, M., Dhondt, S., De Bodt, S., Coppens, F., Skirycz, A., Gonzalez, N., Beemster, G. T. S. and Inzé, D. (submitted). Early leaf development: a not so gradual process

Book chapters

Rymen, B., Coppens, F., Dhondt, S., Fiorani, F. and Beemster, G. T. S. (2010). Kinematic analysis of cell division and expansion, *Methods Mol Biol*, 655, 203–227

Conferences

Coppens, F., Kasschau, K. D., Carrington, J. C., Inzé, D. and Beemster, G. T. S. (2008). Profiling of Small RNAs in *Arabidopsis* Leaf Development

Conferences

Invited Speaker

Profiling of Small RNAs in *Arabidopsis* Leaf Development NIBB Conference 2008, Okazaki, Japan

nCounter®, multiplex transcript quantification using a molecular barcoding detection system MAF user meeting 2009, Gent, Belgium

Attendee

Life, A Nobel Story, 2004, Expo Heizel, Brussels, Belgium

Benelux qPCR symposium 2008, ICC Ghent, Belgium

qPCR symposium 2009, ICC Ghent, Belgium

Plant Growth Biology and Modeling workshop, 2009, Elche, Spain

Advances in Genomics symposium 2010, ICC Ghent, Belgium

Knowledge for Growth 2010, ICC Ghent Belgium

Extra-curricular experience

2000 – 2010 Bestuurslid Koninklijke Harmonie Amicitia, Eeklo

2006 – 2010 Lid Comité voor Verzustering Eeklo

1998 – 2002 Lid Dagelijks Bestuur Jeugdraad Eeklo

Dankwoord

Nu het officiële gedeelte achter de rug ligt, is het hoog tijd om vele mensen te bedanken.

In de eerste plaats wil ik Gerrit bedanken. Hij nam me, ondertussen meer dan zes jaar geleden, als thesisstudent op in zijn onderzoeksgroep. Vanaf het begin wist hij me warm te maken voor wetenschappelijk onderzoek en al snel was het duidelijk dat een doctoraat me wel zou aanspreken. In de loop van de jaren heb ik Gerrits wetenschappelijke ingesteldheid en aanpak enorm geapprecieerd. Zijn enthousiasme en steun hebben er mee toe geleid dat ik een paar weken naar de VS kon gaan werken, meer bepaald in Carrington-lab. Een ervaring die ik iedereen kan aanraden. Iets meer dan twee jaar geleden besloot Gerrit om andere oorden op te zoeken en verhuisde ik naar de groep van Dirk. Bedankt om al onze experimenten en projecten mogelijk te maken.

In de afgelopen jaren heb ik deel uitgemaakt van twee groepen binnen PSB, ik wil ook de medewerkers van deze groepen hartelijk danken. Ooit begon hij als student, maar ondertussen is Stijn een collega geworden op wie ik altijd kan rekenen, zeker bij het bedenken en realiseren van weer een 'wild' idee om het labowerk iets meer te automatiseren. Nathalie, merci beaucoup de m'obliger à parler français, je me souviendrai surtout de nos conversations scientifiques et parfois 'un peu moins scientifiques'. Ik vermoed dat het voor de meesten wel duidelijk is dat bio-informatica 'mijn ding' is, maar het is niet altijd even evident om er je weg in te vinden. Stefanie, bedankt om me er verder wegwijs in te maken, het was steeds fijn samen te werken. Aleksandra, Elena, Hannes C, Hannes V, Jasper, Liesbeth, Lieven, Marieke, Mattias, Megan, Nubia, Pieter, Sergei, Wannes, the 'maize people' Bart, Hilde, Jérôme, Joke and Jolien, I've very much enjoyed working with you guys. Els, Freya, Katrien, Liesbeth en Twiggy, bedankt voor alle hulp bij het uitzaaien, oogsten, tekenen, blad-series-maken,... Ik heb ook een aantal studenten 'onder mijn hoede' gehad: Stijn, Yonca, Joanna en Liesbeth, dankjewel voor al het werk dat jullie verzet hebben.

Ook buiten de eigen groep heb ik de afgelopen jaren vele collega's leren kennen en ik heb veel van hen geleerd. Samen gestudeerd, beiden een masterthesis gedaan op PSB en vervolgens allebei een doctoraat begonnen; Bert, dikke merci voor al die jaren. Steffen en Daniël, bedankt voor de vele hulp en de vlotte samenwerking. Een welkom rustpunt in de dag dankzij onze vaste lunchgroep: Bert, Gert, Giel, Leen,

Marlies, Stijn en Wim. Iedereen die geholpen heeft bij de practica: Annelies, Astrid, Bart, Bert, Carina, Eveline, Hilde, Korneel, Liesbeth, Silvie, Stephen, Stijn, Olivier en Toon, het was aangenaam om samen met jullie de studenten te trotseren. Fabio, abbiamo avuto tante e talvolta lunghe discussioni sulle scienze. Mi hai sempre sorpreso perché, mi hai costretto a riflettere sempre un po' più. In bocca al lupo a Jülich ! Anche Tommaso, ti auguro buona fortuna a Milano!

Op PSB hebben we het geluk dat vele tijdrovende zaken voor ons gedaan worden: het welzijn van onze plantjes wordt verzekerd door Nico en Miguel; Jackie en Kristof zorgen ervoor dat al ons materiaal proper en steriel is; dankzij Nancy kunnen we de meeste buffers en media gewoon van het schap halen en An zorgt ervoor dat de producten die we dagelijks nodig hebben op diezelfde schappen voor het rapen liggen; Carina en Els nemen de transformaties voor hun rekening; Wilson zorgt ervoor dat je zeker weet dat je de juiste sequentie gekloneerd hebt in de vector die Mansour maakte; maar als je geluk had kon je die kloon ook gewoon bij Rebecca halen; Karel verzorgt de professionele fotoshoots voor onze plantjes; mijn partner-in-crime voor qPCR en onze hulp bij robotongevallen, Nino; voor technische problemen of hulp bij weer een nieuwe, ingewikkelde constructie om een 'wild' idee te verwezenlijk kunnen we bij Dirk terecht; publicaties uit de meest exotische journals vergen slechts een mailtje naar Jacques of Martine; bij computerproblemen, om dat wilde idee van daarnet van IT-kant te ondersteunen, toegang tot allerlei servers en netwerken te krijgen of te Skypen tijdens je doctoraatsverdediging, kan je bij Hendrik, Raf, Dany en Luc terecht (binnen de openingsuren uiteraard, dat spreekt voor zich); wil je net die ene bio-informaticatool die niet op de server staat, heb je de MySQL-server gecrasht of heb je je net gerealiseerd dat die ene analyse zo'n drie weken zal duren, dan heb je Frederik die dat installeert, herstart of je leert om de cluster te gebruiken; het administratieve kluwen van VIB, PSB, UGent, allerlei beurzen enz. wordt iets draaglijker door Diane, Christine, Bernard, David, Hilde en Nathalie. Voor alle hulp gedurende die jaren: enorm bedankt!

We considered doing a small RNA experiment in March 2007, got an enthusiastic response from Jim Carrington and September the same year I've spent in Oregon extracting these tiny molecules. Jim and Kristin, thank you for making this a fantastic experience. Taiowa and Brooke, thanks for letting me stay with you during those weeks, I've enjoyed it very much.

Maar uiteraard is niet alleen werk van belang. In de eerste plaats wil ik mijn ouders bedanken voor alle kansen die ze me gegeven hebben, zonder hen zou ik hier vandaag niet staan. Ook aan mijn broers, Tonny en Gino, schoonzussen, Sylvie en Katleen en neefjes en nichtjes Elies, Sien, Wout, Thomas en Liesel, bedankt voor de steun door de jaren. Sinds een paar jaar heb ik er een tweede familie bij gekregen die me enorm gesteund heeft: Hilda en Johan, Veerle, Davy, Els, Guy, Jan, Lies en hun kroost Rune, Lore, Lynn, Dries, Robbe en Roos. Ook een dikke merci aan 'de cel-en-genners', Bert en Karen, Caroline en Koen, Delphine, Els en Kris, het is altijd

plezant om elkaar nog eens terug te zien. We hebben het nu allemaal druk met doctoraten schrijven, huizen bouwen, nieuwe jobs of kindjes, maar ik hoop dat het binnenkort nog eens lukt om af te spreken. Simon, dankjewel voor zowel alle hulp bij het maken van dit boekje als, samen met Sofie, voor de gezellige weekendjes en Catan-avonden. Last but not least: Kathleen, schat, bedankt voor alle geduld, vooral de afgelopen maanden, vanaf nu zal vakantie écht vakantie zijn!

Ik hoop dat het niet nodig is, maar mocht ik toch iemand vergeten zijn: bedankt aan iedereen die me de afgelopen jaren, op welke manier dan ook, geholpen heeft.

Bassevelde, 10 januari 2011
Frederik Coppens