1-1-2011

The Role of dsRNA in Nuclear Differentiation and Remodeling in the Ciliate, Tetrahymena thermophila

Jason Motl
Washington University in St. Louis

WASHINGTON UNIVERSITY

Division of Biology and Biomedical Sciences

Molecular Genetics & Genomics

Dissertation Examination Committee:
Douglas L. Chalker, Chair
Sarah C. R. Elgin
Susana Gonzalo-Hervas
John E. Majors
Tim B Schedl
Sheila A. Stewart

THE ROLE OF dsRNA IN NUCLEAR DIFFERENTIATION AND REMODELING IN THE CILIATE, *TETRAHYMENA THERMOPHILA*

by

Jason Andrew Motl

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

August 2011

Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

The Role of dsRNA in Nuclear

Differentiation and Remodeling

in the Ciliate, *Tetrahymena thermophila*

by

Jason Andrew Motl

Doctor of Philosophy in Biology and Biological Sciences

(Molecular Genetics and Genomics)

Washington University in St. Louis, 2011

Professor Douglas L. Chalker, Chairperson

The ciliate, *Tetrahymena thermophila,* like a handful of other eukaryotes, engages in massive genome reorganization known collectively as chromatin diminution. Part of this process involves large-scale DNA excision known as DNA elimination. Recent data has shown DNA elimination to be dependent on RNA interference (RNAi). Using *T. thermophila*, I have sought to determine the role of non-coding RNA (ncRNA) in RNAi-dependent DNA elimination through studies of DNA sequences that are to be eliminated called **i**nternal **e**liminated **s**equences (IESs) and through a conjugation-specific Dicer protein and its putative tandem dsRNA-binding motif (DSRM) protein partners. Studies of the R IES revealed the requirement of IES DNA for production of long, bidirectional ncRNA early in conjugation. This ncRNA is essential for IES excision in zygotic nuclei later in conjugation. The conjugation-specific Dicer homologue, DCL1, was shown to be required for production of a species of sRNA called scnRNAs from the long,

bidirectional ncRNA from IESs. Knockouts of DCL1 displayed a loss of these scnRNAs as well as an increase in the long, bidirectional ncRNA precursors. A deficiency in these scnRNAs was sufficient to block modification of chromatin associated with IESs and prevent their rearrangement later in conjugation. Failure of DNA elimination caused DCL1 knockout cells to arrest before completion of conjugation. Further studies of the tandem DSRM-containing proteins, DRB2 and DRB1, revealed that neither are solely partners for DCL1 or any other Dicer protein but play other important roles during conjugation. Zygotic expression of DRB2 was shown to be essential for DNA elimination and completion of conjugation. Interaction with the chromo-domain containing protein, Pdd1p, by Drb2p implicates ncRNA or sRNA in later stages of conjugation after scnRNA production. Knockouts of the tandem DSRM-containing DRB1 caused higher numbers of cells to abort conjugation and therefore produce fewer progeny. Localization of this protein to the crescent micronucleus during prophase of meiosis I links DRB1 to a probable role in ensuring proper recombination during meiosis for haploid gamete production. All these studies suggest that ncRNA has many roles in conjugation-specific processes including RNAi-directed DNA elimination.

ACKNOWLEDGEMENTS

I would like to thank all the current members of the Chalker lab, Annie Shieh, Scott Horrell, Christine Carle and Rachel Schwope, for their support when things went wrong, for listening and helping when I had questions and for producing a lab environment conducive to research (even when no experiments were working). In addition I would like to thank Scott Horrell for taking the time to proofread portions of this dissertation. I would also like to thank all the past members of the Chalker lab. In particular, I would like to thank Colin Malone and Alissa Anderson for all their help with the DCL1 knockout project and for making the Chalker lab feel like a home, in which I could do research.

Next I would like to thank my thesis committee for their probing questions and excellent advice to further this research and produce the best data in the finite amount of time I had. In particular I would like to show my deep gratitude to my thesis advisor Doug Chalker for showing the patience and flexibility I needed to do research while trying to balance work with a growing young family. Given a different thesis advisor I do not think I could have completed these research projects and this dissertation. Thank you Doug.

Lastly I would like to thank my wife, Jen, and my children, Kaya and Nathan, for supporting me through graduate school and lifting me up after experiments did not work or produced confounding results. Being a father to Kaya and Nathan is as much, if not more, a part of my identity as being a scientist. I love you both very much. Finally I would like to thank Jen for catching the last few typos I had in this thesis.

TABLE OF CONTENTS

Abstract: **ii**

Acknowledgments: **iv**

Chapter 1: Introduction **1**
I: Epigenetics of Ciliates
1.Ciliate Biology **2**
2. Epigenetic Phenomena in Ciliates **12**
3. RNA-Mediated Epigenetic Mechanisms **20**
4. Small RNA-Mediated DNA Rearrangements **25**
5. Chromosome Fragmentation and Elimination of DNA During During Conjugation in *Oxytricha* **50**

II: Double-Stranded RNA Binding Motif Proteins **53**

III: Scope of the Thesis **55**

Chapter 2: Transcription of dsRNA from the R IES is Required for Subsequent Targeting and DNA Elimination Late in Conjugation in *Tetrahymena thermophila* **86**

Chapter 3: Germ Line Transcripts are Processed by a Dicer-Like Protein that is Essential for Developmentally Programmed Genome Rearrangements of *Tetrahymena thermophila* **106**

Chapter 4: The Role of the Tandem Double-Stranded RNA Binding Motif Protein, DRB2, in DNA Elimination in the Ciliate, *Tetrahymena thermophila* **123**

Chapter 5: The Double-Stranded RNA Binding Motif Protein, DRB1, Plays an Auxiliary Role in Meiosis I in the Ciliate, *Tetrahymena thermophila* **173**

Chapter 6: Future Perspectives **207**

Appendix: Subtraction by Addition: Domesticated Transposases in Programmed DNA Elimination **221**

FIGURES AND TABLES

Chapter 1:	Introduction	**Page**
Figure 1:	Ciliate life cycle	**74**
Figure 2:	Somatic genome rearrangements of ciliates through chromosome fragmentation and DNA elimination	**75**
Table 1:	Histone modifications found in the nuclei of ciliates	**76**
Figure 3:	Cytoplasmic inheritance-mating type determination in *Paramecium*	**77**
Figure 4:	Epigenetic inheritance of mating type in *Paramecium* demonstrated by the mutation of mtF^{E}	**78**
Figure 5:	RNAi-dependent DNA elimination in *P. tetraurelia*	**79**
Figure 6:	RNAi and heterochromatin components of RNAi-dependent DNA elimination in *T. thermophila*	**81**
Figure 7:	Histone methylation triggers DNA elimination of internal eliminated sequences (IESs) in *T. thermophila*	**83**
Figure 8:	Template-guided gene unscrambling, DNA elimination and chromosome fragmentation in *O. trifallax*	**85**
Chapter 2:	Transcription of dsRNA from the R IES is Required for Subsequent Targeting and DNA Elimination Late in Conjugation in *Tetrahymena thermophila*	
Figure 1:	Diagram of R IES micro- and macronuclear loci in wild-type and R IES knockout strains	**100**
Figure 2:	Creation and verification of R IES knockout strains	**101**
Table 1:	Progeny production of R IES mic knockouts in wild-type and knockout matings	**102**
Figure 3:	Failure of plasmid-based R IES rearrangement in homozygous micronuclear R IES knockout strain matings	**103**
Table 2:	Oligonucleotides used in the course of this study	**105**
Chapter 3:	Germ Line Transcripts are Processed by a Dicer-Like Protein that is Essential for Developmentally Programmed Genome Rearrangements of *Tetrahymena thermophila*	
Table 1:	Oligonucleotides used in course of this study	**111**
Figure 1:	*Tetrahymena thermophila* encodes three Dicer-like proteins	**113**
Figure 2:	Germ line knockout of *DCL1*	**114**
Figure 3:	Δ*DCL1* strains arrest late in conjugation	**115**
Figure 4:	Conjugating Δ*DCL1* strains exhibit loss of small RNA production and germ line transcript accumulation	**116**
Figure 5:	The *DCL1* protein is localized to meiotic micronuclei.	**117**
Figure 6:	*DCL1* knockouts fail to excise micronucleus-limited DNA	**118**
Figure 7:	Chromosome breakage does not occur in *DCL1* knockouts	**118**

		Page
Figure 8:	*DCL1* is not required for H3K9 methylation	**119**
Chapter 4:	The Role of the Tandem Double-Stranded RNA Binding Motif Protein, DRB2, in DNA Elimination in the Ciliate, *Tetrahymena thermophila*	
Figure 1:	*T. thermophila* contains two predicted tandem double-stranded RNA binding motif proteins	**150**
Figure S1:	ClustalW alignment of dsRNA binding motifs (DSRMs) of *D. melanogaster* Loqs and R2D2, *C. elegans* RDE-4, *H. sapiens* TRBP2 and *T. thermophila* DRB2 and DRB1	**151**
Figure S2:	Alignment of DRB1 and DRB2 protein sequences outside of the DSRMS shows additional regions of homology	**155**
Figure 2:	Nuclear localization of Drb2p and Drb1p during conjugation	**159**
Figure S3:	Southern blots and RT-PCR during conjugation of DRB1 knockout strains and DRB2 mic knockout strains	**160**
Table 1:	Progeny production of DRB1 knockouts in wild-type and knockout matings	**162**
Figure 3:	Zygotic expression of DRB2 is necessary for completion of conjugation	**163**
Table 2:	Progeny production of DRB2 mic knockouts in wild-type and knockout matings	**164**
Figure 4:	DNA rearrangement of internal eliminated sequences (IESs) and chromosome breakage are impaired in DRB2 mic knockouts	**165**
Figure S4:	Global DNA rearrangement of IESs is impaired in DRB2 mic knockouts	**167**
Figure 5:	Drb2p co-localizes with the essential conjugation chromodomain protein, Pdd1p, in DNA elimination bodies.	**168**
Figure 6:	Failure of DNA elimination bodies to form in DRB2 mic knockouts late in conjugation	**169**
Table 3:	Oligonucleotides used in the course of this study	**171**
Chapter 5:	The Double-Stranded RNA Binding Motif Protein, DRB1, Plays an Auxiliary Role in Meiosis I in the Ciliate, *Tetrahymena thermophila*	
Figure 1:	Somatic knockouts of DRB1 produce normal levels of scnRNAs and are able to undergo DNA elimination normally	**197**
Figure 2:	More DRB1 knockout pairs fail to progress to formation of zygotic nuclei during conjugation compared to wild-type	**199**
Table S1:	Mating progression of wild-type (pICC-CNA1 #10.1 X pICC-CNA1 #16.5) and DRB1 knockout (DRB1 KO pICC-CNA1 #13.1 X DRB1 KO pICC-CNA1 #14) matings during conjugation until terminal arrest	**200**
Figure 3:	Exogenous expression of DRB1 rescues DRB1 knockout	**201**

	backout phenotype during conjugation	**Page**
Table S2:	Mating progression of wild-type matings without exogenous DRB1 (CU427 X CU428) and with exogenous DRB1 (CU427 X pICY-DRB1 #5) compared with DRB1 knockout matings without exogenous DRB1 (DRB1 KO #5.1.3 X DRB1 KO #7.1) and with exogenous DRB1 (DRB1 KO pICY-DRB1 #1 X DRB1 KO pICY-DRB1 #2) during conjugation until terminal arrest	**202**
Figure 4:	DRB1 genomic locus, expression and protein localization at the tips of the crescent micronucleus prior to anaphase of meiosis I during conjugation	**203**
Figure 5:	Co-localization of Drb1p-YFP early in conjugation with the centromere histone, Cna1-CFP, in the crescent micronucleus and with the nucleolar protein, Nopp52p, in the macronucleus	**204**
Table 1:	Oligonucleotides used in the course of this study	**205**
Appendix:	Subtraction by Addition: Domesticated Transposases in Programmed DNA Elimination	
Figure 1:	Pgm-mediated DNA elimination of an IES and the evolution of IESs and a domesticated transposase	**224**

CHAPTER 1

INTRODUCTION

EPIGENETICS OF CILIATES

Encyclopedia of Molecular Cell Biology and Molecular Medicine. 2011. In Press.

I. Epigenetics of Ciliates

Jason A. Motl, Annie W. Shieh, and Douglas L. Chalker

Washington University in St. Louis, St. Louis, Missouri, United States of America

1

Ciliate Biology

1.1

Historical Perspective

The concept of "epigenetics" was largely formulated by Conrad Waddington to provide a framework to describe the development of multi-cellular organisms and explain how cells with the same genetic composition can differentiate into functionally distinct types. During these early days of genetics research, the chromosome theory of inheritance was viewed to bridge the observations of Mendelian inheritance and microscopic description of chromosome behavior in cells. However, this genetic theory was somewhat inadequate to account for development of different tissues within an individual, where all cells had the same chromosomes. It was difficult to envision how the apparently static chromosomes (genes) could by themselves manifest phenotypic differences – the fundamental basis of cellular differentiation.

Epigenetic theory thus arose from the need to bridge the gap between genotype and observed phenotypes that could not be accounted for by the behavior of chromosomes. The gap was quite apparent in single celled organisms, most notably the ciliate *Paramecium* in studies by Tracy Sonneborn [1]. Sonneborn and his colleagues described several examples of phenotypic traits, e.g. serotype and mating type, which did not follow conventional Mendelian inheritance, but instead appeared to be passed on

through cytoplasmic inheritance. Thus while these traits were encoded by genes, clonal lines with identical genotypes arose with persistently different phenotypes. Through these studies, Sonneborn and others revealed that the cytoplasm was an important supplement to chromosomes in transmitting heritable information.

While studies of ciliate genetics largely started with Sonneborn, research using these organisms has continued to provide important understanding of epigenetic phenomena and their underlying mechanisms that help explain unexpected patterns of inheritance. In this review, early examples of non-Mendelian inheritance observed in ciliates are described to provide a historical context, even though the exact mechanisms that account for these phenomena still await discovery. Nonetheless, research efforts aimed at describing the intricate biology of this fascinating group of microbes have provided new ways to think about epigenetics that stretch well beyond ciliates. Fundamental discoveries of the role of chromatin modification in gene regulation and the role of non-coding RNAs in gene silencing have secured the place of ciliates as pioneering model systems for epigenetic studies. Much of the utility of these organisms for this research stems from their unique biology with both germline and somatic copies of the genome maintained in a single cell. Below we first describe the germline and somatic dichotomy of ciliates and the process of their differentiation to provide the necessary background to describe these epigenetic discoveries.

1.2

Life Cycle and Genetics

The ciliated protozoa belong to the superphylum of Alveolates, which is a lineage that diverged from the ancestors of plants and animals more than a billion years ago [2].

They have evolved into a diverse array of species that have adapted to different environments and strategies for life. Members of the phylum *Ciliophora* (i.e. ciliates) are commonly found in fresh water, but can also be found in many water-rich environments as free-living organisms, symbiotes, or even parasites. Ciliates have elaborate cellular architectures, most noticeable in the organized arrays of cilia that cover their exteriors. An anterior oral apparatus or "mouth", constant swimming enabled by their cilia and relatively large size give them animal-like qualities despite being unicellular. Ciliates are capable of both asexual and sexual reproduction. Asexual reproduction or vegetative population growth occurs by binary fission and is the means through which ciliates amplify their populations clonally (Figure 1A). In contrast, sexual reproduction occurs upon conjugation of two cells, which involves the exchange of genetic information between each partner and new somatic genome differentiation without an increase in cell number (Figure 1B).

The most important feature of ciliates to consider in regard to inheritance is their nuclear dimorphism. In each single cell, ciliates organize two copies of their genome in nuclei that are structurally and functionally distinct. These two different genomes serve the analogous roles to that of germline and somatic cells in metazoans. The germline copy of the genome is contained in the smaller nuclear compartment, the micronucleus. Micronuclei are diploid, but interestingly are transcriptionally silent during vegetative growth, serving only to maintain and transmit the genome to progeny cells upon sexual reproduction. The much larger macronuclei, on the other hand, carry the somatic genome and, as such, are responsible for all gene expression necessary for vegetative growth. Macronuclei are polyploid, with different ciliate species having widely different copy

numbers in their somatic genomes. For example, *Tetrahymena* retain ~50 copies of each macronuclear chromosome whereas *Paramecium* macronuclei contain several hundred copies. During sexual reproduction, the macronucleus, like the soma of metazoa, is lost when a new one is formed from a zygotic nucleus, which is derived from the germline genomes of the parental cells after meiosis.

While all ciliates exhibit nuclear dimorphism, the actual number of germline micronuclei and somatic macronuclei in each cell differs between species. In many of the figures, we have elected to illustrate a single micro- and macronucleus per cell to simplify the discussion. The key nuclear events that occur throughout the ciliate life cycle are presented in a generalized representation in figure 1. Vegetative growth involves clonal amplification of the cell's population, during which the micronucleus is duplicated by closed mitosis (i.e. without the dissociation of nuclear envelope) ensuring accurate maintenance of the germline genome (Figure 1A). The polyploid macronucleus divides amitotically splitting its nuclear content into roughly equal halves to partition its centromere-less chromosomes into each progeny cell. How macronuclei maintain the proper copy number of somatic chromosomes is not well understood. Studies in *Oxytricha* and *Stylonychia* indicate that copy number can be regulated epigenetically [3, 4]. Nevertheless, high ploidy and endoreplication of somatic chromosomes appear to maintain proper DNA content and prevent lethal gene loss.

A ciliate's micro- and macronuclei are replaced after each round of sexual reproduction. Conjugation, which can be induced in the laboratory by nutrient starvation, begins with the pairing of two mating compatible cells (See Figure 1B). Micronuclei of the mating partners undergo meiosis where a single haploid meiotic product in each

partner is selected and passed on to their progeny; the non-selected meiotic products are degraded. The chosen haploid nuclei replicate their chromosomes then undergo an additional nuclear division to produce two haploid nuclei with identical genomes, one of which is exchanged with the mating partner. The exchanged haploid nucleus then fuses with the partner's stationary haploid nucleus to form the zygotic nuclei of the mating pair. This nuclear cross-fertilization produces identical heterozygous, diploid genomes in each partner. In the case where mating compatible partners are unavailable, some species will undergo autogamy, a form of self-fertilization, where two genetically identical haploid nuclei fuse with each other, producing a homozygous diploid genome (Figure 1C).

After the haploid "gametic" nuclei fuse (karyogamy) to give rise to the zygotic genome, additional rounds of DNA replication and nuclear division produce the precursors of new micro- and macronuclei. As development proceeds, these progenitors (often called anlagen) differentiate into the new germline and somatic nuclei. While the cross fertilization that occurs during conjugation generates genetically identical progenitor nuclei, the individual progeny cells of a mating pair can differentiate with distinct phenotypes (e.g. different mating types) in non-Mendelian inheritance patterns. In some cases, specific phenotypes can be traced through a particular cytoplasmic lineage. It is important to point out that new somatic nuclei differentiate within the cytoplasms of the two parental cells, thus the DNA is replaced while many existing cellular structures are preserved. This feature of ciliate biology is a major contributor to the non-Mendelian inheritance phenomena described below.

1.3

Differentiation of Somatic and Germline Genomes

Macronuclear differentiation is an extreme example of genome reprogramming as cells start with a genome that is transcriptionally silent and remodel it into one that supports regulated gene expression during vegetative growth. In addition to switching the genome from a silent to an active state, this reprogramming involves a transition from mitotic to amitotic division accompanied by chromosome breakage and extensive DNA rearrangements (Figure 2). Research efforts to understand the differences between the transcriptional activity of micro- and macronuclei have uncovered regulatory systems that have solidified the ciliates' place as major models for elucidating epigenetic mechanisms. Before discussing these discoveries further, it is important to briefly touch upon the structural rearrangements that streamline and selectively amplify the genome in differentiating macronuclei as they are important both for understanding some of the historical examples of non-Mendelian inheritance and have provided through their study new avenues on which to further explore epigenetic regulation.

Ciliates streamline their somatic, macronuclear genome through massive genome rearrangements that fragment the germline-derived chromosomes and eliminate large portions of their genomic complexity (Figure 2). The fraction of the germline genome removed from the macronucleus ranges from 15% to as much as 95% (reviewed in [5]). Fragmentation of the developing macronuclear chromosomes is coupled with *de novo* telomere addition, which stabilizes the newly formed termini. The degree of chromosome fragmentation varies widely among the ciliate lineages. For Stichotrichs such as *Oxytricha* and *Euplotes*, this fragmentation is so extensive that the average macronuclear chromosome is only a few kilobase pairs (kbp) that typically contains a

single gene. At the other end of the spectrum, the Oligohymenophora, which include *Tetrahymena* and *Paramecium*, break their developing macronuclear chromosomes at just tens to hundreds of sites to produce chromosomes that are typically several hundred kbp. After chromosome fragmentation, these small chromosomes are amplified to their final high copy number in the polyploidy macronucleus.

In addition to chromosome fragmentation, ciliates eliminate many DNA segments from internal sites. These germline-limited, **I**nternal **E**liminated **S**equences (IESs) are numerous in all ciliate genomes studied. They are removed from thousands of loci and in some species up to tens of thousands of loci. In some ciliates, such as *Tetrahymena*, nearly all IESs are found within intergenic regions, whereas in most other ciliates studied, IESs are also common within genes. When IESs are present in coding regions, they are precisely excised during macronuclear differentiation. A common class of IESs found in diverse ciliates species is characterized by flanking 5'-TA-3' dinucleotides, one copy of which is retained upon excision (reviewed in [6]). The sequences eliminated from somatic macronuclei represent most of the repetitive sequences residing in the germline genome including transposable elements [7]. The majority of IESs may actually be the remnants of transposons or be otherwise derived from the activity of transposable elements ([8-10], reviewed in [11]). Intriguingly, recent evidence suggests that the excision of IESs utilizes domesticated transposases [12-14]. As we will discuss, the mechanisms that ciliates use to identify IESs are related to RNA interference (RNAi), which is used by many eukaryotes as a surveillance system to limit the activity of transposons in the genome (reviewed in [15]). We will describe these mechanisms in

detail as they reveal important insights into the use of homologous, non-coding RNAs in epigenetic regulation.

1.4

Micro- and Macronuclei: Models for Silent and Active Chromatin

The recognition that micro- and macronuclei of ciliates have opposite activity states promoted the development of these organisms as models with which to examine cellular mechanisms that differentially regulate identical sequences, the very definition of epigenetics. The most significant contributions in this area have been made by researchers investigating the chromatin structure of the different nuclei of *Tetrahymena*. These efforts started about four decades ago and helped establish a number of paradigms of epigenetic control including the importance of histone variants and the role of histone acetylation in transcriptional regulation. A summary of the histone variants and modifications found in the micro- and macronucleus are shown in table 1.

1.4.1

Differential Histone Composition of Micro- and Macronuclei

The core histones make up the largest fraction of chromatin in both the micro- and macronuclei; however, comparison of chromatin proteins that are found in each type of nucleus led to the characterization of some of the first known histone variants. **H**istone **v**ariants, Hv1 and Hv2, were identified as forms of Histone H2A and H3, respectively that are localized specifically within the transcriptionally active macronucleus [16, 17]. These proteins represent the equivalent of the widely conserved variants H2A.Z and H3.3. Hv1 (H2A.Z) is essential in *Tetrahymena* [18]. In addition to its presence in the macronucleus, this variant has been observed in micronuclei during early conjugation

when these nuclei first exhibit transcriptional activity [19-21]. Hv2 (H3.3) has properties consistent with its role as a replacement histone. This variant was shown to be constitutively expressed during the cell cycle, which is in contrast to core histone H3.1, which is only expressed in early S phase [22]. This led to the hypothesis that H3.1 is only deposited into chromatin during DNA replication, while Hv2 is deposited outside of S phase. The presence of these two histone variants exclusively in the macronucleus (or meiotic micronuclei) provided some of the first evidence that specific variants are preferentially associated with transcriptionally active chromatin.

In addition to core histone variants, micro- and macronuclei have distinct linker histones. Neither linker histone is essential [23]. Nevertheless, when the genes for the micronuclear and the macronuclear linker histones were disrupted, the nucleus in which they normally reside increased in volume. These results were interpreted to mean that the chromosomes in the absence of the linker histones exhibited lower degrees of chromatin compaction. In addition, cells lacking the macronuclear linker histone showed altered gene expression profiles, a result providing some of the first evidence that linker histones have roles outside of maintaining general chromosome structure [24].

1.4.2

Differential Histone Modifications of Micro- and Macronuclei

The finding that histones in the macronucleus were hyperacetylated relative to those in the micronucleus provided evidence that corroborated Allfrey's observations that acetylated histones were important for transcriptional activity in animals [25]. The ability to make targeted mutations in *T. thermophila* allowed Martin Gorovsky and colleagues to test whether acetylation of the H2A.Z tail is critical for transcription and to

further assess whether specific sites needed to be acetylated [26]. These researchers found that mutation of all normally acetylated lysines in the H2A.Z tail to arginines, which are not able to be acetylated, was lethal. The mutant phenotype could be rescued by H2A.Z proteins containing a single acetylated lysine. In addition, the Hv1 tail could be substituted for by the core H2A tail demonstrating that the overall histone tail charge density was more important than the modification of particular tail lysine residues [27].

Arguably one of the landmark discoveries in epigenetics research was the cloning of the first nuclear histone acetyltransferase (HAT). C. David Allis and colleagues set out to identify the protein responsible for hyperacetylation of macronuclear chromatin by using an in gel histone acetylation assay [28]. These researchers polymerized histones directly into denaturing protein gels used to fractionate *Tetrahymena* extracts. Then, after renaturing proteins in the polyacrylamide matrix, they incubated the gels with radiolabeled acetyl-CoA. They identified, and then purified, a 55 kD protein that shared significant similarity with the yeast GCN5 transcriptional regulator. This discovery established the paradigm that transcriptional regulators act by modifying chromatin [28, 29].

Other histone modifications enriched in either micro- or macronuclei hinted at their biological function. Histone H3 methylated on lysine 4 was found exclusively in the macronucleus, which provided the early evidence that this modification was associated with active chromatin [30]. This modification is absent from micronuclei, but is rapidly established on the bulk of the genome soon after developing macronuclei are formed. In contrast, methylation of histone H3 on lysine 9 is found exclusively during conjugation on the chromatin of IESs in developing macronuclei [31]. This modification is lost from

macronuclei as IESs are removed from the genome. While methylation of histone H3 on lysine 9 was already known to be associated with silent heterochromatin in *S. pombe* and other eukaryotes, its linkage to IES excision, which was concurrently found to be controlled by a RNA interference related mechanism, provided one of the first examples, along with the work in *S. pombe*, that RNAi-directed transcriptional gene silencing targeted chromatin modifications to specific genomic regions [31-33].

While most chromatin modifications are enriched in macronuclei, phosphorylation of histone H3 on serine 10 was found to be highly enriched in micronuclei undergoing mitosis or meiosis, indicating that this modification may be involved in chromosome condensation [34]. Mutation of serine 10 to alanine resulted in chromosome segregation defects, which further supported the importance of phosphorylation of this position on histone H3 in chromatin compaction during nuclear division [35]. These structural and functional differences between micro- and macronuclei provided a rich biological context to begin to unravel the role of chromatin proteins and their post-translational modifications in controlling epigenetic phenomena in ciliates.

2

Epigenetic Phenomena in Ciliates

Ciliates had been firmly established as genetic models for uncovering epigenetic phenomena long before researchers began using the differentiation of micro- and macronuclei as means to uncover the molecular basis of epigenetic control. In this next section, many of the classical examples of non-Mendelian inheritance and other epigenetic phenomena are described. It is useful to revisit these early observations using

the light of newer molecular studies. These examples of structural and cytoplasmic inheritance have a common feature: the pre-existing phenotypic state of the parent cells is able to somehow "template" the phenotype that emerges in the next generation. These phenomena challenge many of our preconceived ideas of simple genetic inheritance and beg for more study to decipher their underlying mysteries.

2.1

Structural Inheritance

In addition to nuclear dualism, ciliates are characterized by the extraordinary complexity and asymmetry of their cellular structures. The ciliate cortex is comprised of a matrix of cytoskeletal and membranous components. Organized within the cortex are organelles with specialized functions, such as the anteriorly positioned oral apparatus (a mouth-like phagocytic structure) and a posterior cytoproct. The elaborate ciliate body plan is faithfully reconstructed after each round of binary fission. The anterior daughter cell must reform posterior structures and the posterior daughter must generate a new mouth and other anterior components. Genetic and physical manipulation of the cortex has revealed that the cellular structure of ciliates is largely organized by the pre-existing structures, thereby showing that a cell's phenotype is not solely determined by genotype.

These cells' numerous cilia, which are primarily used for locomotion and feeding, project from arrays of cortical units, aligned into rows that are organized along the anteroposterior axis. Each cortical unit assumes a distinct anterior-posterior and left-right orientation crucial for the proper function of the cilia. During each cell cycle, the units are duplicated to ensure that each daughter cell inherits a complete set of structures that assumes the correct orientations. An early scientific question was whether this cortical

organization was determined by the action of genes. What these studies revealed is that the structural organization of daughter cells is not purely established by the cells genotype, but is instead templated by the geometry of the pre-existing units (i.e. inherited through a non-genic mechanism). One of the earliest studies on cortical inheritance was performed using doublet cells. The "doublet" phenotype arises from failure of pair separation at the end of conjugation, leading to fusion of the progeny. This phenotype is fairly stable and can be propagated such that vegetative progeny inherits a duplicated set of cortical structures. Genetic crosses demonstrated that heredity of the doublet phenotype was not determined by genes or the cytoplasm, but rather was communicated through the architecture of the cortex itself [36].

Cortical inversion further illustrates the phenomenon of structural inheritance. Cortical inversion refers to a condition when cells have one or more ciliary rows rotated 180° in the plane of the cell surface. This produces an inverted patch of cilia resulting in cells exhibiting an abnormal "twisting" swimming phenotype. Like the "doublet" phenotype, the progeny of cells with inverted patches inherits the inverted orientation of cilia as the new cortical organization is templated by the parental cortical organization [37]. What this and other experiments tells us is that, in the case of ciliary orientation, genes supply building blocks, but the assembly into a functional organelle is determined by the structure of the pre-existing cortex. The ciliate cortex thus provides an example of structural memory and reveals that genes are not the only cellular component that can pass on heritable information to the next generation.

2.2

Cytoplasmic Inheritance

Inheritance of pre-existing cellular structures is a specialized example of epigenetic influence on phenotype. A more general non-nuclear medium for transmission of heritable information is the cytoplasm. The role of the cytoplasm as a director for epigenetic information is well-documented in ciliates. It has been particularly well studied in sexually reproducing *Paramecium aurelia* and related species. One reason for this is that, unlike some ciliates (e.g. *T. thermophila),* conjugation of *P. aurelia* involves almost no cytoplasmic exchange between the mating pairs. Therefore, while cross-fertilization produces identical zygotic nuclei, these identical genomes develop in the different cytoplasmic environments of their respective parental cells. The interesting observation is that these progeny, which are genetic twins, commonly express different phenotypes as determined by the cytoplasm in which their macronuclei develop.

Cytoplasmic inheritance in ciliates is most easily illustrated by determination of the mating type trait (Figure 3) (reviewed in [38, 39]). *Paramecium* exist as two mating types: Even (E) and Odd (O). When two cells of opposite type mate, the progeny that arise from the E parent almost always assume the E mating type, whereas the ones from O almost always assume O, despite each receiving identical genotypes. This observation suggests that something other than genes directs the determination of the mating phenotype. Comparison of progeny mating types from crosses that do and do not exchange cytoplasm during conjugation further implicated cytoplasm as a key component in mating type determination [40-42]. If cytoplasmic exchange occurred between the mating pairs during conjugation, the progeny of the O cell often switched to the E mating type. Furthermore, injection of cytoplasm from the E mating partner into O partner was found to transform the progeny's mating type from O to E. No effect was observed upon

transfer of O cytoplasm into E cells, which suggested that the cytoplasmic factor(s) must exist in the E cell to determine the E mating type and that the E mating type is dominant over O [43].

Like mating type, the serotype of *Paramecium* progeny can be strongly influenced by the cytoplasm in which a new somatic genome differentiates. Serotype is determined by the specific surface antigen protein that is expressed and displayed on the cell surface. Several genes encode the different antigen proteins; however, only one is expressed in any given cell. Upon conjugation the sexual progeny typically express the parental serotype. For instance, when cells of serotype A are crossed with serotype B, progeny of both types will emerge expressing the serotype of the parent in which their nuclei developed [44]. Inheritance of mating type and serotype are therefore specified by the cytoplasmic environment rather than purely as genetic traits.

2.3

Epigenetic Control of Traits Converge with the Regulation of DNA Rearrangements

As mentioned above, differentiation of a developing somatic macronucleus from its zygotic precursor involves extensive streamlining of its germline-derived genome by removing extraneous "junk" DNA (See Figure 2). Thus the process of genome rearrangement directs major changes to the overall DNA sequence in the somatic macronucleus relative to the input from the germline. As the DNA removed is primarily non-coding, whether or not this DNA reorganization affects gene expression has not been extensively studied. For many ciliates, which have IESs imbedded within coding regions, DNA elimination must occur to generate an expressible protein coding region. It has been postulated, and supported by several experimental observations, that epigenetic

control of these DNA rearrangements may underlie at least some of the examples of non-Mendelian inheritance that have been discovered. The following section discusses studies that highlight how differential elimination of DNA sequences may be one mechanism ciliates use to alter the phenotype expressed by progeny.

A genetic screen aimed at elucidating the molecular basis for mating type expression uncovered an intriguing link between this trait and the control of DNA rearrangement. A genetic mutation, mtF^{E}, was isolated in a cell line that produces only mating type E [45]. As mentioned earlier, *Paramecium* sexual progeny almost always assume the mating type of the parent (i.e. O parent, O progeny; E parent, E progeny). When an E individual that carries the mtF^{E} mutation (mtF^{E}/mtF^{E}) is crossed with a wild-type O individual (mtF^{+}/mtF^{+}), the mating type of F1 progeny still follows the cytoplasmic inheritance typical of a wild-type mating (Figure 4). However, homozygous mtF^{E} F2 progeny produced from autogamy of F1 O individuals (mtF^{+}/mtF^{E}) frequently switch to mating type E (Compare Figures 3 and 4). The gene mutated in mtF^{E} strains has not been identified, but detailed study of the mtF^{E}/mtF^{E} strains demonstrated that this mutation also led to failure to eliminate an IES located in the G surface antigen gene. This observation lead to the hypothesis that the gene mutated in mtF^{E} strains is involved in DNA rearrangement, and by extension, DNA rearrangement may be involved with mating type determination [46].

The mtF^{E} mutation studies did more than just link DNA rearrangement to mating type; they uncovered a means by which epigenetic regulation of DNA rearrangements can alter the expression of specific traits. Further studies, not of mating type, but of G gene expression, revealed that the IES+ state of the G gene, apparently caused by the

mtF^E mutation, became the heritable state of the G gene that was propagated through subsequent generations, even after reintroduction of the wild-type mtF^+ allele. Given the observed cytoplasmic inheritance patterns of both mating type and serotype traits in *Paramecium*, this finding offered a very intriguing connection between alternative rearrangements and altered phenotypes.

The propagation of the IES+ state in the mtF^+ progeny showed that it was not a genetic lesion or other alteration to the germline genome that limited expression of the G gene. It was in fact the IES+ state itself present in the parental macronucleus that elicited the transmittable influence of the 'cytoplasm' during development. This was more conclusively demonstrated by directly injecting the IES+ version of the G coding sequence into maternal macronucleus and showing that this alone was able to block elimination of the homologous IES from the newly developed macronucleus after autogamy [47]. It is important to note that the injected DNA is destroyed along with the maternal macronucleus, so the IES+ state must be communicated to the developing macronucleus through the cytoplasm. Injection of plasmid DNA containing just the one IES without any flanking G gene coding sequence was found to be sufficient to block elimination of this IES, while the remaining IESs within the G gene were excised efficiently. Thus particular IES sequences present in the maternal macronucleus are able to communicate their presence to the zygotic macronucleus and alter the normally efficient removal of the homologous sequence.

Not all IESs were found to be subject to this form of homology-dependent regulation. Ten different IESs were microinjected into parental macronuclei to test their ability to block excision of the homologous sequence. Only four IESs were able to

inhibit DNA rearrangement. While it is unclear why only some IESs in the zygotic macronucleus can sense the presence of homologous copies in the parental macronuclei, the implication is clear that many characteristics can be reproducibly inherited in a non-Mendelian fashion every time a new macronucleus is formed.

The serotype genes of *Paramecium* have proven to be fertile ground for uncovering epigenetic phenomena relating to genome rearrangements. An early and particularly interesting example was revealed by studies of a mutant strain called d48 that lacked the ability to express the surface antigen A gene [48]. Careful genetic studies showed that the d48 micronucleus contained a wild-type copy of the A gene; but the macronucleus was missing the A gene coding region [49]. The remarkable discovery was that the progeny of d48 strains reproducibly eliminated the A gene from their developing macronuclei during conjugation, making these progeny unable to express the A serotype.

A series of microinjection and nuclear transplantation experiments demonstrated that the presence of the A gene in the parental macronucleus was necessary for retention of the A gene in progeny. Microinjection of the A gene into the macronucleus of strains lacking the A gene in both the micro- and the macronuclei was sufficient to restore A gene expression during vegetative growth, however this expression was lost during sexual reproduction when the microinjected parental macronuclei were fragmented and destroyed [50]. On the other hand in the d48 strain, which lacks the A gene only in the macronucleus, microinjection of the A gene was sufficient to rescue A gene expression during vegetative growth both in the parental strain and also in progeny cells following sexual reproduction [51-53]. Strains missing the surface antigen B gene have also been observed and rescued in a similar fashion [54, 55].

Rescue of A gene expression in the *Paramecium* d48 strain was found to be sequence specific. Microinjection of the A gene or an allele of the A gene that has 97% identity resulted in A gene retention in the newly formed macronuclei of progeny. In contrast, introduction of the G surface antigen gene, which shares ~80% similarity with the A gene, failed to rescue the A gene deficiency in the progeny [56, 57]. Thus the DNA sequence of the parental macronucleus was again shown to have the ability to dramatically influence the types of sequences retained during development of new macronuclei.

The observations made with d48 strains share intriguing parallels with both the inheritance of the IES+ state in the mtF progeny and with the examples of cytoplasmic inheritance described above. In each case, the trait (or sequence) propagated is that which was expressed from the parental macronucleus. Thus for ciliates, the regulation of DNA rearrangements allows for somatic states of gene expression to be transmitted to the next generation. Recent studies of the mechanisms that guide DNA rearrangements have revealed that homologous RNAs and chromatin-based regulatory schemes are key components. Studies of ciliate DNA elimination during macronuclear development reveals that non-coding RNAs (ncRNAs) may also be the molecules responsible for many of the cytoplasmic and homology dependent inheritance phenomena observed previously. In the remainder of this review, these mechanisms will be described in detail as they offer many unique insights into how ncRNAs can pattern the genome and influence chromatin structure.

3

RNA-Mediated Epigenetic Mechanisms

3.1

Homology Dependent Gene Silencing

Homology-dependent epigenetic phenomena have been observed widely. Introduction of transgenes into plant cells can often lead to silencing of the endogenous copy. One of the most cited examples resulted from an effort to make petunias that had darker flower petals by adding exogenous copies of the chalcone synthase gene that makes the purple pigment [58]. Instead of producing the expected increase in petal pigmentation, the transgenic petunias showed a decrease in coloration in conjunction with an overall reduction in the mRNA level of chalcone synthase, a phenomenon termed co-suppression. Similarly, introduction of transgenes into the fungi *Neurospora crassa* induces a phenomenon called quelling, a silencing of the homologous endogenous gene [59]. Co-suppression has also been observed in the ciliate, *Paramecium tetraurelia*, upon high-copy microinjection of transgenes that lack 5'and 3' regulatory regions (i.e. lacking either promoters or transcription terminators), resulting in silencing of the endogenous homologous genes [60, 61].

The mysterious mechanism underlying these phenomena was discovered to be RNAi. Mechanistic insight into homology dependent phenomena in ciliates has likewise been provided via connections to RNAi. In general, RNAi refers to a diverse collection of cellular mechanisms that use RNA molecules to regulate the expression of genes (reviewed in [62-64]). The triggering molecule is typically double-stranded RNA (dsRNA) that is recognized by a ribonuclease called Dicer that cleaves dsRNA into ~20-30 nt fragments. These small RNA (sRNA) species serve as the specificity factors that guide an associated protein complex to a target mRNA or gene, where these effector

RNA-protein complexes can promote silencing either transcriptionally or post-transcriptionally.

RNAi appears to be an integral part of a variety of processes in ciliates. Examination of the bulk small RNA species in either *Paramecium* or *Tetrahymena* reveals distinct size classes, suggesting at least two different RNAi pathways [33, 65, 66]. The larger species (~25 nt in *Paramecium* and 27-30 nt in *Tetrahymena*) are produced exclusively during conjugation and have been shown to guide the extensive DNA rearrangements that occur in the differentiating somatic macronucleus. This RNA-guided genome reorganization will be discussed in detail (Section 4).

A second class of ~23 nt RNAs is produced in growing cells as well as during conjugation in *Paramecium* and *Tetrahymena*. This size class mediates post-transcriptional gene silencing and the transgene co-suppression introduced above. In addition, introduction of dsRNA aimed to experimentally induce gene silencing, either through feeding or direct injection into *Paramecium* cells or hairpin RNA expression in *Tetrahymena*, results in the production of these ~23 nt RNAs [61, 66-70]. Thus these small RNAs are similar in function to the small interfering RNAs (siRNAs) discovered initially in plants by the Baulcombe lab, in that they carry out post-transcriptional gene silencing (PTGS) [71]. In *Tetrahymena*, these small RNAs are produced by Dcr2p from presumed pseuodogenes or defective endogenous genes, which triggered the production of dsRNA precursors necessary for siRNA production [65, 72]. They are anti-sense to these predicted open-reading frames (ORFs) and depend upon the activity of an RNA-dependent RNA polymerase (RdRP), Rdr1p, which is found in a common complex with Dcr2p [72]. In *Paramecium* a subclass of these smaller sRNAs is only anti-sense to

mRNA transcripts and produced by a secondary amplification involving the RNA-dependent RNA polymerases, Rdr1p and Rdr2p [66, 73]. RNAi is clearly an important mechanism during the vegetative life of ciliates, but the critical role has yet to be carefully examined. On the other hand, the function of RNAi pathways during development of the zygotic macronucleus has contributed new ways to think about epigenetic programming of the genome.

3.2

RNA-Guided Genome Reorganization

Non-coding RNAs and RNAi-related mechanisms do much more than gene silencing in ciliates as these organisms use RNAs as guides to extensively remodel their genomes during sexual differentiation. Investigations aimed at elucidating the molecular mechanisms associated with the reorganization of the somatic genome of several ciliates have uncovered the involvement of non-coding RNAs [69, 74-76]. The mechanisms revealed have turned out to vary substantially among the different ciliate species studied, thus we will describe the data for *Paramecium*, *Tetrahymena*, and *Oxytricha* separately in subsequent sections. Nevertheless, a common theme has emerged: RNAs serve as potent mediators capable of transmitting sequence specific information between generations. The examples of homology-dependent regulation of phenotypes, particularly those described earlier in *Paramecium* (Section 2.3), hinted that the mechanism(s) guiding genome rearrangements utilized some form of nucleic acid to transmit sequence specific information between the somatic macronucleus of one generation and the developing macronucleus of the next. These phenomena require that the state of the DNA in the

parental macronucleus serve as a "template" for the traits expressed from the genome of the progeny.

Studies of *Paramecium* and *Tetrahymena* DNA rearrangements have identified two types of sequence specific mediator RNAs, one produced from the germline genome and a second from the parental somatic genome [69, 74, 75]. The germline-specific RNAs come in the form of sRNAs, called scan RNAs (scnRNAs), that are produced during meiosis and act to identify the IESs as germline-limited sequences to be eliminated from the developing somatic genome [33, 66]. The second type consists of longer transcripts produced from the parental macronucleus that appear to antagonize the action of the scnRNAs [69, 75]. It is these macronuclear transcripts that are the key epigenetic regulators that may explain the non-Mendelian inheritance of specific traits. In *Oxytricha*, analogous transcripts made from the parental somatic genome are postulated to serve as templates to directly guide the rearrangements, while a role for sRNAs is as yet unknown [76].

Genome scanning is one term used to describe the mechanism by which RNAs from the germline and somatic genomes can communicate the existing genomic content of the parental nucleus to the next generation [33]. Scanning occurs by comparison of the germline-derived scnRNAs with long ncRNA transcripts produced by the parental macronucleus [69, 75]. This scanning assures that scnRNAs made to regions of the genome that are not IESs are removed from the pool of scnRNAs that target specific sequences for elimination. Scanning not only allows proofreading of the sRNA pool to prevent inadvertent elimination of sequences that should be retained, but also permits the retention of IESs that were maintained in the macronucleus of the previous generation

that offered some advantage or specified an alternative phenotype. In the following sections we detail the mechanisms of RNA-guided genome reorganization and genome scanning as these studies reveal the power of homologous RNAs to direct the programming of the somatic genome.

4

Small RNA-Mediated DNA Rearrangements

4.1

RNAi-Dependent DNA Elimination in *Paramecium*

The germline genome of *Paramecium tetraurelia* contains approximately 60,000 IESs ranging from 28 bp to 886 bp [8, 77]. Many are found within coding sequences and must be identified and precisely excised from the developing macronuclear chromosomes. Furthermore, during this genome maturation in *P. tetraurelia* the more than 50 micronuclear chromosomes are fragmented into unknown number mini-chromosomes amplified to 800n [78, 79]. The elimination of IESs occurs during both self-mating and sexual reproduction, during which the parental macronucleus is destroyed and a new zygotic one is generated.

Studies over the past decade have revealed that the IESs are identified through the actions of homologous RNAs via an RNAi-related mechanism in *Paramecium* and support the model shown in figure 5 [69]. A class of ~25nt sRNAs produced only during meiosis has been shown to be necessary and sufficient to trigger DNA elimination of IES sequences [66]. They were found to be homologous to a variety of DNA sequences throughout the genome and likely function in a similar manner to the *T. thermophila* scnRNAs described below [33, 66, 69]. These *Paramecium* scnRNAs have 2 bp 3’

overhangs consistent with cleavage by an RNase III homologue ([66, 80-82], reviewed in [83]). Seven RNase III homologues are present in *P. tetraurelia*, three Dicer (DCR) and four Dicer-like (DCL) homologues [66]. Single knockdown of the DCL genes have no effect on scnRNA production, but double knockdowns of DCL2 and DCL3 abolish it. Localization of Dcl2p in the crescent micronucleus early in meiosis indicates that the production of scnRNAs only takes place there at this early time point of conjugation. Double knockdowns of DCL2 and DCL3 also caused failure of DNA elimination and produced non-viable progeny, which further supported the conclusion that the scnRNAs they produce target IESs for excision.

The scnRNAs produced by Dcl2p and Dcl3p cleavage in the crescent micronucleus are transported by the Piwi homologues, Ptiwi01p and Ptiwi09p, into the parental macronucleus to carry out genome scanning [84]. The scnRNAs that match the parental macronuclear genome are removed from the population that will be transported to the developing macronucleus later in development to participate in genome restructuring. This scanning occurs by comparison of these germline-derived scnRNAs with a second type of regulatory RNA, long, non-coding RNA (ncRNA) transcripts produced in the maternal macronucleus and ensures that scnRNAs made to regions of the genome that are not IESs are not inadvertently excised by removing these scnRNAs from the overall pool of scnRNAs that will guide genome reorganization [69, 75].

Only a few proteins are known to play a role in the genome scanning process in *P. tetraurelia*. Two proteins identified that play a role in this process are, Nowa1p and Nowa2p [85]. Nowa1p and Nowa2p are GW repeat proteins found to initially localize within the parental macronucleus during pre-zygotic development and then move to the

developing macronucleus after its formation. Deletion analysis of Nowa1p found that the N-terminal portion of the protein has nucleic acid binding capability, particularly for RNA/DNA duplexes. Dimerization of Nowa1p with itself or perhaps Nowa2p appears to be essential for nucleic acid binding function. Double knockdown of NOWA1 and NOWA2 causes failure of DNA elimination of a specific class of IESs in *P. tetraurelia* called maternally controlled IESs (mcIESs) [77, 85]. The failure of DNA elimination was complete in some cases and incomplete in others [85]. Double knockdown of NOWA1 and NOWA2 also produced non-viable progeny indicating an essential function for the completion of autogamy or conjugation.

How might the NOWA proteins contribute to the epigenetic control of IES excision? What RNAs might they interact with? Long, ncRNA has been shown to have a role in several epigenetic phenomena in higher eukaryotes including dosage compensation and genomic imprinting [86-92]. Data in *P. tetraurelia* provides strong support for interaction between maternal transcription of long ncRNA and meiotic scnRNAs and reveal exactly why this interaction is likely fundamental to genome programming [69]. RT-PCR of RNA isolated early in autogamy shows production of ncRNA without IESs, which are thought to be transcribed from the parental macronucleus. When a strain of *P. tetraurelia* containing a mcIES in the parental macronucleus are fed bacteria producing dsRNA prior to autogamy, or directly injected with 23 nt siRNAs or 25 nt scnRNAs early during autogamy against this mcIES, the mcIES is removed from the developing macronucleus later in autogamy. This result indicates that genome scanning can be affected by degrading the long ncRNA in the parental macronucleus through bacterial feeding to produce 23 nt siRNAs or direct

injection of 23 nt siRNAs, as well as directly injecting the biologically active 25 nt scnRNAs to allow removal of an mcIES that would normally be retained after the completion of autogamy.

Long ncRNA also plays a role in the developing macronucleus by directing the remaining scnRNAs to sequences of DNA that are to be eliminated. Transport of these remaining scnRNA complexes in *P. tetraurelia* to the developing macronucleus is mediated by the Piwi homologues, Ptiwi01 and Ptiwi09, where production of long, ncRNA containing IESs has been detected by RT-PCR [69, 84]. Injection of the 25 nt scnRNAs in the same *P. tetraurelia* strain containing a mcIES in the parental macronucleus later during autogamy also causes removal of the mcIES, however injection of 23 nt siRNAs at the same time failed to cause DNA elimination [69]. In this case it seems likely that the 23 nt siRNAs actually promote failure of DNA elimination by targeting the long ncRNA needed for DNA elimination for degradation, while the 25 nt scnRNAs are able to recruit the necessary proteins for DNA elimination of this mcIES.

4.2

The Role of a Domesticated *PiggyBac* Transposase in DNA Elimination and Chromosome Breakage in the Developing Somatic Nucleus of *Paramecium*

All the different varieties of RNA seen only during autogamy or conjugation in *P. tetraurelia* are all directed to one goal, elimination of IESs and repetitive sequences. Removal of any of these types of RNA during the reproductive process causes non-viability [66, 69]. To eliminate IESs and repetitive sequences from the genome these scnRNAs must recruit an excisase. Recent data has implicated the domesticated piggyBac transposase, Pgmp, as this excisase [12]. In order to understand the role of Pgmp in DNA

elimination, a brief description of IESs is called for. In *P. tetraurelia* each IES is flanked by terminal inverted repeats, whose consensus sequence is 5'-tggTAYAGYNR-3' [8, 93]. Cleavage occurs between the two guanosines in the consensus sequence producing a 5' 4 bp overhang centered around the TA dinucleotide [94]. Mutations in either the T, A or G in the third, fourth and eighth position, respectively, of the above consensus sequence is sufficient to block cleavage [95-98]. Cleavage of the consensus sequence, 5' TTAA 3', by piggyBac transposases to produce a 5' 4 bp overhang is somewhat similar to the *P. tetraurelia* consensus IES sequence and cleavage product [99, 100]. Analysis of the *P. tetraurelia* genome identified a piggyBac homologue, called PGM [12]. Localization of GFP-Pgmp was found only in the developing macronucleus late in autogamy. Knockdown of PGM late in conjugation resulted in failure to produce viable progeny, failure of IES excision and chromosome breakage, and overexpression of IES-containing ncRNA from the developing macronucleus. These knockdown phenotypes implicate Pgmp as having an essential role in completion of DNA elimination and chromosome breakage in *P. tetraurelia* likely through Pgmp-mediated dsDNA breakage to remove IESs and other repetitive sequences. Repair of these dsDNA breaks is mediated by the DNA ligase IV homologues, LIG4a and LIG4b [101].

Removal of these IESs and other repetitive sequences in *P. tetraurelia* and other ciliates is the ultimate epigenetic action. Unlike most other eukaryotes, which heterochromatize their repetitive and non-coding sequences, ciliates excise and degrade these sequences from their somatic macronucleus and amplify the remaining sequences to create a streamlined genome that allows greater cell size than most other eukaryotes and a growth rate comparable to that of yeast. As briefly discussed earlier, there is removal of

IESs and other repetitive elements at the end of sexual reproduction. The actual removal of these two types of sequences differs slightly and can impact the final state of the genome after sexual reproduction. Two different classes of IESs exist, mcIESs and non-mcIESs, which are small, found throughout the genome and are eliminated precisely [47, 77, 94]. The mcIESs are able to have their excision blocked by insertion of a copy of the mcIES into the parental macronucleus [47, 77]. mcIESs tend to be larger in general and it is hypothesized that their elimination is dependent on chromatin modifications directed by genome scanning [102]. On the other hand, non-mcIESs are smaller with most being shorter than the amount of DNA wrapped around a nucleosome, which would necessitate a different targeting method for DNA elimination. It seems possible that their elimination could take place through directed binding of Pgmp or through guidance of Pgmp via nucleotide modification to their cleavage sequences. Repetitive sequences are removed with much less precision method resulting in either variable cleavage or fragmentation of the chromosome [103]. Both types of DNA elimination depend on the action of Pgmp [12].

Despite all that has been learned about the epigenetic phenomenon of RNAi-directed DNA elimination in *P. tetraurelia* there are still many questions to be answered. How are these ncRNAs produced in any of the nuclei? What is the difference between mcIESs and non-mcIESs and how does that affect their DNA elimination? How does DNA elimination, RNAi and heterochromatin function in related ciliates, and in general how is this biological process related to other epigenetic processes in other eukaryotes? Research into the RNAi-directed DNA elimination process in a related ciliate, *T. thermophila*, has provided additional insight into many of these questions.

4.3

RNAi-Dependent DNA Elimination in *Tetrahymena*

Like *P. tetraurelia*, the ciliate *T. thermophila* also undergoes massive DNA elimination and chromosome breakage during the sexual reproduction or conjugation. During conjugation in *T. thermophila* the developing zygotic macronucleus is fragmented into approximately 200 mini-chromosomes from 5 chromosomes, 15% of the overall DNA content is removed and the remaining DNA content is amplified to 50n ([7, 104-109], reviewed in [5]). Similar to *P. tetraurelia* the mechanism of this process was poorly understood until the discovery of conjugation-specific long, ncRNAs and a class of sRNA, called scnRNAs, which are derived from the ncRNAs, which has led to the model shown in figure 6 [33, 74, 75].

The scnRNA model of RNAi-dependent DNA elimination in *T. thermophila* can effectively be broken into two parts. The first section will deal with production and selection of the scnRNAs by conventional RNAi-associated proteins, while the second part will discuss how the scnRNA signal is transduced into heterochromatin formation, which subsequently triggers DNA elimination of heterochromatic DNA in the developing zygotic macronucleus. Each section will detail the experimental data supporting the model, discuss how each piece of data can be used to further elucidate the mechanism of RNAi-dependent DNA elimination and consider how these results relate to epigenetics in ciliates and other eukaryotes.

4.4

RNAi Apparatus and Genome Scanning in DNA Elimination

The role of RNA during the development of many eukaryotes has been well documented [33, 74, 86-91, 110-115]. *T. thermophila*, like *P. tetraurelia*, has been shown to have two classes of sRNAs that range from 23nt-24nt and from 28nt-30nt [33, 65, 116, 117]. This larger class, the scnRNAs, is restricted to conjugation [33]. They appear to be functionally similar to piRNAs that have been described in a variety of organisms, which are known to act to protect the germline genome in the micronucleus from possible deleterious effects that active transposons can inflict, such as gene inactivation, chromosome translocation and chromosome breakage [120-127]. Unlike piRNAs, which are Dicer-independent, scnRNA production in both *P. tetraurelia* and *T. thermophila* is totally dependent on a group of Dicer-like proteins [66, 116, 117, 121, 124]. If these Dicer-like genes are knocked out or knocked down, scnRNAs are not produced during conjugation, which triggers a developmental arrest [66, 116, 117]. It is intriguing that these scnRNAs in *P. tetraurelia* and *T. thermophila* exhibit properties of both piRNAs and siRNAs. Further study of the scnRNA pathway may contribute to fundamental understanding of how both the piRNA and siRNA pathways arose in higher eukaryotes.

4.4.1

Bidirectional Transcription of Long dsRNAs

The production of scnRNAs depends on the synthesis of long dsRNA precursors [116, 117]. Early in conjugation the micronucleus detaches from a groove in the parental macronucleus and elongates to form a crescent that is approximately two cells lengths [128, 129]. During vegetative growth in *T. thermophila* the micronucleus is transcriptionally silent, however decades ago it had been observed that early during conjugation, beginning after micronuclear detachment from the parental macronucleus

and peaking just prior to full crescent elongation, there was copious transcription from the micronucleus [19, 20]. Later studies in *T. thermophila* found that just prior to this period of micronuclear transcription, the *T. thermophila* H2A.Z homologue was deposited in the micronucleus, which is normally only found in the macronucleus [21, 130]. Other studies also found that RNA polymerase components, including a putative TATA-binding protein, TBP1, and a RNA polymerase II subunit, RPB3, localized to the micronucleus during this burst of transcription [131, 132]. This implicates that the RNA polymerase responsible for this early micronuclear transcription is RNA polymerase II. Research on an IES, the M element, showed that transcription of both strands was markedly increased early in conjugation during the same time period general micronuclear transcription was increased [74]. These transcripts produced early in conjugation were also heterogeneous at the 5' and 3' ends and, unlike RNA polymerase II mRNA transcripts, lacked 3' polyadenylation. Further research on transcription of other known IESs indicated that this is a general characteristic of RNAs produced during this time point in conjugation in *T. thermophila*, meaning that the burst of transcription seen in the micronucleus produces the long, IES-specific dsRNA precursors needed for scnRNA production.

4.4.2

Processing of Long dsRNAs into Scan RNAs (scnRNAs) and their Subsequent Nuclear Localization

The long, IES-specific dsRNA transcripts are scnRNA precursors, which are processed by Dicer proteins [74, 116, 117]. Analysis of the sequence of the *T. thermophila* macronuclear genome indicated the presence of three putative Dicer proteins [116, 117].

Two of these Dicer proteins were expressed throughout the *T. thermophila* life cycle, but one of these Dicer proteins, Dicer-like protein 1 (Dcl1p), was expressed exclusively during conjugation. Dicer-like proteins, like DCL1 in *T. thermophila* and DCL2 and DCL3 in *P. tetraurelia*, lack the conserved RNA helicase domain but have been shown to play an important role in epigenetic phenomena in other organisms besides ciliates, including *A. thaliana* [133]. Knockouts of DCL1 caused a massive increase in these long, IES-specific dsRNA transcripts, while at the same time caused the abrogation of scnRNAs [33, 116, 117]. This verified that the long, IES-specific dsRNA transcripts produced early in conjugation are precursors for scnRNAs [116, 117]. Knockouts of DCL1 also fail to complete conjugation and more importantly fail to undergo DNA elimination similar to the DCL2/DCL3 double knockdown in *P. tetraurelia* [66, 116, 117]. Localization of Dcl1p, like Dcl2p in *P. tetraurelia*, showed that it was exclusively a micronuclear protein, which meant that the long dsRNAs produced in the micronucleus were processed into scnRNAs in the micronucleus itself and not exported for cleavage.

Study of the scnRNA structure itself showed that they were phosphorylated at the 5’ end and also contained a 3’ hydroxyl group, which is consistent with cleavage by the ribonuclease III family member Dcl1p [33, 116, 117, 134-137]. Hybridization of these scnRNAs to micronuclear and macronuclear genomic DNA preparations from early to late in conjugation (2hrs to 10hrs) showed a gradual increase in the ratio of scnRNAs hybridizing to micronuclear DNA when compared to macronuclear DNA indicating the existence of a scnRNA sorting mechanism [33, 138]. At 2 hours, the ratio of micronuclear DNA to macronuclear DNA binding was approximately 3-fold [138]. As conjugation proceeded this ratio gradually increased to a maximum of approximately 30-

fold at 10 hours [33]. Further analysis of some of these scnRNAs also showed that they were homologous to the M and TLR IES sequences, consistent with their production from long, IES-specific dsRNAs [75, 116].

Argonaute proteins have been shown to be essential effector proteins in sRNA pathways [139]. The same is true for *T. thermophila* as an Argonaute homologue, TWI1, was shown to bind scnRNAs [33]. Phylogenetic analysis of Twi1p indicated that it was homologous to the *D. melanogaster* Piwi protein and belonged to the Piwi subfamily of Argonaute proteins. TWI1 was predicted to contain functional PAZ and PIWI domains, which facilitate nucleic acid binding and "Slicer" or ribonuclease activity, respectively. Immunoprecipitation of Twi1p shortly after production of scnRNAs at 5 hours into conjugation demonstrated Twi1p/scnRNAs interaction [138]. Localization of Twi1p showed that the protein was predominately macronuclear with some cytoplasmic localization but excluded completely from the crescent micronucleus, which indicates that scnRNAs would have to undergo active or passive transport into the cytoplasm to interact with Twi1p [33]. Mutation of the DDH motif in the PIWI domain of TWI1 abolishes ribonuclease activity in Twi1p and prevents removal of the passenger strand in Twi1p/scnRNA complexes found in the cytoplasm [140]. Mutation of the DDH motif also blocks import of the Twi1p/scnRNA complexes into the parental macronucleus, which leads to scnRNA instability and degradation over a similar time course when compared to TWI1 knockouts [138, 140].

Like Argonaute proteins in other organisms, Twi1p does not act alone during RNAi-dependent DNA elimination in *T. thermophila*. For import into the parental macronucleus of the Twi1p/scnRNA complexes to occur Twi1p must also interact with

an accessory protein called Giw1p [140]. GIW1 shows no homology to any known domains of any gene, but Giwi1p co-immunoprecipitates with full-length Twi1p interacting with the PAZ and PIWI domains of Twi1p along several discrete protein sequences. Mutation of the DDH motif in Twi1p, which blocks cleavage of the double-stranded scnRNA and also prevents binding of Giw1p to Twi1p, ensures Twi1p/scnRNA complex activation prior to parental macronuclear import. Localization of Giw1p is seen generally in the parental macronucleus and the cytoplasm early in conjugation, where it is capable of participation with Twi1p/scnRNA complexes before importing them into the parental macronucleus. Giw1p also localizes to the developing zygotic macronucleus later in conjugation but its function there at that time is not known. Knockouts of GIW1 cause failure of Twi1p/scnRNA complex import into the parental macronucleus but do not affect scnRNA cleavage or unwinding of the scnRNA passenger strand, which along with the Twi1p/scnRNA complex binding data, indicates activation of Twi1p/scnRNA complexes before Giw1p-dependent import. Like the DCL1 knockout, knockouts of TWI1 and GIW1 as well as the TWI1 PIWI domain mutation fail to complete conjugation and block DNA elimination of IESs [33, 138, 140].

4.4.3

Genome Scanning via Comparison of scnRNA Complexes to the Parental Genome

Localization of the Twi1p/scnRNA complexes into the parental macronucleus sets the stage for one of the unique aspects of DNA elimination in *T. thermophila*. As mentioned earlier there is an increase in hybridization levels of scnRNAs to micronuclear genomic DNA when compared to macronuclear genomic DNA as conjugation proceeds, indicating the presence of a sorting mechanism [33, 138]. The sorting process through which

micronuclear-specific scnRNA enrichment occurs is called genome scanning [33]. Similar to genome scanning in *P. tetraurelia* this involves comparing each Twi1p/scnRNA complex to ncRNA transcribed from the parental macronucleus. Those Twi1p/scnRNA complexes binding to the parental macronuclear ncRNA are removed from the biologically active Twi1p/scnRNA complex pool through unknown means, although a handful of proteins have been identified that play a role in this genome scanning process.

To emphasize the connection of scnRNAs with piRNAs, a homologue of HEN1, the protein known to stabilize piRNAs through methylation, has also been found to have the same role in *T. thermophila* with scnRNAs [141]. The homologue in *T. thermophila*, also called HEN1, is a RNA methyltransferase that adds a methyl group to the terminal 2' hydroxyl group of scnRNAs and has homologues in *A. thaliana*, *D. melanogaster*, and *M. musculus* [141-145]. Hen1p co-localizes with Twi1p in the parental macronucleus early in conjugation during meiosis of the micronucleus, and *in vitro* experiments with recombinant Hen1p and Twi1p showed that Hen1p also co-immunoprecipitates with Twi1p during this period of development [141]. Knocking out HEN1 causes a loss of 2'-O-methylation in scnRNAs and decreases scnRNA stability in a similar fashion to the TWI1 knockout and TWI1 PIWI domain mutant [138, 140, 141]. However unlike TWI1, GIW1 and DCL1 knockouts, knockouts of HEN1 do not show complete failure of conjugation and blockage of DNA elimination [116, 117, 138, 140, 141]. HEN1 knockouts are able to produce only 3% of possible progeny but are able to undergo complete rearrangement of the IESs tested 67.8% (38/56) of the time [141]. It is possible that since scnRNA destabilization is not as extreme as in a TWI1 knockout or PIWI

domain mutant that the sheer number of scnRNAs remaining is able to facilitate DNA elimination of IESs and completion of conjugation.

Several Argonaute proteins that associate with piRNAs in other organisms have been found to associate with RNA helicases as well ([127, 146-148], reviewed in [149]). An RNA helicase in *T. thermophila*, Ema1p, interacts with Twi1p/scnRNA complexes and plays a pivotal role in genome scanning by facilitating the Twi1p/scnRNA/ncRNA interaction [75]. Ema1p co-localizes with Twi1p in the parental macronucleus early in conjugation and later in the developing zygotic macronucleus, where the proteins have also been found to interact through co-immunoprecipitation [33, 75]. Ema1p localization is unaffected in TWI1 or GIW1 knockouts indicating that it is imported into the parental macronucleus by itself or by the same group of proteins that imports Giw1p/Twi1p/scnRNA complexes [75, 140]. Knockouts of EMA1 logically do not inhibit scnRNA cleavage or import of Twi1p/scnRNA complexes into the parental macronucleus since it is never seen to accumulate in the cytoplasm where these processes occur. However, chromatin spreading experiments and RNA immunoprecipitation followed by RT-PCR show that in EMA1 knockouts Twi1p/scnRNA complexes are no longer able to interact with chromatin and ncRNA when compared to wild-type, which is significant since it is thought that chromatin is the site of ncRNA production and the Twi1p/scnRNA/ncRNA interaction required for genome scanning. EMA1 knockouts also display an increase in macronuclear-specific scnRNAs compared to wild-type matings as conjugation proceeds. These data implicate Ema1p is facilitating genome scanning by coupling Twi1p/scnRNA complexes with the ncRNA produced in the parental macronucleus and through an unknown mechanism negatively selecting against those

Twi1p/scnRNA complexes that are able to successfully bind to the ncRNA. Lastly, EMA1 knockouts fail to complete conjugation but curiously only show failure of DNA rearrangement in a select set of IESs, which may point toward the existence of different classes of IESs in *T. thermophila*, as is seen in *P. tetraurelia*, that do not undergo this selection process [47, 75, 77].

Although relatively few proteins are known to play a role in the RNAi-dependent DNA elimination process, there exists a few situations in this process where homologues in one ciliate are found to play the same or similar role in another ciliate [12, 14, 33, 66, 84, 85, 116, 117, 150]. One of these sets of homologues is the GW repeat proteins Nowa1p and Nowa2p in *P. tetraurelia* and Wag1p and CnjBp in *T. thermophila* [85, 150]. GW repeat proteins have been found to interact with Argonaute family proteins in *A. thaliana*, *D. melanogaster* and *H. sapiens* and to play a role in sRNA effector function [151-153]. Although Nowa1p and Nowa2p appear to have RNA-binding capability, it is unclear what the function of their homologues in *T. thermophila*, Wag1p and CnjBp, is [85, 150]. Co-localization and co-immunoprecipitation experiments of Wag1p and CnjBp demonstrated a protein-protein interaction with Twi1p [75, 150]. CnjBp also localizes to the crescent micronucleus during meiosis unlike Twi1p and Wag1p although its role there is currently unknown [150]. Double knockouts of WAG1 and CNJB caused retention of macronuclear-specific scnRNAs compared to wild-type matings as conjugation proceeds in a similar fashion to the EMA1 knockout [75, 150]. Unlike the EMA1 knockout out the double WAG1/CNJB knockout also showed a slight increase in retention of micronuclear-specific scnRNAs which may entail a more general function of these two GW repeat proteins in the genome scanning process for both Twi1p/scnRNA complexes

that need to be sequestered in the parental macronucleus and those complexes that need to eventually be transported to the developing zygotic macronucleus [150]. Although double knockouts of WAG1/CNJB show increased retention of scnRNAs, Twi1p/scnRNA complexes are able to interact with ncRNA through Ema1p normally, indicating that their biological function lies downstream of initial binding of Twi1p/scnRNA complexes with ncRNA. Like many of the proteins involved in RNAi-directed DNA elimination double knockouts of WAG1/CNJB fail to complete conjugation, but fail DNA elimination in a specific set of IESs only like EMA1 knockouts [75, 150]. Curiously this set of IESs is slightly different from those in EMA1 knockouts [150]. Although GW repeat proteins have been shown to affect Argonaute function the actual mechanism is still a mystery [150-153]. Even among ciliates there is no clear mode of action for these GW repeat proteins [85, 150]. As more is discovered about the RNAi-dependent DNA elimination pathway in both *P. tetraurelia* and *T. thermophila*, it will be interesting to see whether Nowa1p and Nowa2p in *P. tetraurelia* function similarly to Wag1p and CnjBp in *T. thermophila* through sorting Argonaute/scnRNA complexes. Likewise proof of RNA binding by Wag1p and CnjBp, which has already been shown in Nowa1p and Nowa2p, could help define a common mode of action for GW repeat proteins in ciliates and possible in other eukaryotes in general.

Long ncRNA has been shown to play a vital role in a variety of epigenetic phenomena as previously mentioned [74, 75, 86-92]. In *P. tetraurelia* and *T. thermophila* there appear to be three sources of long ncRNA during sexual reproduction, the crescent micronucleus, parental macronucleus and developing zygotic macronucleus [74, 75]. The

ncRNA produced in the parental macronucleus is vital to the genome scanning process and was initially detected in *T. thermophila* alongside the bidirectional transcribed long, IES-specific dsRNA scnRNA precursors and the ncRNA produced in the developing macronucleus [74]. PCR based assays looking further into ncRNA transcription during conjugation showed that the long, IES-specific dsRNA scnRNA precursor transcription peaked at 3 hours, ncRNA transcription from the parental macronucleus necessary for genome scanning peaked at 6 hours and ncRNA transcription from the developing zygotic macronucleus for IES targeting peaked at 10 hours [75]. Blocking transcription of parental macronuclear ncRNA by actinomycin D treatment during the peak hours of genome scanning (4 hours to 6 hours into conjugation) causes a significant increase in failure of IES excision and DNA elimination [74]. Besides using actinomycin D treatment to block IES excision, it has also been demonstrated that individual IESs can have their excision blocked by insertion of the IES sequence into the parental macronucleus prior to conjugation similar to blockage of mcIES excision in *P. tetraurelia* [47, 77, 154, 155]. For example, in *T. thermophila* insertion of the M element IES into the parental macronucleus causes a massive increase in M element long dsRNAs with no change in the level of scnRNAs [154]. This indicates that the excess long dsRNAs were not being processed into scnRNA but were likely acting as ncRNA in the parental macronucleus removing M element scnRNA/Twi1p complexes from the biological active pool of Twi1p/scnRNA complexes.

4.5

DNA Elimination of DNA Sequences from the Developing Somatic Nucleus

When initially discovered the phenomenon of DNA elimination in ciliates appeared to be an aberration in the world of biology that was focusing more and more on genetic processes. The rise of epigenetics has facilitated a clearer view of how DNA elimination relates to other biological processes. The link between scnRNAs and piRNAs was discussed in the previous section; however that is not the only biologically relevant link that DNA elimination in ciliates has to other organisms. Just as RNAi was shown to direct heterochromatin formation in *A. thaliana* and *S. pombe*, it was discovered that proper heterochromatin formation in the developing zygotic macronucleus through H3K9 and H3K27 methylation depended on normal function of RNAi components in *T. thermophila* [111, 116, 117, 156-158]. Thus DNA elimination depends on this establishment of heterochromatin to control the glut of repetitive elements in its genome [31, 156, 159]. Like other eukaryotes, initial methylation of histones associated with repetitive elements precipitates heterochromatin formation and compaction of these sequences. *T. thermophila* and other ciliates take the additional step of removing these heterochromatic sequences out of their somatic genome to create a streamlined genome not unlike many simple eukaryotes to optimize fitness. This streamlining process begins once Twi1p/scnRNA complexes are transported to the developing zygotic macronucleus to target H3K9 and H3K27 methylation of IESs [31, 138, 159]. These methylated histones then act to recruit chromodomain and other accessory proteins, which ultimately promote IES excision and DNA elimination by the domesticated piggyBac transposase, Tpb2p [14, 160-164]. This next section will describe in *T. thermophila* the link between RNAi and heterochromatin, IES-specific chromatin modifications, heterochromatin

readers and the nature of IESs and DNA elimination detailing the experiments relevant to each of these steps.

4.5.1

Targeting of scnRNA Complexes and Modification of Chromatin of DNA Sequences to be Eliminated

Like RNAi-directed heterochromatin formation in *A. thaliana* and *S. pombe*, RNAi-dependent DNA elimination in *T. thermophila* requires the production of ncRNA [74, 75, 111, 157, 165]. This ncRNA, produced in the developing zygotic macronucleus, is necessary for targeting IESs and interacts with the remaining Ema1p/Twi1p/scnRNA complexes, which are transported there once the developing macronucleus moves to the anterior of the cell and begins to enlarge [33, 75, 138, 140]. The Twi1p accessory proteins involved in genome scanning, Ema1p, Wag1p and CnjBp, are also transported to the developing macronucleus, although it is not known if this occurs in a greater complex with Twi1p or independently [75, 150]. The Ema1p/Twi1p/scnRNA/ncRNA complex interaction facilitates the binding of this complex with another group of proteins called the Ezl1p complex [75, 166].

In the RNAi-directed heterochromatin formation pathways in *A. thaliana* and *S. pombe*, heterochromatin formation is directed by H3K9me2, which is catalyzed by the Su(var) 3-9 homologues, KYP and Clr4, respectively [167-169]. RNAi-dependent DNA elimination in *T. thermophila* is dependent instead on Ezl1p, an E(z) homologue, and other associated proteins [159, 166]. The Ezl1p complex, which consists of Ezl1p, Esc1p, Rnf1p, Rnf2p and Nud1p, contains homologues from two protein complexes, PRC1 and PRC2, found in higher eukaryotes. These complexes are found to play a fundamental role

in the developmental regulation of heterochromatin through histone methylation and gene silencing in many organisms, which the Ezl1p complex has subsumed in *T. thermophila* [159, 166, 170-172]. Immunoprecipitations of Ezl1p, Nud1p and Rnf1p are able to pull-down Ema1p demonstrating a protein-protein interaction between the Ema1p/Twi1p/scnRNA complex and the Ezl1p complex [166]. Nud1p, Rnf1p, Rnf2 and Esc1p of the Ezl1p complex appear to have no catalytic function themselves, unlike other homologues found in PRC1 and PRC2 complexes, but instead act to enhance targeting of Ezl1p to IESs and Ezl1p methylase activity at the IESs ([166, 173-177], reviewed in [178]). Ezl1p, the effector component of the Ezl1p complex, is an E(z) homologue and contains the SET domain, which is capable of trimethylation of H3K9 and H3K27 [159, 166, 179-182]. Co-immunoprecipitation of Ezl1p is able to pull-down the other members of the Ezl1p complex, Nud1p, Rnf1p, Rnf2 and Esc1p [166]. Reciprocal pulldowns using tagged-Nud1p and -Rnf1p are also able to immunoprecipitate Ezl1p. Co-localization of H3K9me3 and H3K27me3 with Rnf1p of the Ezl1p complex demonstrates that it is capable of histone methylation during conjugation. A knockout of any of the Ezl1p complex components causes disassociation of the complex and loss of H3K9 methylation along with aberrant H3K27 methylation implicating the Ezl1p complex in both H3K9me3 and H3K27me3 during conjugation [159, 166]. Knockouts of the EZL1 complex also result in an increased accumulation of scnRNAs and ncRNAs produced in the developing macronucleus from the M IES indicating the existence of a feedback mechanism controlling both scnRNA and ncRNA production throughout the cell during conjugation [166]. DCL1, TWI1, and EZL1 complex knockouts also form aberrant DNA elimination bodies, which contain a number of proteins including the chromodomain

proteins, Pdd1p and Pdd3p [159, 166]. Like other components of RNAi-directed DNA elimination knockouts of the Ezl1p complex caused failure of DNA elimination [166]. In the case of EZL1 knockouts a failure to complete conjugation has also been observed [159].

Methylation of H3K9 and H3K27 by the Ezl1p complex is an integral part of the RNAi-dependent DNA elimination process [31, 156, 159]. Inhibiting this methylation by the Ezl1p complex through knockout of any component of RNAi-directed DNA elimination upstream or mutation of histone 3 itself is sufficient to block binding of the chromodomain proteins, Pdd1p and Pdd3p, and its association with other proteins to form DNA elimination bodies necessary for DNA elimination [75, 116, 117, 156, 159]. Mutation of H3K9Q directly blocks the site from methylation while mutations of H3S10E and H3S28E created an artificially phospho-switch, which naturally prevents methylation of the lysine directly downstream. All of these histone 3 mutations prevent Pdd1p and Pdd3p association with IESs [156, 159].

4.5.2

Protein Binding of Modified Chromatin, Protein Aggregate Formation and DNA Elimination

The role of chromodomain proteins in RNAi-directed heterochromatin formation and heterochromatin formation in general in eukaryotes is well documented [110, 183-185]. Once H3K9me3 and H3K27me3 modification occurs on histones associated with IESs, the aforementioned chromodomain proteins, Pdd1p and Pdd3p, are able to bind the IES chromatin, which, along with other associated proteins, condense the approximately 6000 IES loci into a handful of cellular foci called DNA elimination bodies [14, 31, 159, 160,

162-164, 186-188]. In these DNA elimination bodies a domesticated piggyBac transposase, Tbp2p, directs endonucleolytic cleavage of IESs at the IES boundaries excising the IES [14]. These double-strand breaks are thought to be repaired through one of the dsDNA break repair pathways but it is currently not known which pathway is responsible for this repair in *T. thermophila*.

Chromodomain proteins are pivotal heterochromatin histone readers. Knockouts of chromodomain proteins cause derepression of heterochromatin [185, 189]. Likewise, knockouts of PDD1 also see a decrease in heterochromatin formation [31, 159]. This implicates that establishment of H3K9me3 and H3K27me3 and binding of the two chromodomain proteins, Pdd1p and Pdd3p, to these marks are interconnected in DNA elimination body formation and DNA elimination (See figure 7A and 7B) [31, 159]. Pdd1p and Pdd3p, along with Pdd2p, were discovered by isolation of proteins enriched in developing zygotic macronuclei late during conjugation and were the first identified proteins shown to play a role in DNA elimination [160, 162, 186, 187]. Pdd1p contains two chromodomains and is capable of binding either H3K9me3 or H3K27me3 peptides *in vitro* and co-localizes with H3K9me3, H3K27me3 modified chromatin and IESs late in conjugation [31, 159, 160, 190].

Pdd1p may play multiple roles during development as it has been shown to localize within crescent micronuclei early during meiosis, within parental macronuclei and developing zygotic macronuclei, as well as in a cytoplasmic body known as the conjusome [160, 186, 191, 192]. Its biological role in the crescent micronucleus and parental macronucleus are unknown, although loss of expression during early developmental stages is sufficient to block DNA elimination indicating that such Pdd1p

localization is biologically relevant [192]. Localization of Pdd1p in the conjusome is thought to reflect the conjusome's role as a distribution center for the parental and developing macronuclei or as a staging ground for Pdd1p transition from the parental macronuclei into the developing zygotic macronuclei later in conjugation [191]. Other proteins that are known to localize to the developing zygotic macronucleus later in conjugation, such as Lia1p, Lia3p, and Lia5p, also appear in the conjusome [163, 164]. To signal the transition from the parental macronucleus to the conjusome and the developing zygotic macronucleus Pdd1p is phosphorylated up to four times [160]. This phosphorylation is lost as conjugation proceeds, which may trigger DNA elimination body formation. Co-localization of Pdd1p with H3K9me3, H3K27me3 and IESs occurs in the developing zygotic macronucleus [31, 159, 160, 190]. Localization of Pdd1p is initially diffuse throughout the entire nucleus but as the developing zygotic macronucleus matures Pdd1p concentrates into approximately 10-1μm foci called DNA elimination bodies [186]. These Pdd1p-containing DNA elimination bodies also contain a number of other proteins including Pdd2p, Pdd3p, Lia1p, Lia3p, Lia4p, Lia5p, Tbp2p and the Ezl1p complex (See figure 7B) [14, 159, 162-164, 166, 190]. Double knockouts of WAG1 and CNJB, LIA1 knockouts, EZL1 complex knockouts, PDD1 somatic knockouts and PDD2 somatic knockouts disrupt DNA elimination body formation [150, 166, 192, 193]. Tethering of Pdd1 to an artificial IES with no native histone methylation is also sufficient to direct DNA elimination indicating that Pdd1p itself is sufficient to recruit its accessory proteins such as Tbp2p and trigger DNA elimination [31].

The other chromodomain protein, Pdd3p, has been shown to bind strongly to H3K9me3 but not H3K27me3 *in vitro* [31, 159]. Pdd3p localization is limited to the

developing zygotic macronucleus, where it, like Pdd1p, is at first diffuse but then condenses into the DNA elimination bodies [162]. The other programmed DNA degradation protein, Pdd2p, is a protein with no known homology [187]. Pdd2p localization differs slightly from that of Pdd1p in that it only localizes to the parental and developing macronuclei [187, 193]. Like Pdd1p and Pdd3p localization in the developing zygotic macronucleus is initially diffuse until DNA elimination body formation. In a similar manner to Pdd1p, Pdd2p is phosphorylated once during the transition from the parental macronucleus to the developing zygotic macronucleus [190]. Likewise this phosphorylation is removed just prior to DNA elimination body formation. PDD2 somatic knockouts are sufficient to cause failure of cells to undergo DNA elimination and to complete conjugation, which, like PDD1 somatic knockouts, may indicate a vital role for early localization in the parental macronucleus [193].

Other proteins have been found to influence DNA elimination body formation. A diverse group of proteins participating in this process were identified by their localization specifically to differentiating macronuclei and were named **L**ocalization **i**n macronuclear **a**nlagen (Lia) proteins [163, 164]. Lia1p, Lia4p and Lia5p, all play a role in DNA elimination body formation. Lia5p contains a PHD Zn Finger and Lia4p contains a putative chromo shadow domain, but otherwise these proteins show no obvious homology to other known proteins. Lia1p is the best characterized and localizes to both the conjusome and developing zygotic macronucleus [163]. Late in conjugation Lia1p is found in association with Pdd1p and IESs in DNA elimination bodies. Knockouts of LIA1 fail to eliminate IESs and complete conjugation like many other proteins in RNAi-directed DNA elimination. Preliminary characterization of Lia3p, Lia4p and Lia5p show

diffuse localization early in the developing zygotic macronucleus and later localization in DNA elimination bodies [164]. LIA3, LIA4 and LIA5 knockouts also fail to undergo DNA elimination and complete conjugation [194]. While the role of these non-chromodomain proteins in RNAi-directed DNA elimination is not clear, perhaps these proteins form a scaffold through which Pdd1p and Pdd3p, interacting with specific classes of IESs, can be brought together to form the foci necessary for DNA elimination by the domesticated piggyBac transposase, Tbp2p.

Domesticated transposases have been shown to play an important role in a variety of eukaryotic organisms, for example RAG1/RAG2 recombinase in VDJ recombination in the human immune system [195, 196]. Ciliates appear to have domesticated transposases in order to facilitate removal of repetitive sequences and IESs during conjugation [12-14]. In *T. thermophila* Tbp2p, a piggyBac transposase homologue, is essential for removing IESs during conjugation (See figure 7C) [14]. Analysis of the TBP2 ORF shows homology with and preservation of the catalytic DDD motif in the domesticated piggyBac transposase in *P. tetraurelia*, PGM, and other piggyBac transposases in *H. sapiens, Xenopus sps.* and the moth, *T. ni.* Tbp2p co-localizes with H3K9m3, H3K27me3 and Pdd1p in the developing zygotic macronucleus before and after DNA elimination body formation. Knockdown of TBP2 using RNA hairpins does not inhibit Pdd1p association with H3K9me3 and H3K27me3 [14, 70]. However, TBP2 knockdown does inhibit DNA elimination body formation, IES removal and completion of conjugation implying an essential function downstream of Pdd1p and Pdd3p binding [14]. *In vitro* analysis of the catalytic DDD motif of Tbp2p, shows that it is capable of cutting the consensus piggyBac cleavage sequence, 5’ TTAA 3’, as well as a variety of

divergent sequences (See figure 7C) [14, 197, 198]. As previously reported, Tbp2p cleavage produces a four base pair 5' overhang, which is not observed in mutants of the Tbp2p DDD catalytic motif [14, 198].

4.6

Chromosome Breakage in the Developing Somatic Nucleus

The epigenetic RNAi-directed DNA elimination process in *T. thermophila* is only a part of the global genome rearrangement that occurs in the developing zygotic macronucleus during conjugation. Chromosome breakage and differential chromosome amplification must also take place for this process to be complete ([104-109, 199-202], reviewed in [5]). This epigenomic process differs between *P. tetraurelia* and *T. thermophila*. In *P. tetraurelia* this process seems to depend on RNAi-dependent DNA elimination machinery, while in *T. thermophila* chromosome breakage during conjugation is prompted by a conserved DNA sequence called the **c**hromosome **b**reakage **s**equence (CBS) [12, 103, 203]. Chromosome breakage and differential chromosome amplification have been shown to be essential for completion of conjugation and are linked to RNAi-directed DNA elimination [33, 116, 150, 159, 193]. The conserved 15bp CBS sequence is sufficient and necessary for chromosome breakage and telomere addition, which is blocked in CBS mutants (See figure 7C) [203-206]. Genomic analysis of the *T. thermophila* genome has shown that with little variation the CBS is present at all sites of chromosome breakage [106, 107]. Like IES excision, chromosome breakage appears to be dependent on the piggyBac transposase, Tbp2p [14].

5

Chromosome Fragmentation and Elimination of DNA During Conjugation in *Oxytricha*

The studies of DNA elimination in *P. tetraurelia* and *T. thermophila* described above have revealed the role of sRNAs and long ncRNAs in remodeling genomes during development. They have also hinted to possible mechanisms that allow phenotypic traits to be propagated to the next generation. DNA elimination and chromosome fragmentation occur throughout the entire ciliate clade (reviewed in [5]). Do RNAs play a similar role in more distantly related ciliates? Research on these processes in a subgroup of ciliates called stichotrichs has provided a definitive answer [76]. Stichotrichs, including the genera *Oxytricha* and *Stylonychia*, undergo DNA elimination and chromosome breakage, but these processes are much more extreme resulting in elimination of greater than 95% of the genome and gene sized mini-chromosomes approximately 2 kb in size ([207-210], reviewed in [5]). Further complicating understanding of these processes in *Oxytricha* and *Stylonychia* was the discovery of scrambled genes in the micronucleus [211-217]. Recent data in *O. trifallax* indicates that parental macronuclear ncRNA is able to direct unscrambling of genes, DNA elimination and chromosome breakage [76].

5.1

Gene Unscrambling and Domesticated Transposases in DNA Elimination and Chromosome Breakage

Like other ciliates, it seems likely that in stichotrichs DNA elimination and by extension gene scrambling in the micronucleus are ways to prevent active transposons from appearing in the somatic macronuclear genome (reviewed in [5]). By scrambling

macronuclear-destined sequences (MDSs) of genes in the germline micronucleus, stichotrichs ensure that DNA elimination must occur during sexual reproduction to generate intact coding regions if progeny are to be viable. Gene scrambling takes several forms with some MDSs having undergone permutation in linear order, while others are even inverted with respect to the other MDSs to complicate the unscrambling process further (See figure 8B) [211-217].

The scrambled genes discovered so far include actin I, α telomere binding protein (αTBP) and DNA polymerase α with many more likely waiting to be discovered. Similar to *P. tetraurelia* the MDSs of *O. trifallax* are bordered by short repeats termed pointers that may help direct gene unscrambling and DNA elimination [218]. However, these repeats are too short to unambiguously accomplish this task. The discovery of parental macronuclear ncRNA during conjugation and its role in gene unscrambling and DNA elimination illuminates how these processes occur in *O. trifallax* and possibly stichotrichs in general [76]. RT-PCR of RNA isolated from conjugating *O. trifallax* early and late in conjugation detected the presence of sense and anti-sense ncRNAs. These ncRNAs, which are longer than mRNAs and contain telomeres, imply that general transcription of all mini-chromosomes initiates at the telomere sequence early during conjugation. RNAi against these ncRNAs during conjugation was sufficient to block rearrangement of the target genes in the developing macronucleus. To validate the role of the parental macronucleus in producing these ncRNAs, the Landweber group injected artificial DNA and RNA transcripts to a known gene, TEBPβ, which contained different permutations of the MDSs. Upon completion of conjugation, some TEBPβ genes containing the alternative MDS order were found in the developing macronucleus, verifying the ability

of artificial DNA in the parental macronucleus to produce ncRNA transcripts and alter DNA elimination in the developing macronucleus.

Similar to *P. tetraurelia* and *T. thermophila*, a family of domesticated transposases has been found to play a role in gene unscrambling and DNA elimination in *O. trifallax* [12-14]. Unlike *P. tetraurelia* and *T. thermophila*, these transposases in *O. trifallax*, TBE1, TBE2 and TBE3, belong to the TBE family of transposons and are not retained in the macronucleus after DNA elimination and chromosome breakage [13, 219, 220]. Triple knockdown of these transposases is sufficient to cause aberrant gene unscrambling and DNA elimination [13].

These pieces of data have allowed a model, shown in figure 8, for gene unscrambling, DNA elimination and chromosome breakage in *O. trifallax* to be proposed [76]. Early in conjugation bidirectional transcription of all mini-chromosomes in the parental macronucleus produces ncRNA. This ncRNA is transported to the developing macronucleus later in conjugation after its appearance and directs gene unscrambling, if necessary, and DNA elimination of IESs by a family of domesticated transposases, TBE1, TBE2 and TBE3, to produce a functional mini-chromosome in the developing macronucleus [13, 76]. Although the presence of sRNAs, a Piwi homologue and heterochromatin marks have been found in the stichotrich, *Stylonychia*, it remains to be seen whether any of these play a role in gene unscrambling and DNA elimination in *Oxytricha*, *Stylonychia* and other stichotrichs similar to *P. tetraurelia* and *T. thermophila* [33, 66, 84, 156, 159, 221]. One thing appears to be consistent in DNA elimination in all ciliates and that is the presence of ncRNA.

II: Double-Stranded RNA Binding Motif Proteins

Many proteins are fundamental for the process of RNAi to be a success, including the RNase III family proteins, Dicer and Drosha [222, 223] . All members of the RNaseIII protein family contain two conserved protein domains, the ribonuclease 3 domain and the dsRNA binding motif domain (DSRM) ([80, 82, 222, 224], reviewed in [83]). A number of Dicer and Drosha homologues also have tandem DSRM-containing protein partners that facilitate optimal sRNA production [225-231]. To date no tandem DSRM-containing protein partner has been found for any ciliate Dicer protein [66, 116, 117]. Initially identified as possible partner proteins for DCL1, I will discuss in chapters 4 and 5 the role of two tandem DSRM-containing proteins in the life cycle of *T. thermophila*. These two proteins, DRB2 and DRB1, do not appear to be canonical Dicer partner proteins but instead are important in other processes during conjugation.

Although much is known about the DSRM-containing proteins belonging to the RNaseIII family and their tandem DSRM-containing partner proteins, the DSRM protein domain is found in a number of other protein families across the eukaryotic lineage (reviewed in [232, 233]). Additional roles for DSRM-containing proteins range from RNA editing in the ADAR family, to translation inhibition in response to viruses by PKR family members, to developmental RNA localization in the Staufen family [234-238].

ADAR family proteins modify RNA sequences containing adenosine (A) nucleotides to create inosine (I) via hydrolytic deamination ([237], reviewed in [239]). ADAR family members contain 1 to 3 DSRMs that bind dsRNA prior to editing ([240], reviewed in [239]). Deletions of these domains or specific amino acid mutations within these domains are sufficient to block dsRNA binding and RNA editing [240, 241]. PKR family proteins are protein kinases that are activated by the presence of dsRNA in the

cytoplasm ([236], reviewed in [242]). dsRNA binding activates the PKR kinase domain, which causes autophosphorylation and dimerization of PKR. Activated PKR then phosphorylates the translation initiator, eIF2α, blocking global translation, and a protein phosphatase subunit B56α, modifying global cell signaling ([236, 243], reviewed in [242]). The DSRMs of PKR are essential for dsRNA binding and subsequent downstream signaling, which are lost by DSRM deletion or specific DSRM amino acid mutations [244, 245]. Another protein family that contains 3-5 DSRM domains is the Staufen family (reviewed in [246]). This was the first identified DSRM containing protein and was shown to be essential for developmental mRNA localization in *D. melanogaster* [234, 235]. A similar role in mRNA transport has been found for mammalian homologues [247, 248]. Deletion of the DSRM domains causes a loss of mRNA binding and is lethal in *D. melanogaster* [247, 249]. Despite the different biological functions of all these DSRM-containing protein families, they all require the DSRM domain for dsRNA binding to fully function. Scores of other DSRM-containing proteins have no identified function. It is likely that many of the newly discovered ncRNAs will uncovered the roles of these proteins, including the *T. thermophila* proteins DRB2 and DRB1. Understanding the interaction of these ncRNAs or their byproducts with these orphan DSRM-containing proteins will expand our knowledge of other important biological processes.

III: Scope of the Thesis

The ciliate, *T. thermophila*, offers a unique opportunity to study the effects of ncRNAs and RNAi on heterochromatin structure and genome organization. To date, only two other model organisms, *S. pombe* and *A. thaliana*, have been able to mechanistically link

RNAi to heterochromatin formation, although RNAi has been found to influence heterochromatin formation in many other model organisms ([111, 157, 158], reviewed in [110]). Unlike *S. pombe* and *A. thaliana*, the developmentally regulated RNAi-directed DNA elimination process in *T. thermophila* allows for straightforward determination of the roles of ncRNAs to direct histone methylation and DNA elimination during conjugation as an undifferentiated macronucleus is remodeled into a functional zygotic macronucleus [31, 116, 117, 159]. The research documented in this thesis has attempted to understand the role of ncRNA throughout RNAi-directed DNA elimination and conjugation through genetic manipulation and microscopy.

As discussed earlier, IESs are targeted for DNA elimination late in conjugation by scnRNAs processed from long, bidirectional ncRNAs transcribed early during conjugation [31, 33, 116, 117, 159]. Experiments have shown that artificial plasmid IESs can be introduced late during conjugation in *T. thermophila* and be successful rearranged [250-253]. However, it was not known if these artificial IESs depended on scnRNAs produced earlier or whether they were capable of producing scnRNAs to direct IES excision themselves. In order to definitively show that ncRNAs produced from endogenous IESs early in conjugation were essential for excision of both types of IESs later in conjugation, I removed the R IES from the micronuclear genome. In chapter 2, I show that without the native R IES, from which to produce scnRNAs, an artificial R IES plasmid was incapable of undergoing DNA rearrangement when introduced into conjugating *T. thermophila*. This demonstrates the importance of long, bidirectional ncRNA transcription from endogenous IESs early during conjugation.

When I began my graduate studies the protein responsible for producing scnRNAs from ncRNAs was not identified in *T. thermophila*. It was suspected that a Dicer homologue was responsible for scnRNA production [33]. I identified the conjugation specific Dicer homologue, DCL1, and demonstrated that it was responsible for scnRNA production. I produced data showing that DCL1 was expressed only during conjugation, where it localized to the crescent micronucleus. I also showed that when scnRNA production in DCL1 knockouts was lost there was also a failure of DNA elimination of IESs. This and additional data resulted in a publication in *Molecular and Cellular Biology* describing the role of DCL1 early in conjugation in scnRNA production from long, bidirectional ncRNAs, which is described in chapter 3 [116].

In an attempt to uncover a tandem DSRM-containing protein partner for DCL1, I identified the two tandem DSRM-containing proteins encoded in the *T. thermophila* genome, DRB2 and DRB1. In chapter 4, I describe the basic characterization of these proteins in *T. thermophila*. Both genes are expressed throughout the life cycle of *T. thermophila* and localize to nuclei. I further show that zygotic expression of DRB2 is essential for excision of IESs, chromosome fragmentation and completion of conjugation. Late in conjugation I also found that Drb2p co-localizes with the essential chromodomain-containing protein, Pdd1p, in DNA elimination bodies and that Pdd1p-containing DNA elimination bodies are disrupted when zygotic expression DRB2 is abolished. This data implies that dsRNAs or their byproducts have additional roles in RNAi-directed DNA elimination later in conjugation through the tandem DSRM-containing protein DRB2.

Further characterization of the other tandem DSRM-containing protein, DRB1, uncovered a potential role in meiosis described in chapter 5. Although DRB1 is expressed constitutively, it is strongly upregulated early in conjugation in *T. thermophila* where it localizes to the macronucleus and to the termini of the crescent micronucleus. I found localization of Drb1p to overlap with nucleolar structures in the macronucleus and with the centromere-specific histone, Cna1p, in the crescent micronucleus. During the basic characterization of DRB1, DRB1 knockouts were only found to have a slight reduction in progeny production. Analysis of the progression of DRB1 knockout pairs during conjugation revealed a high percentage of pairs aborting conjugation before the production of zygotic nuclei similar to strains unable to produce functional haploid gametic nuclei. This phenotype can be rescued by exogenous DRB1 expression, indicating that Drb1p has a role in the micronucleus in haploid gamete production early in conjugation. In this thesis the examination of these two tandem DSRM-containing proteins has revealed likely additional roles for dsRNA during conjugation in addition to the known interaction of ncRNA from IESs with the Dicer protein, DCL1, during RNAi-directed DNA elimination. Further research on DRB2 and DRB1 in *T. thermophila* should open new avenues into understanding the biological function of ncRNAs in ciliates and of the interaction of ncRNAs with DSRM-containing proteins in other eukaryotes.

References:

1. Sonneborn, T.M., *Sex, Sex Inheritance and Sex Determination in Paramecium Aurelia.* Proc Natl Acad Sci U S A, 1937. **23**(7): p. 378-85.
2. Philippe, H., A. Germot, and D. Moreira, *The new phylogeny of eukaryotes.* Curr Opin Genet Dev, 2000. **10**(6): p. 596-601.

3. Nowacki, M., et al., *RNA-mediated epigenetic regulation of DNA copy number.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(51): p. 22140-4.
4. Heyse, G., et al., *RNA-dependent control of gene amplification.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(51): p. 22134-9.
5. Prescott, D.M., *The DNA of ciliated protozoa.* Microbiol Rev, 1994. **58**(2): p. 233-67.
6. Betermier, M., *Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate Paramecium.* Res Microbiol, 2004. **155**(5): p. 399-408.
7. Yao, M.C. and M.A. Gorovsky, *Comparison of the sequences of macro- and micronuclear DNA of Tetrahymena pyriformis.* Chromosoma, 1974. **48**(1): p. 1-18.
8. Klobutcher, L.A. and G. Herrick, *Consensus inverted terminal repeat sequence of Paramecium IESs : resemblance to termini of Tc1-related and Euplotes Tec transposons.* Nucl. Acids Res., 1995. **23**(11): p. 2006-2013.
9. Herrick, G., et al., *Mobile elements bounded by C4A4 telomeric repeats in Oxytricha fallax.* Cell, 1985. **43**: p. 759-768.
10. Ribas-Aparicio, R.M., et al., *Nucleic acid splicing events occur frequently during macronuclear development in the protozoan Oxytricha nova and involve the elimination of unique DNA.* Genes Dev, 1987. **1**: p. 323-336.
11. Jahn, C.L. and L.A. Klobutcher, *Genome remodeling in ciliated protozoa.* Annual review of microbiology, 2002. **56**: p. 489-520.
12. Baudry, C., et al., *PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia.* Genes Dev, 2009. **23**(21): p. 2478-83.
13. Nowacki, M., et al., *A functional role for transposases in a large eukaryotic genome.* Science, 2009. **324**(5929): p. 935-8.
14. Cheng, C.Y., et al., *A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila.* Mol Biol Cell, 2010. **21**(10): p. 1753-62.
15. Hannon, G.J., *RNA interference.* Nature, 2002. **418**(6894): p. 244-51.
16. Allis, C.D., et al., *Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, Tetrahymena thermophila.* Cell, 1980. **20**(3): p. 609-17.
17. Hayashi, T., et al., *Tetrahymena histone H3. Purification and two variant sequences.* J Biochem (Tokyo), 1984. **95**(6): p. 1741-9.
18. Liu, X., B. Li, and GorovskyMa, *Essential and nonessential histone H2A variants in Tetrahymena thermophila.* Mol Cell Biol, 1996. **16**(8): p. 4305-11.
19. Sugai, T. and K. Hiwatashi, *Cytological and autoradiographic studies of the micronucleus at meiotic prophase in Tetrahymena pyriformis.* J. Protozool, 1974. **21**: p. 542-548.
20. Martindale, D.W., C.D. Allis, and P.J. Bruns, *RNA and protein synthesis during meiotic prophase in Tetrahymena thermophila.* J. Protozool., 1985. **32**(4): p. 644-649.

21. Stargell, L.A., et al., *Temporal and spatial association of histone H2A variant hv1 with transcriptionally competent chromatin during nuclear development in Tetrahymena thermophila.* Genes & dev;, 1993. **7**: p. 2641-2651.
22. Yu, L. and M.A. Gorovsky, *Constitutive expression, not a particular primary sequence, is the important feature of the H3 replacement variant hv2 in Tetrahymena thermophila.* Mol Cell Biol, 1997. **17**(11): p. 6303-10.
23. Shen, X., et al., *Linker histones are not essential and affect chromatin condensation in vivo.* Cell, 1995. **82**(1): p. 47-56.
24. Shen, X. and M.A. Gorovsky, *Linker histone H1 regulates specific gene expression but not global transcription in vivo.* Cell, 1996. **86**(3): p. 475-83.
25. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis.* Proc Natl Acad Sci U S A, 1964. **51**: p. 786-94.
26. Ren, Q. and M.A. Gorovsky, *Histone H2A.Z acetylation modulates an essential charge patch.* Mol Cell, 2001. **7**(6): p. 1329-35.
27. Ren, Q. and M.A. Gorovsky, *The nonessential H2A N-terminal tail can function as an essential charge patch on the H2A.Z variant N-terminal tail.* Mol Cell Biol, 2003. **23**(8): p. 2778-89.
28. Brownell, J.E. and C.D. Allis, *An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei.* Proc Natl Acad Sci U S A, 1995. **92**(14): p. 6364-8.
29. Brownell, J.E., et al., *Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation.* Cell, 1996. **84**(6): p. 843-51.
30. Strahl, B.D., et al., *Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena.* Proc Natl Acad Sci U S A, 1999. **96**(26): p. 14967-72.
31. Taverna, S.D., R.S. Coyne, and C.D. Allis, *Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena.* Cell, 2002. **110**(6): p. 701-11.
32. Nakayama, J., et al., *Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly.* Science, 2001. **292**(5514): p. 110-3.
33. Mochizuki, K., et al., *Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena.* Cell, 2002. **110**(6): p. 689-99.
34. Wei, Y., et al., *Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena.* Proc Natl Acad Sci U S A, 1998. **95**(13): p. 7480-4.
35. Wei, Y., et al., *Phosphorylation of histone H3 is required for proper chromosome condensation and segregation.* Cell, 1999. **97**(1): p. 99-109.
36. Sonneborn, T.M., *Does Pre-formed Cell Structure Play an Essential Role in Cell Heredity? ,* in *The Nature of Biological Diversity*, J.M. Allen, Editor 1963, McGraw-Hill: New York. p. 165-221.
37. Beisson, J. and T.M. Sonneborn, *Cytoplasmic Inheritance of the Organization of the Cell Cortex in Paramecium Aurelia.* Proc Natl Acad Sci U S A, 1965. **53**: p. 275-82.

38. Sonneborn, T.M., *Genetics of cellular differentiation: stable nuclear differentiation in eucaryotic unicells.* Annu Rev Genet, 1977. **11**: p. 349-67.
39. Sonneborn, T.M., *Paramecium aurelia*, in *Handbook of genetics : Plants, plant viruses and protists*, R.C. King, Editor 1974, Plenum Press: New York. p. 469-594.
40. Sonneborn, T.M., *Recent advances in the genetics of Paramecium and Euplotes.* Adv. Genet., 1947. **1**: p. 263-358.
41. Nanney, D.L., *Mating-type inheritance at conjugation in variety 4 of Paramecium aurelia.* J. Protozool., 1957. **4**: p. 89-95.
42. Butzel, H.M., *Abnormalities in Nuclear Behavior and Mating Type Determination in Cytoplasmically Bridged Exconjugants of Doublet Paramecium aurelia*.* Journal of Eukaryotic Microbiology, 1973. **20**(1): p. 140-142.
43. Koizumi, S., *The cytoplasmic factor that fixes macronuclear mating type determination in Paramecium aurelia, syngen 4.* Genetics, 1971. **68**(Supplemental): p. 34.
44. Sonneborn, T.M., *The determination of hereditary antigenic differences in genically identical Paramecium cells.* Proc Natl Acad Sci U S A, 1948. **34**(8): p. 413-8.
45. Brygoo, Y. and A.-M. Keller, *Genetic analysis of mating type differentiation in Paramecium tetraurelia. III. A mutation restricted to mating type E and affecting the determination of mating type.* Dev. Genet., 1981. **2**: p. 13-22.
46. Meyer, E. and A.-M. Keller, *A Mendelian mutation affecting mating-type determination also affects developmental genomic rearrangements in Paramecium tetraurelia.* Genetics, 1996. **143**: p. 191-202.
47. Duharcourt, S., A. Butler, and E. Meyer, *Epigenetic self-regulation of developmental excision of an internal eliminated sequence in Paramecium tetraurelia.* Genes and Development, 1995. **9**: p. 2065-2077.
48. Epstein, L.M. and J.D. Forney, *Mendelian and non-Mendelian mutations affecting surface antigen expression in Paramecium tetraurelia.* Mol. Cell. Biol., 1984. **4**: p. 1583-1590.
49. Preer, L.B., G. Hamilton, and J.R. Preer, *Micronuclear DNA from Paramecium tetraurelia : serotype 51A gene has internally eliminated sequences.* J. Protozool., 1992. **39**: p. 678-682.
50. Godiska, R., et al., *Transformation of Paramecium by microinjection of a cloned serotype gene.* Proc Natl Acad Sci U S A, 1987. **84**(21): p. 7590-4.
51. Koizumi, S. and S. Kobayashi, *Microinjection of plasmid DNA encoding the A surface antigen of Paramecium tetraurelia restores the ability to regenerate a wild-type macronucleus.* Mol Cell Biol, 1989. **9**(10): p. 4398-4401.
52. You, Y., et al., *Macronuclear transformation with specific DNA fragments controls the content of the new macronuclear genome in Paramecium tetraurelia.* Mol Cell Biol, 1991. **11**(2): p. 1133-1137.
53. Jessop-Murray, H., et al., *Permanent rescue of a non-Mendelian mutation of Paramecium by microinjection of specific DNA sequences.* Genetics, 1991. **129**(3): p. 727-34.

54. Scott, J.M., C.L. Leeck, and J.D. Forney, *Analysis of the micronuclear B type surface protein gene in Paramecium tetraurelia.* Nucleic Acids Res, 1994. **22**(23): p. 5079-5084.
55. Scott, J.M., et al., *Non-Mendelian inheritance of macronuclear mutations is gene specific in Paramecium tetraurelia.* Mol. Cell. Biol., 1994. **14**(4): p. 2479-2484.
56. Kim, C.S., J.R. Preer, Jr., and B. Polisky, *Identification of DNA segments capable of rescuing a non-mendelian mutant in paramecium.* Genetics, 1994. **136**(4): p. 1325-8.
57. Forney, J.D., F. Yantiri, and K. Mikami, *Developmentally controlled rearrangement of surface protein genes in Paramecium tetraurelia.* J. Eukaryot. Microbiol., 1996. **43**(6): p. 462-467.
58. Napoli, C., C. Lemieux, and R. Jorgensen, *Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans.* Plant Cell, 1990. **2**(4): p. 279-289.
59. Romano, N. and G. Macino, *Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences.* Molecular microbiology, 1992. **6**(22): p. 3343-53.
60. Ruiz, F., et al., *Homology-dependent gene silencing in Paramecium.* Mol. Biol. Cell., 1998. **9**(4): p. 931-943.
61. Galvani, A. and L. Sperling, *Transgene-mediated post-transcriptional gene silencing is inhibited by 3' non-coding sequences in Paramecium.* Nucleic Acids Res, 2001. **29**(21): p. 4387-94.
62. Lejeune, E. and R.C. Allshire, *Common ground: small RNA programming and chromatin modifications.* Current opinion in cell biology, 2011.
63. Malone, C.D. and G.J. Hannon, *Small RNAs as guardians of the genome.* Cell, 2009. **136**(4): p. 656-68.
64. Carthew, R.W. and E.J. Sontheimer, *Origins and Mechanisms of miRNAs and siRNAs.* Cell, 2009. **136**(4): p. 642-55.
65. Lee, S.R. and K. Collins, *Two classes of endogenous small RNAs in Tetrahymena thermophila.* Genes Dev, 2006. **20**(1): p. 28-33.
66. Lepere, G., et al., *Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia.* Nucleic Acids Res, 2009. **37**(3): p. 903-15.
67. Galvani, A. and L. Sperling, *RNA interference by feeding in Paramecium.* Trends Genet, 2002. **18**(1): p. 11-2.
68. Garnier, O., et al., *RNA-mediated programming of developmental genome rearrangements in Paramecium tetraurelia.* Mol Cell Biol, 2004. **24**(17): p. 7370-9.
69. Lepere, G., et al., *Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia.* Genes Dev, 2008. **22**(11): p. 1501-12.
70. Howard-Till, R.A. and M.C. Yao, *Induction of Gene Silencing by Hairpin RNA Expression in Tetrahymena thermophila Reveals a Second Small RNA Pathway.* Mol Cell Biol, 2006. **26**(23): p. 8731-8742.
71. Hamilton, A.J. and D.C. Baulcombe, *A species of small antisense RNA in posttranscriptional gene silencing in plants.* Science, 1999. **286**(5441): p. 950-2.

72. Lee, S.R. and K. Collins, *Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs.* Nat Struct Mol Biol, 2007. **14**(7): p. 604-10.
73. Marker, S., et al., *Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in Paramecium tetraurelia.* Nucleic acids research, 2010. **38**(12): p. 4092-107.
74. Chalker, D.L. and M.C. Yao, *Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila.* Genes Dev, 2001. **15**(10): p. 1287-98.
75. Aronica, L., et al., *Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in Tetrahymena.* Genes Dev, 2008. **22**(16): p. 2228-41.
76. Nowacki, M., et al., *RNA-mediated epigenetic programming of a genome-rearrangement pathway.* Nature, 2008. **451**(7175): p. 153-8.
77. Duharcourt, S., A. Keller, and E. Meyer, *Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in Paramecium tetraurelia.* Mol Cell Biol, 1998. **18**(12): p. 7075-7085.
78. Aury, J.M., et al., *Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia.* Nature, 2006. **444**(7116): p. 171-8.
79. Dippell, R.V., *A preliminary report on the chromosomal constitution of certain variety 4 races of Paramecium aurelia.* Caryologia, 1954. **6**: p. 1109-1111.
80. Blaszczyk, J., et al., *Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage.* Structure, 2001. **9**(12): p. 1225-36.
81. Han, J., et al., *The Drosha-DGCR8 complex in primary microRNA processing.* Genes Dev, 2004. **18**(24): p. 3016-27.
82. Zhang, H., et al., *Single processing center models for human Dicer and bacterial RNase III.* Cell, 2004. **118**(1): p. 57-68.
83. Kim, V.N., J. Han, and M.C. Siomi, *Biogenesis of small RNAs in animals.* Nature reviews. Molecular cell biology, 2009. **10**(2): p. 126-39.
84. Bouhouche, K., et al., *Functional specialization of Piwi proteins in Paramecium tetraurelia from post-transcriptional gene silencing to genome remodelling.* Nucleic Acids Research, 2011. **39**(10): p. 4249-4264.
85. Nowacki, M., W. Zagorski-Ostoja, and E. Meyer, *Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in Paramecium tetraurelia.* Curr Biol, 2005. **15**(18): p. 1616-28.
86. Meller, V.H., et al., *roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system.* Cell, 1997. **88**: p. 445-457.
87. Amrein, H. and R. Axel, *Genes expressed in neurons of adult male Drosophila.* Cell, 1997. **88**: p. 459-469.
88. Brown, C.J., et al., *The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus.* Cell, 1992. **71**: p. 527-542.
89. Brockdorff, N., et al., *The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus.* Cell, 1992. **71**: p. 515-526.

90. Brown, C.J., et al., *A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome.* Nature, 1991. **349**(6304): p. 38-44.
91. Bartolomei, M.S., S. Zemel, and S.M. Tilghman, *Parental imprinting of the mouse H19 gene.* Nature, 1991. **351**(6322): p. 153-5.
92. Ponting, C.P., P.L. Oliver, and W. Reik, *Evolution and functions of long noncoding RNAs.* Cell, 2009. **136**(4): p. 629-41.
93. Klobutcher, L.A. and G. Herrick, *Developmental genome reorganization in ciliated protozoa: the transposon link.* Prog Nucleic Acid Res Mol Biol, 1997. **56**: p. 1-62.
94. Gratias, A. and M. Betermier, *Processing of double-strand breaks is involved in the precise excision of paramecium internal eliminated sequences.* Mol Cell Biol, 2003. **23**(20): p. 7152-62.
95. Mayer, K.M., K. Mikami, and J.D. Forney, *A mutation in Paramecium tetraurelia reveals functional and structural features of developmentally excised DNA elements.* Genetics, 1998. **148**: p. 139-149.
96. Mayer, K. and J. Forney, *A mutation in the flanking 5'-TA-3' dinucleotide prevents excision of an internal eliminated sequence from the Paramecium tetraurelia genome.* Genetics, 1999. **151**(2): p. 597-604.
97. Gratias, A., et al., *Developmentally programmed DNA splicing in Paramecium reveals short-distance crosstalk between DNA cleavage sites.* Nucleic Acids Res, 2008. **36**(10): p. 3244-51.
98. Ruiz, F., et al., *The SM19 gene, required for duplication of basal bodies in Paramecium, encodes a novel tubulin, eta-tubulin.* Curr Biol, 2000. **10**(22): p. 1451-4.
99. Elick, T.A., C.A. Bauser, and M.J. Fraser, *Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase.* Genetica, 1996. **98**(1): p. 33-41.
100. Mitra, R., J. Fain-Thornton, and N.L. Craig, *piggyBac can bypass DNA synthesis during cut and paste transposition.* EMBO J, 2008. **27**(7): p. 1097-109.
101. Kapusta, A., et al., *Highly Precise and Developmentally Programmed Genome Assembly in Paramecium Requires Ligase IV-Dependent End Joining.* PLoS genetics, 2011. **7**(4): p. e1002049.
102. Duharcourt, S., G. Lepere, and E. Meyer, *Developmental genome rearrangements in ciliates: a natural genomic subtraction mediated by non-coding transcripts.* Trends Genet, 2009. **25**(8): p. 344-50.
103. Le Mouel, A., et al., *Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia.* Eukaryot Cell, 2003. **2**(5): p. 1076-90.
104. Altschuler, M.I. and M.C. Yao, *Macronuclear DNA of Tetrahymena thermophila exists as defined subchromosomal-sized molecules.* Nucleic Acids Res, 1985. **13**(16): p. 5817-5831.
105. Woodard, J., E. Kaneshiro, and M.A. Gorovsky, *Cytochemical studies on the problem of macronuclear subnuclei in tetrahymena.* Genetics, 1972. **70**(2): p. 251-60.

106. Cassidy-Hanley, D., et al., *Genome-wide characterization of Tetrahymena thermophila chromosome breakage sites. II. Physical and genetic mapping.* Genetics, 2005. **170**(4): p. 1623-31.
107. Hamilton, E., et al., *Genome-wide characterization of tetrahymena thermophila chromosome breakage sites. I. Cloning and identification of functional sites.* Genetics, 2005. **170**(4): p. 1611-21.
108. Doerder, F.P., J.C. Deak, and J.H. Lief, *Rate of phenotypic assortment in Tetrahymena thermophila.* Developmental genetics, 1992. **13**(2): p. 126-32.
109. Conover, R.K. and C.F. Brunk, *Macronuclear DNA molecules of Tetrahymena thermophila.* Mol Cell Biol, 1986. **6**(3): p. 900-905.
110. Allis, C.D., et al., eds. *Epigenetics*. 2007, Cold Spring Harbor Press: Cold Spring Harbor.
111. Volpe, T.A., et al., *Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi.* Science, 2002. **297**(5588): p. 1833-7.
112. Reinhart, B.J. and D.P. Bartel, *Small RNAs correspond to centromere heterochromatic repeats.* Science, 2002. **297**(5588): p. 1831.
113. Wassenegger, M., et al., *RNA-directed de novo methylation of genomic sequences in plants.* Cell, 1994. **76**: p. 567-576.
114. Brown, D.D. and I.B. Dawid, *Specific gene amplification in oocytes.* Science, 1968. **160**: p. 272-280.
115. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**: p. 806-811.
116. Malone, C.D., et al., *Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila.* Mol Cell Biol, 2005. **25**(20): p. 9151-64.
117. Mochizuki, K. and M.A. Gorovsky, *A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase.* Genes Dev, 2005. **19**(1): p. 77-89.
118. Liu, Y., et al., *Elimination of Foreign DNA during Somatic Differentiation in Tetrahymena thermophila Shows Position Effect and Is Dosage Dependent.* Eukaryot Cell, 2005. **4**(2): p. 421-31.
119. Yao, M.C., P. Fuller, and X. Xi, *Programmed DNA deletion as an RNA-guided system of genome defense.* Science, 2003. **300**(5625): p. 1581-4.
120. Aravin, A.A., G.J. Hannon, and J. Brennecke, *The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race.* Science, 2007. **318**(5851): p. 761-4.
121. Brennecke, J., et al., *Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila.* Cell, 2007. **128**(6): p. 1089-103.
122. Carmell, M.A., et al., *MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline.* Dev Cell, 2007. **12**(4): p. 503-14.
123. Das, P.P., et al., *Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline.* Mol Cell, 2008. **31**(1): p. 79-90.
124. Gunawardane, L.S., et al., *A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila.* Science, 2007. **315**(5818): p. 1587-90.

125. Kuramochi-Miyagawa, S., et al., *DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes.* Genes Dev, 2008. **22**(7): p. 908-17.
126. Saito, K., et al., *Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome.* Genes Dev, 2006. **20**(16): p. 2214-22.
127. Vagin, V.V., et al., *A distinct small RNA pathway silences selfish genetic elements in the germline.* Science, 2006. **313**(5785): p. 320-4.
128. Martindale, D.W., C.D. Allis, and P. Bruns, *Conjugation in Tetrahymena thermophila: a temporal analysis of cytological stages.* Experimental Cell Research, 1982. **140**: p. 227-236.
129. Ray, C., *Meiosis and Nuclear Behavior in Tetrahymena pyriformis*.* Journal of Eukaryotic Microbiology, 1956. **3**(2): p. 88-96.
130. Wenkert, D. and C.D. Allis, *Timing of the appearance of macronuclear-specific histone variant hv1 and gene expression in developing new macronuclei of Tetrahymena thermophila.* J Cell Biol, 1984. **98**(6): p. 2107-17.
131. Stargell, L.A. and M.A. Gorovsky, *TATA-binding protein and nuclear differentiation in Tetrahymena thermophila.* Mol. Cell. Biol., 1994. **14**(1): p. 723-734.
132. Mochizuki, K. and M.A. Gorovsky, *RNA polymerase II localizes in Tetrahymena thermophila meiotic micronuclei when micronuclear transcription associated with genome rearrangement occurs.* Eukaryot Cell, 2004. **3**(5): p. 1233-40.
133. Henderson, I.R., et al., *Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning.* Nat Genet, 2006. **38**(6): p. 721-5.
134. Elbashir, S.M., W. Lendeckel, and T. Tuschl, *RNA interference is mediated by 21- and 22-nucleotide RNAs.* Genes Dev, 2001. **15**(2): p. 188-200.
135. Robertson, H.D. and J.J. Dunn, *Ribonucleic acid processing activity of Escherichia coli ribonuclease III.* J Biol Chem, 1975. **250**(8): p. 3050-6.
136. Schweitz, H. and J.P. Ebel, *A study of the mechanism of action of E. coli ribonuclease 3.* Biochimie, 1971. **53**(5): p. 585-93.
137. Crouch, R.J., *Ribonuclease 3 does not degrade deoxyribonucleic acid-ribonucleic acid hybrids.* J Biol Chem, 1974. **249**(4): p. 1314-6.
138. Mochizuki, K. and M.A. Gorovsky, *Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement.* Genes Dev, 2004. **18**(17): p. 2068-73.
139. Hammond, S.M., et al., *An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells.* Nature, 2000. **404**(6775): p. 293-6.
140. Noto, T., et al., *The Tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus.* Cell, 2010. **140**(5): p. 692-703.
141. Kurth, H.M. and K. Mochizuki, *2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena.* RNA, 2009. **15**(4): p. 675-85.
142. Horwich, M.D., et al., *The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC.* Curr Biol, 2007. **17**(14): p. 1265-72.

143. Saito, K., et al., *Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends.* Genes Dev, 2007. **21**(13): p. 1603-8.
144. Yu, B., et al., *Methylation as a crucial step in plant microRNA biogenesis.* Science, 2005. **307**(5711): p. 932-5.
145. Kirino, Y. and Z. Mourelatos, *The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs.* RNA, 2007. **13**(9): p. 1397-401.
146. Kirino, Y., et al., *Arginine methylation of vasa protein is conserved across phyla.* J Biol Chem, 2010. **285**(11): p. 8148-54.
147. Kuramochi-Miyagawa, S., et al., *Mili, a mammalian member of piwi family gene, is essential for spermatogenesis.* Development, 2004. **131**(4): p. 839-49.
148. Thomson, T., et al., *Isolation of new polar granule components in Drosophila reveals P body and ER associated proteins.* Mech Dev, 2008. **125**(9-10): p. 865-73.
149. Arkov, A.L. and A. Ramos, *Building RNA-protein granules: insight from the germline.* Trends Cell Biol, 2010. **20**(8): p. 482-90.
150. Bednenko, J., et al., *Two GW repeat proteins interact with Tetrahymena thermophila argonaute and promote genome rearrangement.* Mol Cell Biol, 2009. **29**(18): p. 5020-30.
151. He, X.J., et al., *An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein.* Cell, 2009. **137**(3): p. 498-508.
152. Chekulaeva, M., R. Parker, and W. Filipowicz, *The GW/WG repeats of Drosophila GW182 function as effector motifs for miRNA-mediated repression.* Nucleic Acids Res, 2010.
153. Behm-Ansmant, I., et al., *mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes.* Genes Dev, 2006. **20**(14): p. 1885-98.
154. Chalker, D.L., P. Fuller, and M.C. Yao, *Communication between parental and developing genomes during tetrahymena nuclear differentiation is likely mediated by homologous RNAs.* Genetics, 2005. **169**(1): p. 149-60.
155. Chalker, D.L. and M.C. Yao, *Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of Tetrahymena thermophila with germ line-limited DNA.* Mol Cell Biol, 1996. **16**(7): p. 3658-67.
156. Liu, Y., K. Mochizuki, and M.A. Gorovsky, *Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena.* Proc Natl Acad Sci U S A, 2004. **101**(6): p. 1679-84.
157. Onodera, Y., et al., *Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation.* Cell, 2005. **120**(5): p. 613-22.
158. Kanno, T., et al., *Atypical RNA polymerase subunits required for RNA-directed DNA methylation.* Nat Genet, 2005. **37**(7): p. 761-5.
159. Liu, Y., et al., *RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena.* Genes Dev, 2007. **21**(12): p. 1530-45.
160. Madireddi, M.T., et al., *Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena.* Cell, 1996. **87**(1): p. 75-84.

161. Coyne, R., et al., *Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation.* Mol Cell, 1999. **4**(5): p. 865-872.
162. Nikiforov, M.A., M.A. Gorovsky, and C.D. Allis, *A novel chromodomain protein, pdd3p, associates with internal eliminated sequences during macronuclear development in Tetrahymena thermophila.* Mol Cell Biol, 2000. **20**(11): p. 4128-34.
163. Rexer, C.H. and D.L. Chalker, *Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in Tetrahymena thermophila.* Eukaryot Cell, 2007. **6**(8): p. 1320-9.
164. Yao, M.C., et al., *Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in Tetrahymena.* J Cell Sci, 2007. **120**(Pt 12): p. 1978-89.
165. Cam, H.P., et al., *Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome.* Nat Genet, 2005. **37**(8): p. 809-19.
166. Taverna, S.D., et al., *RNAi-dependent Recruitment of a Polycomb Repressive Complex for Developmentally Regulated Heterochromatin Formation in Tetrahymena (Unpublised Data).*
167. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**(6796): p. 593-9.
168. Ivanova, A.V., et al., *The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast.* Nat Genet, 1998. **19**(2): p. 192-195.
169. Jackson, J.P., et al., *Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase.* Nature, 2002. **416**(6880): p. 556-60.
170. Goodrich, J., et al., *A Polycomb-group gene regulates homeotic gene expression in Arabidopsis.* Nature, 1997. **386**(6620): p. 44-51.
171. Fong, Y., et al., *Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of C. elegans.* Science, 2002. **296**(5576): p. 2235-8.
172. van der Lugt, N.M., et al., *Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene.* Genes Dev, 1994. **8**(7): p. 757-69.
173. Cao, R., Y. Tsukada, and Y. Zhang, *Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing.* Mol Cell, 2005. **20**(6): p. 845-54.
174. Wang, H., et al., *Role of histone H2A ubiquitination in Polycomb silencing.* Nature, 2004. **431**(7010): p. 873-8.
175. Lagarou, A., et al., *dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing.* Genes Dev, 2008. **22**(20): p. 2799-810.
176. Gearhart, M.D., et al., *Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets.* Mol Cell Biol, 2006. **26**(18): p. 6880-9.
177. Sanchez, C., et al., *Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor.* Mol Cell Proteomics, 2007. **6**(5): p. 820-34.

178. Simon, J.A. and R.E. Kingston, *Mechanisms of polycomb gene silencing: knowns and unknowns.* Nat Rev Mol Cell Biol, 2009. **10**(10): p. 697-708.
179. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing.* Science. Online, 2002. **298**(5595): p. 1039-43.
180. Czermin, B., et al., *Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites.* Cell, 2002. **111**(2): p. 185-96.
181. Muller, J., et al., *Histone methyltransferase activity of a Drosophila Polycomb group repressor complex.* Cell, 2002. **111**(2): p. 197-208.
182. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.* Genes Dev, 2002. **16**(22): p. 2893-905.
183. Bannister, A.J., et al., *Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain.* Nature, 2001. **410**(6824): p. 120-4.
184. Ekwall, K., et al., *The chromodomain protein Swi6: a key component at fission yeast centromeres.* Science, 1995. **269**(5229): p. 1429-31.
185. Lorentz, A., et al., *Switching gene swi6, involved in repression of silent mating-type loci in fission yeast, encodes a homologue of chromatin-associated proteins from Drosophila and mammals.* Gene, 1994. **143**(1): p. 139-43.
186. Madireddi, M.T., M.C. Davis, and C.D. Allis, *Identification of a novel polypeptide involved in the formation of DNA-containing vesicles during macronuclear development in Tetrahymena.* Dev Biol, 1994. **165**(2): p. 418-31.
187. Smothers, J.F., et al., *Pdd1p associates with germline-restricted chromatin and a second novel anlagen-enriched protein in developmentally programmed DNA elimination structures.* Development, 1997. **124**(22): p. 4537-45.
188. Yao, M.C., et al., *DNA elimination in Tetrahymena: a developmental process involving extensive breakage and rejoining of DNA at defined sites.* Cell, 1984. **36**(2): p. 433-440.
189. Eissenberg, J.C., et al., *Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster.* Proc Natl Acad Sci U S A, 1990. **87**(24): p. 9923-7.
190. Smothers, J.F., et al., *Pdd1p associates with germline-restricted chromatin and a second novel anlagen-enriched protein in developmentally programmed DNA elimination structures.* Development, 1997. **124**: p. 4537-4545.
191. Janetopoulos, C., et al., *The conjusome: a novel structure in Tetrahymena found only during sexual reorganization.* J Cell Sci, 1999. **112 (Pt 7)**: p. 1003-11.
192. Coyne, R.S., et al., *Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation.* Mol Cell, 1999. **4**(5): p. 865-72.
193. Nikiforov, M., et al., *Excision of micronuclear-specific DNA requires parental expression of Pdd2p and occurs independently from DNA replication in Tetrahymena thermophila.* Genes Dev, 1999. **13**(21): p. 2852-2862.
194. Shieh, A.W.-Y., et al., *Unpublished Data.*
195. Agrawal, A., Q.M. Eastman, and D.G. Schatz, *Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system.* Nature, 1998. **394**(6695): p. 744-51.

196. Hiom, K., M. Melek, and M. Gellert, *DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations.* Cell, 1998. **94**(4): p. 463-70.
197. Cary, L.C., et al., *Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses.* Virology, 1989. **172**(1): p. 156-69.
198. Saveliev, S.V. and M.M. Cox, *Developmentally programmed DNA deletion in Tetrahymena thermophila by a transposition-like reaction pathway.* EMBO J., 1996. **15**(11): p. 2858-2869.
199. Doerder, F.P. and L.E. Debault, *Cytofluorimetric analysis of nuclear DNA during meiosis, fertilization and macronuclear development in the ciliate Tetrahymena pyriformis, syngen 1.* J Cell Sci, 1975. **17**(3): p. 471-93.
200. Yao, M.C., *Ribosomal RNA gene amplification in Tetrahymena may be associated with chromosome breakage and DNA elimination.* Cell, 1981. **24**(3): p. 765-774.
201. Yao, M.C. and C.H. Yao, *Repeated hexanucleotide C-C-C-C-A-A is present near free ends of macronuclear DNA of Tetrahymena.* Proc Natl Acad Sci USA, 1981. **78**(12): p. 7436-7439.
202. Yao, M.C., A.R. Kimmel, and M.A. Gorovsky, *A small number of cistrons for ribosomal RNA in the germinal nucleus of a eukaryote, Tetrahymena pyriformis.* Proc Natl Acad Sci USA, 1974. **71**(8): p. 3082-3086.
203. Yao, M.C., K. Zheng, and C.H. Yao, *A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in Tetrahymena.* Cell, 1987. **48**(5): p. 779-788.
204. Fan, Q. and M.-C. Yao, *New telomere formation coupled with site-specific chromosome breakage in Tetrahymena thermophila.* Mol. Cell. Biol., 1996. **16**(3): p. 1267-1274.
205. Kapler, G.M. and E.H. Blackburn, *A weak germ-line excision mutation blocks developmentally controlled amplification of the rDNA minichromosome of Tetrahymena thermophila.* Genes Dev, 1994. **8**(1): p. 84-95.
206. Fan, Q. and M.C. Yao, *A long stringent sequence signal for programmed chromosome breakage in Tetrahymena thermophila.* Nucleic Acids Research, 2000. **28**(4): p. 895-900.
207. Swanton, M.T., J.M. Heumann, and D.M. Prescott, *Gene-sized DNA molecules of the macronuclei in three species of hypotrichs: size distributions and absence of nicks. DNA of ciliated protozoa. VIII.* Chromosoma, 1980. **77**(2): p. 217-27.
208. Swanton, M.T., A.F. Greslin, and D.M. Prescott, *Arrangement of coding and non-coding sequences in the DNA molecules coding for rRNAs in Oxytricha sp. DNA of ciliated protozoa. VII.* Chromosoma, 1980. **77**(2): p. 203-15.
209. Wunning, I.U. and H.J. Lipps, *A transformation system for the hypotrichous ciliate Stylonychia mytilus.* EMBO J, 1983. **2**(10): p. 1753-7.
210. Steinbruck, G., *Over-amplification of genes in macronuclei of hypotrichous ciliates.* 1983. **88**: p. 156-163.
211. Greslin, A.F., et al., *An analysis of the macronuclear actin genes of Oxytricha.* Dna, 1988. **7**(8): p. 529-36.

212. Greslin, A.F., et al., *Reordering of nine exons is necessary to form a functional actin gene in Oxytricha nova.* Proc Natl Acad Sci USA, 1989. **86**(16): p. 6264-6268.
213. Mitcham, J.L., A.J. Lynn, and D.M. Prescott, *Analysis of a scrambled gene: the gene encoding alpha-telomere-binding protein in Oxytricha nova.* Genes Dev, 1992. **6**(5): p. 788-800.
214. Prescott, D.M. and A.F. Greslin, *Scrambled actin I gene in the micronucleus of Oxytricha nova.* Dev Genet, 1992. **13**(1): p. 66-74.
215. DuBois, M. and D.M. Prescott, *Scrambling of the actin I gene in two Oxytricha species.* Proc Natl Acad Sci USA, 1995. **92**(9): p. 3888-3892.
216. Hoffman, D.C. and D.M. Prescott, *Evolution of internal eliminated segments and scrambling in the micronuclear gene encoding DNA polymerase alpha in two Oxytricha species.* Nucleic Acids Res, 1997. **25**(10): p. 1883-9.
217. Prescott, J.D., M.L. DuBois, and D.M. Prescott, *Evolution of the scrambled germline gene encoding alpha-telomere binding protein in three hypotrichous ciliates.* Chromosoma, 1998. **107**(5): p. 293-303.
218. Landweber, L.F., T.-C. Kuo, and E.A. Curtis, *Evolution and assembly of an extremely scrambled gene.* Proc Natl Acad Sci USA, 2000. **97**(7): p. 3298-3303.
219. Doak, T.G., et al., *A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif.* Proc Natl Acad Sci U S A, 1994. **91**(3): p. 942-6.
220. Hunter, D.J., et al., *Precise excision of telomere-bearing transposons during Oxytricha fallax macronuclear development.* Genes Dev, 1989. **3**(12B): p. 2101-12.
221. Juranek, S.A., et al., *snRNA and heterochromatin formation are involved in DNA excision during macronuclear development in stichotrichous ciliates.* Eukaryot Cell, 2005. **4**(11): p. 1934-41.
222. Bernstein, E., et al., *Role for a bidentate ribonuclease in the initiation step of RNA interference.* Nature, 2001. **409**(6818): p. 363-6.
223. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
224. Filippov, V., et al., *A novel type of RNase III family proteins in eukaryotes.* Gene, 2000. **245**(1): p. 213-21.
225. Tabara, H., et al., *The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans.* Cell, 2002. **109**(7): p. 861-71.
226. Liu, Q., et al., *R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway.* Science, 2003. **301**(5641): p. 1921-5.
227. Forstemann, K., et al., *Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein.* PLoS Biol, 2005. **3**(7): p. e236.
228. Saito, K., et al., *Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells.* PLoS Biol, 2005. **3**(7): p. e235.
229. Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex.* Nature, 2004. **432**(7014): p. 231-5.

230. Chendrimada, T.P., et al., *TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing.* Nature, 2005. **436**(7051): p. 740-4.
231. Gregory, R.I., et al., *The Microprocessor complex mediates the genesis of microRNAs.* Nature, 2004. **432**(7014): p. 235-40.
232. Doyle, M. and M.F. Jantsch, *New and old roles of the double-stranded RNA-binding domain.* Journal of structural biology, 2002. **140**(1-3): p. 147-53.
233. Tian, B., et al., *The double-stranded-RNA-binding motif: interference and much more.* Nature reviews. Molecular cell biology, 2004. **5**(12): p. 1013-23.
234. St Johnston, D., D. Beuchle, and C. Nusslein-Volhard, *Staufen, a gene required to localize maternal RNAs in the Drosophila egg.* Cell, 1991. **66**(1): p. 51-63.
235. St Johnston, D., et al., *A conserved double-stranded RNA-binding domain.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(22): p. 10979-83.
236. Meurs, E., et al., *Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon.* Cell, 1990. **62**(2): p. 379-90.
237. Kim, U., et al., *Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(24): p. 11457-61.
238. Green, S.R. and M.B. Mathews, *Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI.* Genes & development, 1992. **6**(12B): p. 2478-90.
239. Hundley, H.A. and B.L. Bass, *ADAR editing in double-stranded UTRs and other noncoding RNA sequences.* Trends in biochemical sciences, 2010. **35**(7): p. 377-83.
240. Lai, F., R. Drakas, and K. Nishikura, *Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts.* The Journal of biological chemistry, 1995. **270**(29): p. 17098-105.
241. Liu, Y., et al., *Functionally distinct double-stranded RNA-binding domains associated with alternative splice site variants of the interferon-inducible double-stranded RNA-specific adenosine deaminase.* The Journal of biological chemistry, 1997. **272**(7): p. 4419-28.
242. Pindel, A. and A. Sadler, *The role of protein kinase R in the interferon response.* Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research, 2011. **31**(1): p. 59-70.
243. Xu, Z. and B.R. Williams, *The B56alpha regulatory subunit of protein phosphatase 2A is a target for regulation by double-stranded RNA-dependent protein kinase PKR.* Molecular and cellular biology, 2000. **20**(14): p. 5285-99.
244. McCormack, S.J., et al., *Mechanism of interferon action motif I of the interferon-induced, RNA-dependent protein kinase (PKR) is sufficient to mediate RNA-binding activity.* Virology, 1994. **198**(1): p. 92-9.
245. McMillan, N.A., et al., *Mutational analysis of the double-stranded RNA (dsRNA) binding domain of the dsRNA-activated protein kinase, PKR.* The Journal of biological chemistry, 1995. **270**(6): p. 2601-6.

246. Saunders, L.R., et al., *Characterization of two evolutionarily conserved, alternatively spliced nuclear phosphoproteins, NFAR-1 and -2, that function in mRNA processing and interact with the double-stranded RNA-dependent protein kinase, PKR.* The Journal of biological chemistry, 2001. **276**(34): p. 32300-12.
247. Wickham, L., et al., *Mammalian staufen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum.* Molecular and cellular biology, 1999. **19**(3): p. 2220-30.
248. Marion, R.M., et al., *A human sequence homologue of Staufen is an RNA-binding protein that is associated with polysomes and localizes to the rough endoplasmic reticulum.* Molecular and cellular biology, 1999. **19**(3): p. 2212-9.
249. Micklem, D.R., et al., *Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation.* The EMBO journal, 2000. **19**(6): p. 1366-77.
250. Wells, J.M., et al., *A small family of elements with long inverted repeats is located near sites of developmentally regulated DNA rearrangement in Tetrahymena thermophila.* Mol Cell Biol, 1994. **14**(9): p. 5939-5949.
251. Kowalczyk, C.A., et al., *The germ line limited M element of Tetrahymena is targeted for elimination from the somatic genome by a homology-dependent mechanism.* Nucleic Acids Res, 2006.
252. Godiska, R. and M.C. Yao, *A programmed site-specific DNA rearrangement in Tetrahymena thermophila requires flanking polypurine tracts.* Cell, 1990. **61**(7): p. 1237-46.
253. Li, J. and R.E. Pearlman, *Programmed DNA rearrangement from an intron during nuclear development in Tetrahymena thermophila: molecular analysis and identification of potential cis-acting sequences.* Nucleic Acids Res., 1996. **24**(10): p. 1943-1949.

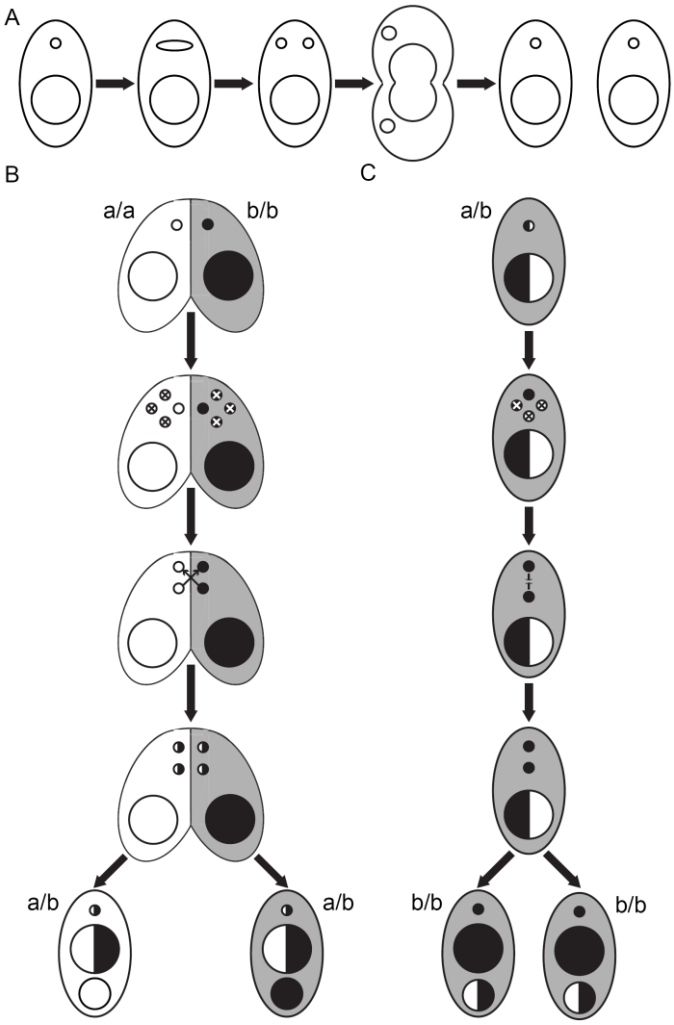


Figure 1. Ciliate life cycle. A. Vegetative Growth. Vegetative growth proceeds in ciliates by closed mitosis of the micronucleus, followed by amitotic division of the macronucleus during cytokinesis to produce two clonal cells.
B. Conjugation. Ciliate conjugation begins by pairing of cells with different mating types. Their micronuclei (○ and●, color denotes linage) then undergo meiosis, during which all but one of the haploid meiotic products in each cell disintegrates. The remaining haploid gamete undergoes one round of mitosis to generate two identical haploid gametes, one of which migrates to the conjugating partner to fuse with the remaining gamete. Such cross-fertilization produces identical heterozygous zygotic nuclei (◐) in each cell, which replicate and differentiate into the micro- and the macronuclei of the progeny, while the old macronucleus is eventually lost (● and ○, color denotes linage). When there is little or no cytoplasmic exchange between mating partners, as is the case in *Paramecium*, nuclei with identical zygotic genomes differentiate in the cytoplasm of their parent cells.
C. Autogamy. Nuclear events of autogamy occur similarly to conjugation, except that instead of cross-fertilization, the mitotic products of the selected gamete fuse with each other to produce a homozygous zygotic nucleus.

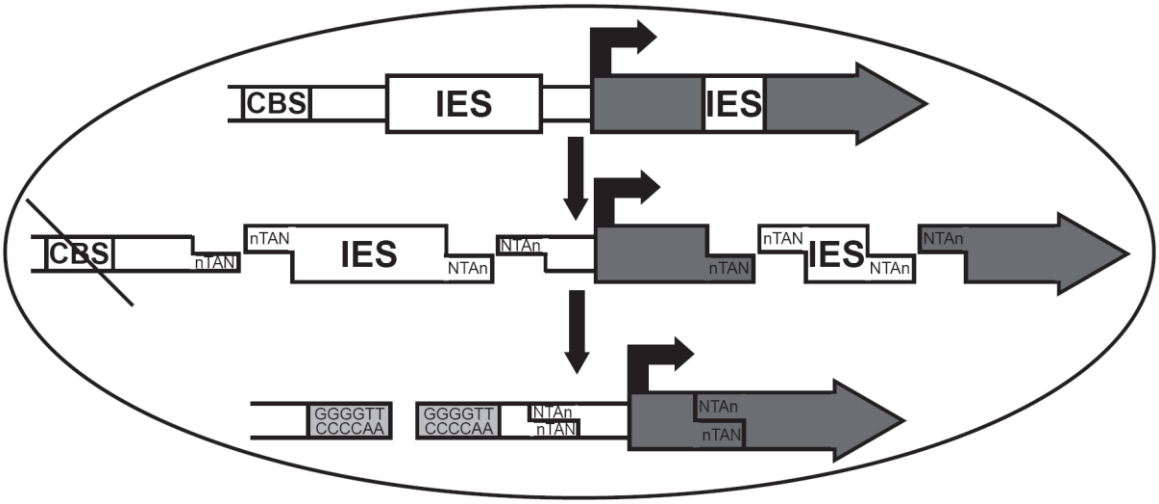


Figure 2. Somatic genome rearrangements of ciliates through chromosome fragmentation and DNA elimination. During sexual reproduction the developing, somatic macronucleus of ciliates undergoes chromosome fragmentation and elimination of a variety of intra- and intergenic sequences. Chromosome fragmentation can occur through a variety of mechanisms but for simplicity is shown here by the chromosome breakage sequence (CBS) found in *T. thermophila*. As sexual reproduction proceeds zygotic, somatic chromosomes in the developing macronucleus undergo double-stranded DNA breaks. Telomere addition to double-stranded DNA breaks at CBSs results in chromosome fragmentation generating hundreds to up to tens of thousands of macronuclear chromosomes. Other double-stranded DNA breaks bound by tandem, direct 5' TA 3' repeats called internal eliminated sequences (IESs) are repaired such that one of the repeats and the intervening sequence are lost. The resulting somatic macronuclear genome contains many-fold more chromosomes and also has a significant percentage of its DNA removed when compared to the germline micronucleus.

Table 1. Histone modifications found in the nuclei of ciliates

Nucleus	Histone Composition	Histone Modifications[a]
Micronucleus	H2A, H2B, H3, H4, micH1	H3K27me, H3S10ph, micH1ph
Macronucleus	H2A, H2B, H3, H4, hv1, hv2, macH1	H3K4me, H3K9me[b], H3K27me, H2Aac, H2Bac, H3ac, H4ac, H2Aph, macH1ph

Table 1. Histone modifications found in the nuclei of ciliates. The histone composition and modifications of the micro- and macronucleus are listed above. Most of these histones and modifications are found throughout the life cycle of *T. thermophila* but one (H3K9me) is restricted to developing macronuclei during conjugation.

a. There is no distinction between mono-, di- and tri-modifications

b. This modification is only found during sexual reproduction in the developing zygotic macronucleus

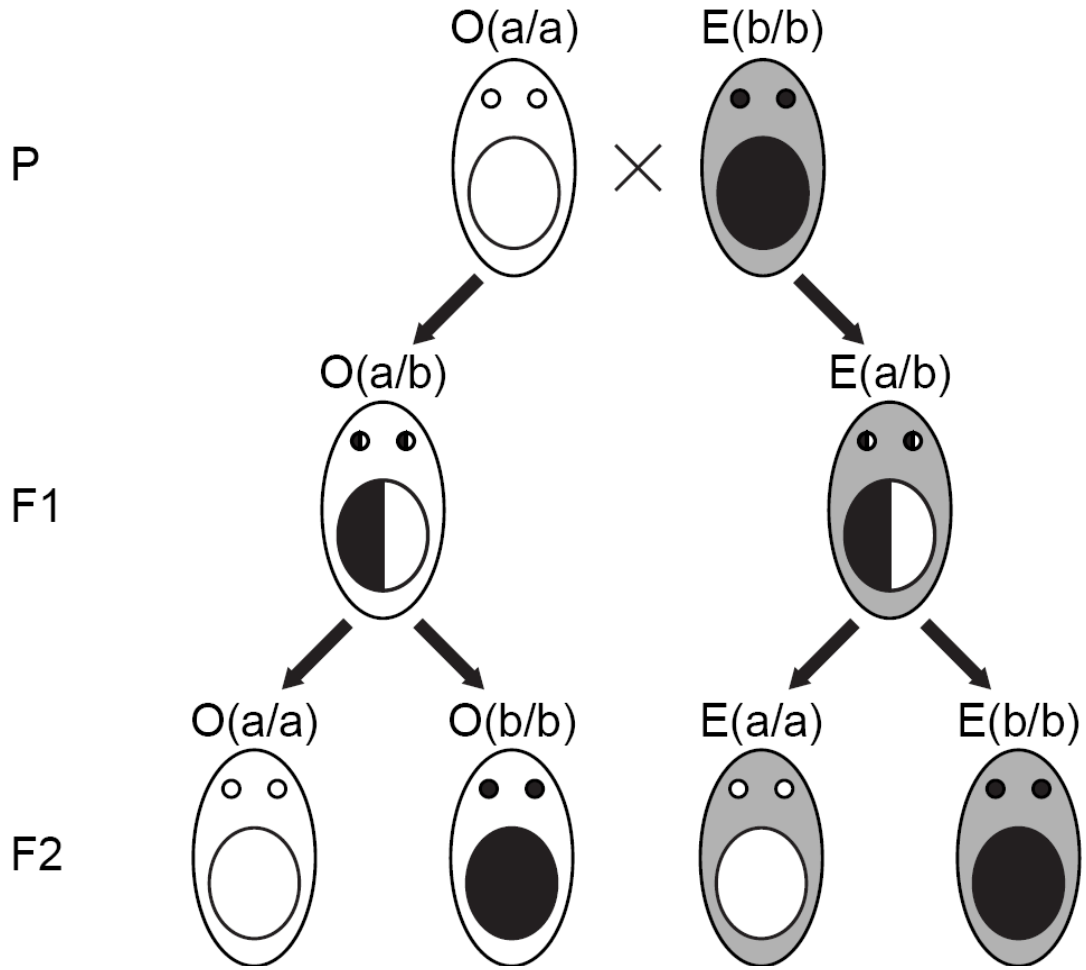


Figure 3. Cytoplasmic inheritance – mating type determination in *Paramecium*. Mating scheme between mating type E and O in *Paramecium* of different genotypes is illustrated. Conjugation between parental cells (P) that are homozygous for genetic allele "a" or "b" produces genetically identical, heterozygous progeny (F1) with the (a/b) genotype. Mating type of the progeny, however, is determined by the mating type of the parental cell, suggesting a mode of inheritance that is independent of the genes inherited from the parents. Likewise, when the conjugants undergo autogamy producing the F2 generation, regardless of 1:1 segregation of the "a" and "b" alleles, progeny mating type consistently follows the mating type of the parental cell (F1).

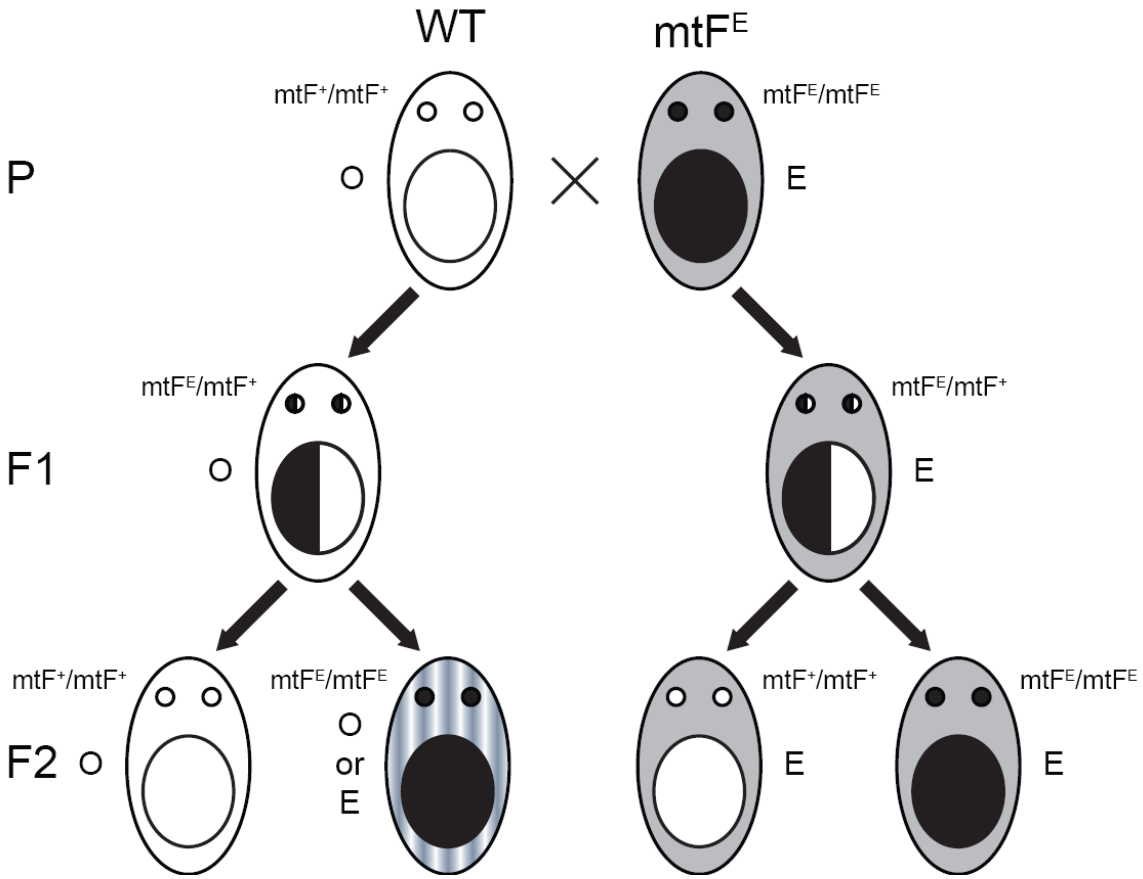


Figure 4. Epigenetic inheritance of mating type in *Paramecium* demonstrated by the mutation mtF^{E}. Matings between a wild-type O strain (mtF^{+}/mtF^{+}) and a mtF^{E} mutant E strain (mtF^{E}/ mtF^{E}) produces O and E strain heterozygotes (F1) (mtF^{+}/ mtF^{E}). Autogamy of the heterozygous E strain generates E strains (F2) that are either homozygous wild-type (mtF^{+}/ mtF^{+}) or homozygous mtF^{E} mutant (mtF^{E}/ mtF^{E}). However, autogamy of the heterozygous O strain generates an O strain (F2) that is homozygous wild-type (mtF^{+}/ mtF^{+}) and another strain (F2) that is typically E that is homozygous mtF^{E} mutant (mtF^{E}/ mtF^{E}) (Compare with figure 3). The last mutant strain has been shown to retain an internal eliminated sequence (IES) in the G surface antigen leading to the hypothesis that mtF is involved in the epigenetic inheritance of mating type in *Paramecium*.

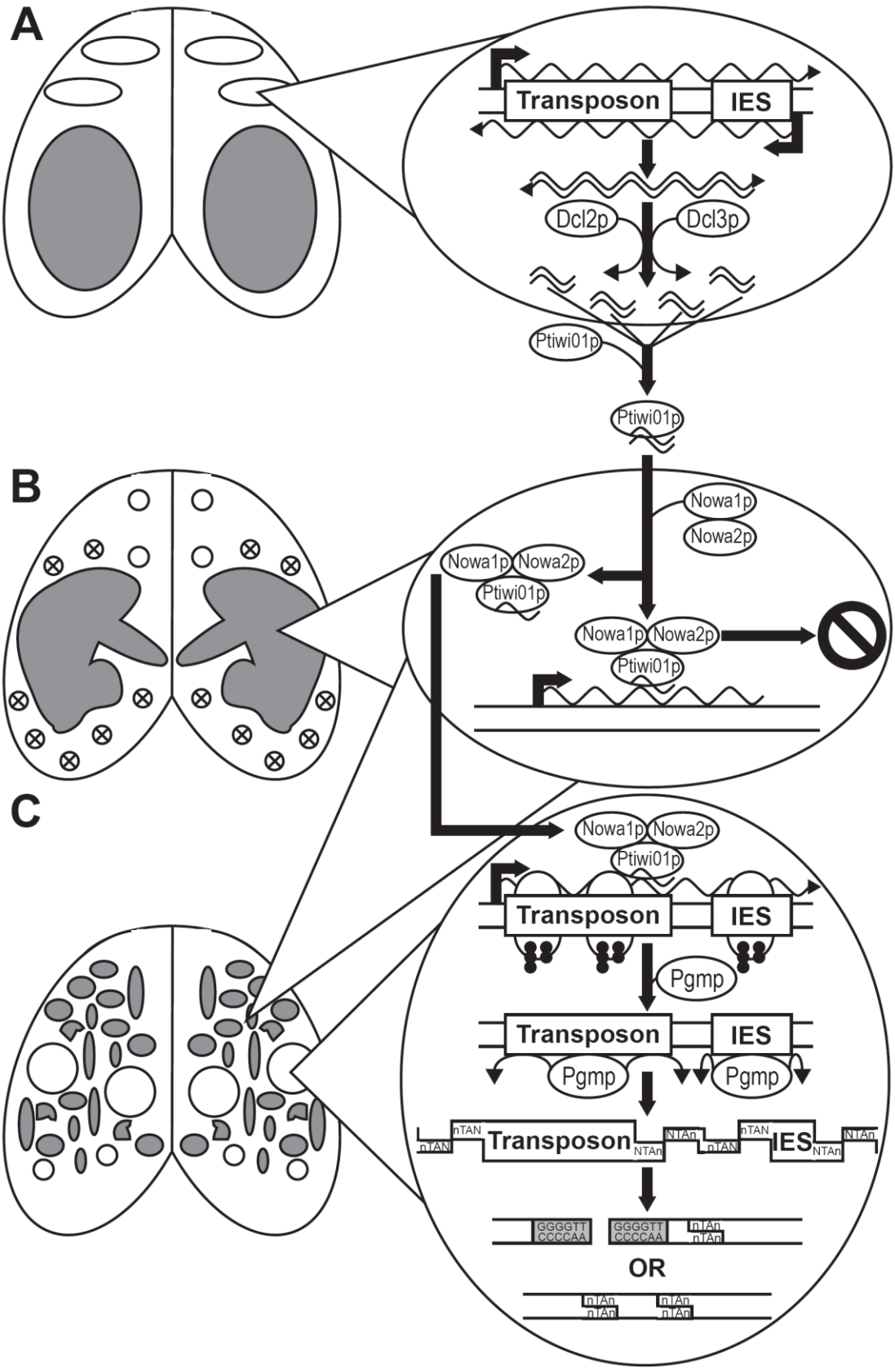

A
Transposon
IES
Dcl2p
Dcl3p
Ptiwi01p
Ptiwi01p
B
Nowa1p
Nowa2p
Nowa1p
Nowa2p
Ptiwi01p
Nowa1p
Nowa2p
Ptiwi01p
C
Nowa1p
Nowa2p
Ptiwi01p
Transposon
IES
Pgmp
Transposon
IES
Pgmp
Pgmp
nTAN
nTAn
Transposon
IES
GGGGTT
CCCCAA
GGGGTT
CCCCAA
OR

Figure 5. RNAi-dependent DNA elimination in *P. tetraurelia*.
A. In the elongated micronuclei (white ovals) early in conjugation (or autogamy) bidirectional transcription of repetitive sequences and internal eliminated sequences (IESs) produce long, dsRNA. Cleavage of the long, dsRNAs by Dicer-like proteins 2 and 3 (Dcl2p and Dcl3p) generates approximately 25 nt scan RNAs (scnRNAs). scnRNAs are transported into the cytoplasm where they are bound by Piwi homologues Ptiwi01 or Ptiwi09 (not shown here) and are transported to the parental macronucleus (gray oval).
B. After import of the Ptiwi01/scnRNA complexes into the parental macronucleus (amorphous gray blob), the Ptiwi01/scnRNA complexes are bound by the RNA-binding GW repeat proteins Nowa1p and Nowa2p. Non-coding RNA (ncRNA) produced throughout the parental macronucleus facilitates genome scanning of the scnRNA population. Those Ptiwi01/scnRNA complexes found to bind the ncRNA are sequestered, while those unable to bind are transported to the developing macronuclei later in conjugation.
C. The remaining Ptiwi01/scnRNA complexes are transported from the parental macronucleus (assorted gray shapes) into the developing macronuclei (large, white circles) along with Nowa1p and Nowa2p. Binding of the Ptiwi01/scnRNA complexes with ncRNA produced in the developing mac allows recruitment of the domesticated PiggyBac transposase, Pgmp. Pgmp creates concomitant, DNA double-stranded breaks (DSBs) centered on a conserved TA dinucleotide of the sequences to be eliminated generating 5’ four basepair overhangs. Ligation of the 5’ overhangs of the macronuclear destined sequences (MDSs) retains one of the TA dinucleotide repeats. Unlike IESs, which are always precisely excised and ligated, repetitive sequences can either be excised and their flanking MDSs ligated or the DSB can be repaired by telomere addition (gray rectangles).

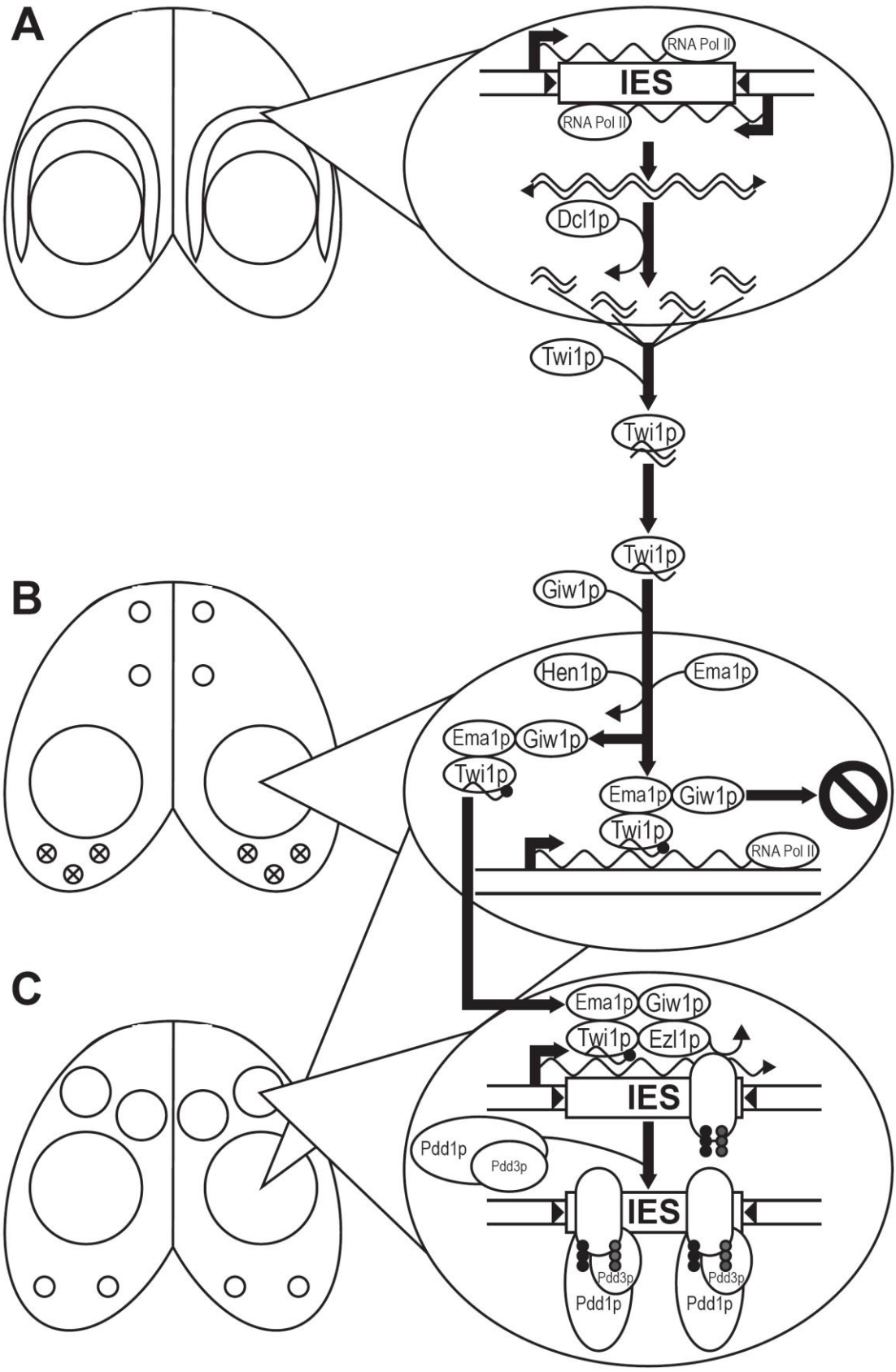

A
RNA Pol II
IES
RNA Pol II
Dcl1p
Twi1p
Twi1p
Twi1p
Giw1p
B
Hen1p
Ema1p
Ema1p
Giw1p
Twi1p
Ema1p
Giw1p
Twi1p
RNA Pol II
C
Ema1p
Giw1p
Twi1p
Ezl1p
IES
Pdd1p
Pdd3p
IES
Pdd3p
Pdd3p
Pdd1p
Pdd1p

Figure 6. RNAi and heterochromatin components of RNAi-dependent DNA elimination in *T. thermophila*. A. Bidirectional transcription of internal eliminated sequences (IESs) in the crescent micronucleus (elongated white oval) early in conjugation by RNA polymerase II produces long, dsRNAs. Cleavage of these long, dsRNAs by Dicer-like protein 1 (Dcl1p) produces scan RNAs (scnRNAs) of approximately 28 nt. scnRNAs are transported into the cytoplasm, where they are bound by the PIWI homologue, Twi1p. Twi1p/scnRNA complexes are activated by cleavage of the double-stranded scnRNAs and unwinding of the scnRNA passenger strand by Twi1p. Activated Twi1p/scnRNA complexes are transported into the parental macronucleus (large, white circle) by Giw1p.
B. In the parental macronucleus (large, white circle) activated scnRNAs are stabilized by 2'O methylation by the HEN1 homologue, Hen1p. Ema1p binds to activated Giw1p/Twi1p/scnRNA complexes and facilitates complex binding with non-coding RNA (ncRNA) transcribed from the parental macronucleus for genome scanning. Ema1p/Giw1p/Twi1p/scnRNA complexes that bind the ncRNA are removed from the general scnRNA population, while those not found to match are transported to the developing mac later in conjugation.
C. The remaining Ema1p/Giw1p/Twi1p/scnRNA complexes are transported from the parental macronucleus (large, white circle) to the developing macronuclei (medium, white circles). ncRNA produced in the developing mac assists in Ema1p/Giw1p/Twi1p/scnRNA complex binding, which triggers binding of the E(Z) homologue, Ezl1p. H3K9 (gray circles) and H3K27 methylation (black circles) by Ezl1p allows histone binding by the chromodomain proteins, Pdd1p and Pdd3p.

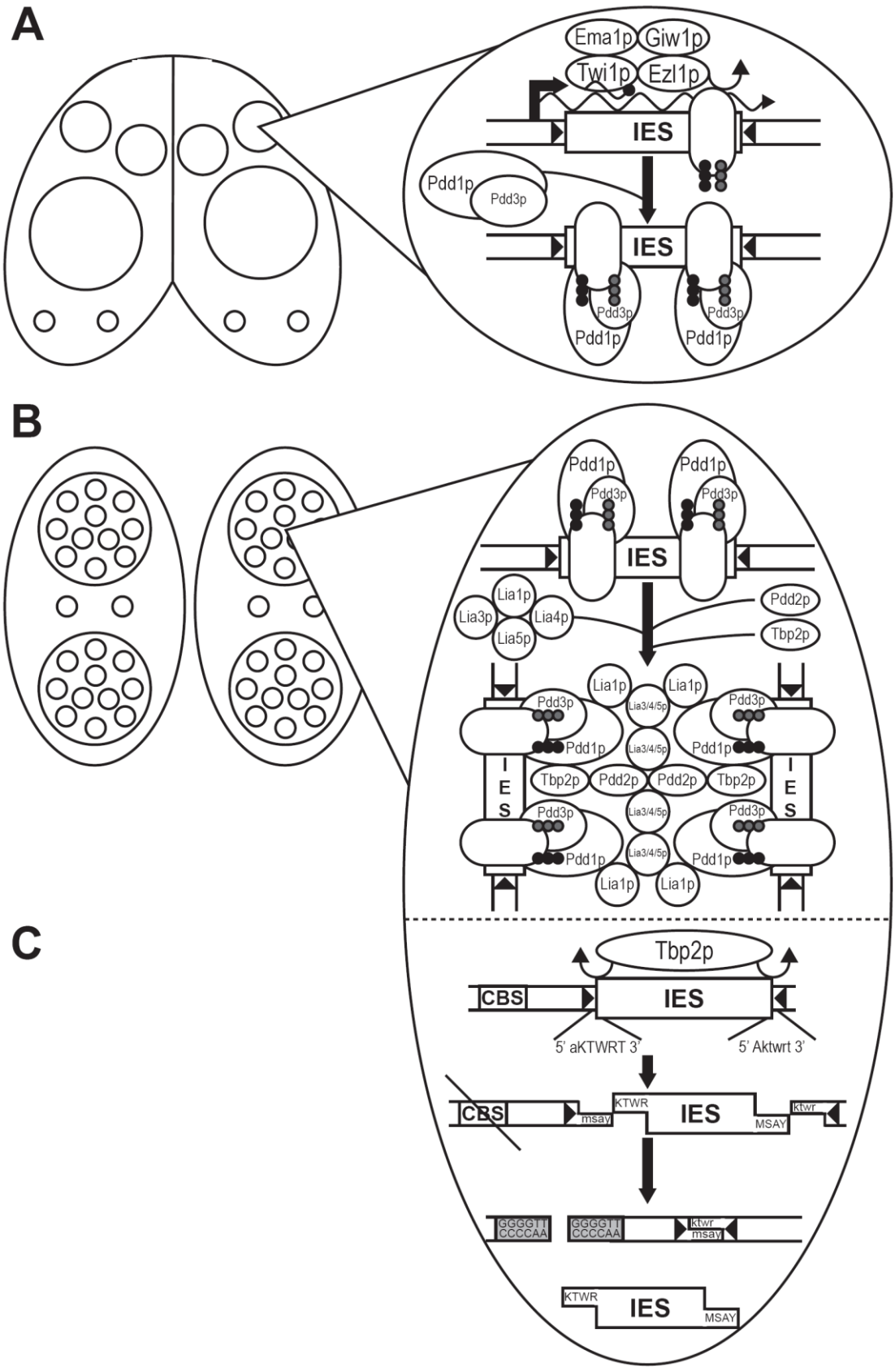

A
Ema1p
Giw1p
Twi1p
Ezl1p
IES
Pdd1p
Pdd3p
IES
Pdd3p
Pdd3p
Pdd1p
Pdd1p
B
Pdd1p
Pdd1p
Pdd3p
Pdd3p
IES
Lia1p
Lia3p
Lia4p
Lia5p
Pdd2p
Tbp2p
Lia1p
Lia1p
Lia3/4/5p
Pdd3p
Pdd3p
Pdd1p
Lia3/4/5p
Pdd1p
I
E
S
Tbp2p
Pdd2p
Pdd2p
Tbp2p
I
E
S
Pdd3p
Pdd3p
Lia3/4/5p
Pdd1p
Lia3/4/5p
Pdd1p
Lia1p
Lia1p
C
Tbp2p
CBS
IES
5' aKTWRT 3'
5' Aktwrt 3'
CBS
msay
KTWR
IES
MSAY
ktwr
GGGGTT
CCCCAA
GGGGTT
CCCCAA
ktwr
msay
KTWR
IES
MSAY

Figure 7. Histone methylation triggers DNA elimination of internal eliminated sequences (IESs) in *T. thermophila*. A. Ema1p/Giw1p/Twi1p/scnRNA complexes in the developing macronuclei (medium white circles) bind ncRNA produced in the developing macronuclei. Ezl1p, an E(Z) homologue, binds the Ema1p/Giw1p/Twi1p/scnRNA complex and methylates H3K9 (gray circles) and H3K27 (black circles) of histones associated with IESs. H3K9 and H3K27 methylation by Ezl1p allows histone binding by the chromodomain proteins, Pdd1p and Pdd3p.
B. Later in conjugation Pdd1p and Pdd3p, along with the proteins Pdd2p, Lia1p, Lia3p, Lia4p, Lia5p and Tbp2p, a domesticated PiggyBac transposase, form protein foci called DNA elimination bodies in the developing mac (small, white circles in large, white circles).
C. Concomitant DNA double-stranded breaks (DSBs) catalyzed by Tbp2p centered on an AT-rich dinucleotide of the IES sequences to be eliminated generate 5' four basepair overhangs between the IESs and IES boundary sequences (black triangles). Ligation of the 5' overhangs of the macronuclear destined sequences (MDSs) retains one of the cleavage sequence repeats, releasing a linearized IES fragment. Concurrent with DNA elimination of IESs, developing macronuclei chromosomes are fragmented at conserved chromosome breakage sequences (CBS). DSBs are repaired by addition of the telomere sequences, G_4T_2.

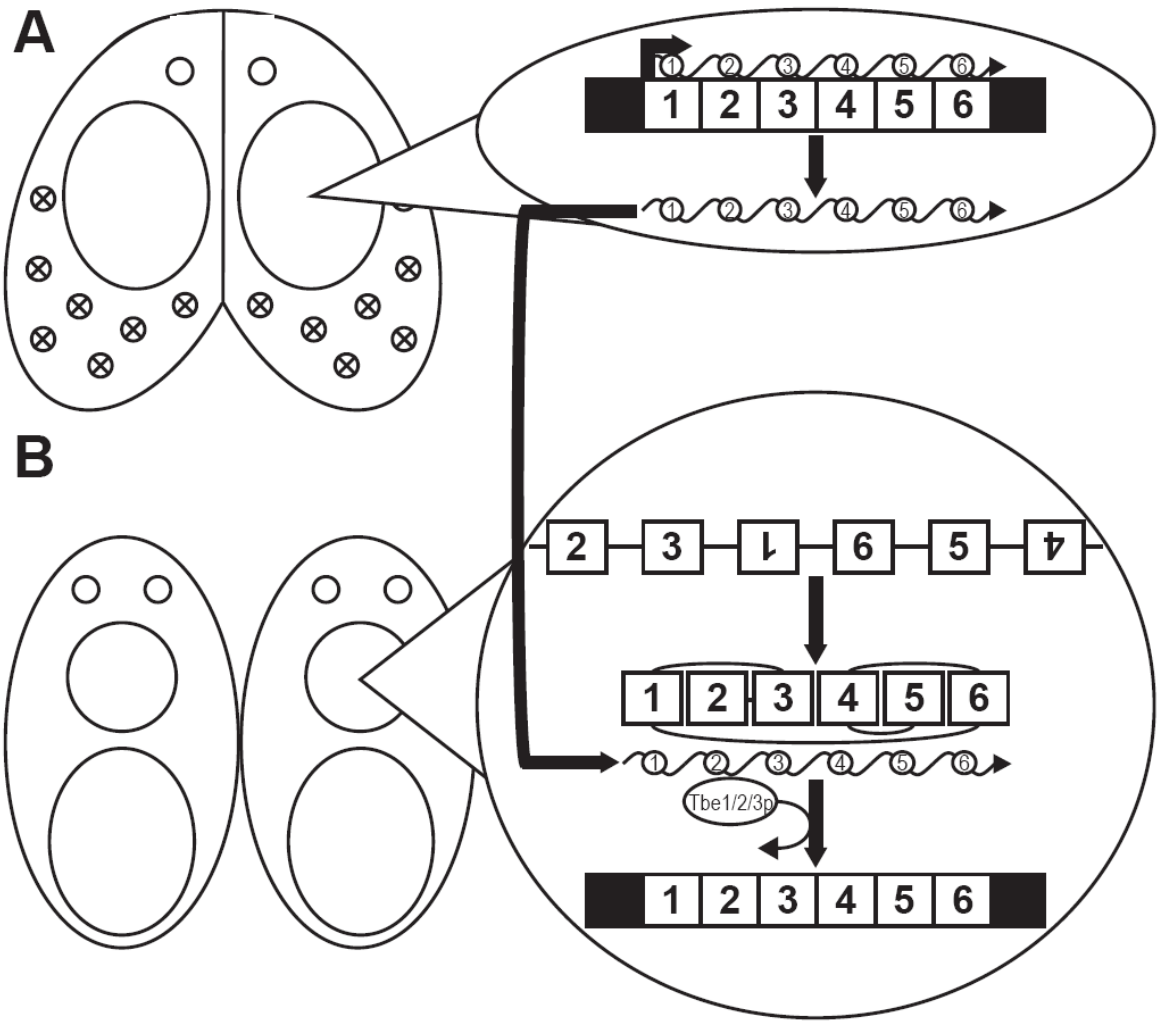


Figure 8. Template-guided gene unscrambling, DNA elimination and chromosome fragmentation in *O. trifallax*.
A. Non-coding RNA (ncRNA) from all mini-chromosomes in the parental macronucleus (oval) is transported to the developing macronucleus later in conjugation.
B. ncRNA from the parental macronucleus triggers rearrangement of scrambled developing macronuclear chromosomes in the developing mac (large circle). Transposases from the TBE1, TBE2 and TBE3 transposon family trigger DNA elimination, chromosome fragmentation and ligation of macronuclear destined sequences (MDSs) to form new mini-chromosomes in the developing mac.

CHAPTER 2

TRANSCRIPTION OF dsRNA FROM THE R IES IS REQUIRED FOR SUBSEQUENT TARGETING AND DNA ELIMINATION LATE IN CONJUGATION IN *TETRAHYMENA THERMOPHILA*

Introduction:

In metazoans variations between cell types typically depend not on differences in genomic content but on chromatin structure and gene expression differences. However, this is not the case in all eukaryotes. Late in the 19th century the cell biologist Theodor Boveri observed during somatic development of the roundworm, *Parascaris univalens*, that a large percentage of chromosomes were lost [1]. Boveri termed this phenomenon chromatin diminution. Over the last hundred years similar examples of large-scale DNA loss have been found in a variety of eukaryotes including sciarid flies, lampreys, copepods, hagfish and ciliated protozoa ([2-4], reviewed in [5, 6]).

In the ciliated protozoan, *Tetrahymena thermophila*, this large-scale DNA loss was first noted when the renaturation kinetics of its two nuclei was analyzed and compared [7]. Like all ciliates, *T. thermophila* contains in each cell two types of nuclei, a germline micronucleus and a somatic macronucleus (reviewed in [5]). Under optimal growth conditions *T. thermophila* reproduce asexually via binary fission. However, mixing two starved strains of opposite mating type induces sexual reproduction or conjugation. During conjugation the parental macronucleus is discarded and new zygotic micro- and macronuclei are generated from the parental micronuclei. Initially the zygotic macronuclei are genomically equivalent to the micronuclei, but as conjugation proceeds the zygotic macronuclei mature and lose approximately 15% of their overall genomic content [7].

In the last decade the mechanism through which this large-scale DNA loss occurs was found to be RNAi-dependent [8]. In the crescent micronucleus early in conjugation, bidirectional transcription produces long, dsRNA from genomic loci targeted for DNA

elimination called **i**nternal **e**liminated **s**equences (IESs) [9]. These long, bidirectional ncRNAs are cleaved into a species of sRNA called scnRNAs by the Dicer homologue, Dcl1p (chapter 3) [10, 11]. In the cytoplasm these scnRNAs are bound by an Argonaute homologue, Twi1p, and are transported to the zygotic macronucleus [8, 12]. There Twi1p/scnRNA complexes target histones associated with IESs for H3K9 and H3K27 methylation by the E(z) homologue, Ezl1p [12-15]. These methylated histones are bound by the chromodomain proteins, Pdd1p and Pdd3p, which triggers removals of IESs through recruitment of the domesticated piggyBac transposase, Tpb2p [13, 15-17].

Analysis of the sequence requirements for IES excision strongly suggests recognition through homologous sRNAs. Deletion analysis of the M IES revealed that no specific DNA segment is required for its elimination [18]. Instead, it was shown that a minimum length of 300 bp of the M IES sequence, as well as intact boundary determinants, were sufficient for proper IES removal. This sequence requirement is likely due to the fact that multiple methylated histones are necessary to mark an IES for DNA elimination. A handful of other IESs have been well characterized and tend to be repetitive in nature. These IESs, along with the M IES, are flanked immediately by short direct repeats, which are flanked by macronuclear-retained boundary determinants that are complex in sequence [19-28]. In mechanistic studies of the M, R and mse2.9 IES boundary determinants, deletions of these sequences were sufficient to inhibit or completely block IES rearrangement [21, 22, 29]. Deletions or insertions of DNA between these boundary determinants and the IES trigger a shift of the IES boundary resulting in retention of a portion of the IES or a loss of macronuclear destined sequence, respectively [21, 22, 30].

Although a general model for RNA-directed DNA elimination is well supported, many questions still remain including understanding the nature of IESs and how they are excised from the macronuclear genome. To determine if histone methylation and DNA elimination of IESs in the zygotic macronucleus were dependent on early IES-specific dsRNA production in the micronucleus, we created strains lacking the R IES in the micronucleus. Removal of the endogenous R IES would allow us to assess how loss of the source of dsRNAs would affect the excision of an ectopically introduced R IES. Here we show that loss of long, bidirectional ncRNA production from the native R IES early in conjugation inhibits later ectopic R IES excision. This further supports that meiotic scnRNAs are essential for RNAi-directed DNA elimination of specific IESs during somatic genome remodeling.

Results:

Formation and Verification of R IES Knockout Strains.

By eliminating dsRNA production early in conjugation we would be able to test whether dsRNAs produced later from an ectopic IES could be used to produce scnRNA/Twi1p complexes that are competent for DNA elimination. Currently, the model for RNA-directed DNA elimination in *T. thermophila* posits that only dsRNAs produced early in the crescent micronucleus can be cleaved by Dcl1p to produce scnRNAs that direct DNA elimination (chapter 3) [10, 11]. Many IESs, including the M IES, are highly repetitive and contain small regions of homology in the macronucleus [18, 31].These macronuclear retained IES-like repeats are potential sources of bidirectional ncRNA and scnRNA, which could direct rearrangement of any ectopic IES. To identify which of the handful of known IESs was solely limited to the micronucleus we did a BLAST analysis of the *T.*

thermophila macronuclear genome with known IES sequences. BLAST analysis of the R IES failed to identify regions of homology greater than 20 bp indicating that the R IES is solely confined to the micronucleus (data not shown). Unlike a knockout of the M IES, which has multiple M IES-like sequences (MLSs) in the macronucleus that could be transcribed to produce long, bidirectional dsRNAs, a micronuclear knockout of the R IES would completely ablate all R IES dsRNAs (K. Gao and D. L. Chalker, unpublished data).

R IES knockout cassettes were constructed by surrounding the selectable marker, MTT1/NEO3, with up- and downstream R IES locus homology. A linearized R IES knockout cassette was biolistically transformed into wild-type conjugating cells (CU428 X B2086) and heterozygous transformants were identified by drug selection (Figs. 1 and 2A). Homozygous micronuclear R IES knockouts were generated by genomic exclusion and verified by Southern blot (Figs. 1, 2A and 2C; data not shown). To certify that the homozygous micronuclear R IES knockout strains had the R IES knockout cassette in the correct locus, Southern blot analysis of progeny from homozygous micronuclear R IES knockout matings with wild-type and with themselves were done (Figs. 2D-2F). The Southern blots showed that a number of progeny had aberrant rearrangements of the R IES knockout cassette. Matings of wild-type strains (CU427 or CU428, whose progeny contain different drug-resistant alleles) with the homozygous micronuclear R IES knockout strains produced abundant progeny resistant to the drug specified by the wild-type allele but these progeny typically lost their resistance to the R IES knockout cassette drug marker, paromomycin (Table 1). Loss of resistance to paromomycin was also seen in matings between the homozygous micronuclear R IES knockout strains. Southern blot

analysis of progeny that had lost drug resistance from CU428 matings with homozygous micronuclear R IES knockout strains showed that these progeny had an extremely low amount of the R IES knockout cassette remaining (Fig. 2D: Lanes 7 and 8). Initially these results surprised us, but other labs have documented that the bacterial neo gene can be excised during conjugation in *T. thermophila* [32, 33]. Excision of the knockout cassette along with transgene silencing, could explain the rearrangements of the R IES knockout cassette and loss of drug resistance in some strains.

Failure of R IES rearrangement in homozygous micronuclear R IES knockout strain matings.

If early scnRNA production is necessary for normal IES excision, the removal of the R IES from the micronucleus should produce strains incapable of DNA rearrangement of an ectopically introduced R IES. To test this we used a plasmid-based assay to determine rearrangement of an ectopic R IES. Use of plasmids to analyze IESs is a well-established assay to monitor rearrangement efficiencies of wild-type and mutant IES sequences [18, 26, 34, 35]. The plasmid, pDLCR6, and a derivative plasmid, pFL, contain a full length R IES with the native boundary elements ([21], C. Montero-Diaz and D.L. Chalker, unpublished data). pDLCR6 has been shown in the plasmid-based rearrangement assay in wild-type matings to have least 50% of the plasmids undergo DNA elimination of the entire R IES, while pFL has been shown to undergo complete rearrangement 76.5% of the time. These plasmids were electroporated into conjugating homozygous micronuclear R IES knockout matings (Fig. 3A). Drug resistant progeny were identified, genomic DNA was isolated and restriction enzyme digested, and run out on a Southern blot (Figs. 3A and 3B). Southern blot analysis showed almost complete

failure of DNA elimination of the R IES in both plasmids (Fig. 3B). 6 out of 8 strains electroporated with pDLCR6 and 4 out of 6 strains electroporated with pFL completely failed to rearrange the R IES. Of the four other strains tested there an average of 12.1% partial and 9.7% complete rearrangement, however one mating [B*VI$^{\Delta R/\Delta R}$ 7 X B*VII$^{\Delta R/\Delta R}$ 2 (pDCLR6) #2] accounted for a supermajority of the partial rearrangement seen (Fig. 3C). Overall, only 5.6% partial and 3.2% complete rearrangement of pDLCR6 and 2.1% partial and 2.7% complete rearrangement of pFL was seen in homozygous micronuclear R IES knockout matings. These levels of rearrangement are significantly lower than those in wild-type matings previously described, indicating that early bidirectional transcription of IES dsRNA strongly enhances proper IES excision late in conjugation ([21], C. Montero-Diaz and D.L. Chalker, unpublished data).

Discussion:

Elimination of the R IES from the micronucleus blocks rearrangement of artificial R IES plasmids introduced later during conjugation. This validates the RNAi-directed DNA elimination model where early dsRNA and scnRNA production in the crescent micronucleus is required for later histone methylation and DNA elimination in the developing zygotic macronucleus. Analysis of scnRNA production from the TLR, M and R IESs has shown a large difference in the amounts of scnRNA produced for each IES (C.D. Malone and D.L. Chalker, unpublished data). The TLR IES has extremely high levels of scnRNA, which is likely due to the fact that it belongs to a family of long, highly repetitive and conserved sequences ([20], C.D. Malone and D.L. Chalker, unpublished data). The M IES has low levels of scnRNA production, while the R IES has even lower levels (C.D. Malone and D.L. Chalker, unpublished data). The levels of IES-

specific scnRNAs appear to inversely correlate with the size of IES necessary for excision using the plasmid-based IES assay ([18], C. Montero-Diaz, J.C.L. Trein, and D.L. Chalker, unpublished data). Therefore, since all plasmid-based R IES excision in the homozygous micronuclear R IES knockout matings is not blocked, this indicates the possible presence of a partial copy of the R IES in the micronucleus. Conversely, the residual rearrangement of the ectopic R IES could also be due to the presence of scnRNAs produced from the ~20bp repeats spread throughout the macronuclear genome. The majority of R IES scnRNAs map to these ~20bp repeats making this possibility very likely (C.D. Malone and D.L. Chalker, unpublished data). In the future this link between scnRNA levels and IES copy number could be used to identify low-repetitive or unique IESs in the micronucleus based on scnRNA sequencing in ciliates. It remains to be seen whether the link between repetitive sequences and chromatin diminution holds true for other eukaryotes besides ciliates but in at least one other organism that undergoes this process, the Japanese hagfish (*Eptatretus okinoseanus*), there does appear to be a link between DNA loss and repetitive sequences [36]. Further research in ciliates and other organisms that partake in chromatin diminution may shed light on how organisms recognize and contain invading exogenous nuclei acid.

Materials and Methods:

***Tetrahymena* strains and growth conditions.** Standard wild-type, laboratory *T. thermophila* strains CU427 (Chx/Chx [VI, cy-s]), CU428 (Mpr/Mpr [VII, mp-s]), B2086 (II), micronucleus-defective strains B*VI (VI) and B*VII (VII) were originally obtained from Peter Bruns (Cornell University, Ithaca, NY). These strains or their transformed progeny were used for biolistic transformations and subsequent analyses. Cells were

grown and maintained as previously described [37, 38]. Strains were starved six hours to overnight in 10 mM Tris (pH 7.5) prior to mixing to initiate conjugation. Optical densities of cell populations were used to estimate cell number prior to mixing equal numbers of mating compatible strains.

Sequence analysis of R IES and generation of R IES knockouts. The size and sequence of the R IES have been previously described and published [25, 39]. R IES sequence was compared to the *T. thermophila* macronuclear genome (http://www.ciliate.org) via BLAST analysis to determine the existence and size of any R IES-like repeats retained in the macronuclear genome.

To create micronuclear knockouts of the R IES, upstream and downstream homology of the R IES was amplified via PCR and recombined into pDONR-P4-P1R (Upstream) and pDONR-P2R-P3 (Downstream) using BP Clonase (Invitrogen) (See Table 2 for primers). The resulting plasmids were identified by lysate PCR and verified through sequencing (See Table 2 for primers). The donor plasmids containing Up- and downstream homology were mixed with equal amounts of pENTR-D-MTT1/NEO3 and the multisite destination vector pDEST-R4-R3 along with LR Clonase Plus (Invitrogen) to create the R IES knockout plasmid, pDEST-B4-RUpstream-B1-MTT1/NEO3-B2-RDownstream-B3. The R IES knockout construct was linearized by restriction digestion with KpnI and introduced into conjugating wild-type cells (CU428 X B2086) between 2 and 3 hours after mixing using a PDS-1000/He particle bombardment system (Bio-Rad) as previously described [40, 41]. Heterozygous micronuclear transformants were identified by their resistance to 80 µg/ml paromomycin with 1 µg/ml $CdCl_2$ and 15µg/ml 6-methylpurine. Heterozygous micronuclear transformants were verified through matings

with CU427 by monitoring segregation of paromomycin resistance conferred by the MTT1/NEO3 knockout cassette among cycloheximide resistant progeny, as well as through PCR screening of *T. thermophila* crude cell lysates (See Table 2 for primers) [42]. Homozygous micronuclear R IES knockout heterokaryons were generated by crossing heterozygous micronuclear transformants with B*VI and B*VII star strains. Homozygous micronuclear knockouts were identified by paromomycin/$CdCl_2$ sensitivity and verified through crosses with CU427 or CU428 for production of progeny completely resistant to 25 µg/ml cycloheximide or 15 µg/ml 6-methylpurine, respectively, and partial resistance to 100 µg/ml paromomycin with 1 µg/ml $CdCl_2$. Progeny of R IES micronuclear strains were also assayed by PCR analysis of *T. thermophila* crude cell lysates for presence of the R IES knockout cassette (See Table 2 for primers) [42].

R IES micronuclear knockout Southern blot analysis. *T. thermophila* genomic DNA was isolated using a Wizard genomic DNA purification kit (Promega). Gel electrophoresis, blotting and hybridization were performed as previously described except Southern blots were washed with 0.5X SSC-1% SDS [10]. Southern blot probe for the R IES was obtained through DNA isolation of SacI restriction digest fragment from pDONR-L4-RUpstream-R1. Examination of homozygous micronuclear R IES knockouts was accomplished by separating BclI-digested CU428 and R IES micronuclear knockouts #1, #2 and #3 (ΔR IES/ΔR IES [VII, +]) genomic DNA on a 0.8% agarose gel prior to blotting. Progeny of R IES micronuclear knockouts #2 and #7 (ΔR IES/ΔR IES [VI, +]) with CU428 had their genomic DNA isolated as above, which, along with control CU428 genomic DNA, was digested with BclI and electrophoresed on a 0.8% agarose gel to assay retention of the R IES knockout cassette. A similar analysis of progeny from R IES

micronuclear knockouts #1 and #2 (ΔR IES/ΔR IES [VII, +]) mated to CU427 was also performed, except their genomic DNA was digested with BglII and EcoRI and the wild-type controls were CU428 and B*VII genomic DNA. Genomic DNA from progeny of R IES micronuclear knockouts #2 and #7 (ΔR IES/ΔR IES [VI, +]) with R IES micronuclear knockouts #1, #2 and #3 (ΔR IES/ΔR IES [VII, +]) was digested with BclI and separated on a 0.8% agarose gel to determine level of aberrant rearrangement of the R IES knockout cassette during DNA elimination.

Full-length R IES plasmids, pDLCR6 [43] and pFL (C. Montero-Diaz and D.L. Chalker, unpublished data), were introduced into matings between R IES micronuclear knockouts #2 X #1 (ΔR IES/ΔR IES [VI/VII, +]) or #7 X #2 (ΔR IES/ΔR IES [VI/VII, +]) through conjugative electroporation as previously described [44]. Progeny were identified by resistance to 100 µg/ml paromomycin and their genomic DNA was isolated as above. Progeny genomic DNA was digested with BamHI and separated on a 1.0% agarose gel. A Southern blot was performed as above and probed with an R IES probe that was previously described [21]. Rearrangement of the full-length R IESs plasmids in R IES micronuclear knockout matings was quantified by phosphorimager analysis using a Personal FX imager and Quantity One software (Bio-Rad).

References:

1. Boveri, T., *Über Differenzierung der Zellkerne während der Fuchung des Eies von Ascaris megalocephala.* Anatomischer Anzeiger, 1887. **2**: p. 688-93.
2. Beermann, S., *The diminution of Heterochromatic chromosomal segments in Cyclops (Crustacea, Copepoda).* Chromosoma, 1977. **60**(4): p. 297-344.
3. Nakai, Y., S. Kubota, and S. Kohno, *Chromatin diminution and chromosome elimination in four Japanese hagfish species.* Cytogenetics and cell genetics, 1991. **56**(3-4): p. 196-8.
4. Smith, J.J., et al., *Programmed loss of millions of base pairs from a vertebrate genome.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(27): p. 11212-7.

5. Prescott, D.M., *The DNA of ciliated protozoa.* Microbiol Rev, 1994. **58**(2): p. 233-67.
6. Goday, C. and M.R. Esteban, *Chromosome elimination in sciarid flies.* BioEssays : news and reviews in molecular, cellular and developmental biology, 2001. **23**(3): p. 242-50.
7. Yao, M.C. and M.A. Gorovsky, *Comparison of the sequences of macro- and micronuclear DNA of Tetrahymena pyriformis.* Chromosoma, 1974. **48**(1): p. 1-18.
8. Mochizuki, K., et al., *Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena.* Cell, 2002. **110**(6): p. 689-99.
9. Chalker, D.L. and M.C. Yao, *Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila.* Genes Dev, 2001. **15**(10): p. 1287-98.
10. Malone, C.D., et al., *Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila.* Mol Cell Biol, 2005. **25**(20): p. 9151-64.
11. Mochizuki, K. and M.A. Gorovsky, *A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase.* Genes Dev, 2005. **19**(1): p. 77-89.
12. Mochizuki, K. and M.A. Gorovsky, *Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement.* Genes Dev, 2004. **18**(17): p. 2068-73.
13. Liu, Y., et al., *RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena.* Genes Dev, 2007. **21**(12): p. 1530-45.
14. Liu, Y., K. Mochizuki, and M.A. Gorovsky, *Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena.* Proc Natl Acad Sci U S A, 2004. **101**(6): p. 1679-84.
15. Taverna, S.D., R.S. Coyne, and C.D. Allis, *Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena.* Cell, 2002. **110**(6): p. 701-11.
16. Cheng, C.Y., et al., *A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila.* Mol Biol Cell, 2010. **21**(10): p. 1753-62.
17. Heinonen, T.Y. and R.E. Pearlman, *A germ line-specific sequence element in an intron in Tetrahymena thermophila.* J Biol Chem, 1994. **269**(26): p. 17428-17433.
18. Kowalczyk, C.A., et al., *The germ line limited M element of Tetrahymena is targeted for elimination from the somatic genome by a homology-dependent mechanism.* Nucleic Acids Res, 2006.
19. Patil, N.S. and K.M. Karrer, *A developmentally regulated deletion element with long terminal repeats has cis-acting sequences in the flanking DNA.* Nucleic acids research, 2000. **28**(6): p. 1465-72.
20. Wuitschick, J.D., et al., *A novel family of mobile genetic elements is limited to the germline genome in Tetrahymena thermophila.* Nucleic Acids Res, 2002. **30**(11): p. 2524-37.

21. Chalker, D.L., et al., *Flanking regulatory sequences of the Tetrahymena R deletion element determine the boundaries of DNA rearrangement.* Mol Cell Biol, 1999. **19**(8): p. 5631-41.
22. Fillingham, J.S., D. Bruno, and R.E. Pearlman, *Cis-acting requirements in flanking DNA for the programmed elimination of mse2.9: a common mechanism for deletion of internal eliminated sequences from the developing macronucleus of Tetrahymena thermophila.* Nucleic Acids Res, 2001. **29**(2): p. 488-98.
23. Yao, M.C., *Ribosomal RNA gene amplification in Tetrahymena may be associated with chromosome breakage and DNA elimination.* Cell, 1981. **24**(3): p. 765-774.
24. Yao, M.C., *Elimination of specific DNA sequences from the somatic nucleus of the ciliate Tetrahymena.* J Cell Biol, 1982. **92**(3): p. 783-789.
25. Austerberry, C.F. and M.C. Yao, *Nucleotide sequence structure and consistency of a developmentally regulated DNA deletion in Tetrahymena thermophila.* Mol Cell Biol, 1987. **7**(1): p. 435-443.
26. Wells, J.M., et al., *A small family of elements with long inverted repeats is located near sites of developmentally regulated DNA rearrangement in Tetrahymena thermophila.* Mol Cell Biol, 1994. **14**(9): p. 5939-5949.
27. Katoh, M., et al., *A micronucleus-specific sequence exists in the 5'-upstream region of calmodulin gene in Tetrahymena thermophila.* Nucleic Acids Res., 1993. **21**(10): p. 2409-2414.
28. Austerberry, C.F. and M.C. Yao, *Sequence structures of two developmentally regulated, alternative DNA deletion junctions in Tetrahymena thermophila.* Mol Cell Biol, 1988. **8**(9): p. 3947-3950.
29. Godiska, R. and M.C. Yao, *A programmed site-specific DNA rearrangement in Tetrahymena thermophila requires flanking polypurine tracts.* Cell, 1990. **61**(7): p. 1237-46.
30. Godiska, R., C. James, and M.C. Yao, *A distant 10-bp sequence specifies the boundaries of a programmed DNA deletion in Tetrahymena.* Genes & development, 1993. **7**(12A): p. 2357-65.
31. Aronica, L., et al., *Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in Tetrahymena.* Genes Dev, 2008. **22**(16): p. 2228-41.
32. Liu, Y., et al., *Elimination of Foreign DNA during Somatic Differentiation in Tetrahymena thermophila Shows Position Effect and Is Dosage Dependent.* Eukaryot Cell, 2005. **4**(2): p. 421-31.
33. Yao, M.C., P. Fuller, and X. Xi, *Programmed DNA deletion as an RNA-guided system of genome defense.* Science, 2003. **300**(5625): p. 1581-4.
34. Godiska, R. and M.C. Yao, *A programmed site-specific DNA rearrangement in Tetrahymena thermophila requires flanking polypurine tracts.* Cell, 1990. **61**: p. 1237-1246.
35. Li, J. and R.E. Pearlman, *Programmed DNA rearrangement from an intron during nuclear development in Tetrahymena thermophila: molecular analysis and identification of potential cis-acting sequences.* Nucleic Acids Res., 1996. **24**(10): p. 1943-1949.

36. Kubota, S., et al., *Germ line-restricted, highly repeated DNA sequences and their chromosomal localization in a Japanese hagfish (Eptatretus okinoseanus).* Chromosoma, 1993. **102**(3): p. 163-73.
37. Gorovsky, M.A., et al., *Isolation of micro- and macronuclei of Tetrahymena pyriformis.* Methods Cell Biol, 1975. **9**(0): p. 311-327.
38. Orias, E., E.P. Hamilton, and J.D. Orias, *Tetrahymena as a laboratory organism: useful strains, cell culture, and cell line maintenance.* Methods Cell Biol, 2000. **62**: p. 189-211.
39. Austerberry, C.F., C.D. Allis, and M.C. Yao, *Specific DNA rearrangements in synchronously developing nuclei of Tetrahymena.* Proc Natl Acad Sci U S A, 1984. **81**(23): p. 7383-7387.
40. Cassidy-Hanley, D., et al., *Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment.* Genetics, 1997. **146**(1): p. 135-47.
41. Bruns, P.J. and D. Cassidy-Hanley, *Biolistic transformation of macro- and micronuclei.* Methods Cell Biol, 2000. **62**: p. 501-12.
42. Chalker, D.L., P. Fuller, and M.C. Yao, *Communication between parental and developing genomes during tetrahymena nuclear differentiation is likely mediated by homologous RNAs.* Genetics, 2005. **169**(1): p. 149-60.
43. Chalker, D.L. and M.C. Yao, *Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of Tetrahymena thermophila with germ line-limited DNA.* Mol Cell Biol, 1996. **16**(7): p. 3658-67.
44. Gaertig, J., et al., *High frequency vector-mediated transformation and gene replacement in Tetrahymena.* Nucleic Acids Res, 1994. **22**(24): p. 5391-8.

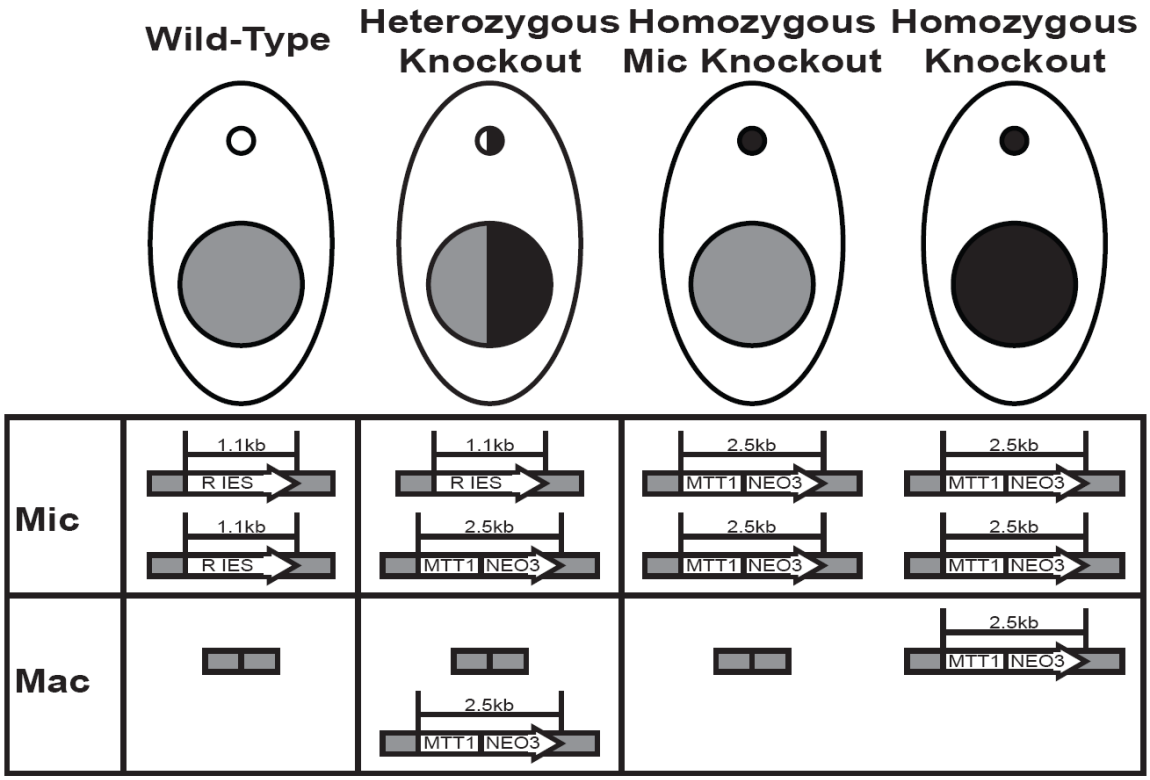


Figure 1: Diagram of R IES micro- and macronuclear loci in wild-type and R IES knockout strains. State of micro- and macronucleus at the R IES locus in wild-type (CU428 or B*VII), heterozygous macro- and micronuclear R IES knockouts, homozygous micronuclear R IES knockouts and homozygous macro- and micronuclear R IES knockouts. Top Panel: Cartoon of *T. thermophila* cells with wild-type and/or R IES knockout cassette in the micro- and macronucleus. White full or half circle, wild-type micronuclear R IES; gray full or half circle, wild-type macronuclear R IES; black full or half circle, R IES knockout cassette. Bottom Panel: Graphic of micro- and macronucleus at R IES locus in wild-type and R IES knockouts. Type of R IES loci in each nucleus is represented by two diagrams in the micronucleus and one or two diagrams in the macronucleus. R IES and R IES knockout cassette are represented by white boxes of 1.1kb and 2.5kb, respectively.

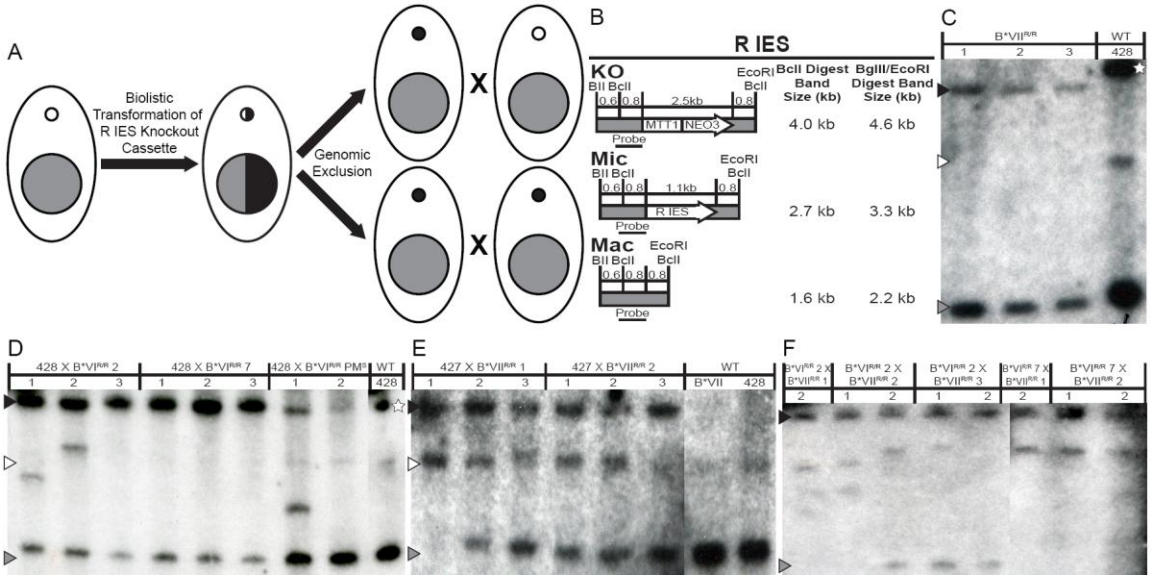


Figure 2: Creation and verification of R IES knockout strains. A. Conjugating wild-type cells (CU428 X B2086) were biolistically transformed with the R IES knockout cassette. Heterozygous macro- and micronuclear transformants were used to generate homozygous micronuclear R IES knockouts. These strains were then mated to wild-type (CU428) or to themselves as further verification of correct R IES knockout cassette incorporation. White full or half circle, wild-type micronuclear R IES; gray full or half circle, wild-type macronuclear R IES; black full or half circle, R IES knockout cassette. B. Diagram of R IES locus in R IES micronuclear knockout, wild-type micronuclear R IES and wild-type macronuclear R IES. Band sizes of each loci from restriction enzyme digestion for Southern blot analysis are listed. BII, BglII restriction enzyme site. C. Southern blot of wild-type (CU428) and homozygous micronuclear R IES knockout genomic DNA digested with BclI. B*VII$^{R/R}$, (ΔR IES/ΔR IES [VII, +]). D and E. Southern blot of wild-type and homozygous micronuclear R IES knockout mating progeny genomic DNA digested with BclI (D) or BglII/EcoRI (E). B*VI$^{R/R}$ or B*VII$^{R/R}$, (ΔR IES/ΔR IES [VI/VII, +]). F. Southern blot of homozygous micronuclear R IES knockout mating progeny genomic DNA digested with BclI to determine level of selection cassette excision. Black arrowhead, R IES knockout cassette fragment; white arrowhead, wild-type R IES micronuclear fragment; gray arrowhead, wild-type R IES macronuclear fragment; white star non-specific band.

Table 1: Progeny production of R IES mic knockout strains in wild-type and knockout matings

Cross	Pair Survival (S/N)[1]	Progeny Production (P/S)[2]	Pm^R Progeny Production (P/S)[2]
CU428 X B*VI$^{\Delta R/\Delta R}$ #2	**69.7 %** (92/132)	**75.0 %** (69/92)	**17.4 %** (12/69)
CU428 X B*VI$^{\Delta R/\Delta R}$ #7	**81.8 %** (108/132)	**73.1 %** (79/108)	**32.9 %** (26/79)
CU427 X B*VII$^{\Delta R/\Delta R}$ #1	**98.9 %** (87/88)	**97.7 %** (85/87)	**23.5 %** (20/85)
CU427 X B*VII$^{\Delta R/\Delta R}$ #2	**94.3 %** (83/88)	**100 %** (83/83)	**21.7 %** (18/83)
B*VI$^{\Delta R/\Delta R}$ #2 X B*VII$^{\Delta R/\Delta R}$ #1	**93.2 %** (123/132)	N/A	**28.5 %** (35/123)
B*VI$^{\Delta R/\Delta R}$ #2 X B*VII$^{\Delta R/\Delta R}$ #2	**97.7 %** (129/132)	N/A	**34.1 %** (44/129)
B*VI$^{\Delta R/\Delta R}$ #2 X B*VII$^{\Delta R/\Delta R}$ #3	**99.2 %** (92/132)	N/A	**25.0 %** (23/92)
B*VI$^{\Delta R/\Delta R}$ #7 X B*VII$^{\Delta R/\Delta R}$ #1	**96.2 %** (127/132)	N/A	**24.4 %** (31/127)
B*VI$^{\Delta R/\Delta R}$ #7 X B*VII$^{\Delta R/\Delta R}$ #2	**95.5 %** (126/132)	N/A	**22.2 %** (28/126)
B*VI$^{\Delta R/\Delta R}$ #7 X B*VII$^{\Delta R/\Delta R}$ #3	**98.5 %** (130/132)	N/A	**41.5 %** (54/130)

Table 1: Progeny production of R IES mic knockouts in wild-type and knockout matings. Table of pair survival and progeny production in micronuclear R IES knockouts crossed with wild-type or with micronuclear R IES knockouts. B*VI$^{\Delta R/\Delta R}$ #2 and #7, (ΔR IES/ΔR IES [VI, +]); B*VII$^{\Delta R/\Delta R}$ #1, #2 and #3, (ΔR IES/ΔR IES [VII, +]). Cells were scored for pair survival, transferred to media containing selective marker and scored for drug resistance later. Pm^R, paromomycin resistance.

1. Pair survival is the % of pairs alive (S) of the total pairs (N) isolated.

2. Progeny production is the % of surviving pairs (S) that successfully completed conjugation and made new macronuclei (P).

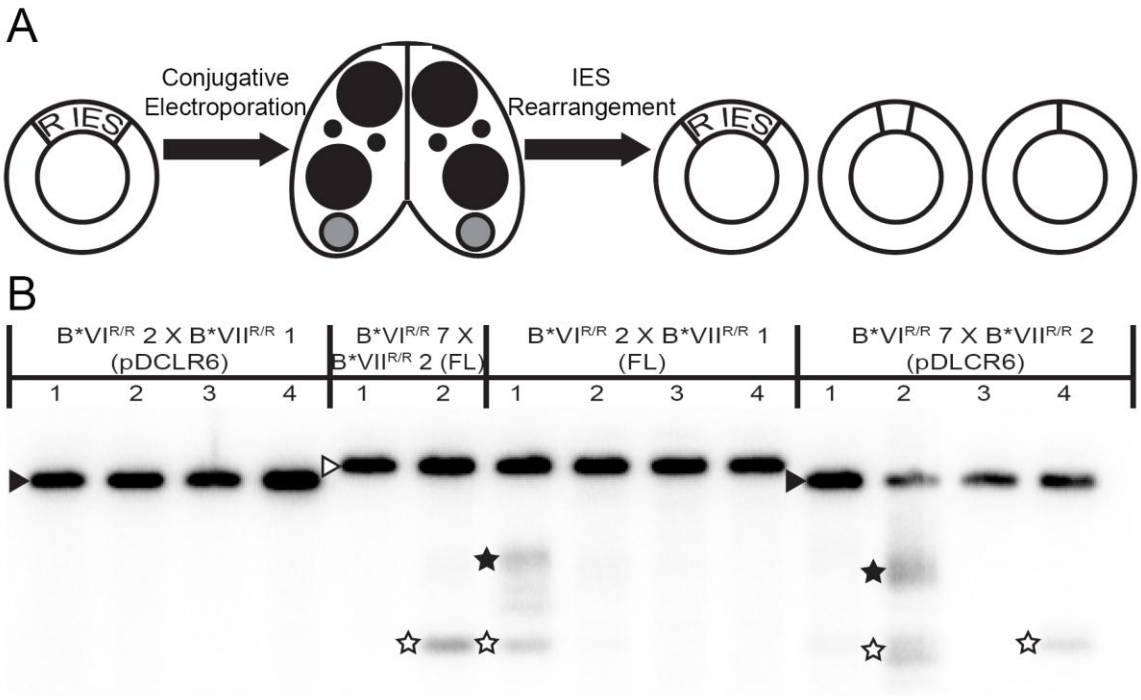


Figure 3: Failure of plasmid-based R IES rearrangement in homozygous micronuclear R IES knockout strain matings. A. Illustration of strategy to test rearrangement of R IES plasmid. Two R IES-containing plasmids were electroporated into conjugating homozygous micronuclear R IES knockout cells prior to DNA elimination. B. Southern blot to test DNA rearrangement efficiency of R IES plasmid. Black arrowhead, unrearranged pDCLR6 R IES; white arrowhead, unrearranged pFL R IES; black star, partially rearranged pDCLR6 or pFL R IES; white star, completely rearranged pDCLR6 or pFL R IES. FL, pFL. C. Chart quantifying levels of R IES plasmid rearrangement.

C

Mating	Intensity (Counts*mm^2)	Area (mm^2)	Percent Rearrangement (Intensity Rearranged/Total Intensity)
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (pDCLR6) #1	66167.12197	8.7600002	0.00%
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (pDCLR6) #2	81193.52242	9.6400003	0.00%
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (pDCLR6) #3	57900.00173	7.7599998	0.00%
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (pDCLR6) #4	122326.9636	12.120001	0.00%
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (FL) #1	64955.96194	7.2400002	0.00%
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (FL) #2	95996.44286	8.5200005	9.97%
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (FL) #2	10635.40032	6.2400002	
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (FL) #1	75798.88226	7.3600001	18.51%
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (FL) #1	11623.88035	10.520000	
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (FL) #1	5599.040167	6.3600001	
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (FL) #2	69846.28208	7.4000001	0.00%
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (FL) #3	72081.40215	7.9200006	0.00%
B*VI$^{R/R}$ 2 X B*VII$^{R/R}$ 1 (FL) #4	68846.68205	8.7200003	0.00%
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (pDCLR6) #1	76352.24228	9.8400002	0.00%
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (pDCLR6) #2	20552.00061	5.0400004	60.54%
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (pDCLR6) #2	23442.60070	15.240001	
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (pDCLR6) #2	8082.840241	7.6400003	
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (pDCLR6) #3	26565.48079	5.2400002	0.00%
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (pDCLR6) #4	34234.60102	6.0799999	10.02%
B*VI$^{R/R}$ 7 X B*VII$^{R/R}$ 2 (pDCLR6) #4	3812.320114	6.7599998	

Table 2: Oligonucleotides used in the course of this study

Purpose and Name	Sequence (5'-3')
Knockout Cassette Generation Upstream:	
#1623-Rup7219AaatB4	**GGGGACAACTTTGTATAGAAAAGTTGGTACCCATCACTTTTGCCATTAG**
#1624-Rup8214attB1	**GGGGACTGCTTTTTTGTACAAACTTGAGCTTTTTGGGTTTTACACTG**
Downstream:	
#1625-Rdown8295attB2	**GGGGACAGCTTTCTTGTACAAAGTGGTCTTATTCTAGAACTATCTTAC**
#1626-Rdown9548AattB3	**GGGGACAACTTTGTATAATAAAGTTGGTACCTTTCTGCATAGCGAACG**
Knockout Cassette Screening and Sequencing	
#3047-M13Forward	**GTAAAACGACGGCCAGT**
#3048-M13Reverse	**TCACACAGGAAACAGCTATGAC**
Knockout PCR Screening	
3' **#1684-R8191**	**AAACAGTGTAAAACCCAAAAAGC**
#1685-R8688	**TTATTCGTTCAATCGATTAGCTT**
#2367-p4T2-3351	**TCGCCTTCTTGACGAGTTCT**

CHAPTER 3

GERM LINE TRANSCRIPTS ARE PROCESSED BY A DICER-LIKE PROTEIN THAT IS ESSENTIAL FOR DEVELOPMENTALLY PROGRAMMED GENOME REARRANGEMENTS OF *TETRAHYMENA THERMOPHILA*

Molecular and Cellular Biology. 2005. 25(20), 9151-9164

Contributions to the Paper:

The discovery that small RNAs are involved in DNA rearrangement suggested that Tetrahymena cells likely contain a Dicer-like ribonuclease that is required for their biogenesis. Initial identification and characterization of the three putative Dicer genes in the *Tetrahymena thermophila* genome was carried out by students in the Biology 3492 class taught by Dr. Douglas Chalker in spring 2004. The studies by these undergraduates provided sufficient preliminary data to indicate that two of these proteins, Dcl1p and Dcr2p, may have essential functions during growth and/or development. Upon beginning my laboratory rotation in Dr. Chalker's lab in June 2004 I carried out a comprehensive analysis of two of the three putative Dicer genes with technical assistance from two laboratory technicians, Colin D. Malone and Alissa M. Anderson. By analyzing strains generated during the class that had the DCR2 gene disrupted, I was able to conclusively demonstrate that Dcr2p is essential for vegetative growth.

In contrast, initial knockouts of the Dicer-like gene, DCL1, generated during the Biology 3492 class proved inadequate for further studies as the disruption also removed an upstream open reading frame (ORF). I therefore created a new DCL1 knockout vector and knocked out this gene. Using these Tetrahymena strains lacking DCL1, I produced the initial data showing that Dcl1p is essential for scnRNA production. Furthermore, loss of DCL1 resulted in developmental arrest during conjugation and failure of M IES DNA elimination.

Complementing the generating and characterizing the DCL1 knockout strains, I examined Dcl1p expression and localization. I verified DCL1 mRNA expression was conjugation-specific using Northern blot hybridization (Fig. 1B). In addition I generated

an N-terminal GFP-tagged DCL1 construct and showed that Dcl1p localized early in conjugation to the crescent micronucleus at the same time as scnRNAs are produced (Fig. 5). After completing these studies, I began another laboratory rotation, during which time Colin Malone and Alissa Anderson completed characterization of the DCL1 knockout strains and refined the initial data to make it suitable for publication. Also, Charles Rexer contributed the chromatin immunoprecipitation data in response to reviewers' comments prior to final publication. Dr. Chalker and colleagues wrote the resulting manuscript and submitted it for publication shortly before I joined the lab permanently in May 2005.

MOLECULAR AND CELLULAR BIOLOGY, Oct. 2005, p. 9151–9164 Vol. 25, No. 20
0270-7306/05/$08.00+0 doi:10.1128/MCB.25.20.9151–9164.2005

Germ Line Transcripts Are Processed by a Dicer-Like Protein That Is Essential for Developmentally Programmed Genome Rearrangements of *Tetrahymena thermophila*

Colin D. Malone,† Alissa M. Anderson,† Jason A. Motl,† Charles H. Rexer, and Douglas L. Chalker*

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

Received 1 April 2005/Returned for modification 4 May 2005/Accepted 19 July 2005

Abundant ~28-nucleotide RNAs that are thought to direct histone H3 lysine 9 (H3K9) methylation and promote the elimination of nearly 15 Mbp of DNA from the developing somatic genome are generated during *Tetrahymena thermophila* conjugation. To identify the protein(s) that generates these small RNAs, we studied three Dicer-related genes encoded within the *Tetrahymena* genome, two that contain both RNase III and RNA helicase motifs, Dicer 1 (DCR1) and DCR2, and a third that lacks the helicase domain, Dicer-like 1 (DCL1). *DCL1* is expressed upon the initiation of conjugation, and the protein localizes to meiotic micronuclei when bidirectional germ line transcription occurs and small RNAs begin to accumulate. Cells in which we disrupted the *DCL1* gene (Δ*DCL1*) grew normally and initiated conjugation as wild-type cells but arrested near the end of development and eventually died, unable to resume vegetative growth. These Δ*DCL1* cells failed to generate the abundant small RNAs but instead accumulated germ line-limited transcripts. Together, our findings demonstrate that these transcripts are the precursors of the small RNAs and that *DCL1* performs RNA processing within the micronucleus. Postconjugation Δ*DCL1* cells die without eliminating the germ line-limited DNA sequences from their newly formed somatic macronuclei, a result that shows that this Dicer-related gene is required for programmed DNA rearrangements. Surprisingly, Δ*DCL1* cells were not deficient in overall H3K9 methylation, but this modification was not enriched on germ line-limited sequences as it is in wild-type cells, which clearly demonstrates that these small RNAs are essential for its targeting to specific loci.

RNA interference (RNAi) describes an array of related mechanisms involved in diverse biological processes including defense against RNA viruses, specification of centromeric heterochromatin structure, and developmental control of gene expression (reviewed in reference 25). These mechanisms share the use of small RNAs to target specific effector protein complexes to homologous sequences via base-pairing interactions. The use of small, homologous RNAs as specificity factors imparts tremendous flexibility of targets on a single protein complex. These targeting RNAs are generated by RNase III enzymes, collectively called Dicer ribonucleases, that cleave longer, double-stranded RNA (dsRNA) into ~20- to 26-nucleotide (nt) species that are incorporated into the effector complexes (3, 24, 27, 30; reviewed in reference 6). The genomes of many eukaryotes encode multiple Dicer-related proteins, and the specific Dicer used to generate the small RNAs can determine the downstream pathway that they enter. For instance, in *Arabidopsis thaliana*, the Dicer-like 3 (Dcl3) gene product is required to produce endogenous short interfering RNAs (siRNAs), *A. thaliana* Dcl2 is necessary for accumulation of siRNAs in response to RNA virus infections, and *A. thaliana Dcl1* is necessary to generate micro-RNAs (miRNAs) involved in the control of flower development (31, 53). Similarly, the *Drosophila melanogaster* Dcr-1 and Dcr-2 genes exhibit distinct roles in siRNA and miRNA regulatory pathways (33). Nevertheless, the diversification of these related pathways remains understood in only the most general ways.

Developmentally programmed genome reorganization of the ciliate *Tetrahymena thermophila* is one of the processes that is directed by homologous, small RNAs (reviewed in reference 43). Ciliates, including *Tetrahymena*, are single-celled organisms that exhibit nuclear dualism, possessing both germ line and somatic genomes that are harbored within distinct nuclei, called micro- and macronuclei, respectively (46). Massive DNA rearrangements are part of the differentiation of somatic macronuclei from germ line micronuclei, which retains the organism's genome intact for future propagation (55). The developmental program during which this nuclear differentiation occurs is initiated by conjugation. Within the first hours of conjugation, the germ line micronucleus within each mating partner undergoes meiosis to produce four haploid pronuclei, one of which is then selected to replicate its DNA and divide to generate one stationary and one migratory gametic nucleus. Nuclear exchange of the migratory nucleus is followed by karyogamy with the partner's stationary nucleus, resulting in the formation of a diploid, zygotic nucleus in each cell. This nucleus proceeds to divide twice, generating the progenitors of the new germ line and somatic nuclei of the progeny from the mating. The parental somatic nuclei begin to degenerate upon formation of these new nuclei.

During nuclear differentiation, the germ line-derived chromosomes within the developing somatic nucleus are broken at 200 to 300 sites. This chromosome breakage is coupled to new

* Corresponding author. Mailing address: Biology Department, Washington University, Campus Box 1137, St. Louis, MO 63130. Phone: (314) 935-8838. Fax: (314) 935-4432. E-mail: dchalker@biology2.wustl.edu.

† C.D.M., A.M.A., and J.A.M. contributed equally.

telomere addition. In addition, ~15 Mbp of DNA is eliminated by specific DNA rearrangements of an estimated 6,000 loci (the DNA segments excised are often called internal eliminated sequences [IESs]). These germ line-limited DNA segments range in size from a few hundred base pairs to more than 20 kbp and are comprised of both unique sequences as well as repetitive elements. While flanking regulatory sequences that demarcate the boundaries of specific deletion events have been identified (8, 17, 21, 22, 45), identification of any consensus sequences that are required to promote these DNA rearrangements has remained elusive. The heterogeneity of the sequences eliminated, together with the lack of a defined consensus sequence, has provided a challenge in describing a simple model for the control of this process.

Recent studies have revealed that *Tetrahymena* DNA rearrangements are guided by an RNAi-related mechanism. Abundant small (28- to 30-nt) RNAs that are enriched in germ line-limited sequences are produced early in development (9, 40). These are suspected to result from the processing of bidirectional transcripts produced in the germ line micronucleus that begin accumulating at the earliest stages of conjugation (11). The finding that the Argonaute homologue Twi1 protein (Twi1p) is required for the accumulation of these small RNAs (named scan RNAs) and for DNA rearrangement provided the first direct link between genome reorganization and RNAi (40). Yao et al. (56) demonstrated that RNA guides DNA rearrangements by injecting dsRNA corresponding to macronuclear regions into conjugating cells and documenting the elimination of the homologous DNA sequence that would normally be retained. Furthermore, it appears that these small RNAs target methylation of lysine 9 of histone H3 (H3K9) (H3K9me) to homologous sequences within the developing somatic macronuclei shortly after they are formed, and this modification is required for DNA rearrangement (34, 50). This modification is presumed to mark specific sequences for elimination by recruiting the machinery that excises the germ line-limited DNA and rejoins the flanking sequence that is retained in the mature somatic genome. This pathway of DNA rearrangement possesses the hallmarks of heterochromatin formation in other eukaryotes, thus making this process an intriguing model with which to study RNAi-directed, genome-wide targeting of this chromatin modification.

Further evidence that DNA rearrangement is controlled by a homology-based recognition system is the observation that germ line-limited sequences introduced into the parental somatic macronucleus block the efficient elimination of their cognate sequences from the developing macronucleus during subsequent nuclear differentiation (10). This sequence-specific inhibition applies even to a cell's wild-type conjugation partner by a method that does not require genetic exchange, a finding that supports the action of homologous RNAs as mediators of this regulation (9). Similar homology-based regulation of DNA rearrangement has been described in the ciliate *Paramecium tetraurelia* (13, 14). Small (23-nt) RNAs have also been implicated in these events (20), indicating that the rearrangement processes of these different ciliates are mechanistically related. The interplay between the germ line and somatic genomes suggests that DNA rearrangement in ciliates is a mechanism of genome surveillance that provides a means to remove foreign sequences from the transcriptionally active genome during development, thus limiting their spread. In support of this, transgenes introduced into the germ line genome can be eliminated from newly formed somatic macronuclei in a process that looks remarkably similar to the process of endogenous genome rearrangements (35, 56).

The bidirectional germ line transcription that occurs early in conjugation provides a source RNA that could be compared between the germ line and somatic genomes and later target sequences found exclusively in the germ line for elimination by the DNA rearrangement machinery. To demonstrate a clear connection between germ line transcription, small RNAs, and DNA rearrangement, we searched the *Tetrahymena* draft genome sequence for candidate Dicer RNase homologues that might encode the enzyme(s) that generates the small RNAs that target elimination. We show that one of three putative Dicer-related genes, *DCL1*, is localized to meiotic micronuclei and is required to process germ line transcripts into these 28- to 30-nt RNAs. Strains lacking *DCL1* are unable to complete development and fail to eliminate germ line-limited sequences from the developing somatic genome. The two other Dicer-related genes, each of which exhibits expression patterns distinct from that of *DCL1*, are not redundant to DCL1; thus, it would appear that even single-celled organisms can differentiate RNAi pathways by the specialization of Dicer function.

MATERIALS AND METHODS

Stocks and growth conditions. *Tetrahymena* cells were grown and maintained in 1× SPP at 30°C (44a). Cells were prepared for mating by washing cells from growth medium into 10 mM Tris-HCl (pH 7.4) and incubation overnight prior to mixing to initiate conjugation. Wild-type, inbred *Tetrahymena thermophila* strains (obtained from Peter Bruns, Cornell University) CU428 (Mpr/Mpr [VII, mp-s]), B2086 (II), and CU427 (Chx/Chx [VI, cy-s]) were used for all expression studies, biolistic transformations, and subsequent analyses. The micronucleus-defective "star" strains B*VI and B*VII were used to convert heterozygous Δ*DCL1* lines to micronuclear homozygosity by genomic exclusion crosses. Δ*TWI1* germ line/somatic knockout lines WG4 and 12-1A were provided by K. Mochizuki (University of Rochester, Rochester, NY).

Sequence identification. Three Dicer homologues were identified by BLAST search of the *Tetrahymena* genome (http://tigrblast.tigr.org/er-blast/index.cgi?project=ttg Assembly 2 [accessed November 2003]) using human Dicer1 (GenBank accession number gi29294651), *Drosophila melanogaster* CG6203-PA (accession number gi19922726), and the *Arabidopsis thaliana* endonuclease Dicer homologue (CARPEL FACTORY protein [accession number gi34922211]). The extents of the coding regions were initially predicted by visual inspection for higher GC content and proper intron/exon splice sites and are as follows: *DCR1*, positions 59218 to 66953 of scaffold CH445757 (accession number gi62422189); *DCR2*, positions 110642 to 117676 of scaffold CH445577 (accession number gi62422369); and *DCL1*, positions 808498 to 804612 of scaffold CH445618 (accession number gi62422328). Current sequence identification numbers from the *Tetrahymena* genome database (http://www.ciliate.org) are as follows: Dcr1p, T000006591; Dcr2p, T000006592; and Dcl1p, T000006590. Partial or full cDNA sequences were deposited in GenBank during the course of this work (gi50897087, gi50897083, and gi50897085) (42). Conserved domains were identified using the Pfam Protein Family Database (http://pfam.wustl.edu/).

Generation of DCL1 knockouts. Upstream *DCL1* sequences plus the first 189 codons of exon 1 (scaffold positions 807496 to 809192) and downstream sequence spanning codons 636 to 1254 including the lone intron (scaffold positions 804682 to 806496) were PCR amplified from genomic DNA and cloned individually into pCR2.1 using the TOPO TA cloning kit (Invitrogen). ApaI-XhoI or BamHI-NotI recognition sites were introduced into the ends of upstream and downstream oligonucleotide primer sets (Table 1), respectively, to facilitate insertion of the fragments into pMNBL flanking the metallothionein 1 (MTT1) promoter-driven *neo3* cassette (*MTT1*-neo) (48). The resulting *DCL1* knockout construct substituted 1.3 kbp of the coding sequence with the *neo3* cassette, effectively removing amino acids 190 to 638 from DCL1. This construct was linearized by digestion with ApaI and NotI and introduced into conjugating B2086 and CU428 cells between 2 and 3 h after mixing using a PDS-1000/He particle bombardment

TABLE 1. Oligonucleotides used in course of this study

Purpose and name	Sequence (5′–3′)
To amplify *DCL1* sequence for generation of knockout construct	
#1358-DCL1-10044B	ATAGGATCCAGTCTTGCTTACAAAAAGAC
#1359-DCL1-11863Nr	ATAGCGGCCGCATCTTAGAAGGCTTTTTTTTCAGC
#1429-DCL1-7349A	ATAGGGCCCACACCTTTATATATCATTTCC
#1430-DCL1-8616Xr	ATACTCGAGGATGATAGGCTTATAGTAG
To screen knockouts during assortment to complete replacement	
#1403-DCL1-9934	ATACCATCAATTTAATCGCCG
#1402-DCL1-10206r	TCTCTAACAATCATGACATCT
#1399-neo3-3351	TCGCCTTCTTGACGAGTTCT
To verify expression knockouts via RT-PCR	
#1471-DCL1-806111	AGGAATTTCAGCGTTTAGAAACGGTC
#1470-DCL1-805866r	CATAAAAGCACCCAACAACTG
#1413-ATU1-1997	TGCTCGATAACGAAGCCATCT
#1412-ATU1-2391r	GTGGCAATAGAAGCGTTGACA
To clone *DCL1* coding sequence for fusion to GFP	
#1445-DCL1-8048X	ATACTCGAGATGAGAAACAAACCTAAAGTTA
#1394-DCL1-11862Ar	ATAGGGCCCATCTTAGAAGGCTTTTTTTTCAGC
To assess enrichment of sequences after chromatin immunoprecipitation	
#1228-BTU-39f	GTACCACCACCGAGGGAGTGGGTG
#1229-BTU1-404r	TAACCAAATTGGTGCTAAGTTCTG
#1240-R-661f	ATGAGGTAAATTGAGGAGGGGAGC
#1241-R-834r	CATGTTTAGCTTGATAATTACTTTTCC
#1242-M-1418f	AAATTGAATAAGGAGACCAGCCTCTC
#1243-M-1635r	TATCAGTTCTCATCAAGTTGTAATGC
#1596-MAC-IR-676r	AGACCCGTAGAAAGCTAACTCCC
#1597-MAC-IR-902r	GAATGAAGGAGACATCGTCTAATA

system (Bio-Rad) as previously described (4, 7). Cells were allowed to complete conjugation in 10 mM Tris overnight before transfer to growth medium. Transformants were selected in 1× SPP containing 1.0 μg/ml $CdCl_2$ and 80 μg/ml paromomycin sulfate (PM) (Sigma) after preinduction of the *neo3* cassette in 1× SPP containing 0.5 μg/ml $CdCl_2$ for 5 to 6 hours at 30°C. Transformants were assessed for disruption of the germ line *DCL1* locus by crossing mature lines with CU427 and testing the cycloheximide-resistant cells (true progeny) for propagation of the *neo3* cassette allowing growth in medium containing $CdCl_2$ and PM (5). The heterozygous germ line knockouts were serially transferred (i.e., subcloned) into increasing concentrations of PM (from a starting concentration of 80 μg/ml to a final concentration of 350 μg/ml; the $CdCl_2$ concentration remained at 1 μg/ml), allowing for random assortment of macronuclear chromosomes until all wild-type alleles had been replaced with a disrupted copy. This assortment to completion of the knockout was monitored by PCR screening (primers are listed in Table 1) of crude cell lysates (9). Lines with complete macronuclear replacement of wild-type *DCL1* were converted to micronuclear homozygosity by crossing with star strain B*VI or B*VII to induce genomic exclusion. Exconjugates from these matings were screened for growth in $CdCl_2$/PM-containing medium to identify the transformant-derived lines and then crossed with CU427 to verify lines that were homozygous for the mutant allele (which produced 100% cycloheximide-, $CdCl_2$-, and PM-resistant progeny).

Southern blot analysis. Total genomic DNA was isolated from vegetative or conjugating cells by gentle lysis using the Promega genomic DNA isolation kit. DNA was digested with appropriate restriction enzymes before standard fractionation on agarose–1× Tris-borate-EDTA gels and subsequent transfer to nylon membranes (Osmonics) by downward capillary blotting in 0.5 M NaOH–1.5 M NaCl. Membranes were hybridized at 65°C with radiolabeled probes in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1 M Tris (pH 7.5)–0.5% sodium dodecyl sulfate (SDS)–2× Denhardt's solution for >16 h and then washed at 65°C in 1× SSC–0.5% SDS to remove nonspecific hybridization. All probes were radiolabeled with [α-^{32}P]dATP, random hexamers, and DNA polymerase I (Klenow fragment). Hybridization was visualized by autoradiography.

To examine the *DCL1* locus in knockout lines before and after genomic exclusion, isolated genomic DNA was digested with HindIII, fractionated by electrophoresis, and hybridized to a radiolabeled fragment corresponding a ~1-kbp region within exon 2. To assess failure of DNA rearrangement or chromosome breakage, total genomic DNA isolated from wild-type or Δ*DCL1* cells after ≥30 h of mating was digested with EcoRI, fractionated, and probed with the following radiolabeled fragments: a 1.9-kbp fragment from pDLCM3 detecting the M-element region (10), a 0.38-kbp fragment upstream of the CaM gene detecting CaM deletion element rearrangement (10, 29), or HhaI fragments B and C of Tt2512 germ line-specific sequence (11, 54). Chromosome breakage was assessed using a 0.8-kbp probe fragment that spans the EcoRI site at position 335013 of chromosomal scaffold CH445662 (GenBank accession number gi62422284). Hybridization was measured using a Personal FX PhosphorImager (Bio-Rad). Membranes were stripped and reprobed with an alpha-tubulin (*ATU1*) probe (11) under the same conditions as described above and quantified as a normalization control. *ATU1* hybridization to DNA from Δ*DCL1* and the wild type was used to measure the relative loading of each lane, and the average hybridization of two Δ*DCL1* samples was arbitrarily set as 1. This factor was used to normalize the quantification of the relative intensities between different samples.

RNA analysis. RNA was isolated from *Tetrahymena* by RNAsol extraction (15). Northern blot analysis was performed as described previously by Ausubel et al. (2). Small RNAs were fractionated on 15% polyacrylamide–urea–1× Tris-borate-EDTA gels, and larger RNAs were fractionated on 1.2% agarose–1× MOPS (morpholinepropanesulfonic acid)–1% formaldehyde gels as previously described (9). Random-primer-labeled *DCR1* and *DCR2* probes were 686-bp and 902-bp fragments corresponding to sequences between scaffold positions 66282 and 66968 of CH445757and positions 116071 and 116973 of CH445577, respectively. Plus- and minus-strand M-element riboprobes were synthesized from pMint7 and pMint2 as previously described (11). *ACT1* and *PDD1* coding region probes (11) were used for control hybridizations.

Reverse transcription-PCR (RT-PCR) was used to examine *DCL1* expression in vegetative cells and confirm its loss in knockout cells. Total RNA (4 μg) isolated at 2 and 4 h of mating from Δ*DCL1* (subclone 18.6) crossed with Δ*DCL1* (subclone 42.4) or wild-type cells was treated with DNase I for 30 min at 37°C, followed by inactivation by addition of EGTA (pH 8.0) to 2 mM and incubation at 65°C for 10 min. Random hexamers were used to prime reverse transcription of 2 μg of the treated RNA with SuperScript II reverse transcriptase (Invitrogen) according to the supplier's instructions. cDNA generated from 200 ng of starting RNA (equivalent to RNA from ~1,000 cells) was used in 34 to 42 cycles of PCR (annealing temperature of 50°C) using primers designed to amplify the *DCL1*

intron-containing region as a 245-bp genomic or 189-bp cDNA fragment (scaffold positions 805866 to 806111) or in 28 cycles (58°C annealing temperature) with *ATU1* primers (Table 1). To quantify the sensitivity of our RT-PCR reactions, 10-fold dilutions (10 pg to 1 fg [1 fg = ~1,000 molecules]) of a 1.8-kb in vitro-transcribed RNA corresponding to the same *DCL1* downstream region in our knockout construct were added to the 2 μg of cellular RNA prior to reverse transcription. PCR products were fractionated on agarose gels and visualized by ethidium bromide staining.

Monitoring of conjugation. Conjugating wild-type or Δ*DCL1* cells were fixed in Schaudin's fixative (2 parts $HgCl_2$ and 1 part 95% ethanol) at 2-h intervals after cells were mixed to initiate mating (52). DNA was then stained with 4′,6′-diamidino-2-phenylindole (DAPI), nuclear configurations were visualized using a Nikon E600 fluorescent microscope, and images were compared to those described previously by Martindale et al. (38) to determine the stage of development.

Localization of *DCL1*. An amino-terminal fusion of green fluorescent protein (GFP) to *DCL1* was created by PCR amplifying the entire coding sequence of *DCL1* from *Tetrahymena* genomic DNA. An XhoI site was added immediately preceding the ATG start codon, and an ApaI site was added downstream of the stop codon. This fragment was inserted in frame and downstream of GFP into the XhoI and ApaI sites within pIGF-1. This plasmid contains the S65T GFP variant expressed from a 1.2-kbp fragment of the MTT1 promoter all inserted into the NotI site of a pD5H8 rRNA gene vector derivative (22) allowing for autonomous replication. Either pIGF-1 or this GFP-*DCL1* fusion vector was introduced into wild-type cells (B2086 × CU428) or germ line *DCL1* knockouts (BVI $DCL1^{+}$ [$DCL1^{-}/DCL1^{-}$] × BVII $DCL1^{+}$ [$DCL1^{-}/DCL1^{-}$]) by conjugative electroporation (19). Mature transformants were starved overnight in 10 mM Tris and mixed to initiate mating. $CdCl_2$ was added to a final concentration of 0.08 to 0.1 μg/ml to induce expression of the fusion protein. Live cells were harvested 2 to 5 h after mixing, DAPI was added to between 1 and 5 μg/ml, and cells were suspended on glass slides in 2% methyl cellulose. GFP and DAPI fluorescence was visualized by epifluorescence microscopy. Images were captured using a Qimaging RetigaEX charge-coupled-device camera (Burnaby, British Columbia, Canada) and Openlab software (Improvision).

Immunoblotting and chromatin immunoprecipitation. Immunoblot analysis was done as previously described (37). B2086 × CU428, Δ*DCL1* × Δ*DCL1*, and Δ*TWI1* × Δ*TWI1* mating cells (2×10^5 cells/ml of each) were harvested at 7.5 h, 9 h, and 10.5 h after mixing of cells and boiled in lysis buffer prior to separation of proteins on 12% SDS-polyacrylamide gels. Proteins were transferred onto nitrocellulose membranes and incubated with modification-specific antibodies (Upstate Biotechnologies, NY). Antibodies were diluted as follows: anti-H3K9me2 (dimethyl), 1:2,000; or anti-H3K4me3 (trimethyl), 1:5,000. Immunoreactivity was detected using a West Pico kit (Pierce) and autoradiography.

Tetrahymena cells were crossed and prepared for chromatin immunoprecipitation with anti-H3K9me2 (dimethyl) antibodies 9 h into conjugation as described previously (50). After recovery of chromatin/antibody complexes using protein A-Sepharose, DNA was extracted using phenol-chloroform (1:1), and 30 ng was used as a template in PCR with primers (Table 1) specific for either the M element, the R element, or the intervening macronuclear retained region (50). PCR products were resolved on a 1.6% agarose gel and stained with 0.5 μg/ml ethidium bromide. Fluorescence intensities of each were quantified using 1D Image Analysis software (Kodak). Primers amplifying the *BTU1* locus were included in each reaction to generate a quantification standard.

RESULTS

The *Tetrahymena* genome encodes three Dicer-like proteins. To further characterize the relationship between germ line transcription, small RNAs, and DNA rearrangement, we searched the *Tetrahymena* genome for Dicer RNase homologues and found three putative Dicer-related coding sequences (Fig. 1). Two of these contain conserved RNA helicase and RNase III domains characteristic of previously described Dicer homologues (reviewed in reference 6); the third lacks the helicase domain but contains two RNase III domains as well as a dsRNA binding domain. While this work was in progress, partial or full cDNA sequences of these genes were deposited in GenBank (accession numbers are in Materials and Methods), and we have adopted the given names Dicer-1 (*DCR1*) and *DCR2* and Dicer-like-1 (*DCL1*), respectively, for our three identified candidates.

The abundant, small (~28-nt) RNAs that likely guide the extensive DNA rearrangements in developing somatic macronuclei are generated early in *Tetrahymena* conjugation (9, 40, 41). To determine whether any of these Dicer-related genes are particularly good candidates to generate these RNA species, we examined the expression of each, anticipating that the expression of one or more may be conjugation specific. *DCR1* and *DCR2* were observed to be expressed at low levels during all life cycle stages as evidenced by ~7-kb and ~6-kb transcripts, respectively, on Northern blots (Fig. 1D and E). *DCR1* appears to be expressed at slightly elevated levels by 6 h of conjugation, whereas *DCR2* appears to be expressed at its highest levels during vegetative growth. In contrast, *DCL1* expression was not detected by Northern blot in vegetative or starved cells (Fig. 1B). We also performed quantitative RT-PCR for which we should detect even one *DCL1* transcript per cell and verified a lack of appreciable expression in growing cells (Fig. 1C), although we did observe some low-level expression in starved cells (data not shown). In contrast, *DCL1* transcription was rapidly induced within the first 2 h of conjugation. Steady-state levels decreased rapidly between 4 and 6 h until accumulation resumed by 8 h of conjugation. The early high-level expression coincides with the initial accumulation of the development-specific small RNAs by 2 h after mixing of cells (40; also see Fig. 5), thus making the *DCL1*-encoded protein (Dcl1p) an attractive candidate to be involved in their generation.

***DCL1* is required for completion of development.** To determine whether this Dicer-related protein is necessary for the production of these small RNAs, we disrupted the *DCL1* gene in both the macronucleus and micronucleus by homology-directed gene replacement (Fig. 2A). We achieved this by introducing a *DCL1* knockout construct (Fig. 2B) into wild-type strains that directed replacement of 1.3 kbp (encoding 448 amino acids) of the *DCL1* gene with the *neo3* selectable cassette (48) that confers resistance to the drug PM. The initial transformants selected had, on average, half of the somatic *DCL1* gene copies disrupted within the polyploid macronucleus. These lines were subcloned successively into medium containing increasingly higher doses of PM that, due to random segregation (assortment) of macronuclear chromosomes, allowed us to generate lines for which all somatic copies of the *DCL1* gene were replaced with the mutant allele. This macronuclear assortment was assessed by PCR (not shown) and by Southern blot analysis (Fig. 2B) to verify complete loss of the wild-type *DCL1* gene copies except those remaining in the germ line micronucleus. We also verified the disruption of germ line copies of the *DCL1* gene within the micronuclei of these strains using genetic crosses and phenotypic analyses described in Materials and Methods. These initial lines were heterozygous [Fig. 2B, Δ*DCL1* (*n3*/+)] for the knockout allele and were converted to micronuclear homozygosity via genomic exclusion by crossing each line to micronucleus-defective "star" strain B*VI or B*VII. The resulting abortive conjugation proceeds through meiosis and the generation of four haploid (gametic) micronuclear products in the knockout lines, one of which is selected to regenerate into a diploid micronucleus due to the failure of this mating partner to receive a

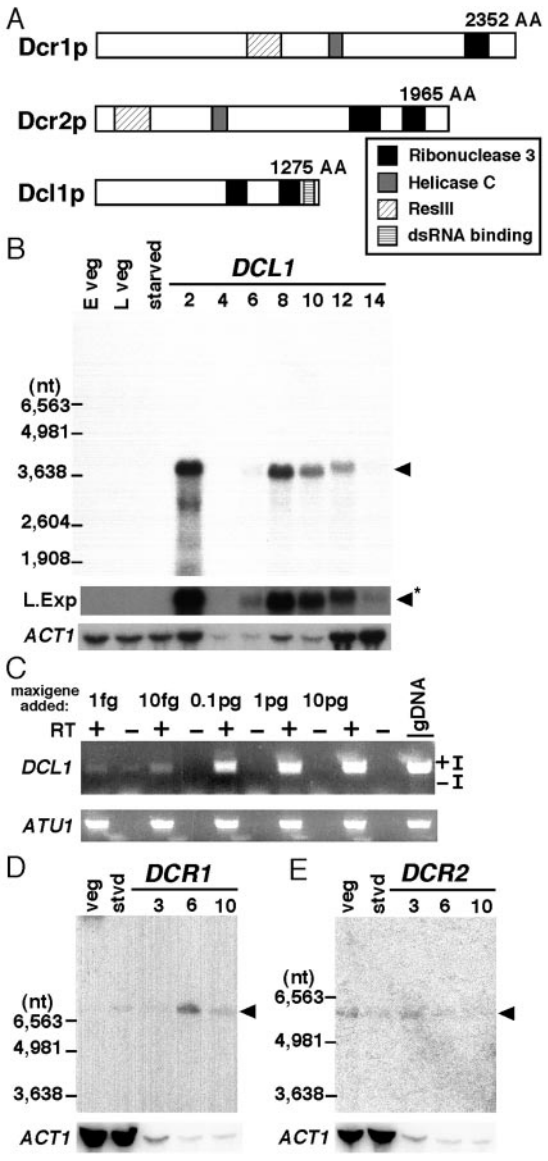


FIG. 1. *Tetrahymena thermophila* encodes three Dicer-like proteins. (A) Total predicted protein length from the *Tetrahymena* genome project is indicated at the right end of each schematic (see Materials and Methods for the locations of each within chromosomal scaffolds). Conserved domains identified by Pfam are indicated by the shaded or hatched boxes (see key). AA, amino acids; dsRNA, small RNA. (B, D, and E) Northern blot analysis was used to examine the expression of Dicer homologues at different life cycle stages. "E veg" refers to early-log-phase vegetative growth, and "L veg" refers to late-log/early-stationary-phase growth. The numbers above each lane denote the hour of mating when RNA was isolated. Arrowheads indicate transcript hybridization. The migration of RNA size markers (Promega) is presented on the left. (B) For *DCL1* expression, a 1-day autoradiogram exposure is shown above a 3-day exposure (L.exp) (arrow with asterisk) that is used to reveal low-level expression at 6 h of conjugation and to highlight the absence of expression in vegetative cells. (C) RT-PCR analysis of RNA isolated from vegetative CU428 cells. The indicated amount of an in vitro-transcribed RNA was added to each 2-μg sample prior to reverse transcription to determine the sensitivity of the assay (1 fg = ~1,000 transcripts). One-tenth (200-ng equivalents) was used in each PCR. *ATU1* amplification was used to confirm cDNA synthesis. gDNA, genomic DNA. (D and E) *DCR1* and *DCR2* expression, respectively, was detected by 5-day exposure of blots to autoradiograph film. To compare loading between samples, each blot was stripped and rehybridized with an actin probe to reveal *ACT1* expression, which is constitutive in vegetative (veg) and starved (stvd) cells but is initially down-regulated early during conjugation before returning to the vegetative level late in development (11 to 12 h).

donor gametic nucleus from the star strain. The cell lines that are now homozygous [Fig. 2B, Δ*DCL1* (*n3*/*n3*)] for the knockout allele in their micronuclei were identified by PCR and verified by genetic crosses with wild-type cells that resulted in 100% propagation of the PM-resistant phenotype to their progeny (data not shown).

Tetrahymena lines lacking all copies of *DCL1* (Δ*DCL1*) exhibited vegetative growth typical of wild-type strains, indicating that this gene is dispensable. This result was not surprising to us, as we could not detect *DCL1* expression in vegetatively growing cells (Fig. 1B). We also generated somatic (macronuclear) knockouts of *DCR1* and *DCR2* (D. L. Chalker, unpublished data). Complete, somatic *DCR1* knockouts exhibited no obvious growth defects, and thus, this gene also appeared to be nonessential. In contrast, after multiple rounds of subcloning, we were unable identify *DCR2* knockout-transformed lines that had reached complete replacement of the wild-type gene with the disrupted allele, a result that suggests that this gene is essential for vegetative growth. These findings indicate that the different RNase III proteins of *Tetrahymena* are not completely overlapping in function.

As *DCL1* expression is induced during conjugation, we crossed two Δ*DCL1* lines to examine the effect of its disruption on development. We verified the loss of *DCL1* expression in these crosses by RT-PCR using oligonucleotide primers that would have detected even low-level transcription of the large carboxy-terminal region remaining in our knockout strains downstream of the *neo3* cassette (Fig. 3A and data not shown). The progression of these mutant cells through conjugation was compared to that of wild-type cells by harvesting cells at 2-h intervals and staining with DAPI to visualize the nuclear configurations that are diagnostic of particular stages of development (Fig. 3B) (38). Δ*DCL1* mating pairs were able to complete most stages of conjugation, although their overall progression was slightly slower than that of wild-type pairs (an observation that was more apparent in some crosses than others but is further evidenced in the crosses shown by the somewhat delayed decrease in *ACT1* expression and accumulation of *PDD1* transcripts [see Fig. 4]). The most dramatic difference observed was that Δ*DCL1* cells failed to eliminate one of the two progenitors of the new micronucleus (Fig. 3C). This corresponds to the last step of conjugation just prior to the return to vegetative growth. This finding suggested that the loss of *DCL1* results in a developmental arrest phenotype. These cells do not appear to fully amplify the genome in the developing macronuclei, as the fluorescence intensity of these nuclei relative to the micronuclei in the same cell upon DAPI staining is lower overall when compared to that of wild-type cells (Fig. 3C and data not shown).

To confirm that Δ*DCL1* cells had arrested late in development, we isolated individual mating pairs of wild-type and Δ*DCL1* cells into separate drops of growth medium and compared their fates. Whereas most wild-type pairs had resumed

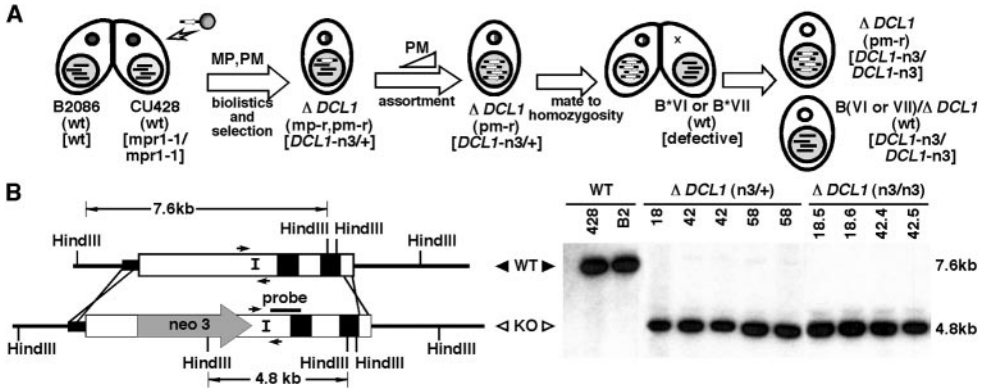


FIG. 2. Germ line knockout of *DCL1*. (A) Knockout strategy. Biolistic transformation was employed to introduce the *DCL1-neo3* (n3) knockout construct into wild-type (wt) strains, and transformant progeny were selected in PM and subsequently in 6-methyl purine (MP) and then assorted to complete replacement in increasing concentrations of PM. The solid lines in the diagram indicate wild-type chromosomes; the lines with white arrows indicate knockout chromosomes. Transformants were converted to homozygosity by genomic exclusion crosses. Exconjugants were separated and assayed for the presence of the knockout construct. Strains homozygous for the construct in the micro- and macronucleus were used for phenotypic analyses, while those homozygous in the micronucleus and wild type in the macronucleus were transformed with the GFP-*DCL1* construct and used in localization studies. Names are given below each strain with the macronuclear phenotype in parentheses and the micronuclear genotype in brackets. pm-r, paromomycin resistant. mp-r, 6-methyl purine resistant. (B) Southern blot analysis was used to verify the genotype Δ*DCL1* strains. Total genomic DNA was isolated, digested with HindIII, and hybridized with the *DCL1* probe shown in the diagram to the left. The region replaced by the *neo3* cassette (shaded arrow) relative to the conserved RNase III domains (solid boxes) and the lone intron (I) is depicted. The wild-type (WT) (closed triangle) and knockout (KO) (open triangle) HindIII fragments are 7.6 kb and 4.8 kb, respectively. Genomic DNA was analyzed from the two original wild-type strains, five somatic Δ*DCL1* strains that are heterozygous in the micronucleus (*n3*/+), and four somatic *DCL1* knockouts that are homozygous in the micronucleus for Δ*DCL1* (*n3*/*n3*) and which were derived from the heterozygotes shown.

vegetative growth as four or more cells were visible in each drop of medium by 24 h after mixing, the majority of Δ*DCL1* cells were still paired at this time. Most Δ*DCL1* pairs eventually separated, but the exconjugates never divided. Our results clearly show that *DCL1* is essential for *Tetrahymena* to complete development and return to vegetative growth.

The observed properties of the Δ*DCL1* strains we created were not entirely congruent with recently reported growth and developmental phenotypes of Δ*DCL1* strains generated by Mochizuki and Gorovsky (42). Most notably, their *DCL1* mutant strains exhibited defects during micronuclear division, resulting in the loss of chromosomal DNA that we did not observe in our knockout lines. In addition, their Δ*DCL1* cells showed significant aberrations during meiosis and progressed through conjugation asynchronously. This asynchrony may simply be a consequence of the observed meiotic defects. Based on these phenotypes, those authors concluded that *DCL1* has distinct roles in micronuclear chromosome segregation, meiotic prophase, and macronuclear development that our study cannot fully support. Conflicting results between their study and ours are not due to the differences in genetic backgrounds, as both studies used the same laboratory strains, but are likely due to the different knockout constructs used (see Discussion).

Δ*DCL1* cells do not generate germ line-specific small RNAs and accumulate nongenic micronuclear transcripts. Disruption of genes (e.g., *PDD1* and *TWI1*) that fail to stabilize small RNAs that have been linked to developmentally programmed DNA rearrangements exhibits developmental arrests very similar to those we observed for Δ*DCL1* cells (12, 40). We therefore asked whether our mutant cell lines fail to generate this specific class of small RNAs. These 28- to 30-nt RNA species are easily visualized on ethidium bromide-stained polyacrylamide gels by 2 h after mixing of wild-type cells and persist throughout conjugation (Fig. 4A and B) (40). In contrast, these small RNAs were undetectable in RNA isolated from Δ*DCL1* mating pairs at any point during development, indicating that this Dicer-related protein is required for their generation. We did not observe a reduction in the small RNA accumulation at any stage of conjugation upon mating of two *DCR1* somatic knockout lines, suggesting that this other Dicer-related protein is unnecessary for their generation (J. A. Motl and D. L. Chalker, unpublished data). In Δ*DCL1* mating cells, a barely perceptible amount of 23- to 24-nt RNAs appeared at later time points (Fig. 4G). These may result from processing of dsRNAs by *DCR1* and/or *DCR2*. Whether these smaller RNAs occur in wild-type cells and play a role in development and/or DNA rearrangement will require further investigation.

Clearly, the bulk of development-specific small RNAs are not produced in cells lacking *DCL1*. To assess whether the generation of small RNAs homologous to specific germ line-limited sequences that undergo DNA rearrangement is also affected in these mutants, we transferred stained RNAs to nylon membranes and hybridized these with strand-specific probes to detect ones homologous to the well-characterized M deletion element. We have previously shown that the M element is bidirectionally transcribed during development of wild-type cells (11) and that its small RNAs accumulate during the first 3 to 4 h of conjugation before they decline to a low steady-state level (9) (Fig. 4C and D). Just as we did not observe the bulk of developmental small RNAs in mating Δ*DCL1* cells, we could not detect small RNAs with probes specific to either strand of the M deletion element. Thus, *DCL1* is required to generate the small RNAs that correspond

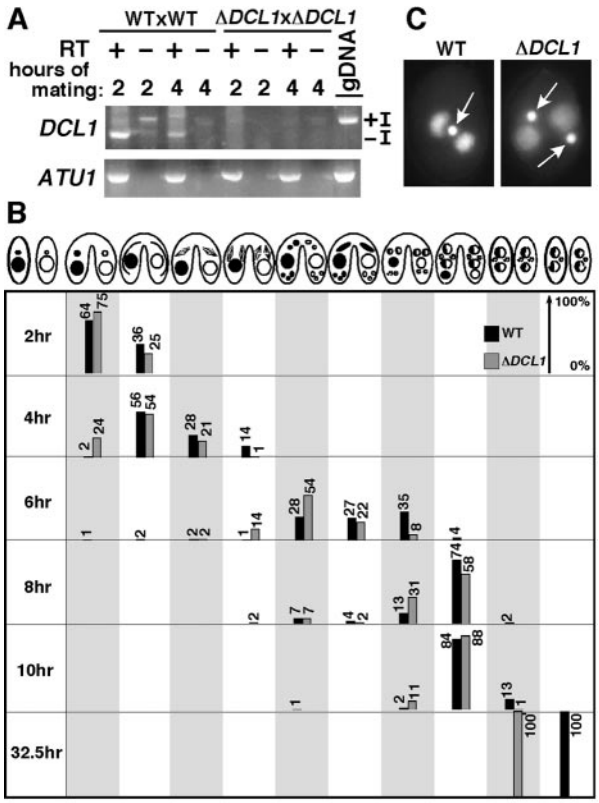


FIG. 3. Δ*DCL1* strains arrest late in conjugation. (A) RT-PCR was used to confirm that *DCL1* was not expressed in knockouts. Total RNA isolated at 2 and 4 h of mating was converted to cDNA to be used as a template for PCR amplification with *DCL1*-specific primers (Table 1), which are indicated as arrows in the knockout construct diagram (Fig. 2B). Identical reactions with *ATU* primers and wild-type (WT) genomic DNA (gDNA) served as positive controls for cDNA conversion and PCR amplification, respectively. Omission of reverse transcriptase (RT) controlled for the possibility of contaminating DNA in the reactions. (B) A diagram of the nuclear configuration diagnostic of individual stages is presented. The progression of wild-type (black bars) and Δ*DCL1* (gray bars) cells at individual time points after mixing was assessed by fluorescence microscopy of DAPI-staining cells. Numbers indicate the percentage of cell pairs at that stage of conjugation. (C) DAPI-stained cells showing the end point of development reached by wild-type and Δ*DCL1* cells at 32.5 h of conjugation. White arrows point to micronuclei.

to germ line-limited sequences that undergo DNA rearrangement.

If the nongenic transcripts of the M element are the precursors of these small RNAs, these larger transcripts should accumulate during conjugation of Δ*DCL1* cells. RNAs isolated from wild-type and Δ*DCL1* mating cells were fractionated on denaturing agarose gels and hybridized with M-element probes. In wild-type cells, transcripts homologous to both strands accumulate to relatively low levels, reaching their peak steady-state abundance approximately 6 h into conjugation (Fig. 4E and F) (11). On the other hand, M-element small RNA abundance peaks earlier, between 3 and 4 h of conjugation (Fig. 4C and D). We have argued that the large transcripts accumulate to their highest levels only after they cease being processed into small RNAs about 4 h into conjugation, when we see their levels begin to decline (9); however, this assumption requires that the large transcripts are indeed precursors of the small RNAs. In developing Δ*DCL1* cells, we observed that RNA species between ~200 and >1 kb corresponding to both M-element strands accumulated to significantly higher levels than in wild-type cells. The peak of accumulation was reached by 4 h into conjugation with lower steady-state levels persisting into the later time points. This peak in accumulation in the Δ*DCL1* cells is consistent with the interpretation that the larger transcripts, while still being synthesized, are processed by Dcl1p primarily during the first few hours of conjugation, when *DCL1* expression peaks. These data provide the first direct evidence that the larger bidirectional transcripts are precursors of the small RNAs that target DNA rearrangement of the M element and that Dcl1p is involved in this processing.

Dcl1p is localized in the micronucleus. The observation that M-element small RNAs cease to accumulate rather early in conjugation also corresponds to the drop in *DCL1* steady-state mRNA observed between 4 and 6 h after mixing (Fig. 1B). To further investigate the relationship between *DCL1* and small RNA generation, we examined the localization of Dcl1p by generating an amino-terminal fusion to GFP. This fusion protein was expressed ectopically under the cadmium-inducible MTT1 promoter (48) and maintained in *Tetrahymena* cells on a high-copy, rRNA gene-based replicating vector. GFP expressed alone from vector pIGF-1 produces bright green cells during either vegetative growth or conjugation, typically within 1 hour of cadmium addition (Fig. 5 and data not shown). The GFP-*DCL1* fusion construct was transformed into both wild-type strains and Δ*DCL1* lines to control for the possibility that localization was affected by the presence of endogenous Dcl1p. Induction of GFP-*DCL1* expression by cadmium addition to vegetatively growing cultures produced very little detectable GFP fluorescence and no specific localization, which is consistent with our inability to detect expression or observe a phenotype upon disruption in growing cells. This suggests that the fusion protein is either poorly translated or rapidly degraded compared to GFP alone. When we crossed GFP-DCL1-containing cells and induced expression at the beginning of conjugation, we observed distinct localization of the fusion protein to meiotic micronuclei in both wild-type cells and those lacking endogenous *DCL1* (Fig. 5). We typically detected GFP-*DCL1* in <10% of mating pairs, which may be indicative of variable expression from the MTT1 promoter in early conjugation or, more likely, that the protein has a short half-life, as GFP requires a period of time after translation to mature before it can fluoresce. We could first detect GFP-*DCL1* in micronuclei just prior to the onset of meiosis, but the fusion protein was most consistently visualized in late prophase, when the micronucleus forms an elongated "crescent" structure (47, 49). Comparison of the GFP localization to both the corresponding bright-field and DAPI-stained images reveals that Dcl1p is present in the nucleoplasm and appears to be primarily excluded from the DNA itself. In some cells, localization was somewhat punctate in crescent micronuclei for which the DNA was less condensed than the image shown, including specific accumulation in the narrow end of these structures. We detected little specific localization of the fusion protein after prophase, an observation that is congruent with the reduction in endogenous *DCL1* transcription and the cessation of small RNA generation. Taken together, our results allow us to con-

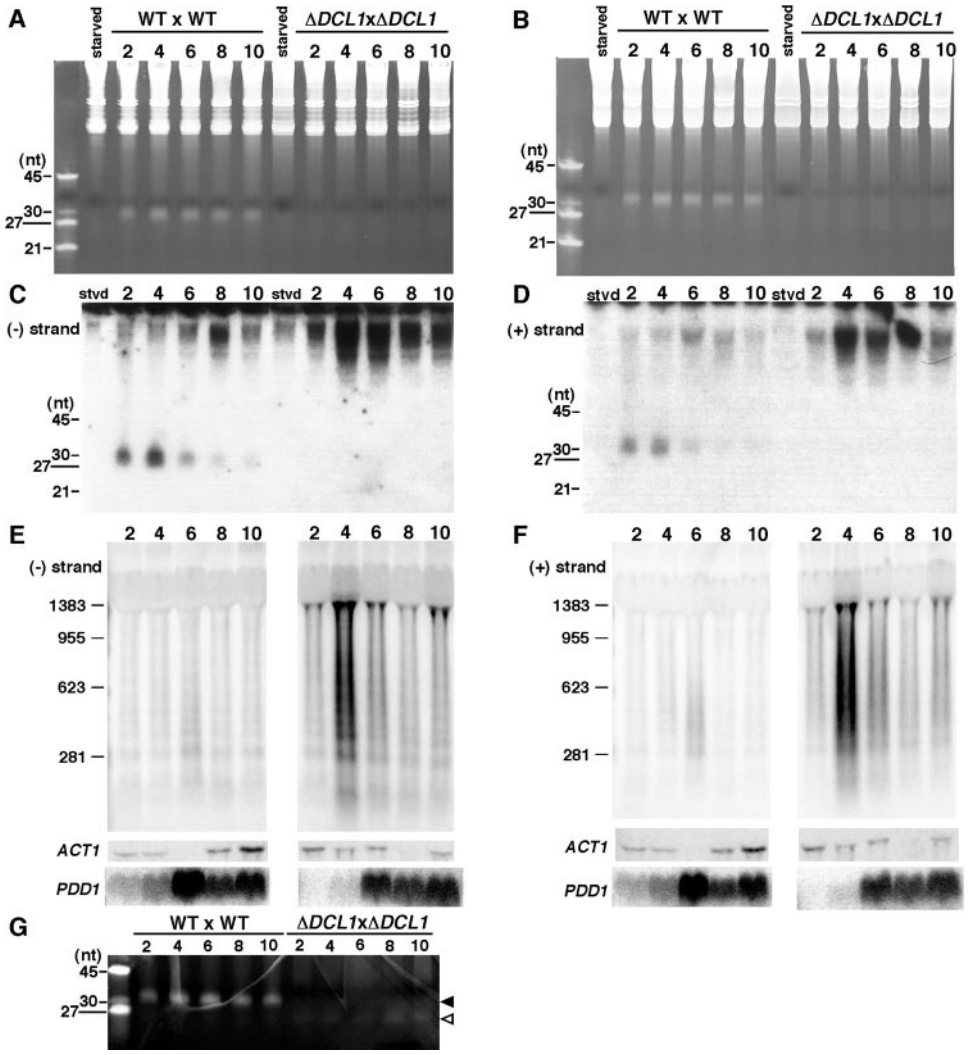


FIG. 4. Conjugating Δ*DCL1* strains exhibit loss of small RNA production and germ line transcript accumulation. RNA isolated at 2-h intervals from the start of conjugation was separated by electrophoresis on either 15% polyacrylamide–urea gels (A to D) or 1.2% agarose–formaldehyde gels (E and F), ethidium bromide stained (A and B) or transferred to membranes (C to F), and hybridized to plus-strand (+)- and minus-strand (−)-detecting M-element riboprobes as indicated. The migration of oligonucleotide (A to D) or RNA size standards (E and F) are indicated to the left of each panel. (A to D) RNA species of 28 to 30 nt (arrowhead) were observed throughout conjugation of wild-type (WT) cells but were undetected in starved (stvd) cells or Δ*DCL1* conjugating strains. (E and F) Northern blot analysis of Δ*DCLl* strains shows an accumulation of M-element bidirectional transcripts. Each filter was rehybridized with *ACT1* and *PDD1* probes for comparison of loading between samples and is shown below the corresponding panel. (G) Stained polyacrylamide gel of RNA isolated from wild-type or Δ*DCL1* cells that reveals smaller species of short RNAs (open arrowhead) migrating below the position of the abundant ~28-nt species (solid arrowhead).

clude that developmental small RNAs are generated by Dcl1p in the micronucleus.

Δ*DCL1* cells fail to eliminate germ line-limited sequences from developing macronuclei. As the transcription of germ line-limited sequences and the generation of small RNAs have been linked to *Tetrahymena* genome rearrangement, the failure of Δ*DCL1* cells to complete conjugation is likely due to a failure in this process. To examine this possibility, we isolated total genomic DNA from populations of wild-type and Δ*DCL1* cells well after the normal completion of macronuclear development (24 to 32 h) and examined the state of rearrangement of several loci by Southern blot analysis (Fig. 6). The 10-kbp genomic region of micronuclear chromosome 4 centered around the M deletion element contains two other germ line-limited sequences, designated the L (left) and R (right) elements (1) (Fig. 6A). Postconjugative wild-type cells had completely eliminated all three germ line-limited DNAs from newly formed somatic macronuclei as the ratio of rearranged to unrearranged chromosomes approached the ratio of macronuclear DNA to micronuclear DNA (15:1 to 20:1) typical for vegetative *Tetrahymena* (Fig. 6B). This was also observed for the germ line-limited sequence located upstream of the *Tetrahymena* calmodulin (CaM) gene (Fig. 6A and C) (29). In contrast, the chromosomes of postconjugative Δ*DCL1* cells retained the four germ line-limited sequences that are normally eliminated from the M-element genomic region and the CaM locus (Fig. 6B and C). The copies of the rearranged forms are likely derived from the 10 to 20% of unmated cells in the

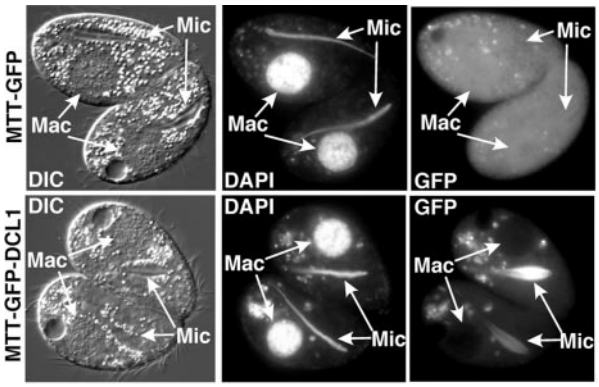


FIG. 5. The *DCL1* protein is localized to meiotic micronuclei. *Tetrahymena* transformed with pIGF-1, which contains GFP only (top panels) or a GFP-*DCL1* construct (bottom panels) was mated to nontransformed wild-type or Δ*DCL1* cells, and expression of the fusion protein was induced by the addition of $CdCl_2$ upon mixing of cells. Differential interference contrast (DIC) light microscopy of single pairs is displayed adjacent to fluorescence imaging of DAPI-stained DNA with and the localization of GFP or the GFP-*DCL1* fusion protein. The *DCL1* protein is observed exclusively in the extrachromosomal space in the elongated, meiotic (prophase) micronuclei (labeled as Mic). The location of the macronucleus (Mac) is also indicated. The mating partner with the brighter GFP fluorescence signal is likely the transformant expressing the GFP fusion that typically shows greater fluorescence despite extensive cytoplasmic exchange within the pair. Background fluorescence apparent in vacuoles is common in DAPI and GFP fluorescence in live *Tetrahymena* and accounts for the cytoplasmic signal observed.

population, and these appear overrepresented due to the polyploidy of the parental macronuclei and the apparent underamplification of developing macronuclei.

To further investigate the extent of failed DNA rearrangement, we examined the fate of the repetitive, germ line-limited sequence represented in clone Tt2512 (54). This >7-kbp sequence is present in the micronucleus at an estimated 50 to 100 loci and is entirely eliminated from the somatic macronucleus. Comparison of the level of hybridization of total genomic DNA from postconjugation wild-type and Δ*DCL1* cells revealed extensive retention of this sequence within the DNA isolated from the mutant strains (Fig. 6D and E). As the Tt2512 probe used hybridized exclusively to germ line-limited sequences, the observation that the majority of hybridizing fragments were much more abundant (quantified as 12- to 15-fold for the two strongest-hybridizing fragments) in the DNA of the mutant cell populations indicates that this repetitive sequence remained within the developing macronuclear genome at most or all its loci. Therefore, for each IES examined, we observed extensive failure of germ line DNA elimination, which demonstrates that Dcl1p is required for *Tetrahymena* genome rearrangements.

We also assessed the occurrence of chromosomal breakage in our *DCL1* mutants (Fig. 7). A membrane containing fractionated EcoRI-digested genomic DNA from postconjugative cells was hybridized with a radiolabeled probe that detected the left end of macronuclear chromosomal scaffold CH445662 (GenBank accession number gi62422284). In DNA from wild-type cells, the predominant hybridizing fragment was a 2.5-kbp species (Fig. 7), which is the size expected after chromosome breakage and addition of 250 to 300 bp of telomeric repeats. This species was absent in DNA recovered from the Δ*DLC1* mating cell population; but instead, the 10.5-kbp micronucleus-specific fragment was in higher abundance relative to the same fragment in wild-type cells. A less abundant population of fragments whose average size was ~2.6 kbp was observed in equal abundance in both populations. We interpret this by suggesting that the smaller, abundant fragments in wild-type cells are the result of new chromosome breakage and new telomere addition, while the majority of the larger fragments are derived from the macronuclei of the remaining unmated cells in the population that on average had longer telomeres. Thus, it appears that chromosome breakage is also perturbed in these *DCL1*-deficient cells.

Histone H3K9 methylation occurs, but is not targeted, in the absence of small RNAs. The chromatin associated with germ line-limited sequences is specifically methylated on lysine 9 of histone H3 prior to DNA rearrangement. Mutant cells lacking the chromodomain-containing protein Pdd1p or the Argonaute homologue Twi1p fail to establish this chromatin mark and eliminate germ line-limited DNA sequences (12, 34, 40, 50). This has led to the model that developmental small RNAs target this chromatin modification specifically to DNA segments that are eliminated from developing macronuclei (see reference 43). We expected that the disruption of *DCL1* that results in failure to generate the small RNAs would also abolish the establishment of the H3K9-methylated chromatin in developing macronuclei. Much to our surprise, when we examined total histone H3K9me2 on Western blots (Fig. 8A), we detected very little change in the overall modification of chromatin in Δ*DCL1* mating cells relative to wild-type cells, whereas our control using Δ*TWI1* cells showed no detectable H3K9me2 as previously reported (34). Similarly, after staining fixed conjugating cells with anti-H3K9me2 antibodies, little difference in the amounts of immunofluorescence was observed between wild-type and Δ*DCL1* cells (data not shown). Therefore, the generation of specific, small RNAs is not required for the establishment of this chromatin modification.

The fact that H3K9 methylation still occurred in our knockouts in the absence of the *DCL1*-generated small RNAs provided us the opportunity to ask whether these RNAs target this modification to specific loci. In wild-type cells, immunoprecipitation of conjugating cell chromatin with anti-H3K9me2 antibodies preferentially recovers germ line-limited sequences as demonstrated by a four- to fivefold enrichment of the adjacent M and R deletion elements but not the macronucleus-retained region between these IESs (Fig. 8B) (50). We did not observe enrichment of these same sequences in a chromatin immunoprecipitation assay of conjugating Δ*DCL1* cells. These data, along with our observation that these cells fail to eliminate numerous germ line-limited sequences, including the two examined here, provide convincing evidence that small RNAs direct this chromatin modification to the proper loci.

DISCUSSION

Distinct roles for Dicer-related proteins in growth and development. Three Dicer-related proteins encoded within the genome of *Tetrahymena thermophila* are each expressed at

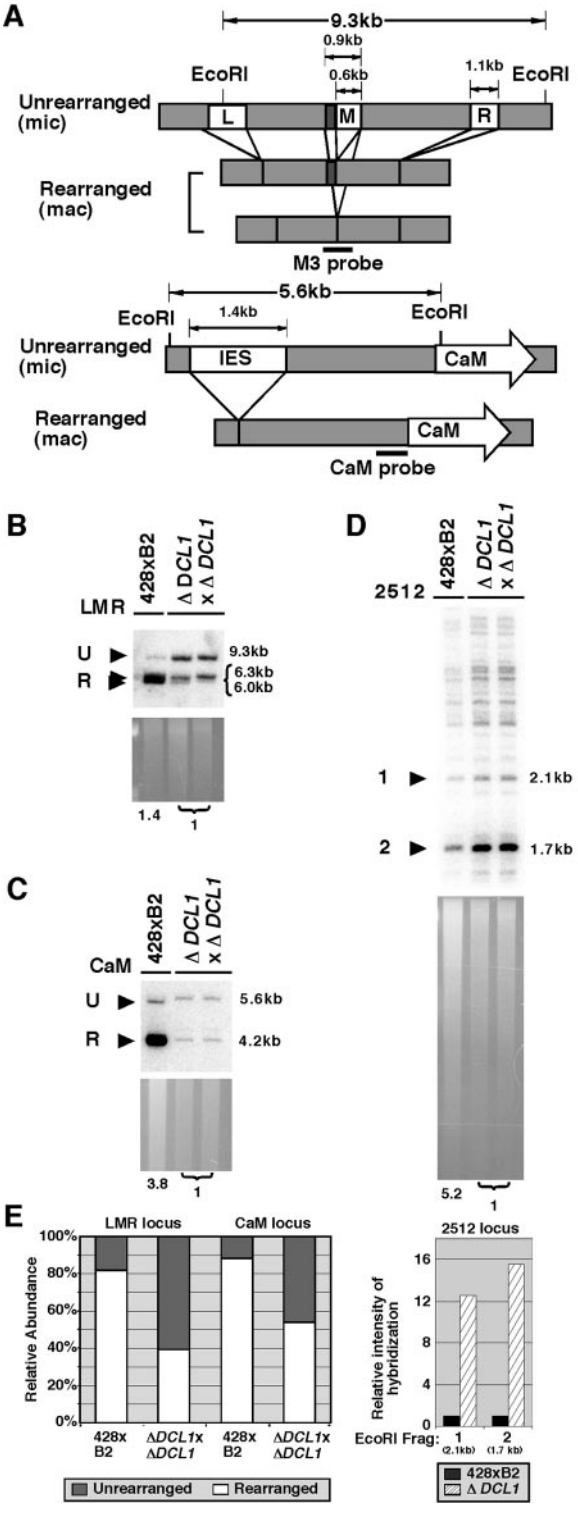


FIG. 6. *DCL1* knockouts fail to excise micronucleus-limited DNA. (A) Diagram of the micronuclear (mic) and macronuclear ¯ L/M/R and CaM loci. IESs are indicated as open boxes, and macronucleus-destined sequences are indicated as shaded boxes. The L/M/R locus contains three IESs named the L (left), M (middle), and R (right) deletion elements. The M element has two alternative, leftward deletion boundaries that generate two rearranged forms of this loci at nearly equally frequencies. The genomic region upstream of the CaM gene contains a 1.4-kbp IES. The locations of the M3 and CaM probes used in Southern blot hybridization are shown (10). (B to D) Southern blot hybridization of DNA from postconjugative wild-type and Δ*DCL1* cells was used to assess IES rearrangement efficiency. For each blot, EcoRI-digested genomic DNA was fractionated, and specific loci were detected with (B) M3-, (C) CaM-, or (D) Tt2512-radiolabeled probes. The stained gel prior to blotting is shown for comparison of loading. Membranes were stripped and probed with the *ATU1* probe (not shown) to measure relative quantities of DNA loaded in each lane, which are reported at the bottom of each lane, with the Δ*DCL1* lanes set to 1 for ease of comparison. (E) Quantification of the rearranged (R) and unrearranged (U) forms in wild-type cells compared to Δ*DCL1* cells for the L/M/R and CaM loci and relative hybridization intensities (adjusted to *ATU1* hybridization) of EcoRI fragment (Frag) 1 (2.1 kb) and fragment 2 (1.7 kb) in the Tt2512 region in wild-type compared to *DCL1* knockout cells. The measured Tt2512 hybridization for wild-type cells was arbitrarily set to 1.

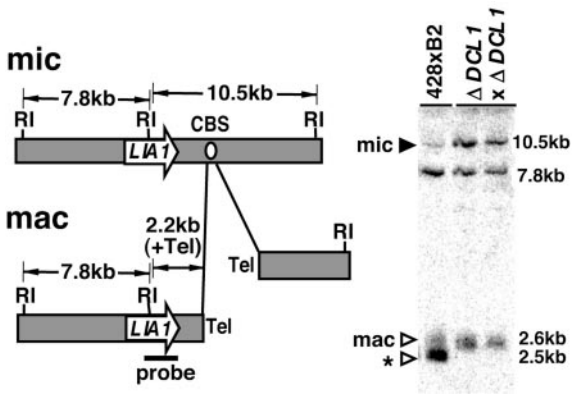


FIG. 7. Chromosome breakage does not occur in *DCL1* knockouts. The diagram shows the left end of macronuclear chromosomal scaffold CH445662, which contains the LIA1 gene within 2.5 kbp from the telomere (Tel), and the deduced ~18-kbp region of the micronuclear (mic) chromosome from which it is derived. The predicted location of the chromosomal breakage sequence (CBS) (white oval) is depicted as well as the relevant EcoRI (RI) restriction sites used for the Southern blot analysis of genomic DNA from postconjugative wild-type and Δ*DCL1* cells used to assess chromosome breakage. The probe spans the central EcoRI site and detects a 7.8-kbp fragment common to both nuclei, which can be used to compare amounts of DNA loaded between each lane, and either the ~10.5-kbp micronucleus-specific fragment (solid arrowhead) or a 2.5- to 2.6-kbp macronucleus-specific fragment (2.2 kbp of unique sequence plus 300 to 400 bp of telomeric DNA) (open arrowheads). The shorter macronuclear fragments marked by the asterisk appear only in wild-type samples and are likely derived from new chromosomal breakage and telomere addition in developing macronuclei, while the larger fragments are presumed to be derived from the macronuclei of unmated cells with, on average, longer telomeres.

different levels and stages of this ciliate's life cycle (Fig. 1). *DCR1* and *DCR2* genes are expressed in vegetative cells, and our inability to completely knock out *DCR2* suggests that it is essential for growth. Neither *DCR1* nor *DCL1* is required for vegetative growth, as we generated complete somatic knockouts of these genes. However, we found that *DCL1*, which is expressed at high levels only during conjugation, is required to complete development. *DCR1* somatic knockouts complete development normally, but as we disrupted only the somatic copies of this gene and not those in the silent, germ line micronucleus, we cannot rule out that it has a role in late stages of conjugation after zygotic expression begins and the wild-

type germ line copies would be expressed. Similar findings were recently reported by Mochizuki and Gorovsky (42). It is clear that these three putative RNase III enzymes have distinct, nonredundant functions in growth and/or development.

The extensive genome rearrangements that occur during *Tetrahymena* development are guided by an RNAi-like mechanism (reviewed in reference 43) which overall exhibits remarkable similarities to the establishment of heterochromatin in other eukaryotes (reviewed in reference 23). In addition to finding that Δ*DCL1* cells arrest late in development, we observed that Δ*DCL1* cells do not generate abundant developmental small RNAs but instead accumulate germ line-limited transcripts homologous to eliminated sequences (Fig. 4). This provides direct evidence that the bidirectional transcripts synthesized during meiotic prophase (11, 39, 49) are the precursors of these abundant RNA species. Furthermore, postconjugative Δ*DCL1* cells failed to eliminate all germ line-limited sequences assayed, providing proof that these RNAs guide genome rearrangement. Δ*DCL1* cells also failed in chromosomal breakage, but we have less evidence to argue that this is a direct effect and not an indirect result of the developmental arrest observed that could occur prior to completion of this process.

The phenotypes of our Δ*DCL1* cells had many similarities but also some marked differences from those described in the *DCL1* study published previously by Mochizuki and Gorovsky (42). One obvious disparate phenotype is their reported defects in micronuclear division during vegetative growth that we do not observe in our Δ*DCL1* cells. We have difficulty attributing this phenotype to the loss of *DCL1* because we cannot detect its expression in vegetative cells using RT-PCR conditions that would have detected even one transcript per cell (Fig. 1C). Nevertheless, we must note that our Δ*DCL1* cells retained additional coding sequence, compared to the cells of the other study, that could potentially rescue some *DCL1* function. To eliminate the possibility of rescue of some *DCL1* function due to undetected expression of a C-terminal polypeptide that retained catalytic activity, we generated new *DCL1* macronuclear knockout lines that removed an additional 464 codons including most of both RNase III domains. These Δ*DCL1* strains exhibited the same developmental arrest phenotype as our original knockout lines, as the majority of cells died with two micronuclei and two developing macronuclei (data not shown) rather than the asynchronous arrest observed by Mochizuki and Gorovsky. Our knockout strains also retained the 189 amino-terminal codons that could theoretically rescue some *DCL1* function, as we did not remove the promoter and can detect transcription of this region in conjugating cells by RT-PCR. However, any rescue must be independent of the RNase activity of this protein. The phenotypes we report above are fully consistent with an exclusive and critical role for Dcl1p in macronuclear development. Nevertheless, if expression of a partial polypeptide within our Δ*DCL1* lines rescues secondary roles of this protein in micronuclear maintenance, the *DCL1* allele that we have created has allowed us to separate the distinct roles of this Dicer-related protein.

An alternative explanation is that the micronucleus-associated phenotypes of the Δ*DCL1* strains reported by Mochizuki and Gorovsky are due to perturbation of another gene in addition to disruption of *DCL1*. One obvious candidate would be the predicted open reading frame immediately upstream of *DCL1*. The last predicted codon of this open reading frame is just under 1 kbp from the *DCL1* start methionine. Disruption of this upstream coding sequence, along with *DCL1*, resulted in suboptimal growth with noticeable cell death, particularly during stationary phase and starvation (D. L. Chalker, unpublished); thus, partial loss of function of this upstream gene could conceivably interfere with proper micronuclear segregation, as was observed previously (42). These double-knockout strains were unable to efficiently initiate conjugation due to starvation defects that inhibited our ability to examine meiotic phenotypes. However, the sequence of this predicted gene was not altered and remains transcribed in their strains (K. Mochizuki, personal communication), so an obvious perturbation is not evident. All tests aimed to resolve the differences between the different *DCL1* mutant strains have proved inconclusive.

While *DCL1* is essential for the accumulation of develop-

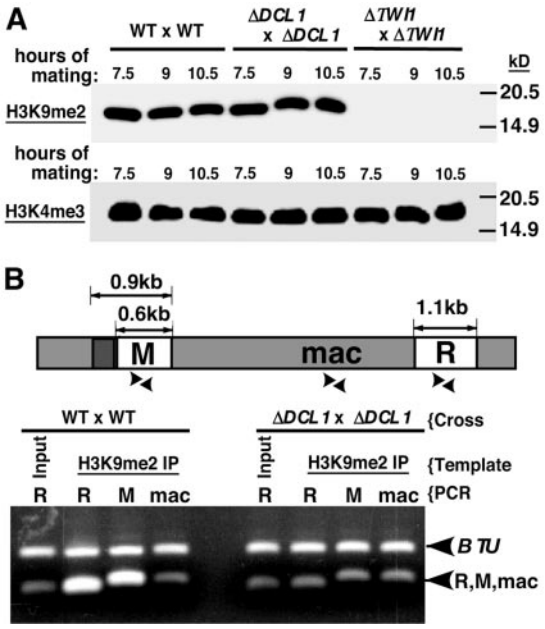


FIG. 8. *DCL1* is not required for H3K9 methylation. (A) Total cell protein extracts isolated from conjugating wild-type (WT), Δ*DCL1*, or Δ*TWI1* cells at the indicated times after mixing were fractionated by SDS-polyacrylamide gel electrophoresis for Western blot analysis. The overall level of modified histone H3 was assessed with antibodies detecting histone H3K9me2 (dimethyl) (top) or histone H3K4me3 (trimethyl) (bottom). The positions of protein molecular mass standards (in kilodaltons [kD]) are given on the right. (B) Representative PCR results after chromatin immunoprecipitation with anti-H3K9me2-specific antibodies from 9-h mating *Tetrahymena* cells. PCR products of specific amplified fragments of the R or M element or the intervening macronucleus-retained sequence were separated by 1.6% agarose gel electrophoresis and stained with ethidium bromide. The locations of amplification primers are shown in the diagram of the analyzed genomic region as arrowheads. Input designates that the DNA template was recovered from chromatin preparations prior to immunoprecipitation (only amplification of input using R-element primers is shown; the upper band corresponds to amplification of a quantification control from the *BTU1* locus, and the lower band corresponds to the IES or macronucleus-retained locus [indicated above each lane]).

mental small RNAs, we cannot discount the possibility that its protein participates in an initial step of processing and that *DCR1* and/or *DCR2* may also be involved in additional RNA processing. Dcl1p could play a role similar to that of Drosha, an RNase III enzyme also lacking a helicase domain that is required for processing primary miRNA transcripts but not the final processing steps that generate mature miRNAs (16, 32). We think this is less likely as *DCR1* and *DCR2* are expressed at relatively low levels early in conjugation and *DCR1* is dispensable for production of these small RNAs (J. A. Motl and D. L. Chalker, unpublished). Further investigation into the role of these other Dicer-related proteins is required to determine their roles, if any, in *Tetrahymena* DNA rearrangement.

Small RNA processing is not required for chromatin modification. Disruption of *DCL1* abolished small RNA production and DNA rearrangement but was not sufficient to eliminate H3K9me2 establishment (Fig. 8). In contrast, Mochizuki and Gorovsky (42) did not detect this chromatin modification occurring in their *DCL1* knockout lines, although it should be noted that their cells exhibited severe developmental abnormalities with <5% even progressing to form new macronuclei. While this subset of cells showed normal localization of Pdd1p, its localization is independent of this chromatin modification (34) and cannot be taken as an indicator that these cells may not have other defects that interfered with this modification. As our Δ*DCL1* cells did not exhibit these early development phenotypes, we were able to more thoroughly examine the knockout phenotype throughout mating, and we observed H3K9 methylation in the absence of small RNAs. However, this modification was no longer enriched on germ line-limited IESs, an observation that convincingly demonstrates that DCL1-generated small RNAs guide H3K9 methylation to the homologous locus.

It is intriguing to us that H3K9me2 chromatin is established in the absence of *DCL1*. This modification was not detected in strains lacking the Argonaute protein Twi1p (Fig. 8; 34) or the chromodomain-containing Pdd1p (50). This difference may indicate that both of these proteins are part of the effector complex that uses small RNAs to target this histone modification to specific loci. Loss of either protein may disrupt the complex and thereby abolish all H3K9 methylation. Both Pdd1p and Twi1p are initially localized in the cytoplasm before relocating to the developing macronucleus, an observation that is consistent with a putative association between them (36, 40). In contrast, Dcl1p is localized to the micronucleus and does not appear to relocalize to the developing nuclei where this modification occurs. In the absence of the abundant small RNAs, this H3K9 methylation complex appears to be directed to chromatin indiscriminately, modifying sequences other than germ line-limited DNA, or to be targeted to some specific regions by an unknown parallel path. In *Schizosaccharomyces pombe*, ATF/CREB proteins can target this modification in the absence of RNAi proteins (28), and a similar pathway may exist in *Tetrahymena*, although it is insufficient to replace all *DCL1* function.

Compartmentalization of the steps in DNA rearrangement. We have previously argued that the temporal separation of small RNA biogenesis early in conjugation from targeting later in development is instrumental to the proper regulation of this process (9). Mechanistically, the localization of *DCL1* within the micronucleus provides an important compartmentalization of this early RNA processing step in DNA rearrangement. To ensure that only the germ line-limited sequences are excised by this irreversible process, *Tetrahymena* cells confine the generation of the targeting small RNAs to the germ line nucleus. Initially, small RNAs homologous to somatic-retained sequences may also be produced, but a proofreading mechanism is postulated to check this initial pool for homology against the sequences within the parental macronucleus (40). This idea is evinced by the failure to eliminate sequences that are usually efficiently excised in the event that the homologous sequence is found within the parental macronucleus (10). This proofreading mechanism is likely mediated by the Twi1p-containing complexes that pick up small RNAs in the cytoplasm and transport them to the parental macronucleus (40, 41). Complexes that interact with a homologous sequence within this nucleus must be inactivated or disassembled and cannot target subsequent DNA rearrangement. We have suggested that these RNA-protein complexes actually interact with homologous transcripts rather than the DNA itself, given that introduced sequences previously shown to block elimination are bidirectionally transcribed (9). The failure of these bidirectional transcripts that are synthesized in the parental macronucleus to be processed into small RNAs further supports a compartmentalization Dicer function (9).

It is unclear whether meiotic micronuclear transcription is concentrated on germ line-limited sequences or extends over most of the genome; however, the *DCL1*-dependent small RNAs are enriched in germ line-limited sequences as early as 2 h into conjugation, which is indicative of some selectivity in this transcription (41). The germ line-limited M and R deletion elements are developmentally transcribed even when placed into the somatic macronucleus, which leads us to argue that sequences that normally undergo DNA rearrangement can be recognized and thus can be preferentially transcribed (9, 11). The increase in levels of nongenic, micronuclear transcripts detected in Δ*DCL1* cells (Fig. 4E and F) is similar to observations made in Dicer-deficient *S. pombe* and DT-40 human-chicken hybrid cells that accumulate transcripts corresponding to the outer centromere repeats and α-satellite sequences, respectively (18, 51). These rather different repetitive sequences all produce noncoding RNAs and exhibit RNAi-directed heterochromatin assembly. These shared properties may be suggestive of a conserved mechanism underlying these rather unconventional transcriptional phenomena. A novel RNA polymerase, polymerase IV, recently discovered in plants, has been shown to be required for RNAi-mediated gene silencing and facultative heterochromatin assembly (26, 44). While polymerase IV may be plant specific, its existence suggests that RNAi-related nongenic transcription may have unique requirements. Further study of the specificity of *Tetrahymena* germ line transcription is sure to shed light on this process.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation research grant MCB-0131421 and by National Institutes of Health research grant GM069593 to D.L.C.

We thank Washington University Biology 3492 students (Eric Archer, Robert Cooper, Laura Ernst, Michael Evenson, Annabel Fu, Xiaoou Pan, and Andrew Schmerling) for technical assistance with the

project. We also thank Martin Gorovsky (University of Rochester) for providing plasmid pMNBL and K. Mochizuki (University of Rochester) for providing the Δ*TWI1* strains, sharing unpublished data, and comments on our manuscript. Preliminary *Tetrahymena* genome sequence data were obtained from The Institute for Genomic Research website (http://www.tigr.org).

REFERENCES

1. **Austerberry, C. F., C. D. Allis, and M. C. Yao.** 1984. Specific DNA rearrangements in synchronously developing nuclei of *Tetrahymena*. Proc. Natl. Acad. Sci. USA **81:**7383–7387.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1990. Current protocols in molecular biology. John Wiley & Sons, New York, N.Y.
3. **Bernstein, E., A. A. Caudy, S. M. Hammond, and G. J. Hannon.** 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature **409:**363–366.
4. **Bruns, P. J., and D. Cassidy-Hanley.** 2000. Biolistic transformation of macro- and micronuclei. Methods Cell Biol. **62:**501–512.
5. **Bruns, P. J., and D. Cassidy-Hanley.** 2000. Methods for genetic analysis. Methods Cell Biol. **62:**229–240.
6. **Carmell, M. A., and G. J. Hannon.** 2004. RNase III enzymes and the initiation of gene silencing. Nat. Struct. Mol. Biol. **11:**214–218.
7. **Cassidy-Hanley, D., J. Bowen, J. H. Lee, E. Cole, L. A. VerPlank, J. Gaertig, M. A. Gorovsky, and P. J. Bruns.** 1997. Germline and somatic transformation of mating *Tetrahymena thermophila* by particle bombardment. Genetics **146:**135–147.
8. **Chalker, D., A. La Terza, A. Wilson, C. Kroenke, and M. Yao.** 1999. Flanking regulatory sequences of the *Tetrahymena* R deletion element determine the boundaries of DNA rearrangement. Mol. Cell. Biol. **19:**5631–5641.
9. **Chalker, D. L., P. Fuller, and M. C. Yao.** 2005. Communication between parental and developing genomes during *Tetrahymena* nuclear differentiation is likely mediated by homologous RNAs. Genetics **169:**149–160.
10. **Chalker, D. L., and M.-C. Yao.** 1996. Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of *Tetrahymena thermophila* with germ line-limited DNA. Mol. Cell. Biol. **16:**3658–3667.
11. **Chalker, D. L., and M. C. Yao.** 2001. Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. Genes Dev. **15:**1287–1298.
12. **Coyne, R. S., M. A. Nikiforov, J. F. Smothers, C. D. Allis, and M. C. Yao.** 1999. Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation. Mol. Cell **4:**865–872.
13. **Duharcourt, S., A. Butler, and E. Meyer.** 1995. Epigenetic self-regulation of developmental excision of an internal eliminated sequence in *Paramecium tetraurelia*. Genes Dev. **9:**2065–2077.
14. **Duharcourt, S., A. Keller, and E. Meyer.** 1998. Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. Mol. Cell. Biol. **18:**7075–7085.
15. **Fan, Q., R. Sweeney, and M.-C. Yao.** 1999. Creation and use of antisense ribosomes in *Tetrahymena thermophila*, p. 533–547. *In* D. J. Asai and J. D. Forney (ed.), *Tetrahymena thermophila*, vol. 62. Academic Press, New York, N.Y.
16. **Filippov, V., V. Solovyev, M. Filippova, and S. S. Gill.** 2000. A novel type of RNase III family proteins in eukaryotes. Gene **245:**213–221.
17. **Fillingham, J. S., D. Bruno, and R. E. Pearlman.** 2001. Cis-acting requirements in flanking DNA for the programmed elimination of mse2.9: a common mechanism for deletion of internal eliminated sequences from the developing macronucleus of *Tetrahymena thermophila*. Nucleic Acids Res. **29:**488–498.
18. **Fukagawa, T., M. Nogami, M. Yoshikawa, M. Ikeno, T. Okazaki, Y. Takami, T. Nakayama, and M. Oshimura.** 2004. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat. Cell Biol. **6:**784–791.
19. **Gaertig, J., L. Gu, B. Hai, and M. A. Gorovsky.** 1994. High frequency vector-mediated transformation and gene replacement in *Tetrahymena*. Nucleic Acids Res. **22:**5391–5398.
20. **Garnier, O., V. Serrano, S. Duharcourt, and E. Meyer.** 2004. RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. Mol. Cell. Biol. **24:**7370–7379.
21. **Godiska, R., C. James, and M. C. Yao.** 1993. A distant 10-bp sequence specifies the boundaries of a programmed DNA deletion in *Tetrahymena*. Genes Dev. **7:**2357–2365.
22. **Godiska, R., and M. C. Yao.** 1990. A programmed site-specific DNA rearrangement in *Tetrahymena thermophila* requires flanking polypurine tracts. Cell **61:**1237–1246.
23. **Grewal, S. I., and J. C. Rice.** 2004. Regulation of heterochromatin by histone methylation and small RNAs. Curr. Opin. Cell Biol. **16:**230–238.
24. **Grishok, A., A. E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D. L. Baillie, A. Fire, G. Ruvkun, and C. C. Mello.** 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell **106:**23–34.
25. **Hannon, G. J.** 2002. RNA interference. Nature **418:**244–251.
26. **Herr, A. J., M. B. Jensen, T. Dalmay, and D. C. Baulcombe.** 2005. RNA polymerase IV directs silencing of endogenous DNA. Science **308:**118–120.
27. **Hutvagner, G., J. McLachlan, A. E. Pasquinelli, E. Balint, T. Tuschl, and P. D. Zamore.** 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science **293:**834–838.
28. **Jia, S., K. Noma, and S. I. Grewal.** 2004. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. Science **304:**1971–1976.
29. **Katoh, M., M. Hirono, T. Takemasa, M. Kimura, and Y. Watanabe.** 1993. A micronucleus-specific sequence exists in the 5′-upstream region of calmodulin gene in *Tetrahymena thermophila*. Nucleic Acids Res. **21:**2409–2414.
30. **Ketting, R. F., S. E. Fischer, E. Bernstein, T. Sijen, G. J. Hannon, and R. H. Plasterk.** 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. Genes Dev. **15:**2654–2659.
31. **Kurihara, Y., and Y. Watanabe.** 2004. Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc. Natl. Acad. Sci. USA **101:**12753–12758.
32. **Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, and V. N. Kim.** 2003. The nuclear RNase III Drosha initiates microRNA processing. Nature **425:**415–419.
33. **Lee, Y. S., K. Nakahara, J. W. Pham, K. Kim, Z. He, E. J. Sontheimer, and R. W. Carthew.** 2004. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell **117:**69–81.
34. **Liu, Y., K. Mochizuki, and M. A. Gorovsky.** 2004. Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in *Tetrahymena*. Proc. Natl. Acad. Sci. USA **101:**1679–1684.
35. **Liu, Y., X. Song, M. A. Gorovsky, and K. M. Karrer.** 2005. Elimination of foreign DNA during somatic differentiation in *Tetrahymena thermophila* shows position effect and is dosage dependent. Eukaryot. Cell **4:**421–431.
36. **Madireddi, M. T., R. S. Coyne, J. F. Smothers, K. M. Mickey, M.-C. Yao, and C. D. Allis.** 1996. Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in *Tetrahymena*. Cell **87:**75–84.
37. **Madireddi, M. T., M. Davis, and D. Allis.** 1994. Identification of a novel polypeptide involved in the formation of DNA-containing vesicles during macronuclear development in *Tetrahymena*. Dev. Biol. **165:**418–431.
38. **Martindale, D. W., C. D. Allis, and P. Bruns.** 1982. Conjugation in *Tetrahymena thermophila*: a temporal analysis of cytological stages. Exp. Cell Res. **140:**227–236.
39. **Martindale, D. W., C. D. Allis, and P. J. Bruns.** 1985. RNA and protein synthesis during meiotic prophase in *Tetrahymena thermophila*. J. Protozool. **32:**644–649.
40. **Mochizuki, K., N. A. Fine, T. Fujisawa, and M. A. Gorovsky.** 2002. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. Cell **110:**689–699.
41. **Mochizuki, K., and M. A. Gorovsky.** 2004. Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. Genes Dev. **18:**2068–2073.
42. **Mochizuki, K., and M. A. Gorovsky.** 2005. A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. Genes Dev. **19:**77–89.
43. **Mochizuki, K., and M. A. Gorovsky.** 2004. Small RNAs in genome rearrangement in *Tetrahymena*. Curr. Opin. Genet. Dev. **14:**181–187.
44. **Onodera, Y., J. R. Haag, T. Ream, P. C. Nunes, O. Pontes, and C. S. Pikaard.** 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell **120:**613–622.

44a. **Orias, E., E. P. Hamilton, and J. D. Orias.** 1999. *Tetrahymena* as a laboratory organism: useful strains, cell culture, and cell line maintenance. Methods Cell Biol. **62:**189–211.

45. **Patil, N., and K. Karrer.** 2000. A developmentally regulated deletion element with long terminal repeats has cis-acting sequences in the flanking DNA. Nucleic Acids Res. **28:**1465–1472.
46. **Prescott, D. M.** 1994. The DNA of ciliated protozoa. Microbiol. Rev. **58:**233–267.
47. **Ray, C. J.** 1956. Meiosis and nuclear behavior in *Tetrahymena pyriformis*. J. Protozool. **3:**88–96.
48. **Shang, Y., X. Song, J. Bowen, R. Corstanje, Y. Gao, J. Gaertig, and M. A. Gorovsky.** 2002. A robust inducible-repressible promoter greatly facilitates gene knockouts, conditional expression, and overexpression of homologous and heterologous genes in *Tetrahymena thermophila*. Proc. Natl. Acad. Sci. USA **99:**3734–3739.
49. **Sugai, T., and K. Hiwatashi.** 1974. Cytological and autoradiographic studies of the micronucleus at meiotic prophase in *Tetrahymena pyriformis*. J. Protozool. **21:**542–548.

50. **Taverna, S. D., R. S. Coyne, and C. D. Allis.** 2002. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. Cell **110:** 701–711.
51. **Volpe, T. A., C. Kidner, I. M. Hall, G. Teng, S. I. Grewal, and R. A. Martienssen.** 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science **297:**1833–1837.
52. **Wenkert, D., and C. D. Allis.** 1984. Timing of the appearance of macronuclear-specific histone variant hv1 and gene expression in developing new macronuclei of *Tetrahymena thermophila*. J. Cell Biol. **98:**2107–2117.
53. **Xie, Z., L. K. Johansen, A. M. Gustafson, K. D. Kasschau, A. D. Lellis, D. Zilberman, S. E. Jacobsen, and J. C. Carrington.** 2004. Genetic and functional diversification of small RNA pathways in plants. PLoS Biol. **2:**E104.
54. **Yao, M. C.** 1982. Elimination of specific DNA sequences from the somatic nucleus of the ciliate *Tetrahymena*. J. Cell Biol. **92:**783–789.
55. **Yao, M. C., S. Duharcourt, and D. L. Chalker.** 2002. Genome-wide rearrangements of DNA in ciliates, p. 730–758. *In* N. Craig, R. Craigie, M. Gellert, and A. Lambowitz (ed.), Mobile DNA II. Academic Press, New York, N.Y.
56. **Yao, M. C., P. Fuller, and X. Xi.** 2003. Programmed DNA deletion as an RNA-guided system of genome defense. Science **300:**1581–1584.

CHAPTER 4

THE ROLE OF THE TANDEM DOUBLE-STRANDED RNA BINDING MOTIF PROTEIN, DRB2, IN DNA ELIMINATION OF THE CILIATE, *TETRAHYMENA THERMOPHILA*

Eukaryotic Cell. Submitted.

Abstract:

Double-**s**tranded **R**NA Binding **M**otif (DSRM)-containing proteins play many roles in the regulation of gene transcription and translation, including some with tandem DSRMs that act in small RNA biogenesis. We report the characterization of **D**ouble-stranded **R**NA **B**inding Proteins 1 and 2 (DRB1 and DRB2), two genes encoding nuclear proteins with tandem DSRMs in the ciliate, *T. thermophila*. Both proteins are expressed throughout growth and development, but exhibit distinct peaks of expression suggesting different biological roles. In support, we show that expression of DRB2 is essential for vegetative growth, while DRB1 expression is not. During conjugation, Drb1p and Drb2p localize to distinct nuclear foci. Cells lacking all DRB1 copies are able to produce viable progeny, although at a reduced rate relative to wild-type cells. In contrast, cells lacking just germline DRB2 copies, which thus cannot express Drb2p zygotically, fail to produce progeny, arresting late into conjugation. This arrest phenotype is accompanied by a failure to execute DNA elimination and chromosome breakage, and does not organize the essential DNA rearrangement protein, Pdd1p, into DNA elimination bodies. These results implicate zygotically expressed Drb2p in the maturation of these nuclear structures, which are necessary for reorganization of the somatic genome.

Introduction:

Proteins containing the **d**ouble-**s**tranded **R**NA (dsRNA) binding **m**otif (DSRM) participate in diverse biological pathways in a wide range of organisms. This motif was first identified in the developmentally essential gene Staufen of *D. melanogaster*, and has since been recognized to be encoded in the genomes in all three domains of living organisms, as well as in viruses ([1], reviewed in [2, 3]). DSRM proteins commonly act

in developmental pathways (e.g. RNA localization by the Staufen family and developmental transcriptional regulation by the DIP1 family) [4-7], but also have ubiquitous roles in transcriptional and translational regulation (e.g. PKR family and PKR-associated proteins) [8-11]. Proteins vital for RNA interference (RNAi) have also been found to contain DSRMs. These include members of the RNase III family (e.g. Dicer and Drosha family proteins) and their tandem DSRM-containing partner proteins (e.g. RDE-4 of *C. elegans*, Pasha, R2D2 and Loqs in *D. melanogaster* and their homologues in *H. sapiens*) [12-20].

In the ciliate, *Tetrahymena thermophila*, the DSRM-containing protein **D**i**c**er-**l**ike 1 (DCL1) has been shown to play a pivotal role in a process linking RNAi to heterochromatin formation and developmentally regulated DNA elimination (chapter 3) [21, 22]. Like all ciliates, *T. thermophila* is unicellular yet contains two distinct types of nuclei, the somatic macronucleus and the germline micronucleus (reviewed in [23, 24]). The polyploid macronucleus (~50C) acts as a transcriptionally active somatic nucleus during vegetative growth, while the diploid, germline micronucleus is transcriptionally silent ([25, 26], reviewed in [23, 24]). In optimal growth conditions *T. thermophila* undergo asexual, binary fission, however when starved *T. thermophila* reproduce through the sexual process of conjugation generating new micronuclei and macronuclei from the parental germline micronucleus (reviewed in [23, 24]). During the maturation of the zygotic macronuclei, the macronuclear chromosomes are fragmented at ~180 sites, lose ~15% of their overall genomic content and are amplified to ~50C [25-31]. The loss of genome complexity is the result of programmed DNA rearrangements that remove

specific DNA sequences, called internal eliminated sequences (IESs), from thousands of chromosomal sites (reviewed in [23, 24]).

DNA elimination has been shown to be guided by an RNAi-related mechanism [21, 22, 32, 33]. Bidirectional transcription of the germline genome in meiotic micronuclei provides an abundant source of IES-specific dsRNA [33, 34]. The resulting ncRNAs are processed into 27nt-30nt sRNAs species called scan RNAs (scnRNAs) by Dcl1p in the meiotic micronucleus [21, 22]. These scnRNAs are exported into the cytoplasm where they are bound by a PIWI homologue, Twi1p [32]. Twi1p/scnRNA complexes are transported into the parental macronucleus, where these complexes ‘ scan’ macronuclear ncRNAs and possibly mRNAs. The Twi1p/scnRNA complexes homologous to the parental macronucleus are removed from the pool of active complexes, and the remaining complexes are transported to the zygotic macronuclei upon their emergence, where they guide H3K9 and H3K27 methylation of IES-associated histones by the E(z) homologue, Ezl1p [32, 35, 36]. Methylated histones in zygotic macronuclei are bound by the chromo-domain containing proteins, Pdd1p and Pdd3p, which along with other associated proteins form large nuclear structures called DNA elimination bodies late in conjugation [35, 37-39]. DNA elimination in these bodies is catalyzed by the domesticated PiggyBac transposase, Tpb2p, resulting in removal of IESs from zygotic macronuclei [40].

As dsRNA has critical roles during genome reorganization, we characterized the two putative tandem DSRM-containing proteins, **D**ouble-stranded **R**NA **b**inding protein 1 and 2 (Drb1p and Drb2p), encoded in *T. thermophila* genome [41, 42]. We show that both are nuclear proteins that exhibit distinct sub-nuclear organization. By knocking out

the gene for each, we found that Drb2p is essential both during vegetative growth and also late in conjugation, where it facilitates DNA elimination body formation and subsequent RNAi-dependent DNA elimination. Drb1p, in contrast, is dispensable, but is nonetheless important for efficient pre-zygotic development. Our data does not support that either protein acts as an essential Dicer partner protein as tandem DSRM proteins do in other eukaryotes, but instead suggest that these proteins have diverse roles during the *T. thermophila* life cycle and exposes a role for dsRNA late in macronuclear development [13-20].

Results:

The *T. thermophila* Macronuclear Genome Encodes Two Proteins with Tandem DSRM.

For optimal sRNA production and protein localization Dicer and Drosha homologues in *C. elegans*, *D. melanogaster* and *H. sapiens* require association with a tandem DSRM-containing protein [13-20, 43]. Bioinformatic analysis (BLAST, Pfam and ClustalW) of the *T. thermophila* macronuclear genome identified two genes, DRB1 and DRB2, encoding tandem DSRM-containing proteins (Fig. 1A). Alignment of their putative DSRMs with other DSRM-containing proteins indicated conservation of key residues essential for DSRM structure and function. The homology of these proteins with other tandem DSRM-containing proteins did not extend beyond these domains (Fig. 1A; Fig. S1A). However, alignment of full length DRB1 and DRB2 revealed additional regions of similarity outside of the DSRMs, one in the N-terminal region (NTR) and two in the C-terminal regions (CT1 and CT2) of each protein (Fig. S2).

RT-PCR and Northern blot analysis demonstrated that DRB1 and DRB2 are both expressed throughout much of the *T. thermophila* life cycle (Fig. 1C and D; data not shown). DRB1 mRNA levels are low in growing and starved cells, but increase significantly during meiosis (2-4 hours into conjugation when scnRNA production occurs) and again after the appearance of the zygotic macronuclei (8 hours), a pattern that parallels DCL1 expression (Fig. 1C) [21, 22]. DRB2 expression is relatively constitutive with decreased expression during starvation and upregulation after the appearance of the zygotic macronuclei (8 hours) (Fig. 1D).

DRB1 and DRB2 Encode Nuclear Proteins that Localize to Distinct Structures.

Ectopic expression of Drb1p and Drb2p tagged with YFP or CFP, respectively, on their C-termini showed that both are nuclear proteins visible in small foci throughout the macronucleus during vegetative growth (data not shown) and in the parental macronucleus early in conjugation (Fig. 2). During early zygotic macronuclear differentiation (10 hours) all Drb1p-YFP and most Drb2p-CPF disappeared from the parental macronucleus and reappear in the zygotic macronuclei (Fig. 2A and 2B, middle row). Whether the foci seen in the parental macronucleus are functionally related to those observed in zygotic macronuclei could not be determined (Fig. 2A and 2B, compare top and middle rows). Near completion of zygotic macronuclear development (14 hours into conjugation), Drb1p-YFP localization is primarily diffuse (Fig. 2A, bottom row). In contrast, the small Drb2p-CFP foci coalesce into larger foci, although low-level diffuse localization remains throughout the entire zygotic macronucleus as well. (Fig. 2B, bottom row).

Upon initial inspection, the size and number of nuclear foci of Drb1p and Drb2p in parental macronuclei appeared rather different. To better compare their localizations, Drb1p-YFP and Drb2p-CFP were co-expressed and visualized 4 hours into conjugation. Their nuclear foci are distinct with only a small degree overlapping localization (Fig. 2C). DRB1 and DRB2 are best reciprocal hits in a BLASTp analysis of the CT2 regions, which could explain the small overlap in localization through partially redundant protein function. Despite this it seems that both Drb1p and Drb2p have distinct primary functions based on distinct localization and divergent protein sequences outside of their DSRMs and CT2.

In addition to its abundant macronuclear localization, Drb1p-YFP also localizes to the micronucleus just prior to and during crescent formation (prophase of meiosis I) (Fig. 2A, top row; data not shown). Drb1p-YFP is observed specifically at the poles of these nuclei, at either one or both ends depending on developmental stage. This micronuclear localization pattern is quite distinct from that of Dcl1p, which is only found throughout the nucleoplasm of the crescent micronucleus, and suggests that Drb1p may not be a critical Dcl1p protein partner [21, 22]. Point localization of Drb1p-YFP is seen early in conjugation once the micronucleus begins to elongate at one end and later after the crescent micronucleus fully elongates it can be seen at both ends of the micronucleus (Fig. 2A, top row; data not shown). Upon anaphase of meiosis I, Drb1p-YFP micronuclear localization is lost. While it is likely that DRB1 and DRB2 arose from an ancient gene duplication, differential localization and expression patterns indicate that each DSRM-containing protein has specific cellular roles.

DRB2, but not DRB1, is Essential for Growth and Development.

We created strains lacking each gene to establish whether and when each protein functions during the *T. thermophila* lifecycle. Constructs containing the NEO3 selectable marker, flanked by up- and downstream homology of either DRB1 or DRB2 were biolistically transformed into cells during conjugation to generate heterozygous micronuclear/macronuclear knockout strains. Taking advantage of the random assortment of alleles during amitotic macronuclear division, we obtained strains for which all wild-type DRB1 gene copies in the macronucleus were replaced with the knockout allele, which revealed that Drb1p is not required for vegetative growth (data not shown; Fig. S3A). In contrast, we were unable to identify strains in which all copies of DRB2 were disrupted, which indicate that vegetative DRB2 expression is essential (Fig. S3C).

To verify that DRB2 but not DRB1 expression was essential for vegetative growth, additional genetic manipulation of the original heterozygous strains produced new strains homozygous for the knockout cassette in the micronucleus while maintaining wild-type copies of DRB2 in macronucleus to support growth (See Materials and Methods for details). These homozygous micronuclear knockout strains were then crossed to generate strains homozygous for the knockout cassette in both the micronucleus and macronucleus, thus eliminating all wild-type DRB2 gene copies. Homozygous micronuclear knockout strains containing the DRB1 knockout cassette produced fertile homozygous DRB1 knockout progeny. Southern blot analysis of these complete DRB1 knockout strains detected only the DRB1 knockout allele (Fig. S3A). RT-PCR of the DRB1 knockout strains during conjugation confirmed loss of all DRB1 expression (Fig. S3B). While these complete DRB1 knockout strains showed no growth defects, matings between two DRB1 knockout strains generated progeny at a reduced rate

relative to crosses of wild-type strains (Table 1). The DRB1 knockout cells that were able to complete conjugation arrested with two new macronuclei and a single micronucleus, as do wild-type conjugant, until they were returned to growth media and started vegetative growth (Fig. 3). The observation that only a fraction of mated DRB1 knockout cells progressed to post-zygotic development suggests that Drb1p is important, but not essential, for pre-zygotic development. The lack of Drb1p during this stage(s) of early conjugation resulted in substantial premature abortion of conjugation (data not shown).

Attempts to make homozygous DRB2 knockouts using DRB2 mic knockout strains proved to be futile. Despite each individual DRB2 mic knockout strain being able to produce progeny when mated to wild-type strains, when these lines were mated to each other no viable progeny emerged (Table 2). Further analysis revealed that DRB2 mic knockout strains are unable to reach the terminal stage of conjugation with 2 macronuclei and 1 micronucleus even 30 hours after pairing, but instead arrest with 2 macronuclei and 2 micronuclei (Fig. 3). Thus not only is DRB2 expression necessary for vegetative growth, but that zygotic DRB2 expression is essential for completion of conjugation (Figs. S3C and 3). As observed in other mutants that arrest at the 2 macronuclei, 2 micronuclei stage, conjugating DRB2 mic knockouts underamplified their macronuclear relative to zygotic macronuclei of wild-type conjugants at their terminal stage prior to refeeding [21, 22, 32, 44]. Although DRB2 mic knockout strains only lack zygotic expression of DRB2, the majority of conjugants arrest at the 2 macronuclei, 2 micronuclei stage, while the remainder arrest after elimination of one of the remaining micronuclei (Fig. 3, bottom). RT-PCR analysis of DRB2 mic knockout matings showed reduced, but not complete loss of, expression after 12 hours of conjugation relative to

wild-type cells, when zygotic DRB2 expression normally should predominate (Fig. S3D). Unmated cells as well as parentally-expressed DRB2 mRNA in the DRB2 mic knockout mating population account for the DRB2 mRNA detected. The residual, parentally-expressed DRB2 transcripts may enable a fraction of cells to proceed further into conjugation and eliminate one micronucleus.

DRB2 Mic Knockouts Fail to Excise IESs and Break Chromosomes Late in Conjugation.

The DRB2 conjugation arrest phenotype is commonly observed in knockouts of genes necessary for genome rearrangement in *T. thermophila* including DCL1, TWI1 and PDD1 [21, 22, 32, 44]. To determine whether the DRB2 mic knockout arrest is accompanied by failure of RNA-directed DNA elimination or due to some other perturbation during conjugation, we monitored the rearrangement of several IESs. Genomic DNA was isolated from mated cell populations 30 hours after initiating conjugation when all genome reorganization should be complete. PCR using primers able to detect both the unrearranged (micronuclear form of the locus) and rearranged (macronuclear form) IESs allowed assessment of the level of excision. Whereas DNA from wild-type mating populations showed predominantly the rearranged locus for each IES, DRB2 mic knockout or control DCL1 knockout matings exhibited accumulation of the unrearranged form of both IES B and the M IES (Figs. 4 and S4; data not shown). IES B is a 327 bp IES found within the LIA2 gene and the M IES is a well-studied intergenic IES that undergoes alternative rearrangement that removes either 0.6 kb or the complete 0.9 kb IES sequence [45, 46]. PCR analysis of IES B clearly showed that the 597 bp product indicative of the micronuclear locus was overrepresented in the DCL1 and DRB2

mic knockout matings relative to wild-type (Fig. 4A). It is important to note that the cell populations monitored included some percentage of unmated cells whose DNA likely contributes much of the template for the 270 bp product representing the rearranged form in the mutant cell lines. The PCR analysis of the M IES utilized three primers for PCR, which we find provides a more quantitative assessment of its rearrangement. Two bands at 1192 bp and 386 bp result from amplification of micronuclear DNA containing the IES, while two other bands at 592 bp and 292 bp are the products of removal of either 0.6 kb or 0.9 kb of the M IES locus. As observed for IES B, the unrearranged form of the M IES was overrepresented in the DCL1 and DRB2 mic knockout mating populations relative to wild-type matings. This difference is less apparent in DRB2 mic knockout matings than in DCL1 mutant, which could be due to persistence of parental Drb2p. Analysis of other IESs further demonstrates that these mutants exhibit substantial failure of RNA-directed DNA elimination (Fig. S4; data not shown).

Assessment of chromosome breakage near the LIA1 locus also shows that DRB2 mic knockout progeny fail to properly fragment chromosomes (Fig. 4C). Before the completion of conjugation the chromosomes in the zygotic macronuclei, which contain 5 chromosomes amplified to between 4 and 8 copies, are fragmented at approximately 180 **c**hromosome **b**reakage **s**ites (CBSs) to produce the shortened macronuclear chromosomes. In knockouts of genes essential for genome rearrangement, including DCL1 and TWI1, chromosome breakage fails as well as IES elimination [21, 32]. In a Southern blot of wild-type progeny, chromosome breakage at the LIA1 locus results in a band of approximately 2.5 kb in the zygotic macronuclei. The copies of this chromosome from the parental macronucleus are visible as a 2.6 kb band as they have longer telomeres

relative to newly fragmented ends. Unbroken micronuclear chromosomes are detected as a 10.5 kb band. The probe also detects a 7.8 kb fragment present in all nuclei. Due to the increased copy number of the locus in the macronucleus in wild-type progeny the 2.5 kb and 2.6 kb bands are more intense than the larger 10.5 kb micronuclear band. In the control DCL1 knockout progeny one sees an increase in the levels of the 10.5 kb unrearranged band and an absence of the 2.5 kb band resulting from *de novo* chromosome breakage (Fig 4C). DRB2 mic knockout progeny also show an increase in the 10.5 kb band and a lack of the 2.5 kb band similar to DCL1 knockout progeny. A previous report on chromosome breakage in a somatic knockout of PDD1 showed normal chromosome fragmentation [44]. Here we report using a homozygous PDD1 knockout that chromosome breakage is impaired like DCL1 and DRB2 mic knockouts, emphasizing the importance of zygotic expression of PDD1 and DRB2 in chromosome breakage (Fig. 4C).

DRB2 Co-localizes with Pdd1p in DNA Elimination Bodies.

Failure of DNA elimination and chromosome breakage in DRB2 mic knockout strain matings indicates that the conjugation arrest phenotype described earlier was a result of failure to complete RNA-directed DNA elimination. The localization of Drb2p-CFP into large foci 14 hours into conjugation, which is when DNA elimination normally occurs, prompted us to ascertain whether Drb2p-CFP was localized into DNA elimination bodies. These nuclear structures are enriched for the essential DNA elimination, chromodomain-containing protein Pdd1p. Strains expressing Drb2p-CFP or Pdd1p-YFP were mated and localization of both proteins was monitored at 8 hours into conjugation very early in zygotic macronuclear differentiation and later at 14 hours into conjugation

when DNA elimination occurs (Fig. 5). As was previously reported, Pdd1-YFP is diffusely localized in the zygotic macronuclei at 8 hours, and as conjugation proceeds toward DNA elimination around 14 hours, Pdd1p-YFP localization gradually becomes unevenly dispersed, forming first small foci and then finally large foci (Fig. 5; [37, 47]). These large foci, called DNA elimination bodies, are thought to be the site of IES removal. Localization of Drb2p-CFP in the zygotic macronuclei at 8 hours into conjugation is not markedly different from Pdd1p-YFP localization with small Drb2p-CFP foci throughout the entire nucleus (Fig. 5). However, at 14 hours into conjugation Drb2p-CFP foci aggregate into larger foci, which co-localize with the Pdd1p-YFP containing DNA elimination bodies, indicating a possible interaction with each other in zygotic macronuclei. Of note, localization of Pdd1-YFP and Drb2p-CFP is not exclusive to the zygotic macronuclei. Residual localization of both proteins is seen in the parental macronucleus as well. At 8 hours into conjugation both proteins form strong, distinct foci in the parental macronucleus with Pdd1p-YFP foci localized to the nuclear periphery and Drb2p-CFP foci found in the nuclear interior. During DNA elimination at 14 hours into conjugation remaining Pdd1p-YFP localization is found throughout the parental macronucleus but away from the interior, while Drb2p-CFP is still seen only in the interior. The significance of this late parental macronuclear localization remains to be explored.

Pdd1p Fails to Form DNA Elimination Bodies in DRB2 mic Knockouts.

To understand if Pdd1p and Drb2p co-localization is relevant to the conjugation arrest phenotype and failure of DNA elimination in DRB2 mic knockouts, we sought to determine how Pdd1p localization was affected in DRB2 mic knockout strain matings.

DRB2 mic knockout strains were transformed with an inducible Pdd1p-YFP expression construct, and the resulting transformants were mated and their Pdd1p-YFP localization was examined. 10 hours into conjugation during zygotic macronuclear differentiation, Pdd1p-YFP localization in both DRB2 mic knockouts crossed to wild-type strains, which rescues loss of DRB2 from the mating partner, and DRB2 mic knockout matings appears mottled throughout the developing zygotic macronucleus without obvious defects (Fig. 6A). However, late in conjugation (14 hours), Pdd1p-YFP fails to form DNA elimination bodies in the zygotic macronuclei in DRB2 mic knockout matings (Fig. 6B). Thus Pdd1p-YFP foci fail to mature into DNA elimination bodies without zygotic DRB2 expression. This data indicates that DRB2 participates in the maturation of DNA elimination bodies and implicates a possible role for uncharacterized dsRNAs in genome reorganization.

Discussion:

Our analyses of DRB1 and DRB2 have revealed that each has unique and important functions. While both are predominantly nuclear proteins, they localize into distinct sub-nuclear foci. Furthermore, disruption of the each gene showed that Drbp2p has essential functions during both growth and development while Drb1p appears to be important for pre-zygotic development. Similarity of these two proteins outside their predicted DSRMs suggests that they may have arisen from an ancestral gene duplication. If that is the case, they have significantly diverged in function since the duplication event.

Upon initial recognition that the *T. thermophila* genome encodes two DSRM-containing proteins, we looked for evidence that would connect them as protein partners for the Dicer homologues encoded by DCL1 and DCR2 [21, 22]. Tandem DSRM-

containing partner proteins for Dicer and Drosha family proteins, including R2D2, Loqs and Pasha in *D. melanogaster*, RDE-4 in *C. elegans* and TRBP2 and DGCR8 in *H. sapiens* and other mammals, play vital roles in RNAi by ensuring proper sRNA delivery and in many cases cleavage of sRNA precursors [14-20]. Our analyses provided little support that these serve as major Dicer partners. Neither protein showed abundant localization in meiotic micronuclei where Dcl1p acts (Fig. 2A and 2B, top row) [21, 22]. We also did not find defects in scnRNA accumulation in complete DRB1 knockouts (data not shown). As Drb2p is essential for growth, we were unable to generate full knockouts with which to examine scnRNA accumulation upon its loss. The *T. thermophila* Dicer protein, Dcr2p, is also essential for growth, but previously published characterization of Dcr2p complexes did not find Drb2p as an interacting protein [48, 49].

While we did not find evidence that these proteins act with Dcl1p, we uncovered a critical role for Drb2p in the RNAi-directed DNA elimination pathway. Loss of zygotic expression was sufficient to block DNA rearrangement, thus Drb2p is needed well downstream of scnRNA biogenesis by Dcl1p (Fig. 4A and 4B). Co-localization of Drb2p with Pdd1p-containing DNA elimination bodies and loss of these DNA elimination bodies in DRB2 mic knockouts implicates zygotically expressed Drb2p in promoting or stabilizing these large nucleo-protein structures (Figs. 5 and 6). This may indicate that Drb2p/RNA complexes mediate the formation of mature DNA elimination bodies through facilitating protein-RNA or protein-protein interactions within these structures. Although the exact mechanism of Drb2p action remains to be discovered, its importance in late stages of genome reorganization suggests an unrecognized role for dsRNA in RNAi-directed DNA elimination.

Drb2p is also required for vegetative growth as we were unable to replace all wild-type DRB2 gene copies with the disrupted allele. We tried extensively to assort DRB2 out of the macronucleus without success (data not shown). Furthermore, when DRB2 partial knockout strains were grown in non-selective medium (without paromomycin), the remaining wild-type DRB2 copies rapidly replaced the DRB2 knockout allele (data not shown). As both Drb2p and Dcr2p are essential for growth it remains possible that they act in the same pathway [48, 49]. We cannot rule out the possibility that these proteins transiently interact as RDE-4 and DCR-1 do in *C. elegans* [14]. Further investigation of the function of Drb2p during growth may provide key insight into the role of this protein during genome reorganization.

While Drb1p is predominantly a macronuclear protein, it also localizes to one or both ends of the crescent micronucleus during prophase of meiosis I (Fig. 2; data now shown). Further investigation of this micronuclear point localization indicates a putative interaction with centromeres and possibly with telomeres ([50-52], J.A. Motl and D.L. Chalker, unpublished data). Knockouts of DRB1 were able to complete conjugation, yet a significant percentage of pairs aborted mating without forming new macronuclei (J.A. Motl and D.L. Chalker, unpublished data). Together, the localization of Drb1p near centromeres and telomeres and the reduction in knockout cells completing pre-zygotic stages of development are consistent with a role for Drb1p in maintaining micronuclear chromosome structure (Fig. 2; data not shown). Thus the analysis of both of these DSRM-containing proteins strongly suggests that they perform critical chromosomal functions.

Although many tandem DSRM-containing proteins have been found to interact with Dicer and Drosha family proteins that is by no means the only job these proteins containing DSRMs undertake ([14-20], reviewed in [2, 3]). Roles for these proteins include cleavage of long non-coding RNAs into sRNAs by RNase III family members, RNA editing by the ADAR family, translation inhibition in response to viruses by PKR family members, and developmental RNA localization by the Staufen family [1, 4, 8, 9, 11-13, 53-56]. Besides the partner proteins for the Dicer and Drosha family, at least one other protein family, the NFAT family, also encodes tandem DSRMs. The NFAT protein family, which contain a DZF protein domain in addition to tandem DSRMs are putative nuclear, nucleotide transferases that participate in DNA repair and RNA transport ([57-60], reviewed in [61]). Further study of DRB1 and DRB2 in *T. thermophila* may reveal new roles for tandem-DSRM-containing proteins. The great evolutionary distance between ciliates and other eukaryotes could also facilitate understanding of how DSRM-containing proteins evolved within the eukaryotic lineage [62]. Much remains to be gleaned about the roles of DSRM-containing proteins in eukaryotes, and we expect further investigation of Drb1p and Drb2p function will provide greater understanding of RNAi-directed DNA elimination and the roles for dsRNA in regulating chromosome structure.

Materials and Methods:

***Tetrahymena* strains and growth conditions.** Standard wild-type, laboratory *T. thermophila* strains CU427 (Chx/Chx [VI, cy-s]), CU428 (Mpr/Mpr [VII, mp-s]), B2086 (II), micronucleus-defective strains B*VI (VI) and B*VII (VII) were originally obtained from Peter Bruns (Cornell University, Ithaca, NY). B*VII$^{CU427/CU427}$ (Chx/Chx [VII, cy-

s] was generated through a genomic exclusion mating between CU427 and B*VII. These strains or their transformed progeny were used for expression studies, biolistic transformations, and subsequent analyses. ΔDCL1 homozygous knockouts strains were described earlier [21]. Cells were grown and maintained as previously described [63, 64]. Strains were starved six hours to overnight in 10 mM Tris (pH 7.5) prior to mixing to initiate conjugation. Optical densities of cell populations were used to estimate cell number prior to mixing equal numbers of mating compatible strains.

Identification of DRB1 and DRB2 sequence. DRB1 and DRB2 sequence were identified by BLAST search of the *T. thermophila* macronuclear genome (http://www.ciliate.org) using *D. melanogaster* R2D2 (Accession number CG7138) and Loqs (Accession number CG6866) DNA sequences. DSRMs of DRB1 and DRB2 were initially identified on the *T. thermophila* macronuclear genome by InterProScan [65]. Further sequence analysis included Pfam analysis of DRB1 and DRB2 coding sequence (http://pfam.janelia.org) and alignment of DSRM sequences of DRB1 and DRB2 with those of tandem DSRM-containing proteins R2D2 (Accession number NP_609152.1) and Loqs (Accession number NP_609646.1) from *D. melanogaster*, RDE-4 (Accession number NP_499265.1) from *C. elegans* and TRBP2 (Accession number NP_599150.1) from *H. sapiens*. Identification of additional homologous regions of DRB1 and DRB2 protein sequence was carried out by protein alignment using Emboss Needle (http://www.ebi.ac.uk).

RT-PCR expression analysis. RNA was isolated from growing, starved and conjugating *T. thermophila* (CU428 X B2086) at 2 hour intervals from 2hr to 14hr by RNAsol extraction [66]. RNA isolation from DRB1 knockouts and DRB2 mic knockouts at 4hr

and 12hr after mixing was also done by RNAsol extraction. RT-PCR of wild-type and knockout matings was done to determine expression and show loss or decrease of expression, respectively, as previously described (See Table 3 for primers) [21].

Cloning of *T. thermophila* genes for protein localization. Oligonucleotide primers (Table 3) were used to amplify the entire DRB1 or DRB2 coding sequences from genomic DNA by PCR. The resulting products were cloned into the Gateway recombination compatible pENTR-D to create pENTR-D-DRB1 and pENTR-D-DRB2, respectively. Plasmids containing the DRB1 and DRB2 coding sequences were identified by lysate PCR and sequenced to verify coding sequence integrity (See Table 3 for primers). LR recombination of pENTR-D-DRB1 with pICY-GTW and of pENTR-D-DRB2 with pICC-GTW using LR clonase II (Invitrogen) fused the coding regions to YFP and CFP in pICY-DRB1 and pICC-DRB2, respectively. Construction of the rDNA plasmids pICY-GTW and pICC-GTW was described earlier [67]. Plasmids pICY-DRB1 or pICC-DRB2 were introduced into mating wild-type cells (CU427 X CU428) by conjugative electroporation [68].

Similarly, the entire PDD1 coding sequence was amplify from genomic DNA (See Table 3 for primers) and cloned into pENTR-D to create donor plasmid pENTR-D-PDD1. LR recombination of pENTR-D-PDD1 with pICY-GTW using LR clonase II (Invitrogen) created pICY-PDD1 which was then introduced into cells by conjugative electroporation [68].

For Pdd1p-YFP localization in DRB2 mic knockout strains, pENTR-D-PDD1 was recombined with pBS2-ICY-GTW using LR clonase II (Invitrogen) to create pBS2-ICY-PDD1. BclI- and SalI-digested pBS-ICY-PDD1 was transformed into starved,

homozygous micronuclear DRB2 knockout strains (B*VI$^{\Delta D2/\Delta D2}$ #1 and #6, B*VII$^{\Delta D2/\Delta D2}$ #1 and #2) using PDS-1000/He particle bombardment system (Bio-Rad) as previously described [69, 70]. Transformants were identified by their resistance to 25µg/ml cycloheximide.

To visualize localization, starved transformed cells were mixed to begin conjugation in 0.08 µg/ml $CdCl_2$ to induce expression of the fusion protein. Live cells were harvested by low speed centrifugation (1,000 X g) at 4hr, 10hr and 14hr post-mixing, stained with DAPI (1 µg/ml) and immobilized in 5µl 2% methylcellulose. DIC, CFP fluorescence, YFP fluorescence and DAPI fluorescence images were captured using a Qimaging RetigaEX charge-coupled-device camera (Burnaby, British Columbia, Canada) and Openlab software (PerkinElmer).

Generation of DRB1 and DRB2 knockouts. Genomic sequences from upstream and downstream of each gene's coding region were amplified by PCR and recombined into pDONR-P4-P1R (Upstream) and pDONR-P2R-P3 (Downstream) using BP Clonase (Invitrogen) (See Table 3 for primers). The resulting plasmids were identified by lysate PCR and verified through sequencing (See Table 3 for primers). These donor plasmids containing up- and downstream regions were mixed with equal amounts of pENTR-D-MTT1/NEO3 and the multisite destination vector, pDEST-R4-R3, along with LR Clonase Plus II (Invitrogen) to create the DRB knockout plasmids, pDEST-B4-DRB1Up-B1-MTT1/NEO3-B2-DRB1Down-B3 and pDEST-B4-DRB2UpN1-B1-MTT1/NEO3-B2-DRB2Down-B3. DRB1 and DRB2 knockout constructs were linearized by digestion with KpnI and transformed into conjugating wild-type cells (CU428 X B2086) between 2 and 3 hours after mixing using a PDS-1000/He particle bombardment system (Bio-Rad) as

previously described [69, 70]. Heterozygous micronuclear transformants were identified by their resistance to 80 µg/ml paromomycin with 1 µg/ml $CdCl_2$ and 15µg/ml 6-methylpurine. Heterozygous micronuclear transformants were verified through matings with CU427 by monitoring segregation of paromomycin resistance conferred by the MTT1/NEO3 knockout cassette among cycloheximide resistant progeny, as well as through PCR screening of *T. thermophila* crude cell lysates (See Table 3 for primers) [71]. Homozygous micronuclear knockout heterokaryons were generated by crossing heterozygous micronuclear transformants with B*VI or B*VII star strains. Homozygous micronuclear knockout heterokaryons were identified by paromomycin/$CdCl_2$ sensitivity and verified through crosses with CU427 that produced progeny resistant to 100 µg/ml paromomycin with 1 µg/ml $CdCl_2$ and 25 µg/ml cycloheximide. Homozygous DRB1 knockout strains were generated by crossing homozygous micronuclear knockouts of compatible mating types and screening for progeny resistant to paromomycin/$CdCl_2$ and verified by PCR detection of the knockout allele (See Table 3 for primers) [71].

Southern blot and PCR analyses. *T. thermophila* genomic DNA was isolated using a Wizard genomic DNA purification kit (Promega). Gel electrophoresis, blotting and hybridization were performed as previously described except blots were washed with 0.5X SSC-1% SDS after hybridization [21]. The probe for DRB2 was a radiolabeled KpnI and BsrGI restriction fragment of pDONR-R2-DRB2Down-L3. The DRB1 probe was a labeled BsrGI and XmnI restriction fragment from pDONR-R2-DRB1Down-L3. Genomic DNA from heterozygous (DRB1 knockout #5), homozygous (DRB1 knockout #5.1.3, #7.1 and #7.7.2) and homozygous micronuclear DRB1 knockout (B*VI$^{\Delta D1/\Delta D1}$ #5 and #12 and B*VII$^{\Delta D1/\Delta D1}$ #1 and #5) was digested with XmnI was separated on a 1.0%

agarose gel. Genomic DNA from heterozygous DRB2 knockouts (DRB2 mic knockout #2, #3, #4, #14 and #15) was digested with ClaI and SacI and fractionated on a 0.8% agarose gel prior to blotting.

Chromosome breakage was assayed in DRB2 mic knockouts using genomic DNA from CU428 X B2086, ΔPDD1 39.1 X ΔPDD1 W3.3, ΔDCL1 #1.8.6 X ΔDCL1 #4.2.4, B*VI$^{\Delta D2/\Delta D2}$ #1 X B*VII$^{\Delta D2/\Delta D2}$ #1 30hr after mixing that was digested with EcoRI and separated on a 0.8% agarose gel. The Southern blot probe for chromosome breakage was created using a 0.8-kbp probe fragment that spans the EcoRI site at position 335013 of chromosomal scaffold CH445662 (GenBank accession number gi62422284). DNA rearrangement of IESs A, B, C and the M IES was assayed by PCR using CU428 X B2086, ΔDCL1 #1.8.6 X ΔDCL1 #4.2.4, B*VI$^{\Delta D2/\Delta D2}$ #1 X B*VII$^{\Delta D2/\Delta D2}$ #1 and B*VI$^{\Delta D2/\Delta D2}$ #6 X B*VII$^{\Delta D2/\Delta D2}$ #2 30hr genomic DNA and primers flanking each IES (See Table 3 for primers).

Analysis of nuclear morphology post-conjugation. Wild-type or indicated knockout cells 30hr into conjugation were harvested by low speed centrifugation (1,000 X g), DAPI stained (1 µg/ml) and immobilized in 5µl 2% methylcellulose. DIC and DAPI fluorescence images were captured using a Qimaging RetigaEX charge-coupled-device camera (Burnaby, British Columbia, Canada) and Openlab software (PerkinElmer). Conjugation stage of each mating at 30hr was determined by comparison of images with previously described wild-type stages of conjugation [34].

References:

1. St Johnston, D., et al., *A conserved double-stranded RNA-binding domain.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(22): p. 10979-83.

2. Doyle, M. and M.F. Jantsch, *New and old roles of the double-stranded RNA-binding domain.* Journal of structural biology, 2002. **140**(1-3): p. 147-53.
3. Tian, B., et al., *The double-stranded-RNA-binding motif: interference and much more.* Nature reviews. Molecular cell biology, 2004. **5**(12): p. 1013-23.
4. St Johnston, D., D. Beuchle, and C. Nusslein-Volhard, *Staufen, a gene required to localize maternal RNAs in the Drosophila egg.* Cell, 1991. **66**(1): p. 51-63.
5. Wickham, L., et al., *Mammalian staufen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum.* Molecular and cellular biology, 1999. **19**(3): p. 2220-30.
6. DeSousa, D., et al., *A novel double-stranded RNA-binding protein, disco interacting protein 1 (DIP1), contributes to cell fate decisions during Drosophila development.* The Journal of biological chemistry, 2003. **278**(39): p. 38040-50.
7. Bondos, S.E., et al., *Hox transcription factor ultrabithorax Ib physically and genetically interacts with disconnected interacting protein 1, a double-stranded RNA-binding protein.* The Journal of biological chemistry, 2004. **279**(25): p. 26433-44.
8. Green, S.R. and M.B. Mathews, *Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI.* Genes & development, 1992. **6**(12B): p. 2478-90.
9. Meurs, E., et al., *Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon.* Cell, 1990. **62**(2): p. 379-90.
10. Saunders, L.R., et al., *Characterization of two evolutionarily conserved, alternatively spliced nuclear phosphoproteins, NFAR-1 and -2, that function in mRNA processing and interact with the double-stranded RNA-dependent protein kinase, PKR.* The Journal of biological chemistry, 2001. **276**(34): p. 32300-12.
11. Pires-daSilva, A., et al., *Mice deficient for spermatid perinuclear RNA-binding protein show neurologic, spermatogenic, and sperm morphological abnormalities.* Developmental biology, 2001. **233**(2): p. 319-28.
12. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
13. Bernstein, E., et al., *Role for a bidentate ribonuclease in the initiation step of RNA interference.* Nature, 2001. **409**(6818): p. 363-6.
14. Tabara, H., et al., *The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans.* Cell, 2002. **109**(7): p. 861-71.
15. Liu, Q., et al., *R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway.* Science, 2003. **301**(5641): p. 1921-5.
16. Forstemann, K., et al., *Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein.* PLoS Biol, 2005. **3**(7): p. e236.
17. Saito, K., et al., *Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells.* PLoS Biol, 2005. **3**(7): p. e235.
18. Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex.* Nature, 2004. **432**(7014): p. 231-5.

19. Chendrimada, T.P., et al., *TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing.* Nature, 2005. **436**(7051): p. 740-4.
20. Gregory, R.I., et al., *The Microprocessor complex mediates the genesis of microRNAs.* Nature, 2004. **432**(7014): p. 235-40.
21. Malone, C.D., et al., *Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila.* Mol Cell Biol, 2005. **25**(20): p. 9151-64.
22. Mochizuki, K. and M.A. Gorovsky, *A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase.* Genes Dev, 2005. **19**(1): p. 77-89.
23. Prescott, D.M., *The DNA of ciliated protozoa.* Microbiol Rev, 1994. **58**(2): p. 233-67.
24. Meyer, E. and D.L. Chalker, *Epigenetics of Ciliates*, in *Epigenetics*, C.D. Allis, et al., Editors. 2007, Cold Spring Harbor Press: Cold Spring Harbor. p. 127-150.
25. Woodard, J., E. Kaneshiro, and M.A. Gorovsky, *Cytochemical studies on the problem of macronuclear subnuclei in tetrahymena.* Genetics, 1972. **70**(2): p. 251-60.
26. Doerder, F.P., J.C. Deak, and J.H. Lief, *Rate of phenotypic assortment in Tetrahymena thermophila.* Developmental genetics, 1992. **13**(2): p. 126-32.
27. Yao, M.C. and M.A. Gorovsky, *Comparison of the sequences of macro- and micronuclear DNA of Tetrahymena pyriformis.* Chromosoma, 1974. **48**(1): p. 1-18.
28. Altschuler, M.I. and M.C. Yao, *Macronuclear DNA of Tetrahymena thermophila exists as defined subchromosomal-sized molecules.* Nucleic Acids Res, 1985. **13**(16): p. 5817-5831.
29. Conover, R.K. and C.F. Brunk, *Macronuclear DNA molecules of Tetrahymena thermophila.* Mol Cell Biol, 1986. **6**(3): p. 900-905.
30. Cassidy-Hanley, D., et al., *Genome-wide characterization of Tetrahymena thermophila chromosome breakage sites. II. Physical and genetic mapping.* Genetics, 2005. **170**(4): p. 1623-31.
31. Hamilton, E., et al., *Genome-wide characterization of tetrahymena thermophila chromosome breakage sites. I. Cloning and identification of functional sites.* Genetics, 2005. **170**(4): p. 1611-21.
32. Mochizuki, K., et al., *Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena.* Cell, 2002. **110**(6): p. 689-99.
33. Chalker, D.L. and M.C. Yao, *Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila.* Genes Dev, 2001. **15**(10): p. 1287-98.
34. Martindale, D.W., C.D. Allis, and P. Bruns, *Conjugation in Tetrahymena thermophila: a temporal analysis of cytological stages.* Experimental Cell Research, 1982. **140**: p. 227-236.
35. Liu, Y., et al., *RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena.* Genes Dev, 2007. **21**(12): p. 1530-45.

36. Mochizuki, K. and M.A. Gorovsky, *Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement.* Genes Dev, 2004. **18**(17): p. 2068-73.
37. Madireddi, M.T., et al., *Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena.* Cell, 1996. **87**(1): p. 75-84.
38. Nikiforov, M.A., M.A. Gorovsky, and C.D. Allis, *A novel chromodomain protein, pdd3p, associates with internal eliminated sequences during macronuclear development in Tetrahymena thermophila.* Mol Cell Biol, 2000. **20**(11): p. 4128-34.
39. Taverna, S.D., R.S. Coyne, and C.D. Allis, *Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena.* Cell, 2002. **110**(6): p. 701-11.
40. Cheng, C.Y., et al., *A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila.* Mol Biol Cell, 2010. **21**(10): p. 1753-62.
41. Eisen, J.A., et al., *Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote.* PLoS Biol, 2006. **4**(9): p. e286.
42. Stover, N.A., et al., *Tetrahymena Genome Database (TGD): a new genomic resource for Tetrahymena thermophila research.* Nucleic Acids Res, 2006. **34**(Database issue): p. D500-3.
43. Parrish, S. and A. Fire, *Distinct roles for RDE-1 and RDE-4 during RNA interference in Caenorhabditis elegans.* RNA, 2001. **7**(10): p. 1397-402.
44. Coyne, R.S., et al., *Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation.* Mol Cell, 1999. **4**(5): p. 865-72.
45. Austerberry, C.F., C.D. Allis, and M.C. Yao, *Specific DNA rearrangements in synchronously developing nuclei of Tetrahymena.* Proc Natl Acad Sci U S A, 1984. **81**(23): p. 7383-7387.
46. Fass, J.N., et al., *Genome-scale analysis of programmed DNA elimination sites in Tetrahymena thermophila (Unpublished Data).* 2011.
47. Madireddi, M.T., M.C. Davis, and C.D. Allis, *Identification of a novel polypeptide involved in the formation of DNA-containing vesicles during macronuclear development in Tetrahymena.* Dev Biol, 1994. **165**(2): p. 418-31.
48. Lee, S.R. and K. Collins, *Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs.* Nat Struct Mol Biol, 2007. **14**(7): p. 604-10.
49. Lee, S.R., K.B. Talsky, and K. Collins, *A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis.* RNA, 2009. **15**(7): p. 1363-74.
50. Loidl, J. and H. Scherthan, *Organization and pairing of meiotic chromosomes in the ciliate Tetrahymena thermophila.* Journal of cell science, 2004. **117**(Pt 24): p. 5791-801.
51. Cui, B. and M.A. Gorovsky, *Centromeric histone H3 is essential for vegetative cell division and for DNA elimination during conjugation in Tetrahymena thermophila.* Mol Cell Biol, 2006. **26**(12): p. 4499-510.

52. Cervantes, M.D., et al., *The CNA1 Histone of the Ciliate Tetrahymena thermophila Is Essential for Chromosome Segregation in the Germline Micronucleus.* Mol Biol Cell, 2006. **17**(1): p. 485-97.
53. Grishok, A., et al., *Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing.* Cell, 2001. **106**(1): p. 23-34.
54. Nicholson, R.H. and A.W. Nicholson, *Molecular characterization of a mouse cDNA encoding Dicer, a ribonuclease III ortholog involved in RNA interference.* Mammalian genome : official journal of the International Mammalian Genome Society, 2002. **13**(2): p. 67-73.
55. Bass, B.L. and H. Weintraub, *An unwinding activity that covalently modifies its double-stranded RNA substrate.* Cell, 1988. **55**(6): p. 1089-98.
56. Kim, U., et al., *Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(24): p. 11457-61.
57. Larcher, J.C., et al., *Ilf3 and NF90 associate with the axonal targeting element of Tau mRNA.* The FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2004. **18**(14): p. 1761-3.
58. Zhao, G., et al., *NF45/ILF2 tissue expression, promoter analysis, and interleukin-2 transactivating function.* Experimental cell research, 2005. **305**(2): p. 312-23.
59. Schumacher, J.M., et al., *Spnr, a murine RNA-binding protein that is localized to cytoplasmic microtubules.* The Journal of cell biology, 1995. **129**(4): p. 1023-32.
60. Schumacher, J.M., K. Artzt, and R.E. Braun, *Spermatid perinuclear ribonucleic acid-binding protein binds microtubules in vitro and associates with abnormal manchettes in vivo in mice.* Biology of reproduction, 1998. **59**(1): p. 69-76.
61. Kuchta, K., et al., *Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human.* Nucleic acids research, 2009. **37**(22): p. 7701-14.
62. Philippe, H., A. Germot, and D. Moreira, *The new phylogeny of eukaryotes.* Curr Opin Genet Dev, 2000. **10**(6): p. 596-601.
63. Gorovsky, M.A., et al., *Isolation of micro- and macronuclei of Tetrahymena pyriformis.* Methods Cell Biol, 1975. **9**(0): p. 311-327.
64. Orias, E., E.P. Hamilton, and J.D. Orias, *Tetrahymena as a laboratory organism: useful strains, cell culture, and cell line maintenance.* Methods Cell Biol, 2000. **62**: p. 189-211.
65. Zdobnov, E.M. and R. Apweiler, *InterProScan--an integration platform for the signature-recognition methods in InterPro.* Bioinformatics, 2001. **17**(9): p. 847-8.
66. Fan, Q., R. Sweeney, and M.-C. Yao, *Creation and use of antisense ribosomes in Tetrahymena thermophila,* in *Tetrahymena thermophila,* D.J. Asai and J.D. Forney, Editors. 1999, Academic Press: New York. p. 533-547.
67. Yao, M.C., et al., *Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in Tetrahymena.* J Cell Sci, 2007. **120**(Pt 12): p. 1978-89.
68. Gaertig, J., et al., *High frequency vector-mediated transformation and gene replacement in Tetrahymena.* Nucleic Acids Res, 1994. **22**(24): p. 5391-8.

69. Cassidy-Hanley, D., et al., *Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment.* Genetics, 1997. **146**(1): p. 135-47.
70. Bruns, P.J. and D. Cassidy-Hanley, *Biolistic transformation of macro- and micronuclei.* Methods Cell Biol, 2000. **62**: p. 501-12.
71. Chalker, D.L., P. Fuller, and M.C. Yao, *Communication between parental and developing genomes during tetrahymena nuclear differentiation is likely mediated by homologous RNAs.* Genetics, 2005. **169**(1): p. 149-60.

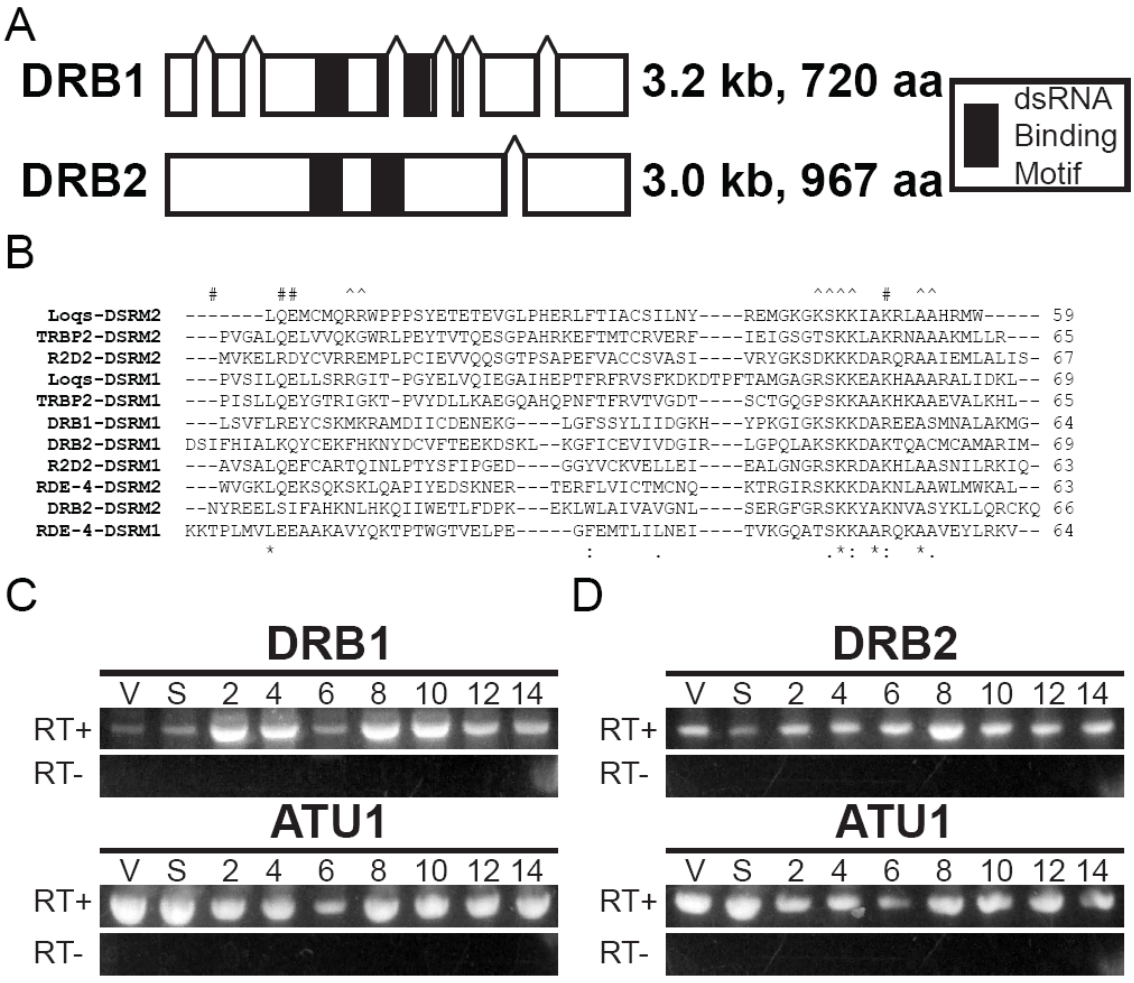


Figure 1: *T. thermophila* contains two predicted tandem double-stranded RNA binding motif proteins. A. Genomic locus, conserved motifs and length of putative tandem double-stranded RNA binding motif proteins, Drb1p and Drb2p). Splice sites are indicated by small connected gaps the gene. Black rectangle, dsRNA binding motif; aa, amino acids. B. ClustalW alignment of DSRMs from Dicer family tandem DSRM-containing partner proteins. #, sites determined to be essential for structure and function of DSRM; ^, sites that have been mutated in DSRM-containing proteins and shown to cause loss of RNA binding. C. and D. RT-PCR analysis of DRB1 and DRB2 expression relative to α-tubulin (ATU1). RNA samples were isolated from CU428 cells growing vegetatively (V), after an 18-hour starvation (S) and from CU428 X B2086 conjugating cells at 2-hour intervals (2, 4, 6, 8, 10, 12 and 14), and were used to monitor the expression of each gene.

A.

```
CLUSTAL 2.1 multiple sequence alignment

              #     ##   ^^                                          ^^^^  #  ^^
 DRB1-DSRM2 -----NYREDVSQLANQVLRLSVSYETLGDQNN--WEVICSIGKS---LKTMGQGLTIKDAKNISAYKMIHLIEQ 65
 DRB2-DSRM2 -----NYREELSIFAHKNLHKQIIWETLFDPKEKLWLAIVAVGN----LSERGFGRSKKYAKNVASYKLLQRCKQ 66
 Loqs-DSRM1 ---PVSILQELLSRRG-ITPGYELVQIEGAIHEPTFRFRVSFKDKDTPFTAMGAGRSKKEAKHAAARALIDKL-- 69
TRBP2-DSRM1 ---PISLLQEYGTRIG-KTPVYDLLKAEGQAHQPNFTFRVTVGDT----SCTGQGPSKKAAKHKAAEVALKHL-- 65
Consen-DSRM ---YKSLLQEYVQKKYKISPSYKLDKEIGPDHDKVFCVELYVGEN---FISNGKGKSKKEAEMRAAEVALKAME- 68
TRBP2-DSRM2 ---PVGALQELVVQKGWRLPEYTVTQESGPAHRKEFTMTCRV-ER---FIEIGSGTSKKLAKRNAAAKMLLR--- 65
 Loqs-DSRM2 -------LQEMCMQRRWPPPSYETETEVGLPHERLFTIACSI-LN---YREMGKGKSKKIAKRLAAHRMW----- 59
 DRB1-DSRM1 ---LSVFLREYCSKMK--RAMDIICDENEKG--LGFSSYLIIDGK---HYPKGIGKSKKDAREEASMNALAKMG- 64
 DRB2-DSRM1 DSIFHIALKQYCEKFH--KNYDCVFTEEKDSKLKGFICEVIVDGI---RLGPQLAKSKKDAKTQACMCAMARIM- 69
 R2D2-DSRM1 ---AVSALQEFCARTQINLPTYSFIPGEDGG----YVCKVELLEI----EALGNGRSKRDAKHLAASNILRKIQ- 63
 R2D2-DSRM2 ---MVKELRDYCVRREMPLPCIEVVQQSGTPSAPEFVACCSVASI----VRYGKSDKKKDARQRAAIEMLALIS- 67
RDE-4-DSRM2 ---WVGKLQEKSQKSKLQAPIYEDSKNERTER---FLVICTMCNQ----KTRGIRSKKKDAKNLAAWLMWKAL-- 63
RDE-4-DSRM1 KKTPLMVLEEAAKAVYQKTPTWG-TVELPEG----FEMTLILNEI----TVKGQATSKKAARQKAAVEYLRKV-- 64
                     .:                          :     .              . : *.  :.

CLUSTAL 2.1 multiple sequence alignment

                #     ##   ^^                                        ^^^^  #  ^^
DRB1-DSRM1   ---LSVFLREYCSKMKRAM-DIICDENEKG---LGFSSYLIIDGKHYP---KGIGKSKKDAREEASMNALAKMG 64
DRB2-DSRM1   DSIFHIALKQYCEKFHKNY-DCVFTEEKDSK-LKGFICEVIVDGIRLG---PQLAKSKKDAKTQACMCAMARIM 69
R2D2-DSRM1   ---AVSALQEFCARTQINLPTYSFIPGEDG----GYVCKVELLEIEA----LGNGRSKRDAKHLAASNILRKIQ 63
Loqs-DSRM1   ---PVSILQELLSRRGITP-GYELVQIEGAIHEPTFRFRVSFKDKDTPFTAMGAGRSKKEAKHAAARALIDKL- 69
                     *::   :            : .     :   : .           .:**::*:  *.   : ::
```

B.

>**DRB1-DSRM1**
LSVFLREYCSKMKRAMDIICDENEKGLGFSSYLIIDGKHYPKGIGKSKKDAREEASMNALAKMG

>**DRB1-DSRM2**
NYREDVSQLANQVLRLSVSYETLGDQNNWEVICSIGKSLKTMGQGLTIKDAKNISAYKMIHLIEQ

>**DRB2-DSRM1**
DSIFHIALKQYCEKFHKNYDCVFTEEKDSKLKGFICEVIVDGIRLGPQLAKSKKDAKTQACMCAMARIM

>**DRB2-DSRM2**
NYREELSIFAHKNLHKQIIWETLFDPKEKLWLAIVAVGNLSERGFGRSKKYAKNVASYKLLQRCKQ

>**R2D2-DSRM1**
AVSALQEFCARTQINLPTYSFIPGEDGGYVCKVELLEIEALGNGRSKRDAKHLAASNILRKIQ

>**R2D2-DSRM2**
MVKELRDYCVRREMPLPCIEVVQQSGTPSAPEFVACCSVASIVRYGKSDKKKDARQRAAIEMLALIS

>**Loqs-DSRM1**
PVSILQELLSRRGITPGYELVQIEGAIHEPTFRFRVSFKDKDTPFTAMGAGRSKKEAKHAAARALIDKL

>**Loqs-DSRM2**
LQEMCMQRRWPPPSYETETEVGLPHERLFTIACSILNYREMGKGKSKKIAKRLAAHRMW

>**RDE-4-DSRM1**
KKTPLMVLEEAAKAVYQKTPTWGTVELPEGFEMTLILNEITVKGQATSKKAARQKAAVEYLRKV

>**RDE-4-DSRM2**
WVGKLQEKSQKSKLQAPIYEDSKNERTERFLVICTMCNQKTRGIRSKKKDAKNLAAWLMWKAL

>**TRBP2-DSRM1**
PISLLQEYGTRIGKTPVYDLLKAEGQAHQPNFTFRVTVGDTSCTGQGPSKKAAKHKAAEVALKHL

>**TRBP2-DSRM2**
PVGALQELVVQKGWRLPEYTVTQESGPAHRKEFTMTCRVERFIEIGSGTSKKLAKRNAAAKMLLR

>**Consensus-DSRM (SMART)**
YKSLLQEYVQKKYKISPSYKLDKEIGPDHDKVFCVELYVGENFISNGKGKSKKEAEMRAAEVALKAME

C.

SeqA	Name	Length	SeqB	Name	Length	Score
1	DRB1-DSRM1	64	2	DRB1-DSRM2	65	7.0
1	DRB1-DSRM1	64	3	DRB2-DSRM1	69	32.0
1	DRB1-DSRM1	64	4	DRB2-DSRM2	66	15.0
1	DRB1-DSRM1	64	5	R2D2-DSRM1	63	25.0
1	DRB1-DSRM1	64	6	R2D2-DSRM2	67	23.0
1	DRB1-DSRM1	64	7	Loqs-DSRM1	69	12.0
1	DRB1-DSRM1	64	8	Loqs-DSRM2	59	13.0
1	DRB1-DSRM1	64	9	RDE-4-DSRM1	64	15.0
1	DRB1-DSRM1	64	10	RDE-4-DSRM2	63	3.0
1	DRB1-DSRM1	64	11	TRBP2-DSRM1	65	23.0
1	DRB1-DSRM1	64	12	TRBP2-DSRM2	65	12.0
1	DRB1-DSRM1	64	13	Consensus-DSRM	68	20.0
2	DRB1-DSRM2	65	3	DRB2-DSRM1	69	3.0
2	DRB1-DSRM2	65	4	DRB2-DSRM2	66	36.0
2	DRB1-DSRM2	65	5	R2D2-DSRM1	63	11.0
2	DRB1-DSRM2	65	6	R2D2-DSRM2	67	16.0
2	DRB1-DSRM2	65	7	Loqs-DSRM1	69	12.0
2	DRB1-DSRM2	65	8	Loqs-DSRM2	59	20.0
2	DRB1-DSRM2	65	9	RDE-4-DSRM1	64	7.0
2	DRB1-DSRM2	65	10	RDE-4-DSRM2	63	17.0
2	DRB1-DSRM2	65	11	TRBP2-DSRM1	65	10.0
2	DRB1-DSRM2	65	12	TRBP2-DSRM2	65	13.0
2	DRB1-DSRM2	65	13	Consensus-DSRM	68	10.0
3	DRB2-DSRM1	69	4	DRB2-DSRM2	66	12.0
3	DRB2-DSRM1	69	5	R2D2-DSRM1	63	4.0
3	DRB2-DSRM1	69	6	R2D2-DSRM2	67	5.0
3	DRB2-DSRM1	69	7	Loqs-DSRM1	69	8.0
3	DRB2-DSRM1	69	8	Loqs-DSRM2	59	3.0
3	DRB2-DSRM1	69	9	RDE-4-DSRM1	64	10.0
3	DRB2-DSRM1	69	10	RDE-4-DSRM2	63	12.0
3	DRB2-DSRM1	69	11	TRBP2-DSRM1	65	10.0
3	DRB2-DSRM1	69	12	TRBP2-DSRM2	65	7.0
3	DRB2-DSRM1	69	13	Consensus-DSRM	68	14.0
4	DRB2-DSRM2	66	5	R2D2-DSRM1	63	15.0
4	DRB2-DSRM2	66	6	R2D2-DSRM2	67	13.0
4	DRB2-DSRM2	66	7	Loqs-DSRM1	69	15.0
4	DRB2-DSRM2	66	8	Loqs-DSRM2	59	22.0
4	DRB2-DSRM2	66	9	RDE-4-DSRM1	64	10.0
4	DRB2-DSRM2	66	10	RDE-4-DSRM2	63	23.0
4	DRB2-DSRM2	66	11	TRBP2-DSRM1	65	20.0
4	DRB2-DSRM2	66	12	TRBP2-DSRM2	65	21.0
4	DRB2-DSRM2	66	13	Consensus-DSRM	68	12.0
5	R2D2-DSRM1	63	6	R2D2-DSRM2	67	25.0
5	R2D2-DSRM1	63	7	Loqs-DSRM1	69	19.0
5	R2D2-DSRM1	63	8	Loqs-DSRM2	59	27.0

5	R2D2-DSRM1	63	9	RDE-4-DSRM1	64	20.0
5	R2D2-DSRM1	63	10	RDE-4-DSRM2	63	23.0
5	R2D2-DSRM1	63	11	TRBP2-DSRM1	65	22.0
5	R2D2-DSRM1	63	12	TRBP2-DSRM2	65	28.0
5	R2D2-DSRM1	63	13	Consensus-DSRM	68	12.0
6	R2D2-DSRM2	67	7	Loqs-DSRM1	69	20.0
6	R2D2-DSRM2	67	8	Loqs-DSRM2	59	32.0
6	R2D2-DSRM2	67	9	RDE-4-DSRM1	64	18.0
6	R2D2-DSRM2	67	10	RDE-4-DSRM2	63	19.0
6	R2D2-DSRM2	67	11	TRBP2-DSRM1	65	16.0
6	R2D2-DSRM2	67	12	TRBP2-DSRM2	65	32.0
6	R2D2-DSRM2	67	13	Consensus-DSRM	68	20.0
7	Loqs-DSRM1	69	8	Loqs-DSRM2	59	33.0
7	Loqs-DSRM1	69	9	RDE-4-DSRM1	64	20.0
7	Loqs-DSRM1	69	10	RDE-4-DSRM2	63	12.0
7	Loqs-DSRM1	69	11	TRBP2-DSRM1	65	47.0
7	Loqs-DSRM1	69	12	TRBP2-DSRM2	65	35.0
7	Loqs-DSRM1	69	13	Consensus-DSRM	68	25.0
8	Loqs-DSRM2	59	9	RDE-4-DSRM1	64	13.0
8	Loqs-DSRM2	59	10	RDE-4-DSRM2	63	32.0
8	Loqs-DSRM2	59	11	TRBP2-DSRM1	65	25.0
8	Loqs-DSRM2	59	12	TRBP2-DSRM2	65	42.0
8	Loqs-DSRM2	59	13	Consensus-DSRM	68	35.0
9	RDE-4-DSRM1	64	10	RDE-4-DSRM2	63	14.0
9	RDE-4-DSRM1	64	11	TRBP2-DSRM1	65	18.0
9	RDE-4-DSRM1	64	12	TRBP2-DSRM2	65	18.0
9	RDE-4-DSRM1	64	13	Consensus-DSRM	68	12.0
10	RDE-4-DSRM2	63	11	TRBP2-DSRM1	65	17.0
10	RDE-4-DSRM2	63	12	TRBP2-DSRM2	65	26.0
10	RDE-4-DSRM2	63	13	Consensus-DSRM	68	15.0
11	TRBP2-DSRM1	65	12	TRBP2-DSRM2	65	27.0
11	TRBP2-DSRM1	65	13	Consensus-DSRM	68	33.0
12	TRBP2-DSRM2	65	13	Consensus-DSRM	68	40.0

SeqA	Name	Length	SeqB	Name	Length	Score
1	DRB1-DSRM1	64	2	DRB2-DSRM1	69	32.0
1	DRB1-DSRM1	64	3	R2D2-DSRM1	63	25.0
1	DRB1-DSRM1	64	4	Loqs-DSRM1	69	12.0
2	DRB2-DSRM1	69	3	R2D2-DSRM1	63	4.0
2	DRB2-DSRM1	69	4	Loqs-DSRM1	69	8.0
3	R2D2-DSRM1	63	4	Loqs-DSRM1	69	19.0

Figure S1: ClustalW alignment of dsRNA binding motifs (DSRMs) of *D. melanogaster* Loqs and R2D2, *C. elegans* RDE-4, *H. sapiens* TRBP2 and *T. thermophila* DRB1 and DRB2. A. Top Panel. ClustalW alignment of *D. melanogaster* Loqs and R2D2, *C. elegans* RDE-4, *H. sapiens* TRBP2 and *T. thermophila* DRB1 and DRB2. Bottom Panel. ClustalW alignment of first DSRM of *D. melanogaster* Loqs and R2D2 and *T. thermophila* DRB1 and DRB2 displaying stronger conservation. #, sites essential for structure and function of DSRM; ^, sites that have been mutated in DSRM-containing proteins and shown to cause loss of RNA binding. B. DSRM protein sequence of consensus DSRM sequence and both DSRMs of *D. melanogaster* Loqs and R2D2, *C. elegans* RDE-4, *H. sapiens* TRBP2 and *T. thermophila* DRB1 and DRB2. C. Top. Scoring table of pairwise comparison of DSRM sequences from consensus DSRM sequence and both DSRMs of *D. melanogaster* Loqs and R2D2, *C. elegans* RDE-4, *H. sapiens* TRBP2 and *T. thermophila* DRB1 and DRB2. Bottom. Scoring table of pairwise comparison of first DSRM sequences of *D. melanogaster* Loqs and R2D2 and *T. thermophila* DRB1 and DRB2.

A:

Sequence	Identity (%)	Similarity (%)	Gaps (%)	Length (nt)	Score
DRB1/DRB2	20.2	39.8	29.0	1007	517.0
DRB1 DSRM1/DRB2 DSRM 2	31.9	52.2	7.2	69	93.5
DRB1 DSRM2/DRB2 DSRM 2	35.8	61.2	4.5	67	100.5
DRB1 NTS/DRB2 NTS	24.5	43.9	13.5	155	95.5
DRB1 CT1/DRB2 CT1	26.0	64.0	6.0	50	54.0
DRB1 CT2/DRB2 CT2	26.8	55.8	8.0	138	141.0

B:

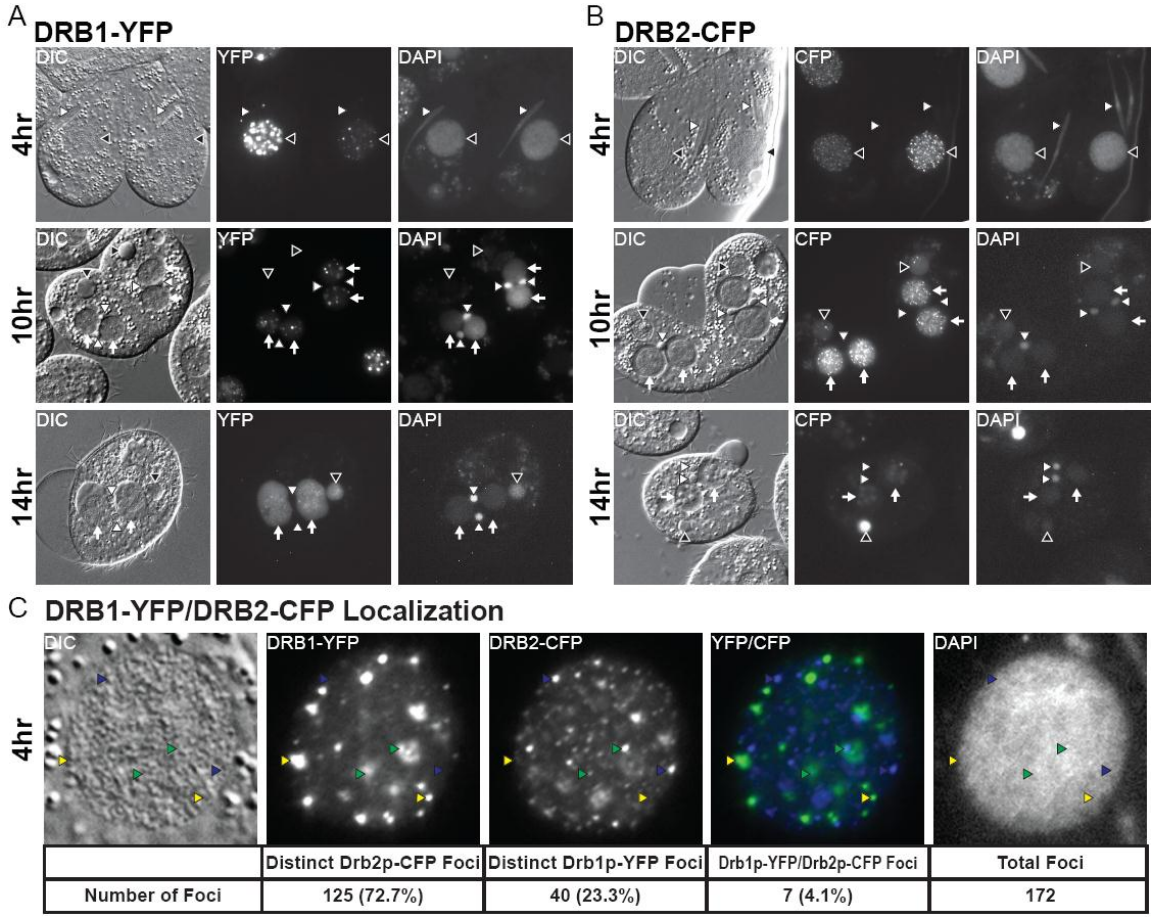
```
Aligned_sequences: 2
1: DRB1
2: DRB2
Matrix: EBLOSUM62
Gap_penalty: 10.0
Extend_penalty: 0.5

Length: 1007
Identity:     203/1007 (20.2%)
Similarity:   401/1007 (39.8%)
Gaps:         292/1007 (29.0%)
Score: 517.0

DRB1               0 --------------------------------------------------      0

DRB2               1 MAQSFRFIDQETQDFLMEFSDLQNNEVVQKLQKDQMKLLQLKNNLNNIAN     50

DRB1               1 -----------------------------------MNSQQELYNLNSF--     13
                                                        .|.|...:|.|.:
DRB2              51 CASINPTKKSQEVIYQNAQQTNNFQQDQQSFNDKNANFQGNNFNGNLYFQ    100

DRB1              14 QNYQNMNQYPQNYPYYSNNQIIHPSQTANMNNSAGMQYINHEQNYPQYQQ     63
                     |||:|.||..|:.    |:..|...:..::....|.|....|....:||.
DRB2             101 QNYKNGNQEQQSI----NDSYIIDKKQNSIRAEQGEQKREKENEIVKYQF    146

DRB1              64 QSFNQ------QNYYHQEAAQYQNQNMFQVPPQQVNPYMGYQNNPPINQN    107
                     :...:      |:....|..::|.|.:.:...:::     :.:|.....
DRB2             147 KELKKKMIQRSQDQIIDEIYKHQKQLILKEATDRLD-----KEDPSQKSE    191

DRB1             108 VN---NLNQNNLLQDQYKRSETQSNFQNSYNYSQQQQHDI-----DPSTI    149
                     .|   .:.|||.|:||.|.:.:.:|.| :.|:.:||:.||     .|:.:
DRB2             192 QNCQQTIQQNNSLEDQDKFNYSVNNGQ-AMNHQRQQKQDILRMLNPPNYL    240

DRB1             150 ---------GIDQQLQT-VKDKEEINLKTLNKK--RKQQEIQECLIDQNK    187
                             ||:.:|.| .:|....||..:.|.  ::::...:.|..:.:
DRB2             241 KQKWNIIQKGIEDRLITHNEDGLPKNLNPIEKSILKEKENGYKLLKPEGE    290

DRB1             188 NDDQSPQKQDKNENNF---------VQNLMDDIQLYFPQKP LSVF---LR    225
                     ..:...|:::::|.|.         :..::::.:...:...|:|   |:
DRB2             291 IGELEQQEEEEDEENMSNEVLITRKIVKILEEFKEVQKRDQ DSIFHIALK    340

DRB1             226 EYCSKMKRAMDIICDE--NEKGLGFSSYLIIDGKHYPKGIGKSKKDAREE    273
                     :||.|..:..|.:..|  :.|..||...:|:||......:.||||||:.:
DRB2             341 QYCEKFHKNYDCVFTEEKDSKLKGFICEVIVDGIRLGPQLAKSKKDAKTQ    390

DRB1             274 ASMNALAKMGEENSSVK--NYFAVMVDKEKKVHQKFKNEKAKSSSSEQKS    321
                     |.|.|:|::.:::.::|  .||...:::.|..:.:.....|::.|...:.
DRB2             391 ACMCAMARIMQQSPNMKPIEYFQKFLEQSKNKYIRLNENIAEAISQHHQQ    440

DRB1             322 KSKDISEEEERALELYDPL-----------HNYREDVSQLANQVLRLSV    359
                     ....:|.:        :||            .||||::|..|::.|...:
DRB2             441 GGVPLSAD--------NPLSDLAIAQEIPENSNYREELSIFAHKNLHKQI    482

DRB1             360 SYETLGD--QNNWEVICSIGKSLKTMGQGLTIKDAKNISAYKMIHLIEQ-    406
                     .:|||.|  :..|..|.::| :|...|.|.:.|.|||:::||::...:|
DRB2             483 IWETLFDPKEKLWLAIVAVG-NLSERGFGRSKKYAKNVASYKLLQRCKQL    531
```

```
DRB1             407 -------SNEKTIKT--------------KQIRPTQSLKELKAQKEKDEQ    435
                             .|.:.:||              :.|:...||.::|...:::||
DRB2             532 KMKDDIFKNPRKVKTQLQFALNLEQNEDEENIKQQNSLIKIKQSGKQNEQ    581

DRB1             436 ------NTQIKEI----------QKEQENLEKENKQHQQSLLNTYSCLID    469
                           .:|::::          |:::.:|:|:.:|.||         |.
DRB2             582 INSKRLRSQVEDLENDLDQEDEEQEQEPDLKKQQQQQQQ---------IQ    622

DRB1             470 SNYEKPQELQKPKIKSQQSII--------------SSCLNKFKQKDVVSE    505
                     ...::|......:||.::.||              ||.||:.|..:..:.
DRB2             623 QQQQQPNTKSVAQIKMKELIISLEEKYIDYYKLELSSNLNQSKINNASNT    672

DRB1             506 EIYIKLRSKFDLLIN-SKQLFQNLG----------LVD--------FCIV    536
                     :::.:| .|:|..|. :|..:|:|.          |:|        ...:
DRB2             673 DVFKQL-LKYDSEIQATKAQYQSLQNMFNEVQKKLLIDQQSQYIKKLIPI    721

DRB1             537 GSHEFRLLRKNKLSIDVI--LNYNKEQLININTGE--DIFYEAQKILKQI    582
                     ||:....:.|....:||:  |..::||..|.....  ||.|  .::...|
DRB2             722 GSYVSNCMNKYSTVMDVVALLMPSQEQGENFEPNRFCDILY--SEVQSTI    769

DRB1             583 LNNH------------------------------TIKLEEK--NQKKYIR    600
                     ||::                             .:.|||.  |::||::
DRB2             770 LNSNIGYFFKMELNKGCVLLVFTHNSQYKIRLIPNLVLEENILNKRKYLQ    819

DRB1             601 VIDD-LEKELEMIIYFSDVSGGSFCNE-IHHAIWFSQVKTKIFDNKDISA    648
                     ..:: |...::.||   |.:...|.:: :.|.||.||.|::|...:.:..
DRB2             820 QTENPLNDSIKTII---DPNESQFYSDGLLHYIWLSQFKSQIHQYESLFR    866

DRB1             649 MLKCLRMWKEASNININTSIIDFIMHTVTHKFTLINAPKNLVSFFEFVAN    698
                     :|:  |::| .....:...:||.::..|:|:|:..:..:|:|.|.:|:|.
DRB2             867 ILR--RVFK-YQGFRVPIDLIDVVVGHVSHQFSSASFQENIVGFVKFMAQ    913

DRB1             699 EGIQDIINLKNKRNIEVTEMHFEMIQKINKEDLDQFVRQCQSIYNSCDYT    748
                     .|:.||.: .|.:.:.:|::|...|:..::.:|.:......|||....|.
DRB2             914 NGLDDITH-DNLQMVSITQLHVNQIKSCSEFELKKITEYFNSIYTKKAYD    962

DRB1             749 FMKSISL    755
                     .:..|
DRB2             963 NLIII--    967

Aligned_sequences: 2
1: DRB1-DSRM1
2: DRB2-DSRM1
Matrix: EBLOSUM62
Gap_penalty: 10.0
Extend_penalty: 0.5

Length: 69
Identity:      22/69 (31.9%)
Similarity:    36/69 (52.2%)
Gaps:           5/69 ( 7.2%)
Score: 93.5

DRB1-DSRM1         1 ---LSVFLREYCSKMKRAMDIICDE--NEKGLGFSSYLIIDGKHYPKGIG     45
                       ..:.|::||.|..:..|.:..|  :.|..||...:|:||.......:.
DRB2-DSRM1         1 DSIFHIALKQYCEKFHKNYDCVFTEEKDSKLKGFICEVIVDGIRLGPQLA     50

DRB1-DSRM1        46 KSKKDAREEASMNALAKMG     64
                     ||||||:.:|.|.|:|::.
DRB2-DSRM1        51 KSKKDAKTQACMCAMARIM     69
```

```
Aligned_sequences: 2
1: DRB1-DSRM2
2: DRB2-DSRM2
Matrix: EBLOSUM62
Gap_penalty: 10.0
Extend_penalty: 0.5

Length: 67
Identity:      24/67 (35.8%)
Similarity:    41/67 (61.2%)
Gaps:           3/67 ( 4.5%)
Score: 100.5

DRB1-DSRM2         1 NYREDVSQLANQVLRLSVSYETLGD--QNNWEVICSIGKSLKTMGQGLTI     48
                     ||||::|..|::.|...:.:|||.|  :..|..|.::| :|...|.|.:.
DRB2-DSRM2         1 NYREELSIFAHKNLHKQIIWETLFDPKEKLWLAIVAVG-NLSERGFGRSK     49

DRB1-DSRM2        49 KDAKNISAYKMIHLIEQ     65
                     |.|||:::||::...:|
DRB2-DSRM2        50 KYAKNVASYKLLQRCKQ     66

Aligned_sequences: 2
1: DRB1-NTS
2: DRB2-NTS
Matrix: EBLOSUM62
Gap_penalty: 10.0
Extend_penalty: 0.5

Length: 155
Identity:      38/155 (24.5%)
Similarity:    68/155 (43.9%)
Gaps:          21/155 (13.5%)
Score: 95.5

DRB1-NTS           1 MNSQQELYNLNSF--QNYQNMNQYPQNYPYYSNNQIIHPSQTANMNNSAG     48
                     .|.|...:|.|.:  |||:|.||..|:.    |:..|...:..::....|
DRB2-NTS           1 ANFQGNNFNGNLYFQQNYKNGNQEQQSI----NDSYIIDKKQNSIRAEQG     46

DRB1-NTS          49 MQYINHEQNYPQYQQQSFNQ------QNYYHQEAAQYQNQNMFQVPPQQV     92
                     .|....|....:||.:...:      |:....|..::|.|.:.:....::
DRB2-NTS          47 EQKREKENEIVKYQFKELKKKMIQRSQDQIIDEIYKHQKQLILKEATDRL     96

DRB1-NTS          93 NPYMGYQNNPPINQNVN---NLNQNNLLQDQYKRSETQSNFQNSYNYSQQ    139
                     :     :.:|......|   .:.|||.|:||.|.:.:.:|.| :.|:.:|
DRB2-NTS          97 D-----KEDPSQKSEQNCQQTIQQNNSLEDQDKFNYSVNNGQ-AMNHQRQ    140

DRB1-NTS         140 QQHDI    144
                     |:.||
DRB2-NTS         141 QKQDI    145

Aligned_sequences: 2
1: DRB1-CT1
2: DRB2-CT1
Matrix: EBLOSUM62
Gap_penalty: 10.0
Extend_penalty: 0.5

Length: 50
Identity:      13/50 (26.0%)
Similarity:    32/50 (64.0%)
Gaps:           3/50 ( 6.0%)
Score: 54.0

DRB1-CT1           1 SNEKTIKTKQIR-PTQSLK-ELKAQKEKDEQNTQIKEIQKEQENLEKENK     48
                     .||: |.:|::| ..:.|: :|..:.|:.||...:|:.|::|:.::::.:
DRB2-CT1           1 QNEQ-INSKRLRSQVEDLENDLDQEDEEQEQEPDLKKQQQQQQQIQQQQQ     49
```

```
Aligned_sequences: 2
1: DRB1-CT2
2: DRB2-CT2
Matrix: EBLOSUM62
Gap_penalty: 10.0
Extend_penalty: 0.5

Length: 138
Identity:      37/138 (26.8%)
Similarity:    77/138 (55.8%)
Gaps:          11/138 ( 8.0%)
Score: 141.0

DRB1-CT2           1 TIKLEEK--NQKKYIRVIDD-LEKELEMIIYFSDVSGGSFCNE-IHHAIW     46
                     .:.|||.  |::||::..:: |...::.||   |.:...|.:: :.|.||
DRB2-CT2           1 NLVLEENILNKRKYLQQTENPLNDSIKTII---DPNESQFYSDGLLHYIW     47

DRB1-CT2          47 FSQVKTKIFDNKDISAMLKCLRMWKEASNININTSIIDFIMHTVTHKFTL     96
                     .||.|::|...:.:..:|:  |::| .....:...:||.::..|:|:|:.
DRB2-CT2          48 LSQFKSQIHQYESLFRILR--RVFK-YQGFRVPIDLIDVVVGHVSHQFSS     94

DRB1-CT2          97 INAPKNLVSFFEFVANEGIQDIINLKNKRNIEVTEMHF    134
                     .:..:|:|.|.:|:|..|:.||.: .|.:.:.:|::|.
DRB2-CT2          95 ASFQENIVGFVKFMAQNGLDDITH-DNLQMVSITQLHV    131
```

Figure S2: Alignment of DRB1 and DRB2 protein sequences outside of the DSRMs shows additional regions of homology. A. Table of DRB1 and DRB2 protein alignments summarizing percent protein identity, percent similarity, gap percentage, length of compared sequences, and score of full length protein, each DSRM, an N-terminal homology region (NTS) and two C-terminal homology regions (CT1 and CT2). B. Alignment of full length, DSRM1, DSRM2, NTS, CT1 and CT2 of DRB1 and DRB2 protein sequences. Bold sequences in full length DRB1 and DRB2 alignment are DSRM1 and DSRM2.



	Distinct Drb2p-CFP Foci	Distinct Drb1p-YFP Foci	Drb1p-YFP/Drb2p-CFP Foci	Total Foci
Number of Foci	125 (72.7%)	40 (23.3%)	7 (4.1%)	172

Figure 2: Nuclear localization of Drb1p and Drb2p during conjugation. A and B. Nuclear localization of Drb1p-YFP (A) and Drb2p-CFP (B) at 4, 10 and 14 hours into conjugation. White arrowhead, micronuclei; black arrowhead, parental macronuclei; white arrow, zygotic macronuclei. C. Simultaneous localization of Drb1p-YFP and Drb2p-CFP in the parental macronucleus 4 hours into conjugation. Top. Drb1p-YFP and Drb2p-CFP foci are predominantly distinct in the macronucleus early during conjugation. Yellow arrowheads, Drb1p-YFP foci only; blue arrowheads, Drb2p-CFP foci only; green arrowheads, Drb1p-YFP and Drb2p-CFP foci. Bottom. Table of number and percentage of Drb1p-YFP foci, Drb2p-CFP foci, Drb1p-YFP/Drb2p-CPF co-localization foci and total number of foci in parental macronucleus above.

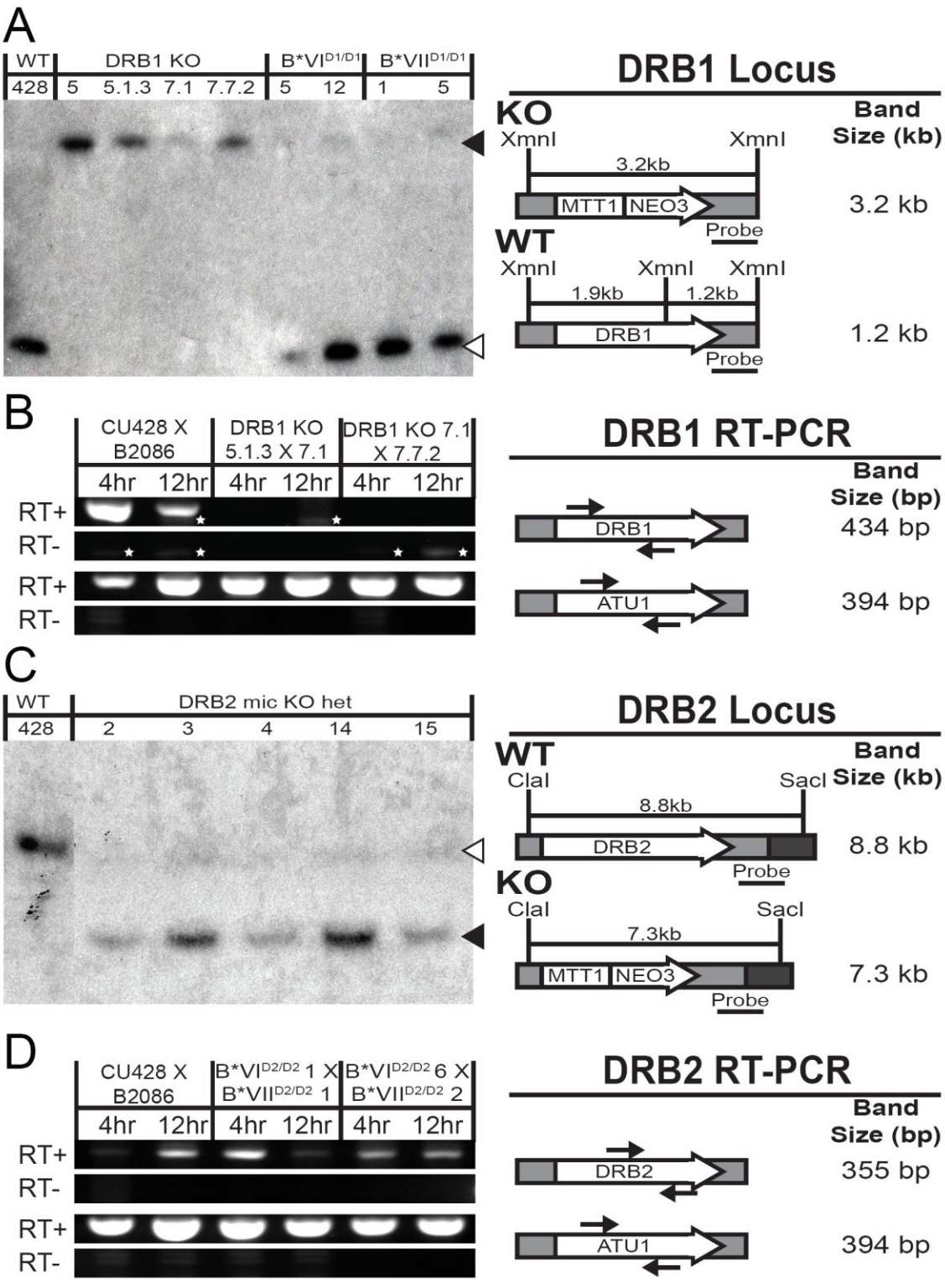

A
WT
DRB1 KO
B*VI D1/D1
B*VII D1/D1
428
5
5.1.3
7.1
7.7.2
5
12
1
5
DRB1 Locus
KO
XmnI
XmnI
3.2kb
MTT1
NEO3
Probe
Band Size (kb)
3.2 kb
WT
XmnI
XmnI
XmnI
1.9kb
1.2kb
DRB1
Probe
1.2 kb
B
CU428 X B2086
DRB1 KO 5.1.3 X 7.1
DRB1 KO 7.1 X 7.7.2
4hr
12hr
4hr
12hr
4hr
12hr
RT+
RT-
RT+
RT-
DRB1 RT-PCR
Band Size (bp)
DRB1
434 bp
ATU1
394 bp
C
WT
DRB2 mic KO het
428
2
3
4
14
15
DRB2 Locus
WT
ClaI
SacI
8.8kb
DRB2
Probe
Band Size (kb)
8.8 kb
KO
ClaI
SacI
7.3kb
MTT1
NEO3
Probe
7.3 kb
D
CU428 X B2086
B*VI D2/D2 1 X B*VII D2/D2 1
B*VI D2/D2 6 X B*VII D2/D2 2
4hr
12hr
4hr
12hr
4hr
12hr
RT+
RT-
RT+
RT-
DRB2 RT-PCR
Band Size (bp)
DRB2
355 bp
ATU1
394 bp

Figure S3: Southern blots and RT-PCR during conjugation of DRB1 knockout strains and DRB2 mic knockout strains. A. Southern blot of DRB1 knockout strains. Left. Southern blot of DRB1 knockout strains with wild-type CU428 control. WT 428, wild-type CU428; DRB1 KO 5, (ΔDRB1/+ [V, ΔDRB1/+]); DRB1 KO 5.1.3, 7.1, 7.2; (ΔDRB1/ ΔDRB1 [V/IV/III, ΔDRB1]); B*VI$^{\Delta D1/\Delta D1}$ 5 and 12 and B*VII$^{\Delta D1/\Delta D1}$ 1 and 5, (ΔDRB1/ ΔDRB1 [VI/VII, +]). Black arrowhead, DRB1 knockout band; white arrowhead, wild-type DRB1 band. Right. Diagram of DRB1 locus with wild-type DRB1 or DRB1 MTT1-NEO3 knockout cassette with expected band sizes from probing of XmnI genomic DNA digest. B. RT-PCR of DRB1 knockout mating cDNA during conjugation to verify loss of DRB1 expression. Left. RT-PCR of DRB1 knockout mating cDNA at 4 hours (parental expression) and 12 hours (zygotic expression) into conjugation. RNA samples were isolated from wild-type, DRB1 knockout cells at 4 hours and 12 hours into conjugation. RNA loading controls were done by reverse-transcribing the α-tubulin gene (ATU1). Star, non-specific DRB1 RT-PCR band. Right. Diagram of DRB1 and ATU1 locus for RT-PCR. Forward and reverse PCR primers represented by black arrows. C. Southern blot of DRB2 mic knockout strains. Left. Southern blot of DRB2 mic knockout strains with wild-type CU428 control. WT 428, wild-type CU428; DRB2 mic KO 2, 3, 4, 14, 15 het, (ΔDRB2/+ [II/V/IV, ΔDRB2/+]). Black arrowhead, DRB2 knockout band; white arrowhead, wild-type DRB2 band. Right. Diagram of DRB2 locus with wild-type DRB2 or DRB2 MTT1-NEO3 knockout cassette with expected band sizes from probing of ClaI and SacI genomic DNA digest. D. RT-PCR of DRB2 mic knockout mating cDNA during conjugation to verify reduction of zygotic DRB2 expression. Left. RT-PCR of DRB2 mic knockout mating cDNA at 4 hours (parental expression) and 12 hours (zygotic expression) into conjugation. RNA samples were isolated from wild-type and DRB2 mic knockout cells at 4 hours and 12 hours into conjugation. B*VI$^{\Delta D2/\Delta D2}$ 1 and 6 and B*VII$^{\Delta D2/\Delta D2}$ 1 and 2 are (ΔDRB2/ ΔDRB2 [VI/VII, +]) and derived from DRB2 mic KO 15. RNA loading controls were done by reverse-transcribing the α-tubulin gene (ATU1). Right. Diagram of DRB2 and ATU1 locus for RT-PCR. Forward and reverse PCR primers represented by black arrows.

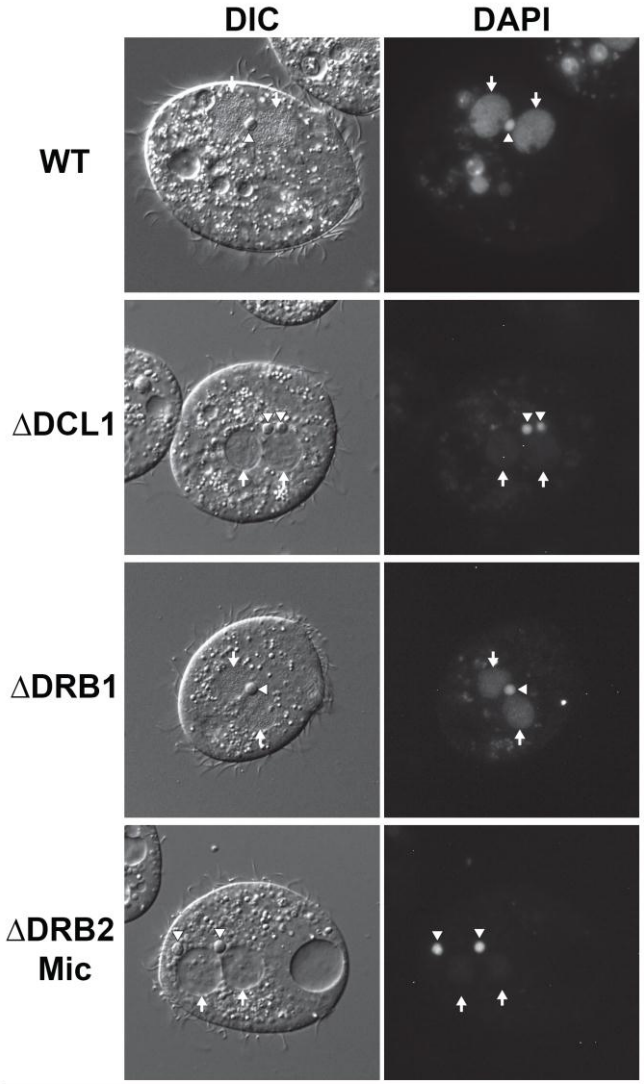
Table 1: Progeny production of DRB1 knockouts in wild-type and knockout matings

Cross	Pair Survival (S/N)[1]	Progeny Production (P/S)[2]
CU427 X DRB1 KO #5.1.3	**97.2 %** (171/176)	**96.5 %** (165/171)
CU427 X DRB1 KO #6.1.6	**98.8 %** (87/88)	**98.9 %** (86/87)
CU427 X DRB1 KO #6.1.12.1	**96.0 %** (169/176)	**98.2 %** (166/169)
CU427 X DRB1 KO #6.1.12.2	**98.9 %** (174/176)	**97.1 %** (169/174)
CU427 X DRB1 KO #7.1	**97.7 %** (129/132)	**98.4 %** (127/129)
CU427 X DRB1 KO #7.7.2	**99.2 %** (131/132)	**94.7 %** (124/131)
DRB1 KO #5.1.3 X #6.1.12.1	**93.5 %** (247/264)	**33.3 %** (6/18)
DRB1 KO #5.1.3 X #6.1.12.2	**94.7 %** (250/264)	**51.5 %** (35/68)

Table 1: Progeny production of DRB1 knockouts in wild-type and knockout matings. Table of pair survival and progeny production in DRB1 knockouts crossed with wild-type or with DRB1 knockouts. DRB1 KO #5.1.3, DRB1 KO #6.1.6, DRB1 KO #6.1.12.1, DRB1 KO #6.1.12.2, DRB1 KO #7.1, DRB1 KO #7.7.2; (ΔDRB1/ ΔDRB1 [V/IV/III, ΔDRB1]). Cells were scored for pair survival, transferred to media containing selective marker and scored for drug resistance later. Progeny production in DRB1 knockout matings with other DRB1 knockouts was determined by assaying mating competence and anlagen production.

1. Pair survival is the % of pairs alive (S) of the total pairs (N) isolated.

2. Progeny production is the % of surviving pairs (S) that successfully completed conjugation and made new macronuclei (P).



Strain	2 Mic/2 Mac	2 Mic/1 Mac	Total
WT	0.0%	99.5%	200
ΔDCL1	96.0%	0.3%	400
ΔDRB1	2.0%	98.0%	600
ΔDRB2 Mic	93.5%	6.4%	1900

Figure 3: Zygotic expression of DRB2 is necessary for completion of conjugation. Top. Terminal arrest phenotype of wild-type (WT), ΔDCL1, ΔDRB1 and ΔDRB2 mic cells 30 hours into conjugation. WT, ΔDCL1, ΔDRB1 and ΔDRB2 mic cells were mated and harvested after 30 hours into conjugation. Cells were then DAPI stained and DIC (Right Column) and DAPI (Left Column) images were taken. White arrowhead, micronuclei; white arrow, zygotic macronuclei. Bottom. Table of count of terminal arrest phenotype of WT, ΔDCL1, ΔDRB1 and ΔDRB2 mic cells 30 hours into conjugation.

Table 2: Progeny production of DRB2 mic knockouts in wild-type and knockout matings

Cross	Pair Survival (S/N)[1]	Progeny Production (P/S)[2]
B*VII^{427} X B*VI$^{\Delta D2/\Delta D2}$ #1	**99.6 %** (263/264)	**98.8 %** (260/263)
B*VII^{427} X B*VI$^{\Delta D2/\Delta D2}$ #6	**99.2 %** (262/264)	**99.2 %** (260/262)
CU427 X B*VII$^{\Delta D2/\Delta D2}$ #1	**95.1 %** (251/264)	**99.6 %** (250/251)
CU427 X B*VII$^{\Delta D2/\Delta D2}$ #2	**97.3 %** (257/264)	**100 %** (257/257)
B*VI$^{\Delta D2/\Delta D2}$ #1 X B*VII$^{\Delta D2/\Delta D2}$ #1	**2.8 %** (5/176)	**0.0 %** (0/5)
B*VI$^{\Delta D2/\Delta D2}$ #1 X B*VII$^{\Delta D2/\Delta D2}$ #2	**1.7 %** (3/176)	**0.0 %** (0/3)
B*VI$^{\Delta D2/\Delta D2}$ #6 X B*VII$^{\Delta D2/\Delta D2}$ #1	**0.0 %** (0/176)	**0.0 %** (0/0)
B*VI$^{\Delta D2/\Delta D2}$ #6 X B*VII$^{\Delta D2/\Delta D2}$ #2	**1.1 %** (2/176)	**0.0 %** (0/2)

Table 2: Progeny production of DRB2 mic knockouts in wild-type and knockout matings. Table of pair survival and progeny production in DRB2 mic knockouts cells mated to wild-type or to DRB2 mic knockout cells. Cells were scored for pair survival, transferred to media containing selective marker and scored for drug resistance later. B*VI$^{\Delta D2/D2}$ #1 and #6 and B*VII$^{\Delta D2/D2}$ #1 and #2; (ΔDRB2/ ΔDRB2 [VI/VII, +]).

1. Pair survival is the % of pairs alive (S) of the total pairs (N) isolated.

2. Progeny production is the % of surviving pairs (S) that successfully completed conjugation and made new macronuclei (P).

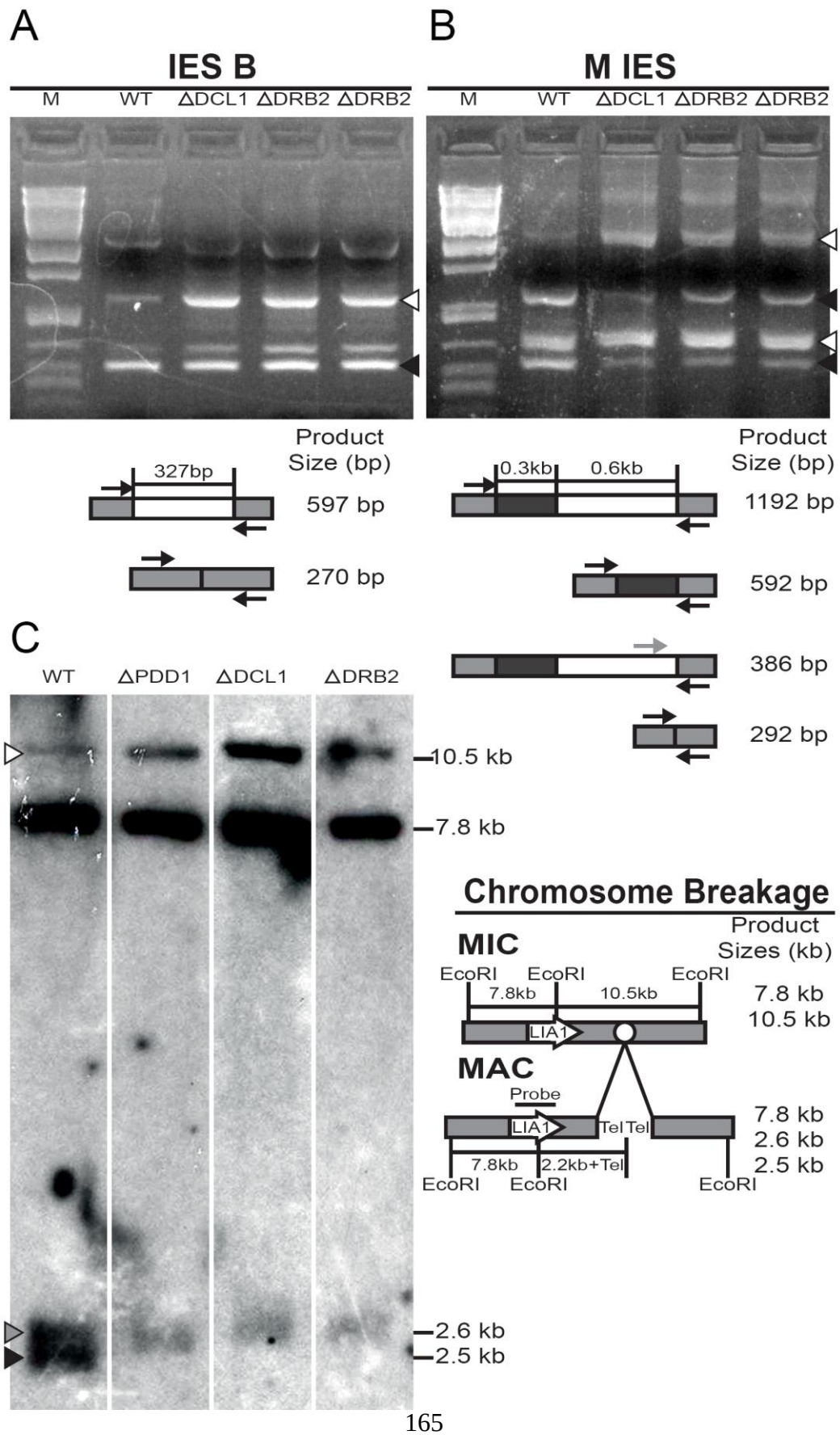

A
IES B
M
WT
ΔDCL1
ΔDRB2
ΔDRB2
Product Size (bp)
327bp
597 bp
270 bp
B
M IES
M
WT
ΔDCL1
ΔDRB2
ΔDRB2
Product Size (bp)
0.3kb
0.6kb
1192 bp
592 bp
386 bp
292 bp
C
WT
ΔPDD1
ΔDCL1
ΔDRB2
10.5 kb
7.8 kb
2.6 kb
2.5 kb
Chromosome Breakage
MIC
Product Sizes (kb)
EcoRI
EcoRI
EcoRI
7.8kb
10.5kb
7.8 kb
10.5 kb
LIA1
MAC
Probe
LIA1
Tel Tel
7.8 kb
2.6 kb
2.5 kb
7.8kb
2.2kb+Tel
EcoRI
EcoRI
EcoRI

Figure 4: DNA rearrangement of internal eliminated sequences (IESs) and chromosome breakage are impaired in DRB2 mic knockouts. A and B. Rearrangement of IES B (A) and the M IES (B) was assessed by two or three primer PCR, respectively, in genomic DNA isolated from WT, ΔDCL1 and ΔDRB2 mic cells post-conjugation. White arrowheads denote the unrearranged/ micronuclear form; black arrowheads, the rearranged/ macronuclear form. Diagrams of each IES locus are shown below the gel image. IES, white and dark grey boxes; flanking DNA, grey boxes; PCR primers, black and grey arrows. The M IES undergoes alternative rearrangement through elimination of the 0.6 kb (white box) or the 0.9 kb (white and dark grey boxes) sequence. The expected PCR product size is provide beside each form. C. Chromosome breakage fails in DRB2 mic knockouts. Left. Southern blot hybridization of total genomic DNA isolated from WT or mutant cells post-conjugation. White arrowhead, micronuclear specific fragment; grey arrowhead, parental macronuclear specific fragment; black arrowhead, zygotic macronuclear specific fragment. Right. Diagram of CBS near LIA1 locus in the micro- and macronucleus. Southern blot band sizes are listed next to each locus diagram. CBS, white circle, LIA1 gene, white arrow; Tel, telomere.

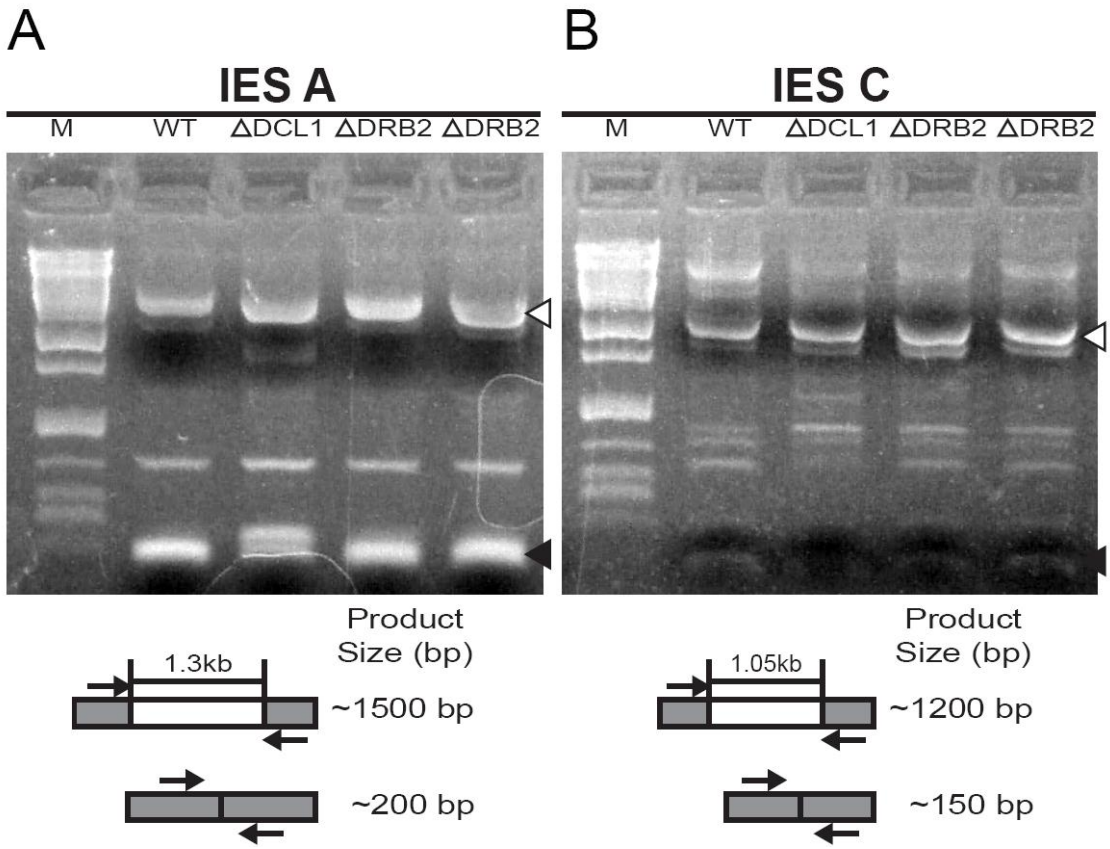


Figure S4: Global DNA rearrangement of IESs is impaired in DRB2 mic knockouts. A and B. Zygotic expression of DRB2 is essential for rearrangement of IES A (A) and IES C (B). Top. Total genomic DNA isolated from WT, ΔDCL1 and ΔDRB2 mic cells 30 hours into conjugation was used in a PCR-based assay to assay rearrangement of IES A (A; 2 primer) or IES C (B; 2 primer). Black arrowheads, unrearranged/ micronuclear band; white arrowheads, rearranged/ macronuclear band. Bottom. Left. Diagram of IES A locus in the micro- and macronucleus. IES A is represented by a white box and forward and reverse PCR primers are represented by black arrows. Right. Diagram of IES C locus in the micro- and macronucleus. IES C is represented by a white box and forward and reverse are represented by black arrows.

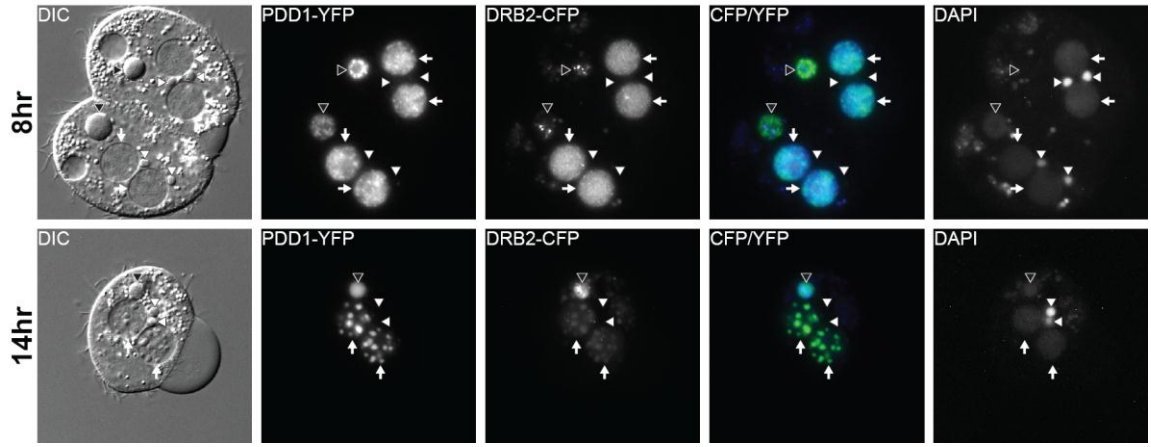


Figure 5: Drb2p co-localizes with the essential conjugation chromodomain protein, Pdd1p, in DNA elimination bodies. Cells expressing Drb2p-CFP were mated with cells expressing Pdd1p-YFP. Both proteins localized to the developing zygotic macronucleus (8hr) and in DNA elimination bodies (14hr). White arrowhead, micronuclei; black arrowhead, parental macronuclei; white arrow, zygotic macronuclei.

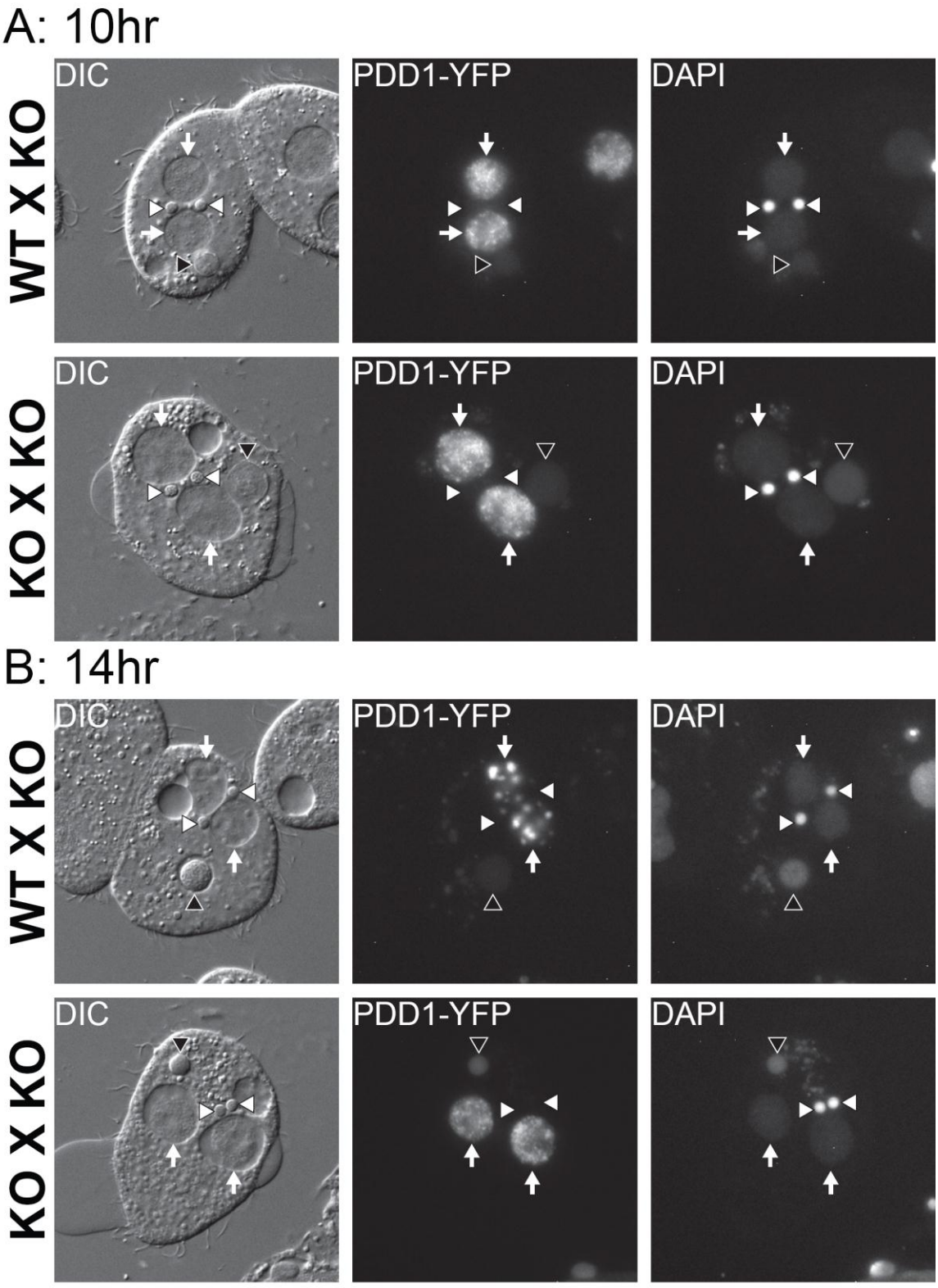

A: 10hr
WT X KO
DIC
PDD1-YFP
DAPI
KO X KO
DIC
PDD1-YFP
DAPI
B: 14hr
WT X KO
DIC
PDD1-YFP
DAPI
KO X KO
DIC
PDD1-YFP
DAPI

Figure 6: Failure of DNA elimination bodies to form in DRB2 mic knockouts late in conjugation. A. Normal zygotic macronuclear localization of Pdd1p in DRB2 mic knockout matings midway through conjugation. DRB2 mic knockouts ectopically expressing Pdd1p-YFP were mated with wild-type or with DRB2 mic knockouts. Pdd1p-YFP localizes to the developing zygotic macronucleus in both matings. White arrowhead, micronuclei; black arrowhead, parental macronuclei; white arrow, zygotic macronuclei. B. Ectopically expressed Pdd1p fails to form DNA elimination bodies in DRB2 mic knockouts. DRB2 mic knockouts ectopically expressing Pdd1p-YFP were mated as in A. In DRB2 mic knockouts mated to DRB2 mic knockouts Pdd1p-YFP fails to form DNA elimination bodies in the developing zygotic macronucleus. White arrowhead, micronuclei; black arrowhead, parental macronuclei; white arrow, zygotic macronuclei.

Table 3: Oligonucleotides used in the course of this study

Purpose and Name	Sequence (5'-3')
RT-PCR to Determine Expression	
DRB1: **#1688-Loq1-836**	**CGAAAAGGGGTTAGGGTTTTCTAGC**
#1689-Loq1-1325r	**CCCTTATCCCATCGTTTTCAG**
DRB2: **#1692-Loq3-1875**	**GCAATAGCCAAACACAAAGAGTGTAGC**
#1879-Loq3-2230r	**GCATCAATAAGGCTACAACATCC**
ATU1: **#3364-ATU1-2391r**	**GTGGCAATAGAAGCGTTGACA**
#3365-ATU1-1997	**TGCTCGATAACGAAGCCATCT**
Gene Amplification of Coding Sequence	
DRB1: **#1701-Loq1X**	**CACCCTCGAGAAAATGAATTCTTAGCAAG**
#1732-Loq1rH-Short	**AAGCTTTAGACTTATACTTTTCATGAAAG**
DRB2: **#1887-Loq3X-Modified**	**CACCCTCGAGAAAATGGCGCAATCTTTTAGATTTATAG**
#1911-Loq3rP-Full length	**CTGCAGCCCATTACAAATAATTATTAAGTTATCATAAGC**
Lysate PCR of Coding Sequence Plasmids	
DRB1: **#1689-Loq1-1325r**	**CCCTTATCCCATCGTTTTCAG**
#1701-Loq1X	**CACCCTCGAGAAAATGAATTCTTAGCAAG**
DRB2: **#1692-Loq3-1875**	**GCAATAGCCAAACACAAAGAGTGTAGC**
#1879-Loq3-2230r	**GCATCAATAAGGCTACAACATCC**
Sequencing of Coding Sequence Plasmids	
DRB1: **#1689-Loq1-1325r**	**CCCTTATCCCATCGTTTTCAG**
#1721-Loq1-5105	**GTCTCAGTAAAGCATTATAAGTAGCTG**
#1722-Loq1-5215r	**GAGTTTATCAGCAAATCGAATTTAGACC**
#3047-M13Forward	**GTAAAACGACGGCCAGT**
#3048-M13Reverse	**TCACACAGGAAACAGCTATGAC**
DRB2: **#1889-Loq3-2500r**	**CTATTATGGTTTTAATTGAGTCATTTAAAGG**
#1919-Loq3-4050	**CAGATAGGCTTGATAAAGAAGATCC**
#1920-Loq3-5343r	**GGTTCTTATTCTTACTCTTCATCTTC**
#2049-DRB2-2782r	**CAAAGATTAATACTAGGCTTTTGTTGC**
#3047-M13Forward	**GTAAAACGACGGCCAGT**
#3048-M13Reverse	**TCACACAGGAAACAGCTATGAC**
Knockout Cassette Generation	
DRB1: Upstream	
#1761-Loq1-2321AattB4	**GGGGACAACTTTGTATAGAAAAGTTGGTACCGGGATTACA TAAAGATTTGATTCC**
#1762-Loq1-3366rattB1	**GGGGACTGCTTTTTTGTACAAACTTGCACAATTCAATCAAA AGTGCG**
Downstream	
#1763-Loq1-5746attB2	**GGGGACAGCTTTCTTGTACAAAGTGGCACTCTCATTAATGC CCCC**
#1764-Loq1-7093AattB3	**GGGGACAACTTTGTATAATAAAGTTGGTACCAGTAAAGAG CCTAAATCAAGG**

Table 3 (Cont.): Oligonucleotides used in the course of this study

Purpose and Name	Sequence (5'-3')
Knockout Cassette Generation	
DRB2: Upstream	
#2370-DRB2-1226AattB4	**GGGGACAACTTTGTATAGAAAAGTTGGTACCGAAAGCCTATGGGAGAGCAAG**
#2372-DRB2-2595rattB1	**GGGGACTGCTTTTTTGTACAAACTTGCACTTTTAGGAAATAATGAATGTGTCAC**
Downstream	
#2322-DRB2-6530attB2	**GGGGACAGCTTTCTTGTACAAAGTGGGTTGTGTTTAAAAAGAAGGTGTGTGTTATG**
#2323-DRB2-7708ArattB3Ext	**GGGGACAACTTTGTATAATAAAGTTGGTACCTTCACTTAAACCGCACCCAG**
Lysate PCR and Sequencing of Knockout Cassettes	
#3047-M13Forward	**GTAAAACGACGGCCAGT**
#3048-M13Reverse	**TCACACAGGAAACAGCTATGAC**
Knockout PCR Screening	
DRB1:	
5' **#1679-MTT1-11484r**	**ATTTGGAATTAAGTACTTATTTCCAAAC**
#1946-DRB1-1086	**CGCGCACTTTTGATTGAATTGTG**
3' **#1866-Neo KO 2**	**CGTGATATTGCTGAAGAGCTTG**
#1867-Loq1-5353	**CAGGGGAAGATATATTTTATGAAGC**
#1868-Loq1-5764r	**GGGGGCATTAATGAGAGTG**
DRB2:	
5' **#2477- DRB2-2195**	**CAATTTATCTATTAAAATACCTTTACTTAC**
#2478- DRB2-2805r	**AAAATCTGTAATTGAGAAGAAACAAAAAC**
#3001-LIA4MTTLR	**AACATTCAAACATTGTGCACTAAATA**
3' **#2367-p4T2-3351**	**TCGCCTTCTTGACGAGTTCT**
#2391- DRB2-6323	**GCTTAGATGATATTACACATGATAATC**
#2392- DRB2-6721r	**AAAGAGAGTGAGTTTTTCTTTTTGG**
Assay Rearrangement of IESs	
M: **#1439-M808**	**ATATTGTGTGGTACAATAGGTTGTCGTAG**
#3111-M002	**AGCTTAAACAAATGCCATATTGAG**
#3114-M1194	**GTGGGGAGGGAGAAGGATTCAAC**
A: **#3244-IES1_MDSL-110**	**TGGAAGATCTACTTCAAAGCGAAT**
#3245-IES1_MDSR-31	**CCAGCTAGACACCCTGTATCAA**
B: **#3246-IES7_MDSL-112**	**GGATTGATTGGCATAAATGGA**
#3247-IES7_MDSR-158	**AAGCCCAGAATACCGCAGTTC**
C: **#3248-IES11_MDSL-42**	**GGCCACAATATACTAAGGCAATTT**
#3249-IES11_MDSR-34	**GGCCACCTTGATACCAGTTT**

CHAPTER 5

THE DOUBLE-STRANDED RNA BINDING MOTIF PROTEIN, DRB1, PLAYS AN AUXILIARY ROLE IN MEIOSIS I IN THE CILIATE, *TETRAHYMENA THERMOPHILA*

Abstract:

Double-stranded RNA Binding Motif (DSRM)-containing proteins and non-coding RNAs (ncRNAs) affect a variety of biological processes in eukaryotes. Here we propose that a tandem DSRM-containing gene in *Tetrahymena thermophila*, **D**ouble-stranded **R**NA **B**inding Protein 1 (DRB1), plays a supplementary role in ensuring proper generation of haploid gametes during meiosis, possibly through ncRNA. Drb1p localizes to ends of crescent micronuclei, where it is found to co-localize with the centromeric histone, Cna1p. Strains lacking Drb1p produce progeny at a lower rate and also abort conjugation at a higher rate. In addition ectopic reintroduction of DRB1 into DRB1 knockout strains reconstitutes wild-type progeny production and progression through conjugation. These data implicate early Drb1p function during conjugation in ensuring proper passage through meiosis I and II.

Introduction:

In the last decade long non-coding RNAs (ncRNAs) have been shown to play a vital role in a multitude of biological processes in eukaryotes (Reviewed in [1]). Many eukaryotes produce these long ncRNAs across large portions of their genomes but the biological role for much of this transcription remains unclear. In metazoans a number of these ncRNAs have been found to be vital for gene imprinting and dosage compensation. Gene imprinting has been found exclusively in mammals and involves differential gene expression depending on the sex of the parent from whom the chromosome was inherited. Two prime examples of gene imprinting are the Igf2 and the Igf2r clusters, which are regulated by the ncRNAs, H19 and Air, respectively [2-4]. Dosage compensation has also been found to be directed by ncRNAs such as Xist, which silences one X chromosome in

female mammals, and roX, which causes enhanced transcription from the X chromosome in male *D. melanogaster* [5-13]. Long ncRNA has also shown to play pivotal roles in single cell eukaryotes, where it is crucial for genome remodeling in the ciliate, *Oxytricha trifallax* [14].

Long ncRNA have also proven to be vital precursors for another essential process in eukaryotes, RNA interference (RNAi) (Reviewed in [15-17]). Typically, long ncRNA is processed by an RNaseIII family protein (i.e. Dicer or Drosha homologues) to produce sRNA species ranging from 20nt-30nt ([18-21], reviewed in [17]). These sRNAs then interact with effector Argonaute protein complexes to regulate gene expression transcriptionally or post-transcriptionally [22-24]. In the ciliate, *Tetrahymena thermophila*, RNAi has been linked to DNA elimination during genome remodeling, which occurs in the zygotic macronucleus during sexual reproduction [25-28].

All ciliates, including *T. thermophila*, contain in one cell two distinct types of nuclei, a macronucleus and a micronucleus (reviewed in [29, 30]). The polyploid macronucleus (~50C) is transcriptionally active during vegetative growth, while the diploid, germline micronucleus is transcriptionally silent ([31, 32], reviewed in [29, 30]). Starvation of *T. thermophila* initiates sexual reproduction or conjugation, generating new micro- and macronuclei from the parental germline micronucleus (reviewed in [29, 30]). During the maturation of the zygotic macronuclei, the macronuclear chromosomes are fragmented at ~200 sites, amplified to ~50C and lose ~15% of their overall genomic content [28, 31-36]. The loss of ~15% of the genomic content in the zygotic macronucleus is the result of RNAi-directed DNA elimination (reviewed in [29, 30].

One of the genes required for RNAi-directed DNA elimination is the Dicer homologue, DCL1 (chapter 3) [25, 27]. Typically for full sRNA production Dicer and Drosha homologues interact with a tandem double-stranded RNA (dsRNA) binding motif (DSRM) protein partner [37-43]. None have been identified for any of the Dicer proteins in *T. thermophila* including Dcl1p. Two putative tandem DSRM-containing proteins, DRB2 and DRB1, were identified as possible partner proteins for the Dicer proteins in *T. thermophila* (chapter 4). Characterization of these protein revealed that neither are exclusively Dicer protein partners for DCL1 or any of the other Dicer homologues. However, zygotic expression of the tandem DSRM-containing DRB2 was found to be essential for completion of conjugation and DNA elimination.

Despite the fact that the other tandem DSRM-containing protein, DRB1, had a very similar expression profile to that of DCL1, knockouts of DRB1 did not support that it was a critical Dcl1p partner protein (Fig. 1B) (chapters 3 and 4) [25]. However, DRB1 knockouts produced fewer progeny when mated with each other than when they were mated to wild-type strains or compared to wild-type control matings. Here we present an analysis of the progression of the DRB1 knockouts through conjugation and find that a greater number of cells quit conjugation before production of the zygotic nuclei. We also find that Drb1p associates with centromere and telomere sequences in the crescent micronucleus and with nucleolar structures in the parental macronucleus. These pieces of data implicate a role for DRB1 early in meiosis, possibly during homologous recombination, which is essential for production of functional haploid gametic nuclei.

Results:

Somatic knockouts of DRB1 are insufficient to cause loss of scnRNA production and DNA elimination.

Despite the similarity between the expression profiles of the conjugation specific Dicer, DCL1, and the tandem DSRM-containing gene, DRB1, knockouts of DRB1 did not arrest at the end of conjugation like DCL1 (Fig. 1A and 1B) (chapters 3 and 4) [25]. However, DRB1 did exhibit several intriguing characteristics. DRB1 knockouts produced fewer progeny than wild-type matings and also had distinct nuclear localization in both the micronucleus and the macronucleus (chapter 4). Why DRB1 knockouts produced fewer progeny was unclear, but perhaps since DRB1 is strongly upregulated during conjugation a non-essential portion of RNAi-directed DNA elimination or other conjugation-specific process is affected (Fig. 1B). To assay whether portions of RNAi-directed DNA elimination could be affected, we surveyed the levels of scnRNAs produced early during conjugation and DNA excision of a couple of IESs later in conjugation (Figs. 1C-1E). A sRNA gel of RNA isolated throughout conjugation from wild-type and DRB1 mac knockout matings showed no difference in the levels or timing of scnRNA production (Fig. 1C). In depth analysis of M IES-specific scnRNA production through a sRNA northern blot showed no significant change as well (Fig. 1D). Southern blots assessing the level of excision of the M and CAM IESs from wild-type and DRB1 mac knockout mating progeny demonstrated no difference in DNA elimination between the matings (Fig. 1E, data not shown). These data indicate that at most, DRB1 has a limited role in DNA elimination, possibly through genome scanning or another process that is difficult to currently detect.

Pairs from knockout DRB1 matings abort conjugation prematurely in higher numbers than wild-type matings, which can be alleviated by reintroduction of DRB1.

To further examine DRB1 knockout progeny production we analyzed pairs of DRB1 knockout cells as they proceeded through all the cytological stages of conjugation. Conjugation can be effectively split into two portions: pre-zygotic and post-zygotic [44, 45]. Pre-zygotic stages of conjugation in *T. thermophila* consist of the detachment of the parental micronucleus from the parental macronucleus (1), cell pairing (2), prophase of meiosis I (3), the reminder of meiosis I (4), meiosis II (5) and pre-zygotic mitosis (6) to produce two haploid gametes in each cell of the mating pair (See Fig. 2A for diagram). One gamete from both cells of the pair is exchanged and these two nuclei fuse to produce the zygotic nucleus. After two rounds of mitosis (7), the zygotic macronuclei undergo three cytological stages of development (I, 8; II, 9, III; 10). Then the pairs separate and eliminate one of the remaining zygotic micronuclei (11) to complete the post-zygotic stages of conjugation. We sought to determine if lower progeny production in DRB1 knockout matings was related to difficulty progressing through any particular stage of conjugation. DRB1 knockout mating pairs were examined for their progression through these stages and it was found that a lower percentage of DRB1 knockouts enter the post-zygotic stages compared to wild-type (Fig. 2). Compared to wild-type pairs, DRB1 knockout pairs proceeded through conjugation at a slightly faster rate when assayed at 2.5hr and 4hr. As the cells proceeded to produce zygotic nuclei beginning at 6hr, a number of DRB1 knockout pairs appeared to stall at the pre-zygotic/post-zygotic transition (stage 6 to stage 7). This is particularly noticeable at 8hr into conjugation when

35 pairs from DRB1 knockout matings are at the terminal pre-zygotic stage compared to 18 wild-type pairs (Table S1). At 30hr into conjugation the number of cells at the final stage of conjugation in both matings is comparable. However, the number of cells that have aborted mating and returned to the first conjugation stage is significantly higher in DRB1 knockout matings (24) compared to wild-type matings (8) (Compare stage 1 cells at 2.5hr and 30hr for both matings in Fig. 2 and Table S1). Cessation of conjugation midway through is indicative of an inability to produce or exchange haploid, gametic nuclei, which is used as a genetic tool in strains with known defective micronuclei [46-48].

To determine if this phenotype was caused by a deficiency of Drb1p, we introduced pICY-DRB1 into conjugating DRB1 knockout strains and observed the progression through conjugation of progeny pairs with exogenous Drb1p-YFP expression compared to DRB1 knockout strain pairs. In contrast to DRB1 knockout strain pairs, DRB1 knockout strain pairs expressing exogenous Drb1p-YFP produce wild-type levels of zygotic nuclei during conjugation (Fig. 3). Throughout all time points measured, wild-type pairs with or without exogenous Drb1p-YFP expression proceeded at a similar rate compared to DRB1 knockout strains expressing exogenous Drb1p-YFP. DRB1 knockout pairs proceeded a little more slowly through the conjugation stages than the other three strains, but were again most notably different at 8hr into conjugation at the pre-zygotic/post-zygotic transition. At the final pre-zygotic stage 8hr into conjugation, there was, yet again, sizable differences in the number of DRB1 knockout pairs (15) compared to the other three matings (3,5,4) (Table S2). Taking all this mating progression data

together it seems to indicate some role for Drb1p in producing functional haploid gametic nuclei. How Drb1p accomplishes this remains to be determined.

Drb1p Localizes to the Tips of the Crescent Micronucleus during Prophase of Meiosis I.

Drb1p had previously been shown to have distinct nuclear localization in the micro- and macronucleus early in conjugation. This observation may offer insight into the lower progeny production and inability to produce functional haploid gametes in DRB1 knockout matings (chapter 4). Further examination of the localization in the crescent micronucleus revealed a particular interesting pattern. Almost immediately after the micronucleus detaches itself from the parental macronucleus and begins prophase of meiosis I, localization of Drb1p-YFP is seen at the narrow end of the early stage II crescent micronucleus (Fig. 4A). This localization persists through meiosis I prophase stages II and III (data not shown). Additional point localization of Drb1p-YFP is seen at the opposite end of the micronucleus beginning in late prophase stage III and stage IV micronuclei (Fig. 4B). Although this secondary point localization is more dispersed than the initial point localization of Drb1p-YFP at the other end of the crescent micronucleus, it also forms small and fairly tight foci not unlike the Drb1p-YFP localization seen earlier. Point localization of Drb1p-YFP is lost shortly after completion of prophase of meiosis I sometime during metaphase (data not shown).

Primary Crescent Micronuclear Localization of Drb1p Overlaps with the Centromere Histone Protein, Cna1p.

Prior research in *T. thermophila* had shown that the centromeres and telomeres localize to the ends of the crescent micronucleus during prophase of meiosis I [49-51]. To determine

if the Drb1p-YFP localization seen in the crescent micronucleus could overlap with either of these structures, we first created a C-terminal CFP fusion of the centromere histone, CNA1. Cna1p shows point localization at one end of the crescent micronucleus beginning with early stage II nuclei of prophase of meiosis I which persists until completion of prophase [49, 50]. Co-localization of Drb1p-YFP with Cna1p-CFP confirmed that the earliest point localization of Drb1p-YFP did indeed co-localize with the centromeres (Fig. 5A). During prophase of meiosis I in early stage II micronuclei, Cna1p-CFP localization is seen in same site as that of Drb1p-YFP (Fig. 5A, top). This continues until the dispersion of Drb1p-YFP shortly after the completion of prophase of meiosis I (Fig. 5A, bottom). Unincorporated Cna1p-CFP persists throughout the nucleoplasm of micronuclei until DNA replication when it is then incorporated. This is likely an artifact of centromere histone overexpression as described previously in *T. thermophila* and other organisms (data not shown, [49, 52, 53]).

Attempts to determine if the secondary point localization of Drb1p-YFP co-localizes with telomere sequences or proteins have so far been unsuccessful. Co-localization of Drb1p-YFP with the *T. thermophila* single-strand telomere end binding homologue, Pot1ap, was inconclusive due to poor Pot1ap antibody staining and loss of Drb1p-YFP micronuclear localization during the cell fixation process (data not shown).

Macronuclear localization of Drb1p is found to associate with nucleolar structures. As discussed earlier Drb1p has distinct localization not only in the micronucleus but also in the macronucleus. Drb1p-YFP foci distribution in the macronucleus had many properties in common with the localization of an abundant nucleolar protein, Nopp52p [54, 55]. To ascertain if the Drb1p-YFP localization pattern seen during conjugation in

macronuclei was nucleolar in origin, Drb1p-YFP expressing cells were fixed and stained with α-Nopp52p. Co-localization of Drb1p-YFP and Nopp52p reveals a high degree of overlap (Fig. 2B). Taken together with the data on progeny production and conjugation progression in DRB1 knockout matings, Drb1p localization in the micronucleus with centromeres and telomeres and in the macronucleus at nucleolar structures implicates a role for Drb1p during meiosis influencing the developmental fate of all of these structures.

Discussion:

Previously published data has shown that the two tandem DSRM-containing proteins in *T. thermophila*, DRB2 and DRB1, do not solely interact with any Dicer protein (chapter 4). However, Drb2p was shown to be a nuclear protein that co-localizes with the chromodomain containing protein, Pdd1p. Zygotic expression of DRB2 was also found to be essential for DNA elimination body formation, DNA elimination of IESs, chromosome breakage and completion of conjugation.

The initial characterization of DRB1 did not produce any obvious knockout phenotype or indication of its role in the life cycle of *T. thermophila* (chapter 4). DRB1 was found to be highly upregulated during conjugation and showed strong macronuclear localization in discrete foci (Figs. 1B and 4) (chapter 4). Drb1p was also found to localize to the ends of the crescent micronucleus throughout prophase of meiosis I (Fig. 4) (chapter 4). Studies in other organisms have shown telomere localization at the ends of the meiotic nuclei, which also holds true for *T. thermophila* ([51, 56], reviewed in [57]). In *T. thermophila* centromere histone localization has also been observed at the opposite end of the crescent micronucleus [49, 50]. This telomere and centromere data compelled

us to examine whether Drb1p could also be found with the telomere and/or centromere sequences at the ends of the crescent micronucleus. Co-localization experiments with Drb1p and Cna1p, the *T. thermophila* CENP-A centromere histone homologue, has shown that the earliest Drb1p point localization seen in the crescent micronucleus is in the same location as the centromeres (Fig. 5A) [49, 50]. Additional point localization of Drb1p is seen only after complete extension of the crescent micronucleus (late stage III or stage IV) and is likely to co-localize and possible interact with the telomere sequences [51]. Attempts to verify co-localization between these Drb1p foci and the telomere have been unsuccessful thus far (data not shown). What the role of these Drb1p foci at the centromere and telomere in the crescent micronucleus remains to be seen.

Although Drb1p has point localization in the micronucleus, it predominately forms foci in the parental and zygotic macronucleus (Figs. 4A and 4B) (chapter 4). During vegetative growth the nucleoli are found at the macronuclear periphery, but during conjugation they become dispersed throughout the macronucleus ([54, 55], D.L. Chalker, unpublished data). Analysis of the macronuclear localization of Drb1p and the nucleolar protein, Nopp52p, demonstrates co-localization, implying that the macronuclear Drb1p foci during conjugation are in the nucleoli (Fig. 5B) [55]. Since Drb1p-YFP appears to interact with telomeres in the micronucleus and a major component of the nucleolus is rDNA chromosomes, which have a higher copy number (~9000C) in the macronucleus compared to other chromosomes (~45C), Drb1p-YFP macronuclear nucleolar localization could be due to Drb1p interaction with rDNA telomeres [31, 32, 58]. The relevant biological role of Drb1p in the macronucleus at the nucleolus or at the rDNA telomeres remains unclear.

Despite the fact that Drb1p is not generally a tandem DSRM-containing protein partner for Dcl1p, localization at the centromere and likely the telomere indicates that it could be involved in generating a small subset of scnRNAs at those sites. Neither centromeres nor micronuclear telomeres are present in the macronucleus since the telomeres in macronucleus have a different sequence [59, 60]. Although somatic knockouts of DRB1 do not show a decrease in overall or M IES-specific scnRNA levels, centromere and telomere scnRNAs would likely account for a minuscule number of overall scnRNAs and could still require Drb1p for their formation and/or stabilization (Figs. 1C and 1D). Once these specialized scnRNAs are generated in the crescent micronucleus, they could be transported to Drb1p foci in both the parental and, later, the zygotic macronuclei. These Drb1p macronuclear nucleolar foci could act as a staging ground after centromere and telomere scnRNA biogenesis for genome scanning in the parental macronucleus and targeting of these sequences for DNA elimination in the developing zygotic macronucleus. Alternatively, these Drb1p macronuclear nucleolar foci could also be sites for genome scanning. In knockouts of DRB1 this scanning could be disrupted causing essential genes to be eliminated in the zygotic macronucleus later in development, which could also explain the lower progeny production in DRB1 knockout matings.

However, DRB1 knockout matings not only produce fewer progeny, they also abort conjugation at the pre-zygotic/post-zygotic transition (Figs. 2 and 3) (chapter 4). It seems unlikely that loss of centromere and telomere scnRNA production or failure of genome scanning would produce a large number of backouts during conjugation. Production of an increased number of faulty haploid gametic nuclei would be able to

explain lower progeny production and higher backout numbers in DRB1 knockout matings. (Figs. 2 and 3). Therefore, it seems likely that localization of Drb1p to the centromere and possibly the telomere in the crescent micronucleus are instead important for homologous recombination or other chromosomal events occurring during prophase of meiosis I. Since *T. thermophila* has no synaptonemal complex, it is possible that other nonconventional proteins, feasibly including Drb1p, have taken over this function to ensure proper recombination and crossover resolution ([51, 61], reviewed in [57]). Further investigation of the role of Drb1p at these micro- and macronuclear foci will include purification of affinity-tagged Drb1p to verify Drb1p interaction with the telomere and centromere proteins in the crescent micronucleus and to determine what role Drb1p truly has in the parental and zygotic macronuclei. Affinity-tagged Drb1p will also allow the identification of the class of RNAs that associate with Drb1p via RIP.

In the fission yeast, *S. pombe*, ncRNAs, siRNAs and associated RNAi proteins have been shown to be essential for heterochromatin formation and centromere function through the RITS complex [62-70]. Long ncRNA, scnRNAs and RNAi proteins play a similar function through heterochromatin formation and DNA elimination in *T. thermophila* [25-27, 71]. Is it possible that dsRNAs in *T. thermophila* also act in a process similar to the one that occurs at the centromere in *S. pombe* to ensure proper homologous recombination or crossover resolution during meiosis to generate mature gametes as well? Tandem DSRM-containing proteins are predominantly known for their interaction with Dicer and Drosha family proteins, but that is by no means the only role proteins containing DSRMs perform ([37-43], reviewed in [72, 73]). The RNase III family, including Dicer and Drosha, ADAR family, PKR family and Staufen family

proteins all contain at least one DSRM [20, 21, 74-82]. One other protein family, the NFAT family, also has tandem DSRMs. Little is known about this protein family, but they are thought to be putative nuclear, nucleotide transferases that can play a role in DNA repair and RNA transport ([83-86], reviewed in [87]). Many of the DSRM-containing proteins have no known function and much remains to be learned about the biological role of these proteins in eukaryotes. Further study of DRB1 in *T. thermophila* may expand what we know about ncRNAs and what processes they can affect in ciliates and in eukaryotes in general.

Materials and Methods:

***Tetrahymena* strains and growth conditions.** Standard wild-type, laboratory *T. thermophila* strains CU427 (Chx/Chx [VI, cy-s]), CU428 (Mpr/Mpr [VII, mp-s]), B2086 (II), micronucleus-defective strains B*VI (VI) and B*VII (VII) were originally obtained from Peter Bruns (Cornell University, Ithaca, NY). B*VII$^{CU427/CU427}$ (Chx/Chx [VII, cy-s] was generated through a genomic exclusion mating between CU427 and B*VII. These strains or their transformed progeny were used for expression studies, biolistic transformations and subsequent analyses. Cells were grown and maintained as previously described [88, 89]. Strains were starved six hours to overnight in 10 mM Tris (pH 7.5) prior to mixing to initiate conjugation. Optical densities of cell populations were used to estimate cell number prior to mixing equal numbers of mating compatible strains.

Identification of DRB1 sequence. The DRB1 sequence was identified by BLAST search of the *T. thermophila* macronuclear genome (http://www.ciliate.org) using *D. melanogaster* R2D2 (Accession number CG7138) and Loqs (Accession number CG6866) DNA sequences. DSRMs of DRB1 were initially identified on the *T. thermophila*

macronuclear genome by InterProScan [90]. Further sequence analysis included PFAM analysis of the DRB1 coding sequence (http://pfam.janelia.org) and alignment of DSRM sequences of DRB1 with those of tandem DSRM-containing proteins R2D2 (Accession number NP_609152.1) and Loqs (Accession number NP_609646.1) from *D. melanogaster*, RDE-4 (Accession number NP_499265.1) from *C. elegans* and TRBP2 (Accession number NP_599150.1) from *H. sapiens*.

Generation of DRB1 knockouts. Genomic sequences from upstream and downstream of DRB1 were amplified by PCR and recombined into pDONR-P4-P1R (Upstream) and pDONR-P2R-P3 (Downstream) using BP Clonase (Invitrogen) (See Table 1 for primers). pDONR-L4-DRB1Up-R1 and pDONR-R2-DRB1Down-L3 were identified by lysate PCR and verified through sequencing (See Table 1 for primers). The resulting donor plasmids containing up- and downstream regions were mixed with equal amounts of pENTR-D-MTT1/NEO3 and the multisite destination vector, pDEST-R4-R3, along with LR Clonase Plus II (Invitrogen) to create the DRB1 knockout plasmid, pDEST-B4-DRB1Up-B1-MTT1/NEO3-B2-DRB1Down-B3. The DRB1 knockout construct was linearized by restriction digestion with KpnI and transformed into conjugating wild-type cells (CU428 X B2086) between 2 and 3 hours after mixing using a PDS-1000/He particle bombardment system (Bio-Rad) as previously described [91, 92]. Heterozygous macronuclear DRB1 transformants were identified by their resistance to 80 µg/ml paromomycin with 1 µg/ml $CdCl_2$ solely. Macronuclear DRB1 transformants were assorted to completion through growth in increasing concentrations of paromomycin. Strains lacking all wild-type DRB1 copies were verified by PCR of *T. thermophila* crude cell lysates (See Table 1 for primers). Heterozygous micronuclear DRB1 transformants

were identified by their resistance to 80 µg/ml paromomycin with 1 µg/ml $CdCl_2$ and 15µg/ml 6-methylpurine. Heterozygous micronuclear DRB1 transformants were verified through matings with CU427 by monitoring segregation of paromomycin resistance conferred by the MTT1/NEO3 knockout cassette among cycloheximide resistant progeny, as well as through PCR screening of *T. thermophila* crude cell lysates (See Table 1 for primers) [93]. Homozygous micronuclear knockout heterokaryons of DRB1were generated by crossing heterozygous micronuclear DRB1 transformants with B*VI or B*VII star strains. Homozygous micronuclear DRB1 knockout heterokaryons were identified by paromomycin/$CdCl_2$ sensitivity and verified through crosses with CU427 that produced progeny resistant to 100 µg/ml paromomycin with 1 µg/ml $CdCl_2$ and 25 µg/ml cycloheximide. Homozygous DRB1 knockout strains were generated by crossing homozygous micronuclear knockouts of compatible mating types and screening for progeny resistant to paromomycin/$CdCl_2$ and verified by PCR detection of the knockout allele (See Table 1 for primers) [93].

RNA analysis of DRB1. RNA was isolated from growing, starved and conjugating *T. thermophila* (CU428 X B2086) at 2 hour intervals from 2hr to 14hr by RNAsol extraction [94]. RNA isolation from DRB1 mac knockouts at 2hr intervals from 0hr to 14hr was also done by RNAsol extraction. RT-PCR of wild-type matings was done to determine expression of DRB1 as previously described (See Table 1 for primers) [25].

Small RNAs isolated from wild-type and DRB1 mac knockouts matings were separated on a 15% polyacrylamide-urea-1X Tris-borate-EDTA gel as previously described [93]. The small RNA Northern blotting was performed as previously described using the Minus-strand M-element riboprobe from pMint2 [71, 93].

DNA elimination analysis in DRB1 mac knockouts. *T. thermophila* genomic DNA was isolated from CU428 X B2086 and DRB1 mac KO #8.1 X DRB1 mac KO #14.1 matings at 30hr after mixing using a Wizard genomic DNA purification kit (Promega). Gel electrophoresis, blotting and hybridization were performed as previously described except blots were washed with 0.5X SSC-1% SDS after hybridization [25]. M IES rearrangement was assayed using CU428 X B2086 and DRB1 mac KO #8.1 X DRB1 mac KO #14.1 30hr genomic DNA that was digested with EcoRI and separated on a 0.8% agarose gel. The Southern blot probe was created using a 1.9 kb fragment from pDLCM3 [95].

Monitoring progression of conjugation in wild-type matings with or without exogenous Drb1p and DRB1 knockout matings with or without exogenous Drb1p. Conjugating CU427 X CU428 and DRB1 KO #5.1.3 X DRB1 KO #7.1 were transformed with pICY-DRB1 or pICC-CNA1 through conjugative electroporation (For construction of pICC-CNA1 and pICY-DRB1 see nuclear localization and co-localization of Drb1p) [96]. Transformants were identified by resistance to 80 µg/ml paromomycin, grown, starved and mixed. pICC-CNA1 #10.1 X pICC-CNA1 #16.5 and DRB1 KO pICC-CNA1 #13.1 X DRB1 KO pICC-CNA1 #14 cells at 2.5hr, 4hr, 6hr, 8hr, 24hr and 30hr after mixing were DAPI stained and their stage of conjugation determined by comparison of images with previously described stages of conjugation [45]. Progression of conjugation was monitored as above in CU427 X CU428, CU427 X pICY-DRB1 #5, DRB1 KO #5.1.3 X DRB1 KO #7.1, DRB1 KO pICY-DRB1 #1 X DRB1 KO pICY-DRB1 #2 at 2hr, 4hr, 6hr, 8hr, 10hr, 12hr and 24hr after mixing.

Nuclear localization and co-localization of Drb1p. Oligonucleotide primers (Table 1) were used to amplify the entire DRB1 and CNA1 coding sequences from genomic DNA by PCR. The resulting products were cloned into the Gateway recombination compatible pENTR-D to create pENTR-D-DRB1 and pENTR-D-CNA1, respectively. Plasmids containing the DRB1 and CNA1 coding sequences were identified by lysate PCR and sequenced to verify coding sequence integrity (See Table 1 for primers). LR recombination of pENTR-D-DRB1 with pICY-GTW and of pENTR-D-CNA1 with pICC-GTW using LR clonase II (Invitrogen) fused the coding regions to YFP and CFP in pICY-DRB1 and pICC-CNA1, respectively. Construction of the rDNA plasmids pICC-GTW and pICY-GTW was described earlier [97].

Strains containing pICC-CNA1 were grown in media containing 1 µg/ml $CdCl_2$ and starved in media containing 0.08 µg/ml $CdCl_2$ prior to mating to ensure proper Cna1p incorporation at the centromere. To visualize localization, starved transformed cells were mixed to begin conjugation in 0.08 µg/ml $CdCl_2$ to induce expression of the fusion protein(s). Live cells were harvested by low speed centrifugation (1,000 X g) at 2hr, 3hr and 4hr post-mixing, stained with DAPI (1 µg/ml) and immobilized in 5µl 2% methylcellulose. DIC, CFP fluorescence, YFP fluorescence and DAPI fluorescence images were captured using a Qimaging RetigaEX charge-coupled-device camera (Burnaby, British Columbia, Canada) and Openlab software (PerkinElmer).

For co-localization of Drb1p-YFP with Nopp52p, CU427 and pICY-DRB1 were mixed together and induced as above. After harvesting the cells at 3hr after mixing, the cells were fixed in 2% paraformaldehyde washed in 1X TBS + 0.1% TWEEN 20 (2X) and 1X TBS + 0.01% TWEEN 20 + 1% BSA (1X). Cells were then incubated at 4°C

overnight in 1X TBS + 0.01% TWEEN 20 + 1% BSA with the rabbit primary antibody anti-Nopp52p (1:25,000). The following day the cells were washed in 1X TBS + 0.1% TWEEN 20 (2X) and 1X TBS + 0.01% TWEEN 20 + 1% BSA (2X) and then probed with the secondary antibody, Alexa Fluor 594-conjugated goat anti-rabbbit antibody (1:1000; Invitrogen), for 1 hr at room temperature. Next the cells were washed in 1X TBS + 0.1% TWEEN 20 (2X) and 1X TBS + 0.01% TWEEN 20 + 1% BSA + 500 ng/ml DAPI (1X) for visualization. DIC, YFP fluorescence, Alexa Fluor (Texas Red filter) and DAPI fluorescence images were acquired as described above.

References:

1. Ponting, C.P., P.L. Oliver, and W. Reik, *Evolution and functions of long noncoding RNAs.* Cell, 2009. **136**(4): p. 629-41.
2. Bartolomei, M.S., S. Zemel, and S.M. Tilghman, *Parental imprinting of the mouse H19 gene.* Nature, 1991. **351**(6322): p. 153-5.
3. Lyle, R., et al., *The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1.* Nature genetics, 2000. **25**(1): p. 19-21.
4. Sleutels, F., R. Zwart, and D.P. Barlow, *The non-coding Air RNA is required for silencing autosomal imprinted genes.* Nature, 2002. **415**(6873): p. 810-3.
5. Brown, C.J., et al., *A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome.* Nature, 1991. **349**(6304): p. 38-44.
6. Brockdorff, N., et al., *The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus.* Cell, 1992. **71**: p. 515-526.
7. Brown, C.J., et al., *The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus.* Cell, 1992. **71**: p. 527-542.
8. Amrein, H. and R. Axel, *Genes expressed in neurons of adult male Drosophila.* Cell, 1997. **88**: p. 459-469.
9. Meller, V.H., et al., *roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system.* Cell, 1997. **88**: p. 445-457.
10. Kelley, R.L., et al., *Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin.* Cell, 1999. **98**(4): p. 513-22.
11. Meller, V.H. and B.P. Rattner, *The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex.* The EMBO journal, 2002. **21**(5): p. 1084-91.

12. Lee, J.T., et al., *A 450 kb transgene displays properties of the mammalian X-inactivation center.* Cell, 1996. **86**(1): p. 83-94.
13. Penny, G.D., et al., *Requirement for Xist in X chromosome inactivation.* Nature, 1996. **379**(6561): p. 131-7.
14. Nowacki, M., et al., *RNA-mediated epigenetic programming of a genome-rearrangement pathway.* Nature, 2008. **451**(7175): p. 153-8.
15. Lejeune, E. and R.C. Allshire, *Common ground: small RNA programming and chromatin modifications.* Current opinion in cell biology, 2011.
16. Malone, C.D. and G.J. Hannon, *Small RNAs as guardians of the genome.* Cell, 2009. **136**(4): p. 656-68.
17. Carthew, R.W. and E.J. Sontheimer, *Origins and Mechanisms of miRNAs and siRNAs.* Cell, 2009. **136**(4): p. 642-55.
18. Hamilton, A.J. and D.C. Baulcombe, *A species of small antisense RNA in posttranscriptional gene silencing in plants.* Science, 1999. **286**(5441): p. 950-2.
19. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**: p. 806-811.
20. Bernstein, E., et al., *Role for a bidentate ribonuclease in the initiation step of RNA interference.* Nature, 2001. **409**(6818): p. 363-6.
21. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
22. Hammond, S.M., et al., *Argonaute2, a link between genetic and biochemical analyses of RNAi.* Science, 2001. **293**(5532): p. 1146-50.
23. Verdel, A., et al., *RNAi-mediated targeting of heterochromatin by the RITS complex.* Science, 2004. **303**(5658): p. 672-6.
24. Aravin, A.A., et al., *Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline.* Current biology : CB, 2001. **11**(13): p. 1017-27.
25. Malone, C.D., et al., *Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila.* Mol Cell Biol, 2005. **25**(20): p. 9151-64.
26. Mochizuki, K., et al., *Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena.* Cell, 2002. **110**(6): p. 689-99.
27. Mochizuki, K. and M.A. Gorovsky, *A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase.* Genes Dev, 2005. **19**(1): p. 77-89.
28. Conover, R.K. and C.F. Brunk, *Macronuclear DNA molecules of Tetrahymena thermophila.* Mol Cell Biol, 1986. **6**(3): p. 900-905.
29. Prescott, D.M., *The DNA of ciliated protozoa.* Microbiol Rev, 1994. **58**(2): p. 233-67.
30. Meyer, E. and D.L. Chalker, *Epigenetics of Ciliates*, in *Epigenetics*, C.D. Allis, et al., Editors. 2007, Cold Spring Harbor Press: Cold Spring Harbor. p. 127-150.
31. Woodard, J., E. Kaneshiro, and M.A. Gorovsky, *Cytochemical studies on the problem of macronuclear subnuclei in tetrahymena.* Genetics, 1972. **70**(2): p. 251-60.
32. Doerder, F.P., J.C. Deak, and J.H. Lief, *Rate of phenotypic assortment in Tetrahymena thermophila.* Developmental genetics, 1992. **13**(2): p. 126-32.

33. Yao, M.C. and M.A. Gorovsky, *Comparison of the sequences of macro- and micronuclear DNA of Tetrahymena pyriformis.* Chromosoma, 1974. **48**(1): p. 1-18.
34. Altschuler, M.I. and M.C. Yao, *Macronuclear DNA of Tetrahymena thermophila exists as defined subchromosomal-sized molecules.* Nucleic Acids Res, 1985. **13**(16): p. 5817-5831.
35. Cassidy-Hanley, D., et al., *Genome-wide characterization of Tetrahymena thermophila chromosome breakage sites. II. Physical and genetic mapping.* Genetics, 2005. **170**(4): p. 1623-31.
36. Hamilton, E., et al., *Genome-wide characterization of tetrahymena thermophila chromosome breakage sites. I. Cloning and identification of functional sites.* Genetics, 2005. **170**(4): p. 1611-21.
37. Tabara, H., et al., *The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans.* Cell, 2002. **109**(7): p. 861-71.
38. Forstemann, K., et al., *Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein.* PLoS Biol, 2005. **3**(7): p. e236.
39. Saito, K., et al., *Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells.* PLoS Biol, 2005. **3**(7): p. e235.
40. Liu, Q., et al., *R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway.* Science, 2003. **301**(5641): p. 1921-5.
41. Chendrimada, T.P., et al., *TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing.* Nature, 2005. **436**(7051): p. 740-4.
42. Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex.* Nature, 2004. **432**(7014): p. 231-5.
43. Gregory, R.I., et al., *The Microprocessor complex mediates the genesis of microRNAs.* Nature, 2004. **432**(7014): p. 235-40.
44. Ray, C., *Meiosis and Nuclear Behavior in Tetrahymena pyriformis*.* Journal of Eukaryotic Microbiology, 1956. **3**(2): p. 88-96.
45. Martindale, D.W., C.D. Allis, and P. Bruns, *Conjugation in Tetrahymena thermophila: a temporal analysis of cytological stages.* Experimental Cell Research, 1982. **140**: p. 227-236.
46. Allen, S.L., *Genomic exclusion: a rapid means for inducing homozygous diploid lines in Tetrahymena pyriformis, syngen 1.* Science, 1967. **155**(762): p. 575-7.
47. Allen, S.L., *Cytogenetics of genomic exclusion in Tetrahymena.* Genetics, 1967. **55**(4): p. 797-822.
48. Allen, S.L., S.K. File, and S.L. Koch, *Genomic exclusion in tetrahymena.* Genetics, 1967. **55**(4): p. 823-37.
49. Cervantes, M.D., et al., *The CNA1 Histone of the Ciliate Tetrahymena thermophila Is Essential for Chromosome Segregation in the Germline Micronucleus.* Mol Biol Cell, 2006. **17**(1): p. 485-97.
50. Cui, B. and M.A. Gorovsky, *Centromeric histone H3 is essential for vegetative cell division and for DNA elimination during conjugation in Tetrahymena thermophila.* Mol Cell Biol, 2006. **26**(12): p. 4499-510.

51. Loidl, J. and H. Scherthan, *Organization and pairing of meiotic chromosomes in the ciliate Tetrahymena thermophila.* Journal of cell science, 2004. **117**(Pt 24): p. 5791-801.
52. Ahmad, K. and S. Henikoff, *The histone variant H3.3 marks active chromatin by replication- independent nucleosome assembly.* Mol Cell, 2002. **9**(6): p. 1191-200.
53. Jager, H., M. Rauch, and S. Heidmann, *The Drosophila melanogaster condensin subunit Cap-G interacts with the centromere-specific histone H3 variant CID.* Chromosoma, 2005. **113**(7): p. 350-61.
54. McGrath, K.E., et al., *An abundant nucleolar phosphoprotein is associated with ribosomal DNA in Tetrahymena macronuclei.* Mol Biol Cell, 1997. **8**(1): p. 97-108.
55. Smothers, J.F., et al., *Programmed DNA degradation and nucleolar biogenesis occur in distinct organelles during macronuclear development in Tetrahymena.* J. Euk. Microbiol., 1997. **44**(2): p. 79-88.
56. Chikashige, Y., et al., *Telomere-led premeiotic chromosome movement in fission yeast.* Science, 1994. **264**(5156): p. 270-3.
57. Zickler, D. and N. Kleckner, *The leptotene-zygotene transition of meiosis.* Annual review of genetics, 1998. **32**: p. 619-97.
58. Yao, M.C., A.R. Kimmel, and M.A. Gorovsky, *A small number of cistrons for ribosomal RNA in the germinal nucleus of a eukaryote, Tetrahymena pyriformis.* Proc Natl Acad Sci USA, 1974. **71**(8): p. 3082-3086.
59. Blackburn, E.H. and J.G. Gall, *A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena.* J. Mol. Biol., 1978. **120**: p. 33-35.
60. Kirk, K.E. and E.H. Blackburn, *An unusual sequence arrangement in the telomeres of the germ-line micronucleus in Tetrahymena thermophila.* Genes & development, 1995. **9**(1): p. 59-71.
61. Wolfe, J., B. Hunter, and W.S. Adair, *A cytological study of micronuclear elongation during conjugation in Tetrahymena.* Chromosoma, 1976. **55**(4): p. 289-308.
62. Ekwall, K., et al., *Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres.* Cell, 1997. **91**(7): p. 1021-32.
63. Volpe, T.A., et al., *Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi.* Science, 2002. **297**(5588): p. 1833-7.
64. Volpe, T., et al., *RNA interference is required for normal centromere function in fission yeast.* Chromosome Res, 2003. **11**(2): p. 137-46.
65. Reinhart, B.J. and D.P. Bartel, *Small RNAs correspond to centromere heterochromatic repeats.* Science, 2002. **297**(5588): p. 1831.
66. Noma, K., et al., *RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing.* Nat Genet, 2004. **36**(11): p. 1174-80.
67. Cam, H.P., et al., *Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome.* Nat Genet, 2005. **37**(8): p. 809-19.

68. Motamedi, M.R., et al., *Two RNAi Complexes, RITS and RDRC, Physically Interact and Localize to Noncoding Centromeric RNAs.* Cell, 2004. **119**(6): p. 789-802.
69. Sadaie, M., et al., *A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast.* The EMBO journal, 2004. **23**(19): p. 3825-35.
70. Partridge, J.F., et al., *cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site.* Current biology : CB, 2002. **12**(19): p. 1652-60.
71. Chalker, D.L. and M.C. Yao, *Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila.* Genes Dev, 2001. **15**(10): p. 1287-98.
72. Doyle, M. and M.F. Jantsch, *New and old roles of the double-stranded RNA-binding domain.* Journal of structural biology, 2002. **140**(1-3): p. 147-53.
73. Tian, B., et al., *The double-stranded-RNA-binding motif: interference and much more.* Nature reviews. Molecular cell biology, 2004. **5**(12): p. 1013-23.
74. Grishok, A., et al., *Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing.* Cell, 2001. **106**(1): p. 23-34.
75. Nicholson, R.H. and A.W. Nicholson, *Molecular characterization of a mouse cDNA encoding Dicer, a ribonuclease III ortholog involved in RNA interference.* Mammalian genome : official journal of the International Mammalian Genome Society, 2002. **13**(2): p. 67-73.
76. St Johnston, D., D. Beuchle, and C. Nusslein-Volhard, *Staufen, a gene required to localize maternal RNAs in the Drosophila egg.* Cell, 1991. **66**(1): p. 51-63.
77. St Johnston, D., et al., *A conserved double-stranded RNA-binding domain.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(22): p. 10979-83.
78. Green, S.R. and M.B. Mathews, *Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI.* Genes & development, 1992. **6**(12B): p. 2478-90.
79. Meurs, E., et al., *Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon.* Cell, 1990. **62**(2): p. 379-90.
80. Bass, B.L. and H. Weintraub, *An unwinding activity that covalently modifies its double-stranded RNA substrate.* Cell, 1988. **55**(6): p. 1089-98.
81. Kim, U., et al., *Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(24): p. 11457-61.
82. Pires-daSilva, A., et al., *Mice deficient for spermatid perinuclear RNA-binding protein show neurologic, spermatogenic, and sperm morphological abnormalities.* Developmental biology, 2001. **233**(2): p. 319-28.
83. Larcher, J.C., et al., *Ilf3 and NF90 associate with the axonal targeting element of Tau mRNA.* The FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2004. **18**(14): p. 1761-3.

84. Zhao, G., et al., *NF45/ILF2 tissue expression, promoter analysis, and interleukin-2 transactivating function.* Experimental cell research, 2005. **305**(2): p. 312-23.
85. Schumacher, J.M., et al., *Spnr, a murine RNA-binding protein that is localized to cytoplasmic microtubules.* The Journal of cell biology, 1995. **129**(4): p. 1023-32.
86. Schumacher, J.M., K. Artzt, and R.E. Braun, *Spermatid perinuclear ribonucleic acid-binding protein binds microtubules in vitro and associates with abnormal manchettes in vivo in mice.* Biology of reproduction, 1998. **59**(1): p. 69-76.
87. Kuchta, K., et al., *Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human.* Nucleic acids research, 2009. **37**(22): p. 7701-14.
88. Gorovsky, M.A., et al., *Isolation of micro- and macronuclei of Tetrahymena pyriformis.* Methods Cell Biol, 1975. **9**(0): p. 311-327.
89. Orias, E., E.P. Hamilton, and J.D. Orias, *Tetrahymena as a laboratory organism: useful strains, cell culture, and cell line maintenance.* Methods Cell Biol, 2000. **62**: p. 189-211.
90. Zdobnov, E.M. and R. Apweiler, *InterProScan--an integration platform for the signature-recognition methods in InterPro.* Bioinformatics, 2001. **17**(9): p. 847-8.
91. Cassidy-Hanley, D., et al., *Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment.* Genetics, 1997. **146**(1): p. 135-47.
92. Bruns, P.J. and D. Cassidy-Hanley, *Biolistic transformation of macro- and micronuclei.* Methods Cell Biol, 2000. **62**: p. 501-12.
93. Chalker, D.L., P. Fuller, and M.C. Yao, *Communication between parental and developing genomes during tetrahymena nuclear differentiation is likely mediated by homologous RNAs.* Genetics, 2005. **169**(1): p. 149-60.
94. Fan, Q., R. Sweeney, and M.-C. Yao, *Creation and use of antisense ribosomes in Tetrahymena thermophila*, in *Tetrahymena thermophila*, D.J. Asai and J.D. Forney, Editors. 1999, Academic Press: New York. p. 533-547.
95. Chalker, D.L. and M.C. Yao, *Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of Tetrahymena thermophila with germ line-limited DNA.* Mol Cell Biol, 1996. **16**(7): p. 3658-67.
96. Gaertig, J., et al., *High frequency vector-mediated transformation and gene replacement in Tetrahymena.* Nucleic Acids Res, 1994. **22**(24): p. 5391-8.
97. Yao, M.C., et al., *Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in Tetrahymena.* J Cell Sci, 2007. **120**(Pt 12): p. 1978-89.

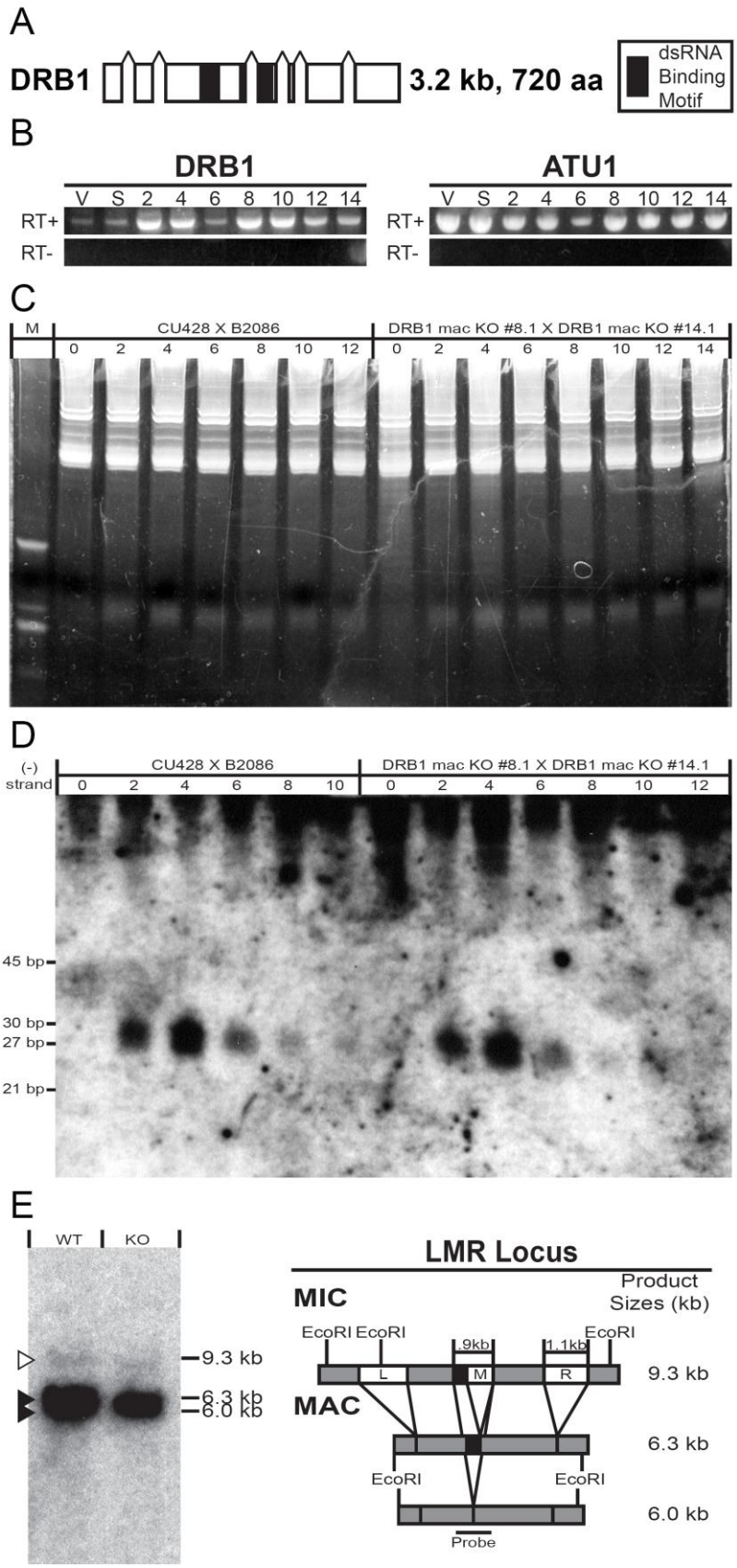

A
DRB1
3.2 kb, 720 aa
dsRNA Binding Motif
B
DRB1
ATU1
V S 2 4 6 8 10 12 14
RT+
RT-
C
M
CU428 X B2086
DRB1 mac KO #8.1 X DRB1 mac KO #14.1
0 2 4 6 8 10 12
0 2 4 6 8 10 12 14
D
(-) strand
CU428 X B2086
DRB1 mac KO #8.1 X DRB1 mac KO #14.1
0 2 4 6 8 10
0 2 4 6 8 10 12
45 bp
30 bp
27 bp
21 bp
E
WT
KO
9.3 kb
6.3 kb
6.0 kb
LMR Locus
MIC
Product Sizes (kb)
EcoRI EcoRI
.9kb
1.1kb
EcoRI
L
M
R
9.3 kb
MAC
6.3 kb
EcoRI
EcoRI
6.0 kb
Probe

Figure 1: Somatic knockouts of DRB1 produce normal levels of scnRNAs and are able to undergo DNA elimination normally. A. Genomic locus, conserved motifs and mRNA and protein length of putative tandem double-stranded RNA binding motif-(DSMR)-containing protein, **D**ouble-Stranded **R**NA **B**inding protein 1 (Drb1p). The DSRM as identified by BLASTN and PFAM analysis (See Material and Methods) is represented by a black rectangle. Splice sites are indicated by small carets above the gene. aa, amino acids. B. RT-PCR analysis of DRB1 during the *T. thermophila* life cycle. RNA samples were isolated from CU428 cells growing vegetatively (V), CU428 cells after an 18-hour starvation (S) and CU428 X B2086 cells at 2-hour intervals starting at 2 hours until 14 hours during the course of conjugation (2, 4, 6, 8, 10, 12 and 14). The α-tubulin gene (ATU1) was monitored to control for RNA loading. C. sRNA gel of RNA isolated from wild-type (CU428 X B2086) and DRB1 mac knockout (DRB1 mac KO #8.1 X DRB1 mac KO #14.1) matings during conjugation show no difference. RNA was harvested from wild-type and DRB1 mac knockout matings at 2-hour intervals starting at 0 hours until 14 hours during the course of conjugation (2, 4, 6, 8, 10, 12, 14). Oligonucleotide markers of 45bp, 30bp, 27bp and 21bp are indicated. D. sRNA northern blot of RNA isolated from wild-type and DRB1 mac knockout matings during conjugation show no difference in IES-specific scnRNAs. Mating wild-type and DRB1 mac knockout RNA was isolated at 2-hour intervals from 0 hours to10 hours during the course of conjugation (2, 4, 6, 8, 10, 12). E. Southern blot assaying DNA rearrangement of the LMR locus in wild-type and DRB1 mac knockout matings show normal IES rearrangement. Left. WT, CU428 X B2086 30hr genomic DNA; KO, DRB1 mac KO #8.1 X DRB1 mac KO #14.1 30hr genomic DNA. Right. Diagram of micro- and macronuclear LRM loci. During DNA elimination the M IES is able to undergo an alternative rearrangement that removes only the later 0.6 kb (white box) or can eliminate the entire 0.9 kb sequence (black and white boxes) (bottom).

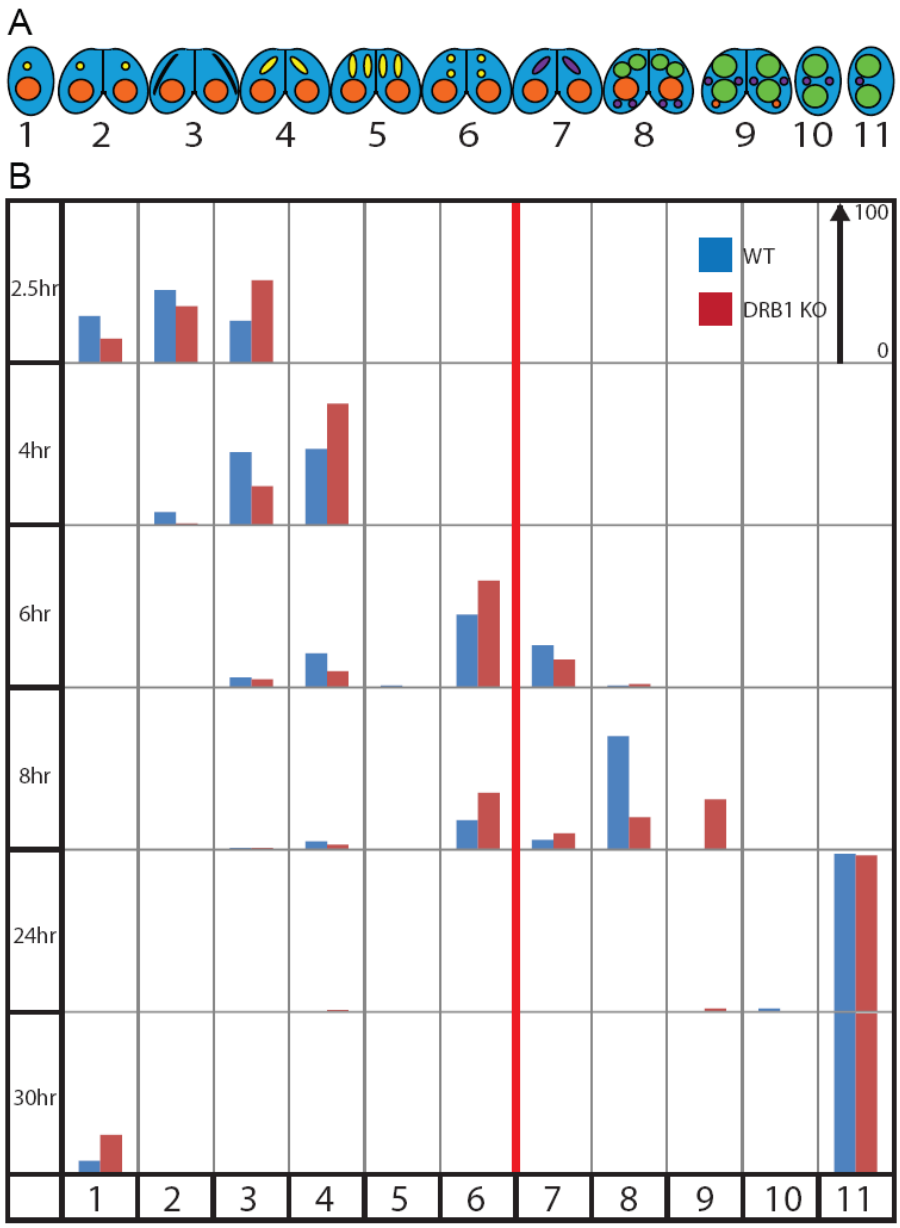


Figure 2. More DRB1 knockout pairs fail to progress to formation of zygotic nuclei during conjugation compared to wild-type. A. Nuclear diagrams of stages of conjugation. 11 stages of conjugation are depicted with the first six being pre-zygotic and the final five being post-zygotic. Yellow shapes, parental micronuclei; orange circles, parental macronuclei; purple shapes, zygotic micronuclei; green circles, developing zygotic macronuclei. B. Progression of wild-type (pICC-CNA1 #10.1 X pICC-CNA1 #16.5) and DRB1 knockout (DRB1 KO pICC-CNA1 #13.1 X DRB1 KO pICC-CNA1 #14) matings assayed at 2.5hr, 4hr, 6hr, 8hr, 24hr and 30hr into conjugation. A red line separates pre-zygotic from post-zygotic stages. Note the increase number of stage 6 cells at 8hr and stage 1 cells at 30hr into conjugation in DRB1 knockout matings. The number of cells at each stage during each time point assayed in both matings is shown in table S1. WT, wild-type mating.

Table S1: Mating progression of wild-type (pICC-CNA1 #10.1 X pICC-CNA1 #16.5) and DRB1 knockout (DRB1 KO pICC-CNA1 #13.1 X DRB1 KO pICC-CNA1 #14) matings during conjugation until terminal arrest.

A.

Mating[1]	Stage[2]											
pICC-CNA1 #10.1 X pICC-CNA1 #16.5	1	2	3	4	5	6	7	8	9	10	11	Total
2.5hr	29	45	26	0	0	0	0	0	0	0	0	100
4hr	0	8	45	47	0	0	0	0	0	0	0	100
6hr	0	0	6	21	1	45	26	1	0	0	0	100
8hr	0	0	1	5	0	18	6	70	0	0	0	100
24hr	0	0	0	0	0	0	0	0	0	2	98	100
30hr	8	0	0	0	0	0	0	0	0	0	100	100

B.

Mating[1]	Stage[2]											
DRB1 KO pICC-CNA1 #13.1 X DRB1 KO pICC-CNA1 #14	1	2	3	4	5	6	7	8	9	10	11	Total
2.5hr	15	35	51	0	0	0	0	0	0	0	0	101
4hr	0	1	24	75	0	0	0	0	0	0	0	100
6hr	0	0	5	10	0	66	17	2	0	0	0	100
8hr	0	0	1	3	0	35	10	20	31	0	0	100
24hr	0	0	0	1	0	0	0	0	2	0	97	100
30hr	24	0	0	0	0	0	0	0	1	0	99	100

Table S1. Mating progression of wild-type (pICC-CNA1 #10.1 X pICC-CNA1 #16.5) and DRB1 knockout (DRB1 KO pICC-CNA1 #13.1 X DRB1 KO pICC-CNA1 #14) matings during conjugation until terminal arrest. The mating, followed by the stages of conjugation as in figure 4A and the total number of cells counted are shown for each time point (2.5hr, 4hr, 6hr, 8hr, 24hr and 30hr) for both pICC-CNA1 #10.1 X pICC-CNA1 #16.5 (A) and DRB1 KO pICC-CNA1 #13.1 X DRB1 KO pICC-CNA1 #14 (B) matings.

1. Hours after mixing cells.

2. Number of cells at each stage post-mixing.

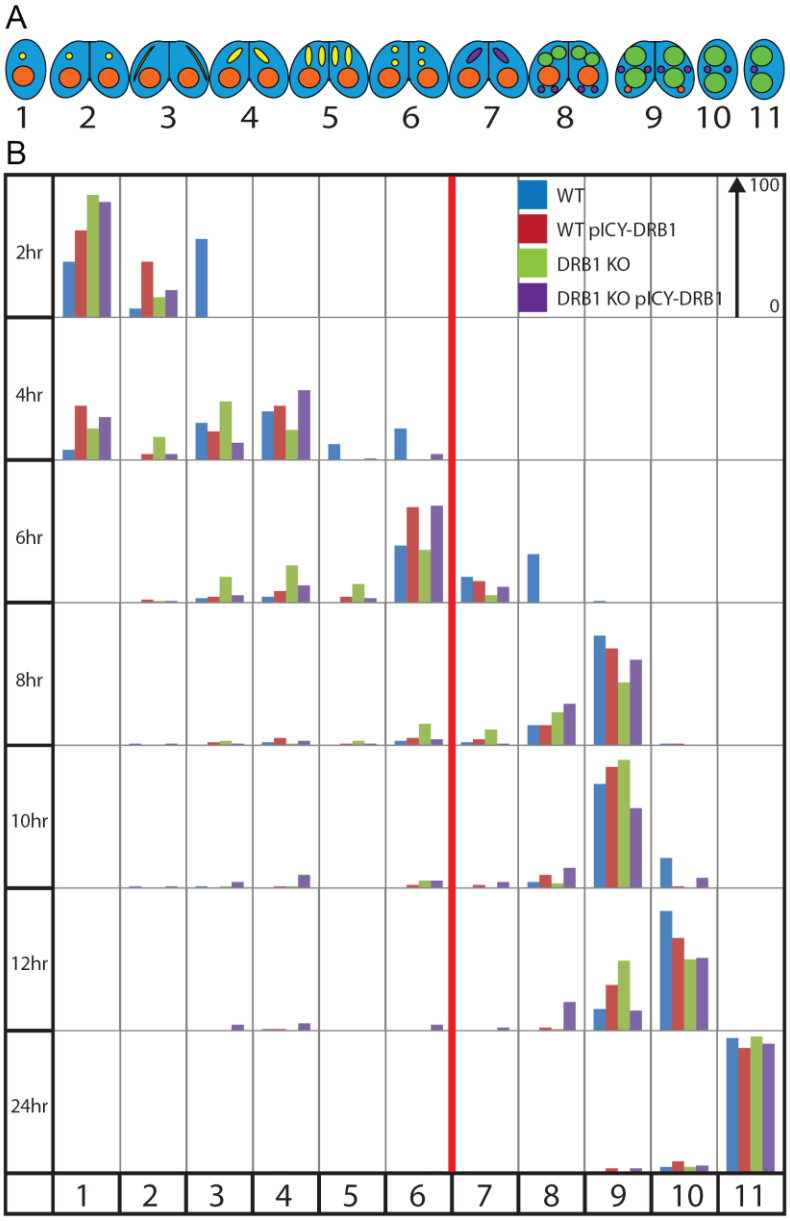


Figure 3. Exogenous expression of DRB1 rescues DRB1 knockout backout phenotype during conjugation. A. Nuclear diagrams of stages of conjugation as in figure 2A. Yellow shapes, parental micronuclei; orange circles, parental macronuclei; purple shapes, zygotic micronuclei; green circles, developing zygotic macronuclei. B. Progression of wild-type matings without exogenous DRB1 (CU427 X CU428) and with exogenous DRB1(CU427 X pICY-DRB1 #5) and DRB1 knockout matings without exogenous DRB1 (DRB1 KO #5.1.3 X DRB1 KO #7.1) and with exogenous DRB1 (DRB1 KO pICY-DRB1 #1 X DRB1 KO pICY-DRB1 #2) were assayed at 2hr, 4hr, 6hr, 8hr, 10hr, 12hr and 24hr into conjugation. A red line separates pre-zygotic from post-zygotic stages. Note the increase number of stage 6 cells at 8hr into conjugation in DRB1 knockout matings. The number of cells at each stage during each time point assayed in all matings is shown in table S2. WT, wild-type mating.

Table S2: Mating progression of wild-type matings without exogenous DRB1 (CU427 X CU428) and with exogenous DRB1 (CU427 X pICY-DRB1 #5) compared with DRB1 knockout matings without exogenous DRB1 (DRB1 KO #5.1.3 X DRB1 KO #7.1) and with exogenous DRB1 (DRB1 KO pICY-DRB1 #1 X DRB1 KO pICY-DRB1 #2) during conjugation until terminal arrest.

A.

Mating[1]	Strain[2]											
CU427 X CU428	1	2	3	4	5	6	7	8	9	10	11	Total
2hr	39	6	55	0	0	0	0	0	0	0	0	100
4hr	7	0	26	34	11	22	0	0	0	0	0	100
6hr	0	0	3	4	0	40	18	34	1	0	0	100
8hr	0	1	0	2	0	3	2	14	77	1	0	100
10hr	0	1	1	0	0	0	0	4	73	21	0	100
12hr	0	0	0	1	0	0	0	0	15	84	0	100
24hr	0	0	0	0	0	0	0	0	1	4	95	100

B.

CU427 X pICY-DRB1 #5	1	2	3	4	5	6	7	8	9	10	11	Total
2hr	61	39	0	0	0	0	0	0	0	0	0	100
4hr	38	4	20	38	0	0	0	0	0	0	0	100
6hr	0	2	4	8	4	67	15	0	0	0	0	100
8hr	0	0	2	5	1	5	4	14	68	1	0	100
10hr	0	0	0	1	0	2	2	9	85	1	0	100
12hr	0	0	0	1	0	0	0	2	32	65	0	100
24hr	0	0	0	1	0	0	0	0	3	8	88	100

C.

DRB1 KO #5.1.3 X DRB1 KO #7.1	1	2	3	4	5	6	7	8	9	10	11	Total
2hr	86	14	0	0	0	0	0	0	0	0	0	100
4hr	22	16	41	21	0	0	0	0	0	0	0	100
6hr	0	1	18	26	13	37	5	0	0	0	0	100
8hr	0	0	3	1	3	15	11	23	44	0	0	100
10hr	0	0	1	1	0	5	0	3	90	0	0	100
12hr	0	0	0	0	0	0	0	1	49	50	0	100
24hr	0	0	0	0	0	0	0	0	0	4	96	100

D.

DRB1 KO pICY-DRB1 #1 X DRB1 KO pICY-DRB1 #2	1	2	3	4	5	6	7	8	9	10	11	Total
2hr	81	19	0	0	0	0	0	0	0	0	0	100
4hr	30	4	12	49	1	4	0	0	0	0	0	100
6hr	0	1	5	12	3	68	11	0	0	0	0	100
8hr	0	1	1	3	1	4	1	29	60	0	0	100
10hr	0	1	4	9	0	5	4	14	56	7	0	100
12hr	0	0	4	5	0	4	2	20	14	51	0	100
24hr	0	0	0	0	0	1	0	0	3	5	91	100

Table S2. Mating progression of wild-type matings without exogenous DRB1 (CU427 X CU428) and with exogenous DRB1 (CU427 X pICY-DRB1 #5) compared with DRB1 knockout matings without exogenous DRB1 (DRB1 KO #5.1.3 X DRB1 KO #7.1) and with exogenous DRB1 (DRB1 KO pICY-DRB1 #1 X DRB1 KO pICY-DRB1 #2) during conjugation until terminal arrest. The mating, followed by the stages of conjugation as in figure 2A and the total number of cells counted are shown for each time point (2hr, 4hr, 6hr, 8hr, 10hr and 24hr) for CU427 X CU428 (A), CU427 X pICY-DRB1 #5 (B), DRB1 KO #5.1.3 X DRB1 KO #7.1 (C) and DRB1 KO pICY-DRB1 #1 X DRB1 KO pICY-DRB1 #2 (D) matings.

1. Hours after mixing cells.

2. Number of cells at each stage post-mixing.

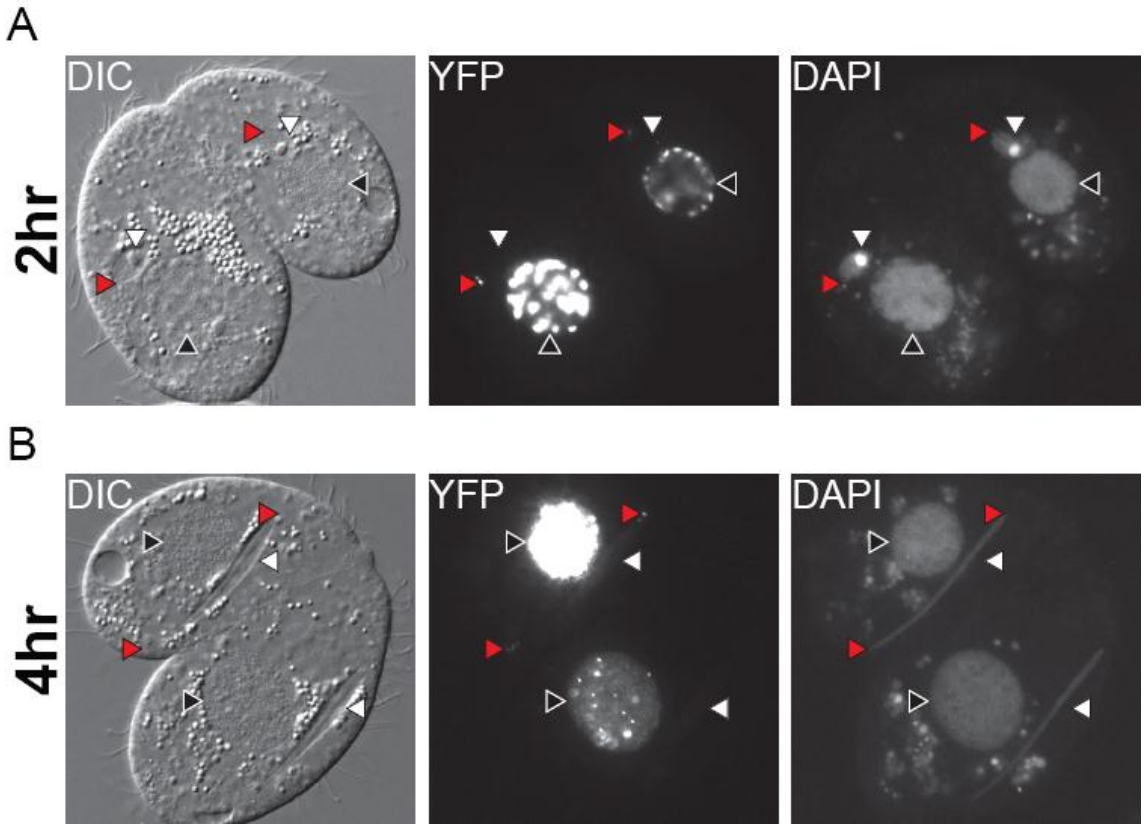


Figure 4: Drb1p localization at the tips of the crescent micronucleus prior to anaphase of meiosis I during conjugation. A and B. Localization of the Drb1p-YFP 2 hours into conjugation in stage II micronuclei and 3 hours into conjugation in stage IV micronuclei, respectively. Drb1p-YFP localizes strongly to the ends of the crescent micronucleus. White arrowhead, micronuclei; black arrowhead, parental macronuclei; red arrowhead, Drb1p-YFP micronuclear point localization.

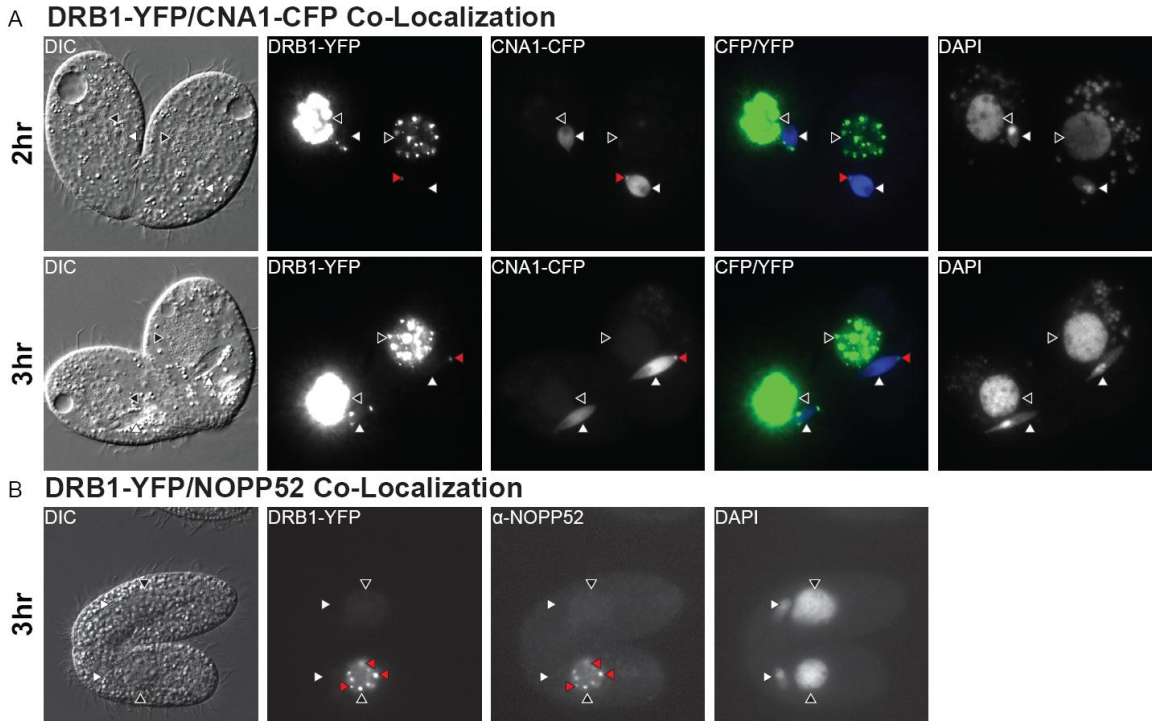


Figure 5: Co-localization of Drb1p-YFP early in conjugation with the centromere histone, Cna1p-CFP, in the crescent micronucleus and with the nucleolar protein, Nopp52p, in the macronucleus. A. Co-localization of Drb1p-YFP and Cna1p-CFP at 2 and 3 hours into conjugation. Drb1p-YFP co-localizes with centromere-incorporated Cna1p. White arrowhead, micronuclei; black arrowhead, parental macronuclei; red arrowhead, Drb1p-YFP/Cna1p-CFP/centromere co-localization. B. Co-localization of Drb1p-YFP and Nopp52p at 3 hours into conjugation. Drb1p-YFP strongly co-localizes with the nucleolus.

Table 1: Oligonucleotides used in the course of this study

Purpose and Name	Sequence (5'-3')
Knockout Cassette Generation	
DRB1: Upstream	
#1761-Loq1-2321AattB4	**GGGGACAACTTTGTATAGAAAAGTTGGTACCGGGATTACATAAAGATTTGATTCC**
#1762-Loq1-3366rattB1	**GGGGACTGCTTTTTTGTACAAACTTGCACAATTCAATCAAAAGTGCG**
Downstream	
#1763-Loq1-5746attB2	**GGGGACAGCTTTCTTGTACAAAGTGGCACTCTCATTAATGCCCCC**
#1764-Loq1-7093AattB3	**GGGGACAACTTTGTATAATAAAGTTGGTACCAGTAAAGAGCCTAAATCAAGG**
Lysate PCR and Sequencing of Knockout Cassettes	
#3047-M13Forward	**GTAAAACGACGGCCAGT**
#3048-M13Reverse	**TCACACAGGAAACAGCTATGAC**
Knockout PCR Screening	
DRB1:	
5' **#1679-MTT1-11484r**	**ATTTGGAATTAAGTACTTATTTCCAAAC**
#1946-DRB1-1086	**CGCGCACTTTTGATTGAATTGTG**
3' **#1866-Neo KO 2**	**CGTGATATTGCTGAAGAGCTTG**
#1867-Loq1-5353	**CAGGGGAAGATATATTTTATGAAGC**
#1868-Loq1-5764r	**GGGGGCATTAATGAGAGTG**
RT-PCR to Determine Expression	
DRB1: **#1688-Loq1-836**	**CGAAAAGGGGTTAGGGTTTTCTAGC**
#1689-Loq1-1325r	**CCCTTATCCCATCGTTTTCAG**
ATU1: **#3364-ATU1-2391r**	**GTGGCAATAGAAGCGTTGACA**
#3365-ATU1-1997	**TGCTCGATAACGAAGCCATCT**
Gene Amplification of Coding Sequence	
DRB1: **#1701-Loq1X**	**CACCCTCGAGAAAATGAATTCTTAGCAAG**
#1732-Loq1rH-Short	**AAGCTTTAGACTTATACTTTTCATGAAAG**
CNA1: **#2274-CNA1_-6gtw**	**CACCTTGCAAATGGCTAGGAAAGC**
#2275-CNA1_561rRV	**GATATCTTTTTTAGTAGGGATAAATATACCTG**
Lysate PCR of Coding Sequence Plasmids	
DRB1: **#1689-Loq1-1325r**	**CCCTTATCCCATCGTTTTCAG**
#1701-Loq1X	**CACCCTCGAGAAAATGAATTCTTAGCAAG**
CNA1: **#3047-M13Forward**	**GTAAAACGACGGCCAGT**
#3048-M13Reverse	**TCACACAGGAAACAGCTATGAC**

Table 1 (Cont.): Oligonucleotides used in the course of this study

Purpose and Name	Sequence (5'-3')
Sequencing of Coding Sequence Plasmids	
DRB1: **#1689-Loq1-1325r**	**CCCTTATCCCATCGTTTTCAG**
#1721-Loq1-5105	**GTCTCAGTAAAGCATTATAAGTAGCTG**
#1722-Loq1-5215r	**GAGTTTATCAGCAAATCGAATTTAGACC**
#3047-M13Forward	**GTAAAACGACGGCCAGT**
#3048-M13Reverse	**TCACACAGGAAACAGCTATGAC**
CNA1: **#3047-M13Forward**	**GTAAAACGACGGCCAGT**
#3048-M13Reverse	**TCACACAGGAAACAGCTATGAC**

CHAPTER 6

FUTURE PERSPECTIVES

The initial focus of this thesis was to investigate the biogenesis of sRNAs homologous to IESs (i.e. R IES; chapter 2). I found that early during conjugation in *T. thermophila* long, bidirectional IES-specific ncRNAs are processed by Dcl1p (chapter 3) and further speculated that putative tandem DSRM-containing proteins, Drb2p and Drb1p, (chapters 4 and 5) would directly assist in RNAi-directed DNA elimination of IESs. However, the work in this thesis, particularly the analysis of DRB2 and DRB1, has shown that dsRNAs have a greater role throughout RNAi-directed DNA elimination and conjugation than initially anticipated.

Prior to my graduate studies it was known that ncRNAs and scnRNAs played an important role in directing DNA elimination during conjugation [1, 2]. The function of ncRNA-derived scnRNAs associated with the Argonaute protein, Twi1p, seemed to be linked to heterochromatin formation on IESs, but how this exactly occurred was unknown [2, 3]. The work in this thesis has further established that genome remodeling in *T. thermophila* is an RNAi-directed process. IESs are transcribed to produce ncRNAs, which are processed into scnRNAs by Dcl1p (chapter 3) [1, 4, 5]. The resulting scnRNAs are bound by Twi1p and transported into the parental macronucleus for genome scanning [2, 3]. In the parental macronucleus, Twi1p/scnRNA complexes found to match ncRNA transcribed in the parental macronucleus are sequestered, while the remaining complexes are transferred to the zygotic macronucleus [2, 3, 6]. After Twi1p/scnRNA complexes bind zygotic macronuclear transcribed ncRNA, another protein in the complex, Ema1p, directs the E(z) homologue, Ezl1p, to methylate H3K9 and H3K27 of histones associated with IESs [6-9]. H3K9 and H3K27 methylation are then bound by the chromodomain-containing proteins, Pdd1p and Pdd3p, which in association with other proteins, form

DNA elimination bodies [7-16]. Later in conjugation in these DNA elimination bodies, the IESs are excised by the domesticated piggyBac transposase, Tpb2p [16].

The Role of ncRNA Early in RNAi-Directed DNA Elimination

I have shown through further study of the R IES (chapter 2), a source of ncRNAs, and DCL1 (chapter 3), the protein that processes these IES-specific ncRNAs, that ncRNAs play a fundamental role in RNAi-directed DNA elimination. Knockouts of the R IES prevent DNA elimination of artificial R IES plasmids, indicating an essential role for ncRNAs produced from the native R IES locus early in conjugation. This micronuclear ncRNA transcription from the R IES and other IESs is cleaved by the conjugation-specific Dicer homologue, Dcl1p, to produce scnRNAs (chapter 3) [4]. Knockouts of DCL1 show a loss of scnRNAs with a concomitant increase in the amount of the long, bidirectional ncRNA precursors. Loss of scnRNAs prevents H3K9 methylation of IESs and subsequent DNA elimination, chromosome breakage and completion of conjugation.

Prior research indicated that ncRNAs are transcribed from IESs early in conjugation by RNA polymerase II [1, 17, 18]. Although there appears to be promiscuous, genome-wide transcription, the majority of scnRNAs are homologous to IESs. How RNA polymerase II is targeted to IESs remains a mystery. Certain IESs (i.e. the TLR IES) appear to have higher levels of scnRNAs compared to others (C.D. Malone and D.L. Chalker, unpublished data). This begs the question as to whether there exists a mechanism to direct RNA polymerase II to IESs or, whether there are a greater number of IESs throughout the micro- and macronuclear genomes that act as templates for IES-specific dsRNA production than are currently believed to exist. Does the number of complete or partial copies in the micro- and macronucleus or the size of the IES also

affect general ncRNA production? Is there preferential cleavage of certain long, bidirectional ncRNAs over others? Given no evidence of the existence of a conserved sequence among IESs it seems difficult to imagine what mechanism could induce preferential cleavage of certain ncRNAs over others. This suggests that scnRNA levels likely correlate with overall ncRNA production; however, this does not resolve whether precursor ncRNA levels are determined by preferential transcription or are dependent on IES copy number. Given that there is a family of TLR IESs in the micronucleus, this lends credence to the theory that ncRNA levels and hence scnRNA levels are dictated by IES copy number [19]. Data regarding the M IES complicates this theory since it has been shown that there are a number of additional M-like sequences of ~200bp that are scattered throughout the genome ([6, 20], K. Gao and D.L. Chalker, unpublished data). Transcription in the micronucleus could produce ncRNA from both the complete M IES and these M-like sequences, which could be processed into high levels of scnRNAs to target the M IES for elimination. However, the M-like sequences retained in the parental macronucleus could sequester many of these scnRNAs blocking M IES and M-like sequence rearrangement [6, 20, 21]. The data I present here and other's data regarding the R IES seems to indicate that it is a single copy IES that produces low levels of ncRNAs and hence low scnRNA levels (chapter 2) (C.D. Malone and D.L. Chalker, unpublished data). Although a knockout of the R IES should completely fail to rearrange the artificial R IES plasmid, this is not the case. This indicates that the macronuclear-retained ~20bp repeats of the R IES are able to contribute to minimum R IES rearrangement. Therefore, it seems that partial IES-like sequences in the macronucleus are also able to direct some DNA elimination. Although the R IES data does not answer the question definitively as

to whether IESs are preferentially transcribed or not, it seems to indicate that ncRNA levels correlate to the number of identified and unidentified partial or complete IESs in the micro- and macronuclear genomes. Completion of the *T. thermophila* micronuclear genome sequencing project and identification and characterization of other IESs would allow the copy number of each full and partial IES to be determined. This data could be compared to observed scnRNA levels during conjugation to definitely ascertain whether IES copy number correlates to scnRNA levels and what contribution these macronuclear-retained IES-like repeats have to ncRNA and scnRNA levels as well.

While DCL1 has been found to be essential for scnRNA production early in conjugation, there is also upregulation of DCL1 at the same time as the zygotic macronuclei begin to appear. Is this increase in DCL1 transcription a result of the promiscuous zygotic macronuclear genome transcription or do these RNA transcripts synthesize biologically relevant Dcl1p later in conjugation? Thus far Dcl1p has been shown to only localize to the crescent micronucleus early in conjugation, but given that another DSRM-containing protein is required zygotically (DRB2), it is reasonable to postulate a secondary role for Dcl1p in RNAi-directed DNA elimination. As mentioned above, Dcl1p localizes to the crescent micronucleus during the time of scnRNA production from long, bidirectional ncRNAs. The Argonaute protein that binds them, Twi1p, localizes predominantly to the parental macronucleus at this time [2, 22]. Low levels of Twi1p can be found in the cytoplasm, but it is completely excluded from the crescent micronucleus. Therefore, it appears that after the scnRNAs are cleaved by Dcl1p they either freely diffuse or are escorted by a protein chaperone into the cytoplasm to be bound by Twi1p. The former seems unlikely since the stability of these scnRNAs has

been shown to be low [2, 22]. If there is active transport, what protein(s) is responsible for transporting these scnRNAs from the micronucleus to Twi1p in the cytoplasm? Initially the DSRM-containing protein, Drb1p, was hypothesized to play this role due to its nuclear localization early in conjugation but DRB1 knockouts have no detectable change in scnRNA levels. A number of putative nucleotide transferases along with an RNA-dependent RNA polymerase have been found to associate with the vegetative essential Dicer, Dcr2p, indicating that further biochemical study of epitope-tagged Dcl1p may yield the identity of any other proteins necessary for scnRNA biogenesis and trafficking [23, 24]. Biochemical protein analysis using an epitope-tagged Dcl1p late in conjugation would also be able to determine if any functional Dcl1p is produced and what role it could likely play.

The Role of ncRNA Later During RNAi-Directed DNA Elimination and Early in Conjugation

Although the tandem DSRM-containing proteins, Drb2p and Drb1p, were preliminarily proposed to be essential partner proteins for Dcl1p during scnRNA biogenesis, work on these two proteins has expanded the possible roles for dsRNA later in RNAi-directed DNA elimination and in conjugation in general. Both proteins are constitutively expressed but show upregulation during different times of conjugation. Localization of Drb2p and Drb1p is found in discrete macronuclear foci with Drb1p localizing to the ends of the crescent micronucleus during prophase of meiosis I as well. Drb2p was also found late in conjugation to co-localize with the chromodomain protein, Pdd1p, in DNA elimination bodies, which are essential for DNA elimination. In matings between DRB2 zygotic knockouts, Pdd1p fails to localize into these DNA elimination bodies. This result

likely explains why DRB2 zygotic knockouts also fail to undergo DNA elimination and chromosome breakage upon completion of conjugation. Zygotic knockouts of DRB2 also arrest before completion of conjugation similar to knockouts of DCL1, indicating a role late in conjugation for Drb2p. This is perhaps the most intriguing piece of data from this set of experiments because it points to ncRNAs playing an essential role late in conjugation. What exactly this role is offers many tantalizing possibilities. Does Drb2p play a role in directing methylation of IES-associated histones on newly synthesized DNA in developing zygotic macronuclei after they undergo one or two rounds of DNA replication to facilitate Pdd1p binding before DNA elimination? It is also possible that instead of directing *de novo* methylation that it promotes overall maintenance of these methylation marks through similar associations with ncRNAs or scnRNAs and their Argonaute proteins later in conjugation in order to maintain Pdd1p binding at IESs [25, 26]. A number of these Argonaute proteins besides TWI1 (ex. TWI2, TWI8, TWI9, TWI10, TWI11 and TWI12) are upregulated during conjugation in *T. thermophila* and bind a variety of sRNAs [26, 27]. The possibility that Drb2p/RNA complexes promote formation of mature DNA elimination bodies directly through protein-RNA or protein-protein interactions with other DNA elimination body proteins also exists. All of these possibilities suggest a further role for ncRNA in RNAi-directed DNA elimination, which can be answered by studies with epitope-tagged Drb2p. Such biochemical studies should clarify what role Drb2p plays late in conjugation and could also be similarly applied during vegetative growth to determine its essential function then.

Further study of the nuclear localization of Drb1p found Drb1p to co-localize in the macronucleus with a nucleolar component, Nopp52p, and to co-localize in the

micronucleus with the centromere histone protein, Cna1p. DRB1 knockout matings also produce fewer progeny and abort conjugation at the pre-zygotic/post-zygotic transition in higher numbers when compared to wild-type matings. Analysis of the progeny production data and the conjugation progression data indicates that Drb1p production early during conjugation is important in haploid gametic nuclei creation. Formation of the crescent micronuclei is the result of an elaborate elongation process that takes place during prophase of meiosis I and is essential for homologous recombination [28-30]. Therefore, the localization of Drb1p at one end of the crescent micronucleus at the centromere and likely Drb1p localization at the telomere at the other end of the crescent micronucleus indicates that this protein has an important function during homologous recombination. It would not be unthinkable that Drb1p could have this role, since *T. thermophila* has no obvious synaptonemal complex and would need other proteins to take over this function to ensure proper recombination and crossover resolution ([30, 31], reviewed in [32]). How Drb1p would accomplish this role is perplexing. Since IES-specific ncRNAs are transcribed at the same time as homologous recombination, it is possible that Drb1p binds to centromere- and telomere-specific ncRNAs. This Drb1p/ncRNA complex could interact with other proteins to either anchor the micronuclear chromosome to the ends of the crescent micronucleus ensuring chromosome stability during homologous recombination or possible interact with a different set of proteins that are necessary for crossover resolution. In either scenario macronuclear Drb1p localized at the nucleolus could also bind telomere ncRNA produced from the rDNA mini-chromosomes. What the function of this hypothetical Drb1p binding remains to be determined. It is also possible that Drb1p acts as a tandem

DSRM-containing protein partner with Dcl1p for centromere- and telomere-specific scnRNA generation and could subsequently transport these scnRNAs into the parental and zygotic macronuclei to direct genome scanning and DNA elimination of these sequences, respectively [33, 34]. The inability to detect changes of scnRNA levels in DRB1 somatic knockouts does not discount this hypothesis since centromere- and telomere-specific scnRNAs would likely only account for a very small percentage of overall scnRNAs. In this scenario the Drb1p foci in the parental macronucleus would act as a staging ground for genome scanning and later DNA elimination in the zygotic macronucleus. A third possibility is that these Drb1p macronuclear nucleolar foci are sites for genome scanning, which is disrupted in DRB1 knockouts causing essential genes to be excised from the zygotic macronucleus later in development. This could also explain the lower progeny production in DRB1 knockout matings. Conversely, neither of these later roles seems likely since loss of centromere and telomere scnRNA production or failure of genome scanning would not likely produce a large number of backouts during conjugation. Perhaps Drb1p has multiple roles where it is involved in homologous recombination in the micronucleus and either centromere- and telomere-specific scnRNA transport or facilitating genome scanning in the parental macronucleus. Examining these possibilities could be facilitated by immunoprecipitation studies with epitope-tagged Drb1p. RNA immunoprecipitation (RIP) would allow identification of the class of RNAs that associate with Drb1p, which in turn could pinpoint the role of Drb1p with any other proteins it might interact with during conjugation.

The work in this thesis on the role of ncRNAs during RNAi-directed DNA elimination and conjugation has opened up new possible research directions. It has

previously been shown mechanistically in *S. pombe* and *A. thaliana* that ncRNAs, siRNAs and associated RNAi proteins are essential for heterochromatin formation, which is similar to RNAi-directed DNA elimination in *T. thermophila* ([1, 2, 4, 5, 7, 8, 11, 13, 16, 35-49], reviewed in [50]). While it was initially thought that the two tandem DSRM-containing proteins, Drb2p and Drb1p, were potential Dcl1p protein partners, it has now been shown that they likely have other unexpected roles in RNAi-directed DNA elimination and in conjugation. respectively. DSRM-containing proteins are also essential for a number of other biological functions including sRNA production, RNA editing, translation inhibition in response to the presence of dsRNA and mRNA localization [51-61]. Many DSRM-containing proteins have no identified function, and through research on Drb2p and Drb1p and the ncRNAs or ncRNA products that they bind, a great deal can be learned about the role of these RNAs and the biological processes they affect.

References:

1. Chalker, D.L. and M.C. Yao, *Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila.* Genes Dev, 2001. **15**(10): p. 1287-98.
2. Mochizuki, K., et al., *Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena.* Cell, 2002. **110**(6): p. 689-99.
3. Mochizuki, K. and M.A. Gorovsky, *Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement.* Genes Dev, 2004. **18**(17): p. 2068-73.
4. Malone, C.D., et al., *Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila.* Mol Cell Biol, 2005. **25**(20): p. 9151-64.
5. Mochizuki, K. and M.A. Gorovsky, *A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase.* Genes Dev, 2005. **19**(1): p. 77-89.
6. Aronica, L., et al., *Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in Tetrahymena.* Genes Dev, 2008. **22**(16): p. 2228-41.
7. Taverna, S.D., R.S. Coyne, and C.D. Allis, *Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena.* Cell, 2002. **110**(6): p. 701-11.

8. Liu, Y., et al., *RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena.* Genes Dev, 2007. **21**(12): p. 1530-45.
9. Taverna, S.D., et al., *RNAi-dependent Recruitment of a Polycomb Repressive Complex for Developmentally Regulated Heterochromatin Formation in Tetrahymena (Unpublised Data).*
10. Madireddi, M.T., M.C. Davis, and C.D. Allis, *Identification of a novel polypeptide involved in the formation of DNA-containing vesicles during macronuclear development in Tetrahymena.* Dev Biol, 1994. **165**(2): p. 418-31.
11. Madireddi, M.T., et al., *Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena.* Cell, 1996. **87**(1): p. 75-84.
12. Smothers, J.F., et al., *Pdd1p associates with germline-restricted chromatin and a second novel anlagen-enriched protein in developmentally programmed DNA elimination structures.* Development, 1997. **124**(22): p. 4537-45.
13. Nikiforov, M.A., M.A. Gorovsky, and C.D. Allis, *A novel chromodomain protein, pdd3p, associates with internal eliminated sequences during macronuclear development in Tetrahymena thermophila.* Mol Cell Biol, 2000. **20**(11): p. 4128-34.
14. Rexer, C.H. and D.L. Chalker, *Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in Tetrahymena thermophila.* Eukaryot Cell, 2007. **6**(8): p. 1320-9.
15. Yao, M.C., et al., *Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in Tetrahymena.* J Cell Sci, 2007. **120**(Pt 12): p. 1978-89.
16. Cheng, C.Y., et al., *A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila.* Mol Biol Cell, 2010. **21**(10): p. 1753-62.
17. Mochizuki, K. and M.A. Gorovsky, *RNA polymerase II localizes in Tetrahymena thermophila meiotic micronuclei when micronuclear transcription associated with genome rearrangement occurs.* Eukaryot Cell, 2004. **3**(5): p. 1233-40.
18. Stargell, L.A. and M.A. Gorovsky, *TATA-binding protein and nuclear differentiation in Tetrahymena thermophila.* Mol. Cell. Biol., 1994. **14**(1): p. 723-734.
19. Wuitschick, J.D., et al., *A novel family of mobile genetic elements is limited to the germline genome in Tetrahymena thermophila.* Nucleic Acids Res, 2002. **30**(11): p. 2524-37.
20. Kowalczyk, C.A., et al., *The germ line limited M element of Tetrahymena is targeted for elimination from the somatic genome by a homology-dependent mechanism.* Nucleic Acids Res, 2006.
21. Chalker, D.L., P. Fuller, and M.C. Yao, *Communication between parental and developing genomes during tetrahymena nuclear differentiation is likely mediated by homologous RNAs.* Genetics, 2005. **169**(1): p. 149-60.
22. Noto, T., et al., *The Tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus.* Cell, 2010. **140**(5): p. 692-703.

23. Lee, S.R. and K. Collins, *Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs.* Nat Struct Mol Biol, 2007. **14**(7): p. 604-10.
24. Lee, S.R., K.B. Talsky, and K. Collins, *A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis.* RNA, 2009. **15**(7): p. 1363-74.
25. Allis, C.D. and D.K. Dennison, *Identification and purification of young macronuclear anlagen from conjugating cells of Tetrahymena thermophila.* Developmental biology, 1982. **93**(2): p. 519-33.
26. Couvillion, M.T., et al., *Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of Tetrahymena thermophila.* Genes & development, 2009. **23**(17): p. 2016-32.
27. Couvillion, M.T., R. Sachidanandam, and K. Collins, *A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo.* Genes & development, 2010. **24**(24): p. 2742-7.
28. Sugai, T. and K. Hiwatashi, *Cytological and autoradiographic studies of the micronucleus at meiotic prophase in Tetrahymena pyriformis.* J. Protozool, 1974. **21**: p. 542-548.
29. Martindale, D.W., C.D. Allis, and P. Bruns, *Conjugation in Tetrahymena thermophila: a temporal analysis of cytological stages.* Experimental Cell Research, 1982. **140**: p. 227-236.
30. Loidl, J. and H. Scherthan, *Organization and pairing of meiotic chromosomes in the ciliate Tetrahymena thermophila.* Journal of cell science, 2004. **117**(Pt 24): p. 5791-801.
31. Wolfe, J., B. Hunter, and W.S. Adair, *A cytological study of micronuclear elongation during conjugation in Tetrahymena.* Chromosoma, 1976. **55**(4): p. 289-308.
32. Zickler, D. and N. Kleckner, *The leptotene-zygotene transition of meiosis.* Annual review of genetics, 1998. **32**: p. 619-97.
33. Blackburn, E.H. and J.G. Gall, *A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena.* J. Mol. Biol., 1978. **120**: p. 33-35.
34. Kirk, K.E. and E.H. Blackburn, *An unusual sequence arrangement in the telomeres of the germ-line micronucleus in Tetrahymena thermophila.* Genes & development, 1995. **9**(1): p. 59-71.
35. Ekwall, K., et al., *Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres.* Cell, 1997. **91**(7): p. 1021-32.
36. Volpe, T.A., et al., *Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi.* Science, 2002. **297**(5588): p. 1833-7.
37. Volpe, T., et al., *RNA interference is required for normal centromere function in fission yeast.* Chromosome Res, 2003. **11**(2): p. 137-46.
38. Reinhart, B.J. and D.P. Bartel, *Small RNAs correspond to centromere heterochromatic repeats.* Science, 2002. **297**(5588): p. 1831.

39. Noma, K., et al., *RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing.* Nat Genet, 2004. **36**(11): p. 1174-80.
40. Cam, H.P., et al., *Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome.* Nat Genet, 2005. **37**(8): p. 809-19.
41. Motamedi, M.R., et al., *Two RNAi Complexes, RITS and RDRC, Physically Interact and Localize to Noncoding Centromeric RNAs.* Cell, 2004. **119**(6): p. 789-802.
42. Sadaie, M., et al., *A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast.* The EMBO journal, 2004. **23**(19): p. 3825-35.
43. Partridge, J.F., et al., *cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site.* Current biology : CB, 2002. **12**(19): p. 1652-60.
44. Liu, Y., K. Mochizuki, and M.A. Gorovsky, *Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena.* Proc Natl Acad Sci U S A, 2004. **101**(6): p. 1679-84.
45. Coyne, R.S., et al., *Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation.* Mol Cell, 1999. **4**(5): p. 865-72.
46. Onodera, Y., et al., *Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation.* Cell, 2005. **120**(5): p. 613-22.
47. Kanno, T., et al., *Atypical RNA polymerase subunits required for RNA-directed DNA methylation.* Nat Genet, 2005. **37**(7): p. 761-5.
48. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**(6796): p. 593-9.
49. Jackson, J.P., et al., *Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase.* Nature, 2002. **416**(6880): p. 556-60.
50. Allis, C.D., et al., eds. *Epigenetics*. 2007, Cold Spring Harbor Press: Cold Spring Harbor.
51. Bernstein, E., et al., *Role for a bidentate ribonuclease in the initiation step of RNA interference.* Nature, 2001. **409**(6818): p. 363-6.
52. Grishok, A., et al., *Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing.* Cell, 2001. **106**(1): p. 23-34.
53. Nicholson, R.H. and A.W. Nicholson, *Molecular characterization of a mouse cDNA encoding Dicer, a ribonuclease III ortholog involved in RNA interference.* Mammalian genome : official journal of the International Mammalian Genome Society, 2002. **13**(2): p. 67-73.
54. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* Nature, 2003. **425**(6956): p. 415-9.
55. St Johnston, D., D. Beuchle, and C. Nusslein-Volhard, *Staufen, a gene required to localize maternal RNAs in the Drosophila egg.* Cell, 1991. **66**(1): p. 51-63.

56. St Johnston, D., et al., *A conserved double-stranded RNA-binding domain.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(22): p. 10979-83.
57. Green, S.R. and M.B. Mathews, *Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI.* Genes & development, 1992. **6**(12B): p. 2478-90.
58. Meurs, E., et al., *Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon.* Cell, 1990. **62**(2): p. 379-90.
59. Bass, B.L. and H. Weintraub, *An unwinding activity that covalently modifies its double-stranded RNA substrate.* Cell, 1988. **55**(6): p. 1089-98.
60. Kim, U., et al., *Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(24): p. 11457-61.
61. Pires-daSilva, A., et al., *Mice deficient for spermatid perinuclear RNA-binding protein show neurologic, spermatogenic, and sperm morphological abnormalities.* Developmental biology, 2001. **233**(2): p. 319-28.

APPENDIX

SUBTRACTION BY ADDITION: DOMESTICATED TRANSPOSASES IN PROGRAMMED DNA ELIMINATION

Genes and Development. 2009. 23: 2455-24

Contributions to the Paper:

In response to a request from the editors of Genes and Development Doug and I produced a paper discussing the implications of a paper identifying a domesticated transposase, PGM, and its role in DNA elimination in the ciliate, *Paramecium tetraurelia* [1]. Doug and I outlined our approach to the paper together. I produced the initial draft of the paper along with the final version of the figure. Doug edited the initial draft of the paper and revised it for publication.

References:

1. Baudry, C., et al., *PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia.* Genes Dev, 2009. **23**(21): p. 2478-83.

Subtraction by addition: domesticated transposases in programmed DNA elimination

Jason A. Motl and Douglas L. Chalker[1]

Biology Department, Washington University in St. Louis, St. Louis, Missouri 63130, USA

The ciliate *Paramecium tetraurelia* must eliminate ~60,000 short sequences from its genome to generate uninterrupted coding sequences in its somatic macronucleus. In this issue of *Genes & Development*, Baudry and colleagues (pp. 2478–2483) identify the protein that excises these noncoding sequences: a domesticated *piggyBac* transposase that has been adapted to remove what are likely the remnants of transposon insertions. This new study reveals how addition of a transposase to small RNA-directed silencing machinery can guide major genome reorganization.

In the midst of studying chromosomal behavior during cell division, a body of work that provided support for his chromosomal theory of inheritance, Boveri (1887) described a peculiar genomic phenomenon of massive DNA elimination that he called "chromatin diminution". Carefully watching embryonic development of the roundworm *Parascaris univalens*, Boveri (1887) observed that the chromosomes of the developing somatic cells became highly fragmented during early cleavages. The central fragments segregated to the poles, becoming the cells' somatic chromosomes, while the remaining chromatin remained in the cytoplasm and was eventually degraded (see Goday and Pimpinelli 1984; Muller et al. 1996; Muller and Tobler 2000). This is not the way one expects an organism to treat its genome. Nevertheless, several examples of large-scale DNA elimination events have since been documented in a taxonomically diverse array of organisms, including sciarid flies, lampreys, copepods, hagfish, and ciliated protozoa (Beerman 1977; Kubota et al. 1993; for review, see Prescott 1994; Goday and Esteban 2001; Smith et al. 2009), which suggests such genome-downsizing must offer some advantage when it occurs.

Another thing that these unorthodox phenomena reveal is that large blocks of genomes are expendable, at least in somatic cells. The fact that such DNA elimination events are developmentally programmed indicates that the cells in which they occur have regulated mechanisms to recognize superfluous DNA and remove it. The eliminated DNA, in many cases, consists largely of repetitive sequences, which are often thought of as junk DNA. DNA elimination appears to be these cells' means to clear the junk out of the attic. More conventional eukaryotes simply store their repetitive DNA in the attic of silent heterochromatin. The results of several recent studies suggest that the recognition of DNA to be packaged into heterochromatin may be a precursor to genome-wide DNA elimination. A report from the research group of Betermier (Baudry et al. 2009) in this issue of *Genes & Development* reveals that a domesticated *piggyBac* transposase has been recruited to carry out programmed DNA elimination in the ciliate *Paramecium tetraurelia*. Previous work in *Paramecium* and related ciliates has shown that an RNAi-related machinery is also essential for these DNA rearrangements, providing a key link between DNA elimination and heterochromatin formation (for review, see Yao and Chao 2005; Duharcourt et al. 2009). Thus, the simple addition of a transposase to an existing silencing pathway provides the means to subtract unwanted DNA from the genome.

Large amounts of DNA are eliminated from the somatic macronucleus (MAC)

Ciliates, such as *Paramecium*, are single-celled protozoa that carry two different copies of their genome, each of which is contained in a functionally distinct nucleus (for review, see Prescott 1994). Micronuclei (MICs) harbor the transcriptionally silent germline genome, ready to supply their genetic material to the next generation; MACs carry the somatic genome, providing for all gene expression in the current generation. The somatic genome is a streamlined version of the germline, having undergone extensive elimination of DNA sequences and fragmentation of its chromosomes during its differentiation. This differentiation occurs during sexual reproduction upon conjugating with a partner cell or, in some ciliates, a self-fertilization process called autogamy (see Meyer and Chalker 2007). During conjugation or autogamy, the parental MAC DNA is eventually discarded and the

[*Keywords*: Transposase domestication; IES excision; chromosome fragmentation; oligohymenophorean ciliates]
[1]**Corresponding author.**
E-MAIL dchalker@biology2.wustl.edu; FAX (314) 935-4432.
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.1864609.

germline MIC provides copies of the genome for both the new MICs and the MACs of the progeny cells. While both the MIC and MAC progenitors start with identical copies of the genome, the differentiation of the somatic macronucleus removes large portions of this DNA (up to 95% of the germline genome is some ciliate species).

It seems reasonable to classify much of the DNA that is eliminated from the *Paramecium* MAC as junk; it is discarded, after all. Most of the repetitive sequences, such as DNA satellites and transposon-like elements, as well as numerous noncoding single-copy sequences, are the target of genome reorganization (for review, see Betermier 2004). The repetitive elements are eliminated through a somewhat imprecise process that deletes hundreds to thousands of base pairs. If the broken ends generated during the excision of a particular germline-limited sequence fail to be joined, the chromosome fragments are stabilized by de novo telomere addition (Le Mouel et al. 2003). The single-copy sequences removed are short (26–886 base pairs [bp]), AT-rich, noncoding sequences, referred to as internal eliminated sequences (IESs), which number from 50,000 to 65,000 per haploid genome. IESs are largely located within the body of genes and must be precisely excised to create uninterrupted coding sequences in the MAC, making their efficient removal an essential process. Both IESs and the repetitive sequences are bound by TA dinucleotides, for which only one copy is retained upon their removal from the genome (see Fig. 1A). For IESs, the TA dinucleotides are imbedded in a larger consensus sequence of 5′-tgg**TA**YAGYNR-3′ that is similar to the terminal inverted repeats (TIRs) of germline-limited Tec elements of the distantly related ciliate *Euplotes crassus*, which are members of the widespread Tc/mariner family of transposons. This finding provided support for the idea that IESs are derived from transposon insertions whose sequence subsequently degenerated until only abbreviated TIRs are recognizable (see Fig. 1B; Klobutcher and Herrick 1995, 1997).

PiggyMac: the domestication of a transposase

The idea that IESs are the remnants of transposons has always been an attractive hypothesis, but whether transposons actively participate in programmed DNA elimination

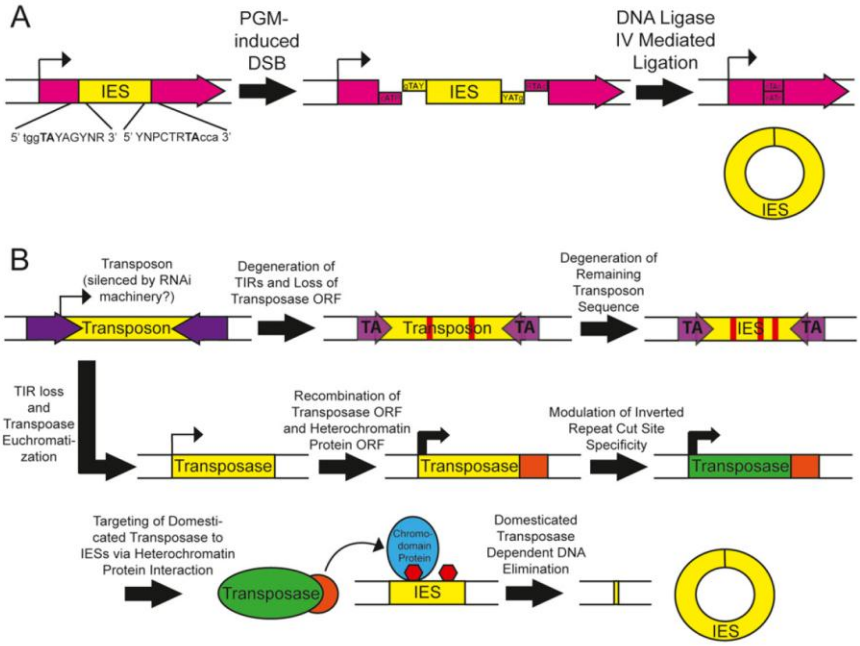


Figure 1. Pgm-mediated DNA elimination of an IES and the evolution of IESs and a domesticated transposase. (*A*) IESs in *P. tetraurelia* contain TIR containing a conserved TA dinucleotide repeat. Concomitant DNA DSBs of the IES generate 4-bp 5′ overhangs, which are paired at the central TA and are repaired by DNA ligase IV and other host factors. (*B*) Domestication of an IES transposase. (*Top* line) A generic transposon undergoes mutation that causes loss of a functional transposase ORF and degeneration of inverted repeats (purple arrow) retaining only the TA dinucleotide and the minimally needed recognition sequence. Transposon undergoes further degeneration (red boxes) to create an IES. (*Bottom* lines) An ancestral *piggyBac* transposon becomes fixed in the euchromatin of the MAC during macronuclear development by loss of its TIRs. Acquisition of heterochromatin protein-binding capacity by the retained transposon creates a developmentally regulated chimeric protein that is able to be recruited to silent heterochromatin. Further modulation of the inverted repeat cut site allows more promiscuous recognition of IES boundary elements. Now, upon RNAi-directed heterochromatin formation in ciliates, these heterochromatic sequences are removed from the MAC during development to create a transposon-free somatic MAC.

was not known. This new report from Baudry et al. (2009) shows that *Paramecium* uses what was once a transposase from a *piggyBac* element to reorganize its somatic genome. When this putative endogenous transposase was fused to green fluorescent protein, it localized to developing MACs where DNA elimination occurs, leading Baudry et al. (2009) to call this gene *PiggyMac* (*PGM*). To further assess the possibility that Pgm participated in DNA elimination, the group knocked down *PGM* levels by feeding *Paramecium* bacterial cells expressing *PGM*-specific dsRNA prior to inducing autogamy. This treatment resulted in a 20-fold reduction in progeny production and thus revealed that the gene product has a critical function during development (Baudry et al. 2009).

If *PGM* actually encodes the enzyme responsible for IES excision, a process that occurs quite late in development, disrupting late-stage expression should be sufficient to perturb production of progeny. To examine the timing of Pgm function, Baudry et al. (2009) exploited the fact that mating cell pairs separate and begin refeeding before the developing MACs carry out DNA elimination (Berger 1973). Feeding cells *PGM* dsRNA early (before conjugation starts) or late (after pairs separated) interfered with normal development. In control experiments, RNAi knockdown of the meiosis-specific Spo11 homolog (Pt*SPO11*) only elicited a phenotype if feeding occurred prior to meiosis. Thus, the critical time window when *PGM* must be active is the period when IES removal from the genome occurs.

Of course, if Pgm is the enzyme that carries out DNA elimination in the *Paramecium* developing MAC, RNAi knockdown should lead to failed excision. To test this possibility, Baudry et al. (2009) used ligation-mediated PCR to assay the ends of a known IES for the appearance of double-strand breaks (DSBs) that normally occur during DNA elimination. In the PGM knockdown cells, they failed to detect breaks that were readily observed in wild-type cells. This result showed that Pgm is required to remove IESs from the genome. All the data taken together strongly support the idea that Pgm is a domesticated transposase that has been recruited to perform programmed DNA elimination.

Not only has the discovery of this domesticated transposase provided needed insight into the mechanism of DNA elimination in *Paramecium*, it has helped clarify some apparently disparate observations reported in previous work. For example, the consensus sequence flanking *Paramecium* IESs is reminiscent of Tc/mariner TIRs; however, the type of DSB characterized at the end of each IES did not match that expected for this family of elements (see Fig. 1A; Gratias and Betermier 2003). The Betermier group (Gratias and Betermier 2003) had shown previously that IESs are removed by a mechanism that generates a 4-base staggered cut, leaving 5′ overhangs; Tc transposases generate 2-base 3′ overhangs (van Luenen et al. 1994). Like the cleavage observed at IES ends, *piggyBac* transposases generate 4-base, 5′ overhangs of the sequence TTAA (Mitra et al. 2008). Note that these overhangs are centered around a TA dinucleotide. The IES TIRs may look like Tc ends, and may even be derived from such transposons, but it is *piggyMac* that is able to recognize and cleave at IES boundaries (Fig. 1).

It is interesting to note that Baudry et al. (2009) report that *piggyBac*-like transposases are encoded in the genome of the related ciliate *Tetrahymena thermophila*. The IESs of this ciliate do not share a common TIR sequence or even possess conserved TA dinucleotides at their ends. In fact, mechanistic studies in *Tetrahymena* revealed that the sequences at the boundaries of IESs were not essential. Instead, sequences located a short distance (45–50 bp) outside the excision boundaries, which are in the DNA that remains in the somatic genome, were shown to determine the boundaries of excision (Godiska et al. 1993; Chalker et al. 1999; Patil and Karrer 2000; Fillingham et al. 2001). These regulatory sequences that specify the excision boundaries share little, if any, similarity with one another, a fact that has always been hard to reconcile with a transposon origin. The one common observation is that both ciliates generated 4-base, staggered cuts upon IES excision. As the Baudry et al. (2009) study suggests, the proteins that excise both *Tetrahymena* and *Paramecium* IESs may share a common *piggyBac* origin. In support of this, a *piggyBac* protein encoded in the *Tetrahymena* genome has been found to be essential for DNA elimination (CY Cheng and MC Yao, pers. comm.).

Going from heterochromatin to DNA elimination: one small step for ciliate-kind?

The discovery of these domesticated *piggyBac* transposases identifies key players in programmed DNA rearrangement that have been missing in the understanding of these events. Recent research has focused on elucidating the role of RNAi-directed heterochromatin formation in DNA elimination (for review, see Yao and Chao 2005; Duharcourt et al. 2009). The exciting discovery that small RNAs, which are homologous to IESs, associate with a Piwi family protein that is essential for DNA elimination in *Tetrahymena* took the emphasis away from the possible role of transposons in mediating their own excision (Mochizuki et al. 2002). These small RNAs were shown to direct heterochromatin-associated modifications, histone H3 Lys 9 and Lys 27 methylation, to the IESs, marking them for elimination (Taverna et al. 2002; Liu et al. 2007). Further characterization of these RNAi-directed mechanisms demonstrated that both *Tetrahymena* and *Paramecium* begin early in their development to decipher what sequences should be eliminated, well before the transposases are expressed. At the start of meiosis, the germline MIC genome is bidirectionally transcribed (Chalker and Yao 2001), and the resulting dsRNA transcripts are processed into 25- to 30-nucleotide (nt) small RNAs (depending on the species) by Dicer-like ribonucleases (Malone et al. 2005; Mochizuki and Gorovsky 2005; Lepere et al. 2008, 2009). Later in development, these meiosis-specific small RNAs move to the developing somatic MACs and, at least in *Tetrahymena*, direct heterochromatin modifications to IESs. It is not clear whether small RNAs direct chromatin modifications

in *Paramecium*, as their histone H3 proteins possess divergent N-terminal tails, making assessment of modifications challenging. These findings suggest that DNA elimination is driven by the host cells trying to silence transposons (and other junk DNA), not simply transposons moving themselves around.

Massive DNA elimination does not seem nearly so unorthodox when one realizes that the mechanisms ciliates use to mark sequences for removal are analogous to those that many organisms use to silence transposons in the germline and elsewhere. Piwi-associated RNAs (piRNAs) of the fruit fly *Drosophila melanogaster* represent a well-studied case of small RNAs acting in germ cells to keep transposons in check (for review, see Saito et al. 2006; Vagin et al. 2006; Malone and Hannon 2009). Plants also use their RNAi machinery to silence transposons in pollen (Slotkin et al. 2009) and somatic tissues (for review, see Matzke et al. 2009). These mechanisms are obviously quite ancient, as they are found in animals, plants, and protists. The discovery of a domesticated transposase as the mediator of wholesale DNA elimination of the sequences recognized by the RNAi machinery provides a vision of how such a phenomenon might evolve.

With the identification of Pgm as an essential player in DNA elimination, ideas as to the origins and excision mechanisms of IESs have come full circle. IESs were often postulated to be degenerate transposons, and this is still the most likely scenario (Fig. 1B). The ancestral cells would be expected to be equipped with an RNAi mechanism to silence these invaders and limit their spread. Along the way, an invading *piggyBac* element in the genome lost its TIRs and could no longer mobilize itself. This change or other alterations allowed it to escape silencing, and it eventually became a fixture within euchromatin (Fig. 1B). At this point, it may still be able to mobilize other elements in *trans*, which could have led to the large expansion of IESs. If this semidomesticated transposase was responsible for spreading sequences around the genome, it should also retain the specificity to remove them, as *piggyBac* elements are cut-and-paste transposons.

It is not hard to imagine how this transposase could become an "excisase" favoring DNA elimination over mobilization. Alteration of the enzyme's affinity for TIRs could decrease its ability to both cut and paste. To compensate for this decreased ability to recognize TIRs of the future IESs, the protein would need to acquire the ability to interact with the silent chromatin associated with these undesirable sequences. Pgm has a C-terminal extension that Baudry et al. (2009) speculate could give it the ability to interact with chromatin modified by the RNAi machinery. The *Tetrahymena* protein also has an extended C terminus, which could suggest that that addition occurred in a common ancestor. This adaptation could also allow the transposase to recognize a larger variety of IES end sequences, as protein–protein interactions would be the main force recruiting the excisase to its sites of action.

It is somewhat paradoxical that *Paramecium* has recruited a transposase to eliminate transposons and other extra DNA from its somatic genome. Originally, *PGM* could have been considered part of the junk DNA, but it is junk no longer. It has taken on an essential role in the major events shaping the somatic genome during its differentiation from the germline. The possibility that a transposase is being used to remove undesirable sequences like other transposons should not be overtly surprising, as several examples of transposon protein domestication have been described (for review, see Volff 2006). Perhaps the most well known is the RAG1/RAG2 recombinase of jawed vertebrates that initiates VDJ recombination in B and T cells (Agrawal et al. 1998; Hiom et al. 1998). Pgm is just the most recent example of the hijacking of a transposase to carry out programmed DNA rearrangement. However, domesticated transposases are by no means limited to acting as transposases as they did in their undomesticated state. For example, a likely former hAT transposase, Daysleeper, acts as a transcriptional activator in *Arabidopsis thaliana* (Bundock and Hooykaas 2005). Furthermore, the former pogo-like transposase CenP-B is an important centromeric protein conserved in diverse organisms, from fission yeast to mammals (Smit and Riggs 1996; Casola et al. 2008).

Given that ciliates go to a lot of work to eliminate transposons and other nonessential DNA from their somatic genome, they must gain something from it. One simple gain is the removal of the transposons from the transcriptionally active nucleus. This should be an effective way to limit their spread and prevent possible deleterious mutations that might arise due to transposition. Possibly a more important force derives from the polyploid state of the somatic genome. By having polyploid MACs, ciliates possess a potential for high-level gene expression that comes with having tens to thousands of copies of each gene. Thus, these organisms are able to obtain a relatively large size for single cells that in turn allows them to have rather big mouths, thus increasing their range of food options. The high level of gene expression still allows them to divide quickly relative to their size, which is important to survive as a species when lots of other creatures see them as lunch as they are scavenging for theirs. By streamlining their polyploidy somatic genome, they have that much less DNA to replicate and can divide that much faster.

This discovery of *piggyMac* provides an interesting glimpse into the evolution of genome-wide DNA elimination. Some common ancestor of *Paramecium* and *Tetrahymena* was likely less drastic in its dealing with the junk in its genome, simply packaging it up into heterochromatin as most eukaryotes do. The maintenance of a separate germline genome and the domestication of a transposase allowed this silenced DNA to be discarded from the soma. All of the ciliates studied exhibit some degree of programmed DNA elimination, so it would be reasonable to assume that the DNA elimination mechanism arose early in this lineage. Even if that is the case, it has been shown recently in *Oxytricha trifallax* that the non-*piggyBac* TBE transposase is necessary for the elimination of the several thousand TBE elements and at least some IESs (Nowacki et al. 2009).

Thus, there have been at least two separate transposon domestication events associated with DNA elimination in ciliates. Extending this observation beyond ciliates, one cannot help but wonder whether other chromatin diminution events (e.g., the fragmentation of chromosomes in *Parascaris* observed by Boveri [1887]) are the result of domesticated transposons acting on silent heterochromatin. Once cells have a means (e.g., RNAi) to recognize the foreign DNA in their genome, picking up a transposase is a handy way to eliminate them altogether.

Acknowledgments

We thank Dr. D. Frank for comments on the manuscript. Both J.A.M. and D.L.C. are supported by NIH grant GM069593 and NSF grant MCB-0642162.

References

Agrawal A, Eastman QM, Schatz DG. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394:** 744–751.

Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Bétermier M. 2009. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes & Dev* (this issue). doi: 10.1101/gad.547309.

Beerman S. 1977. The diminution of heterochromatic chromosomal segments in Cyclops (*Crustacea, Copepoda*). *Chromosoma* **60:** 297–344.

Berger JD. 1973. Nuclear differentiation and nucleic acid synthesis in well-fed exconjugants of *Paramecium aurelia*. *Chromosoma* **42:** 247–268.

Betermier M. 2004. Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res Microbiol* **155:** 399–408.

Boveri T. 1887. Uber Differenzierung der Zellkerne wahrend der Furchung des Eies von *Ascaris megalocephala*. *Anat Anz* **2:** 688–693.

Bundock P, Hooykaas P. 2005. An *Arabidopsis* hAT-like transposase is essential for plant development. *Nature* **436:** 282–284.

Casola C, Hucks D, Feschotte C. 2008. Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol Biol Evol* **25:** 29–41.

Chalker DL, Yao MC. 2001. Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. *Genes & Dev* **15:** 1287–1298.

Chalker D, La Terza A, Wilson A, Kroenke C, Yao M. 1999. Flanking regulatory sequences of the *Tetrahymena* R deletion element determine the boundaries of DNA rearrangement. *Mol Cell Biol* **19:** 5631–5641.

Duharcourt S, Lepere G, Meyer E. 2009. Developmental genome rearrangements in ciliates: A natural genomic subtraction mediated by non-coding transcripts. *Trends Genet* **25:** 344–350.

Fillingham JS, Bruno D, Pearlman RE. 2001. *Cis*-acting requirements in flanking DNA for the programmed elimination of mse2.9: A common mechanism for deletion of internal eliminated sequences from the developing macronucleus of *Tetrahymena thermophila*. *Nucleic Acids Res* **29:** 488–498.

Goday C, Esteban MR. 2001. Chromosome elimination in sciarid flies. *Bioessays* **23:** 242–250.

Goday C, Pimpinelli S. 1984. Chromosome organization and heterochromatin elimination in *Parascaris*. *Science* **224:** 411–413.

Godiska R, James C, Yao MC. 1993. A distant 10-bp sequence specifies the boundaries of a programmed DNA deletion in *Tetrahymena*. *Genes & Dev* **7:** 2357–2365.

Gratias A, Betermier M. 2003. Processing of double-strand breaks is involved in the precise excision of paramecium internal eliminated sequences. *Mol Cell Biol* **23:** 7152–7162.

Hiom K, Melek M, Gellart M. 1998. DNA transposition by the RAG1 and RAG2 proteins: A possible source of oncogenic translocations. *Cell* **94:** 463–470.

Klobutcher LA, Herrick G. 1995. Consensus inverted terminal repeat sequence of *Paramecium* IESs: Resemblance to termini of Tc1-related and *Euplotes* Tec transposons. *Nucleic Acids Res* **23:** 2006–2013.

Klobutcher LA, Herrick G. 1997. Developmental genome reorganization in ciliated protozoa: The transposon link. *Prog Nucleic Acid Res Mol Biol* **56:** 1–62.

Kubota S, Kuro-O M, Mizuno S, Kohno S. 1993. Germ-line restricted, highly repeated DNA sequences and their chromosomal loacalization in a japanese hagfish (*Eptatretus okinoseanus*). *Chromosoma* **102:** 163–173.

Le Mouel A, Butler A, Caron F, Meyer E. 2003. Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. *Eukaryot Cell* **2:** 1076–1090.

Lepere G, Betermier M, Meyer E, Duharcourt S. 2008. Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes & Dev* **22:** 1501–1512.

Lepere G, Nowacki M, Serrano V, Gout JF, Guglielmi G, Duharcourt S, Meyer E. 2009. Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res* **37:** 903–915.

Liu Y, Taverna SD, Muratore TL, Shabanowitz J, Hunt DF, Allis CD. 2007. RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in *Tetrahymena*. *Genes & Dev* **21:** 1530–1545.

Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136:** 656–668.

Malone CD, Anderson AM, Motl JA, Rexer CH, Chalker DL. 2005. Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of *Tetrahymena thermophila*. *Mol Cell Biol* **25:** 9151–9164.

Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ. 2009. RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol* **21:** 367–376.

Meyer E, Chalker DL. 2007. Epigenetics of ciliates. In *Epigenetics* (eds. CD Allis et al.), pp. 127–150. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mitra R, Fain-Thornton J, Craig NL. 2008. piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J* **27:** 1097–1109.

Mochizuki K, Gorovsky MA. 2005. A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes & Dev* **19:** 77–89.

Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA. 2002. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell* **110:** 689–699.

Muller F, Tobler H. 2000. Chromatin diminution in the parasitic nematodes *Ascaris suum* and *Parascaris univalens*. *Int J Parasitol* **30:** 391–399.

Muller F, Bernard V, Tobler H. 1996. Chromatin diminution in nematodes. *Bioessays* **18:** 133–138.
Nowacki M, Higgins BP, Maquilan GM, Swart EC, Doak TG, Landweber LF. 2009. A functional role for transposases in a large eukaryotic genome. *Science* **324:** 935–938.
Patil N, Karrer K. 2000. A developmentally regulated deletion element with long terminal repeats has *cis*-acting sequences in the flanking DNA. *Nucleic Acids Res* **28:** 1465–1472.
Prescott DM. 1994. The DNA of ciliated protozoa. *Microbiol Rev* **58:** 233–267.
Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes & Dev* **20:** 2214–2222.
Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136:** 461–472.
Smit AF, Riggs AD. 1996. Tiggers and DNA transposon fossils in the human genome. *Proc Natl Acad Sci* **93:** 1443–1448.
Smith JJ, Antonacci F, Eichler EE, Amemiya CT. 2009. Programmed loss of millions of base pairs from a vertebrate genome. *Proc Natl Acad Sci* **106:** 11212–11217.
Taverna SD, Coyne RS, Allis CD. 2002. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. *Cell* **110:** 701–711.
Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313:** 320–324.
van Luenen HG, Colloms SD, Plasterk RH. 1994. The mechanism of transposition of Tc3 in *C. elegans*. *Cell* **79:** 293–301.
Volff JN. 2006. Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* **28:** 913–922.
Yao MC, Chao J. 2005. RNA-guided DNA deletion in *Tetrahymena*: An RNAi-based mechanism for programmed genome rearrangements. *Annu Rev Genet* **39:** 537–559.