

Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in *Tetrahymena*

Ursula E. Schoeberl,¹ Henriette M. Kurth,^{1,2} Tomoko Noto, and Kazufumi Mochizuki³

Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), A-1030 Vienna, Austria

The ciliated protozoan *Tetrahymena* undergoes extensive programmed DNA elimination when the germline micronucleus produces the new macronucleus during sexual reproduction. DNA elimination is epigenetically controlled by DNA sequences of the parental macronuclear genome, and this epigenetic regulation is mediated by small RNAs (scan RNAs [scnRNAs]) of ~28–30 nucleotides that are produced and function by an RNAi-related mechanism. Here, we examine scnRNA production and turnover by deep sequencing. scnRNAs are produced exclusively from the micronucleus and nonhomogeneously from a variety of chromosomal locations. scnRNAs are preferentially derived from the eliminated sequences, and this preference is mainly determined at the level of transcription. Despite this bias, a significant fraction of scnRNAs is also derived from the macronuclear-destined sequences, and these scnRNAs are degraded during the course of sexual reproduction. These results indicate that the pattern of DNA elimination in the new macronucleus is shaped by the biased transcription in the micronucleus and the selective degradation of scnRNAs in the parental macronucleus.

[Keywords: DNA elimination; *Tetrahymena*; epigenetics; small RNA]

Supplemental material is available for this article.

Received May 16, 2012; revised version accepted June 21, 2012.

The phenotypes of a parent can be epigenetically transmitted to offspring through the germline in many different eukaryotes. Epigenetic information can be inherited as chromatin modifications that are incompletely reprogrammed in gametes and thereby transmitted to descendants. For example, flower shape in plants and coat color in mice can be transmitted transgenerationally by variations in DNA methylation patterns (Cubas et al. 1999; Morgan et al. 1999). Alternatively, some diffusible factors can transmit epigenetic information from parents to progeny through the germline, such as the inheritance of transposon silencing states by maternally supplied Piwi-associated small RNAs in fruit flies (Brennecke et al. 2008). Transgenerational epigenetic inheritance potentially allows organisms to transmit acquired characteristics to the next generation and thus could increase the fitness of progeny in rapidly changing environments. However, the detailed molecular mechanism of transgenerational epigenetic inheritance is largely unclear. Several lines

of evidence suggest that the patterns of programmed DNA elimination in ciliated protozoans are epigenetically and transgenerationally inherited (for review, see Mochizuki 2010). Therefore, DNA elimination in ciliated protozoans serves as a model to understand how transgenerational epigenetic inheritance can be achieved at a molecular level.

Like most ciliated protozoans, *Tetrahymena thermophila* exhibits nuclear dimorphism. Each cell contains a germline micronucleus (Mic) and a somatic macronucleus (Mac) (Fig. 1). Most, if not all, gene expression occurs in the Mac. In the sexual process of conjugation (Fig. 1A–G), the Mic undergoes meiosis (Fig. 1A), and one of the four meiotic products is selected and mitotically divides to form two haploid pronuclei (Fig. 1B). One of the pronuclei is reciprocally exchanged between the two conjugating cells (Fig. 1C). The migratory and stationary pronuclei then fuse to create a zygotic diploid nucleus (Fig. 1D) that divides twice mitotically (Fig. 1E) to produce the next generation of new Macs and Mics (Fig. 1F). The parental Mac is then destroyed (Fig. 1G).

Programmed DNA elimination occurs during the development of the new Mac. Previous studies have estimated that ~6000 internal eliminated sequences (IESs) are removed from the new Mac (for review, see Yao et al.

¹These authors contributed equally to this work.

²Present address: Viollier AG, Spalenring 145/147, 4002 Basel, Switzerland.

³Corresponding author

E-mail kazufumi.mochizuki@imba.oew.ac.at

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.196493.112>.

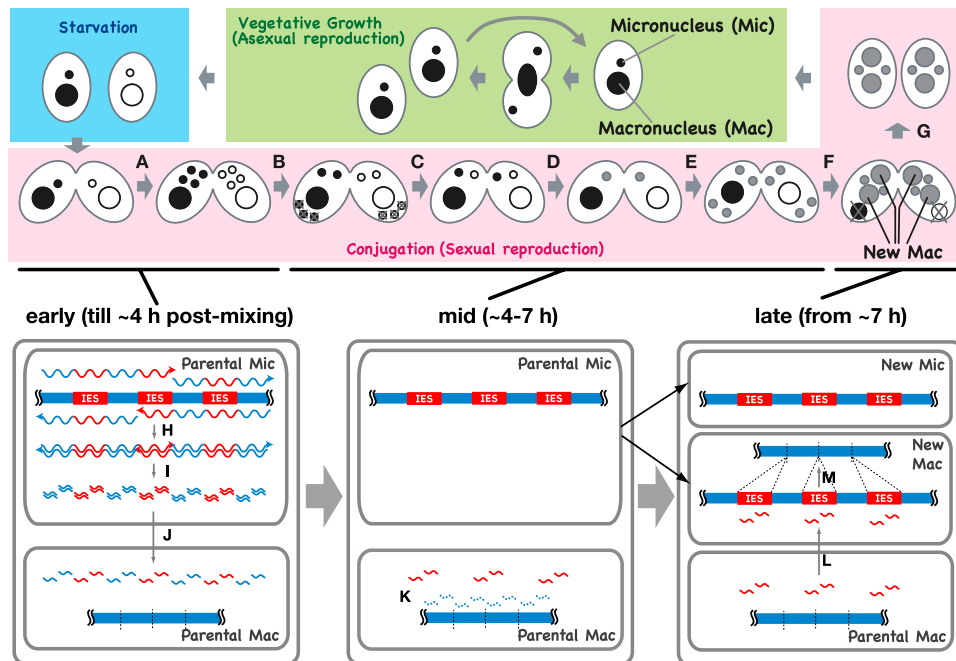


Figure 1. Life cycle of *Tetrahymena* and a model for scnRNA-directed DNA elimination. A single cell of *T. thermophila* contains two different types of nuclei: a Mac and a Mic. When sufficient nutrients are available, *Tetrahymena* grows by binary fission, and the Mac and the Mic are divided independently (vegetative growth, green background). After prolonged starvation (blue background), two cells of complementary mating types fuse to begin the sexual reproduction process (conjugation, pink background). Their Mics undergo meiosis (A), and one of the meiotic products survives and divides mitotically, giving rise to two pronuclei, one stationary and one migratory (B). The migratory gametic nucleus crosses the conjugation bridge (C) and fuses with the stationary nucleus to produce the zygotic nucleus (D). The zygotic nucleus divides twice (E), and two of the products differentiate as Macs, while the other two differentiate as Mics (F). The parental Mac is degraded, and the pair is dissolved (G). (Bottom left) In the early conjugation stages, the Mic genome, including IESs (red boxes) and MDSs (blue bars), is transcribed bidirectionally, and these transcripts (wavy arrows) form dsRNA (H). The dsRNAs are processed into small RNAs (scnRNAs) (I), and the scnRNAs are transferred to the parental Mac (J). (Bottom middle) In the mid-conjugation stages, all scnRNAs complementary to DNA in the parental Mac are degraded (K). (Bottom right) In the late conjugation stages, the scnRNAs that are not degraded in the parental Mac (those complementary to IESs) are transferred to the developing new Mac (L), where they target IESs for elimination by base pairing (M).

2002). The Mic sequences that remain in the future Mac are called Mac-destined sequences (MDSs). IESs in *Tetrahymena* vary in size (~0.5–20 kb) and sequence, and recent Mic genome sequencing indicates that they account for about one-third of the Mic genome (*Tetrahymena* Comparative Sequencing Project, Broad Institute of Harvard and Massachusetts Institute of Technology [http://www.broadinstitute.org]). Some IESs are moderately repeated in the Mic genome (Yao and Gorovsky 1974) and are related to transposable elements (Wuitschick et al. 2002; Fillingham et al. 2004). IESs are primarily located in nongenic regions, although some are located in genic regions (Fass et al. 2011). The excision of IESs can occur reproducibly at a specific site or with a limited number of alternative boundaries (Austerberry et al. 1989; Fass et al. 2011). However, despite this precision, the sites of DNA elimination do not seem to be specified by the primary DNA sequences of the Mic genome because no obvious consensus sequence has been identified in or around IESs, and the *in vitro* DNA specificity of the enzyme responsible for the DNA excision process is ambiguous (Cheng et al. 2010). Instead, DNA elimination in *Tetrahymena* and some other ciliates is epigenetically regulated by the

DNA sequence of the parental Mac such that the elimination of a new Mac DNA is inhibited by the presence of the same sequence in the parental Mac (Koizumi and Kobayashi 1989; You et al. 1991; Duhaucourt et al. 1995; Chalker and Yao 1996). Because the chromosomes of the parental and the new Macs do not directly interact, some sequence-specific diffusible signal must be transmitted from the parental to the new Macs through the cytoplasm to mediate the epigenetic effect (Chalker et al. 2005).

The injection of dsRNAs complementary to a MDS locus into an early conjugating cell induces ectopic DNA elimination at the MDS locus (Yao et al. 2003), indicating that an RNAi-related mechanism plays a role in DNA elimination. Indeed, the Dicer homolog Dcl1p and the Argonaute protein Twi1p are essential for DNA elimination (Mochizuki et al. 2002; Malone et al. 2005). Bidirectional Mic noncoding transcripts (Chalker and Yao 2001), which are probably transcribed by RNA polymerase II (Mochizuki and Gorovsky 2004b), are processed by Dcl1p (Malone et al. 2005; Mochizuki and Gorovsky 2005) to small (~28- to 30-nucleotide [nt]) RNAs, named scan RNAs (scnRNAs). scnRNAs are then complexed with Twi1p (Mochizuki and Gorovsky 2004a). The fact that

the Twilp–scnRNA complex first localizes to the parental Mac and then to the new Mac (Mochizuki et al. 2002; Noto et al. 2010) makes scnRNA a strong candidate to be the mediator of the epigenetic signal.

Consistent with the production of scnRNAs from the Mic, scnRNAs complementary to MDSs and those complementary to IESs have been detected in the early stages of conjugation. However, scnRNAs complementary to IESs become enriched as conjugation proceeds (Mochizuki and Gorovsky 2004a; Aronica et al. 2008). Moreover, most of the scnRNAs cloned from the late stage of conjugation are complementary to IESs (Lee and Collins 2006). Thus, scnRNAs are selected for IES specificity after they are diced. Based on these and other observations, a model has been proposed to explain how scnRNAs mediate the epigenetic regulation of DNA elimination (Fig. 1H–M). In the model, scnRNAs are produced from the entire Mic genome, including both IESs and MDSs, in the early stages of conjugation (Fig. 1H,I), and scnRNAs that share homology with any Mac genome sequences (MDSs) are degraded in the parental Mac in the middle stages of conjugation (Fig. 1J,K). This process is called scnRNA selection. It has been further proposed that these IES-specific scnRNAs move to the developing Mac in the late stages of conjugation (Fig. 1L), where they identify homologous sequences as IESs and target them for heterochromatin formation (Coyne et al. 1999; Taverna et al. 2002; Liu et al. 2007), followed by DNA elimination (Fig. 1M).

Three issues with this proposed model have not been resolved: (1) Are scnRNAs produced exclusively from the Mic? (2) Are scnRNAs produced homogeneously from both IESs and MDSs at a local level and from the entire Mic genome at a global level? (3) Does scnRNA selection occur at a genome-wide level? To address these points, we analyzed the production and turnover of scnRNAs by deep sequencing.

Results

scnRNAs are predominantly produced from the Mic

It has been proposed that scnRNAs are produced exclusively from the Mic because some fraction of scnRNAs is complementary to the Mic-limited sequences (Mochizuki et al. 2002; Mochizuki and Gorovsky 2004a; Aronica et al. 2008), and the Dicer homolog Dcl1p, which is responsible for the production of scnRNAs, is detected only in the Mic (Malone et al. 2005; Mochizuki and Gorovsky 2005). However, another fraction of scnRNAs is complementary to MDSs (Mochizuki et al. 2002; Mochizuki and Gorovsky 2004a; Aronica et al. 2008), which are shared in the Mic and the Mac; therefore, the production of scnRNAs from the Mac has not been completely excluded. To determine whether the Mac produces scnRNAs, we compared scnRNAs from wild-type cells with those from chromosome 4 (Chr4)–nullisomic (Nulli4) strains. In Nulli4 strains, Chr4 is completely missing in the Mic, while the Mac chromosomes derived from Chr4 are maintained (Bruns et al. 1983). Therefore, if any Chr4-derived scnRNA

is detected in Nulli4 strains, it must be transcribed from the Mac genome.

Total RNA was extracted from wild-type and Nulli4 strains at 3 h post-mixing, when the production of scnRNAs is expected to be almost complete but scnRNA selection is expected to have just initiated (see below). The RNA was gel-fractionated, and a small RNA population, including scnRNAs (26–32 nt), was used to construct the cDNA library for deep sequencing. To determine from which Mic chromosome each scnRNA was derived, we chose two ~100-kb Mic genome loci for which chromosomal origins have been identified. One, named the Cal locus, contains the calmodulin gene and 12 IESs, including the well-studied Cal-IES element (Katoh et al. 1993); the other, named the LMR locus, contains 13 IESs, including the well-studied L-IES, M-IES, and R-IES elements (Austerberry et al. 1984). The Cal and LMR loci are on Chr2 and Chr4, respectively (E Hamilton and E Orias, pers. comm.).

The numbers of scnRNAs that uniquely mapped to the loci (unique mappers) were calculated. The first 25-nt sequences of scnRNAs (26–32 nt) that did not match to rRNAs, tRNAs, snRNAs, and the mitochondrial genome (filtered reads) (Supplemental Table S1) were used for mapping. The number of scnRNAs from wild-type and Nulli4 cells that mapped to the Cal locus were similar (Fig. 2A,B). In contrast, the number of scnRNAs that mapped to the LMR locus was significantly lower in Nulli4 strains (Fig. 2D) than in wild-type strains (Fig. 2C). We believe that the small number of scnRNAs from Nulli4 strains that mapped to the LMR locus were transcribed from non-Chr4 Mic chromosomes because some repeated sequences likely failed to be assembled into the current draft Mic genome, and a small proportion of repeated sequences that were present on both Chr4 and a different chromosome could have been falsely recognized as unique to Chr4 in this study. Consistent with this view, most of the scnRNAs from Nulli4 strains that mapped to the LMR locus were complementary to IESs (Fig. 2D, IESs are marked in red) that did not exist in Nulli4 strains. Therefore, we believe that Nulli4 strains do not produce scnRNAs from the LMR locus and conclude that scnRNAs are produced predominantly, if not exclusively, from the Mic.

scnRNAs are preferentially produced from IESs

The current model of scnRNA-directed DNA elimination (Mochizuki and Gorovsky 2004c; Aronica et al. 2008) suggests that scnRNAs are produced homogeneously from both IESs and MDSs in the Mic genome. To determine whether this occurs at the genome-wide level, we first annotated the draft Mic genome. All possible 25-nt sequences from the draft Mic genome were sorted to three classes: sequences found only in MDSs, sequences found only in IESs, or sequences found in both MDSs and IESs (see the Materials and Methods for details), which correspond to 98.8 Mb, 33.1 Mb, and 3.0 Mb of genomic sequences, respectively (Fig. 3A, left). We used the sum of these classified sequences as the annotated Mic genome

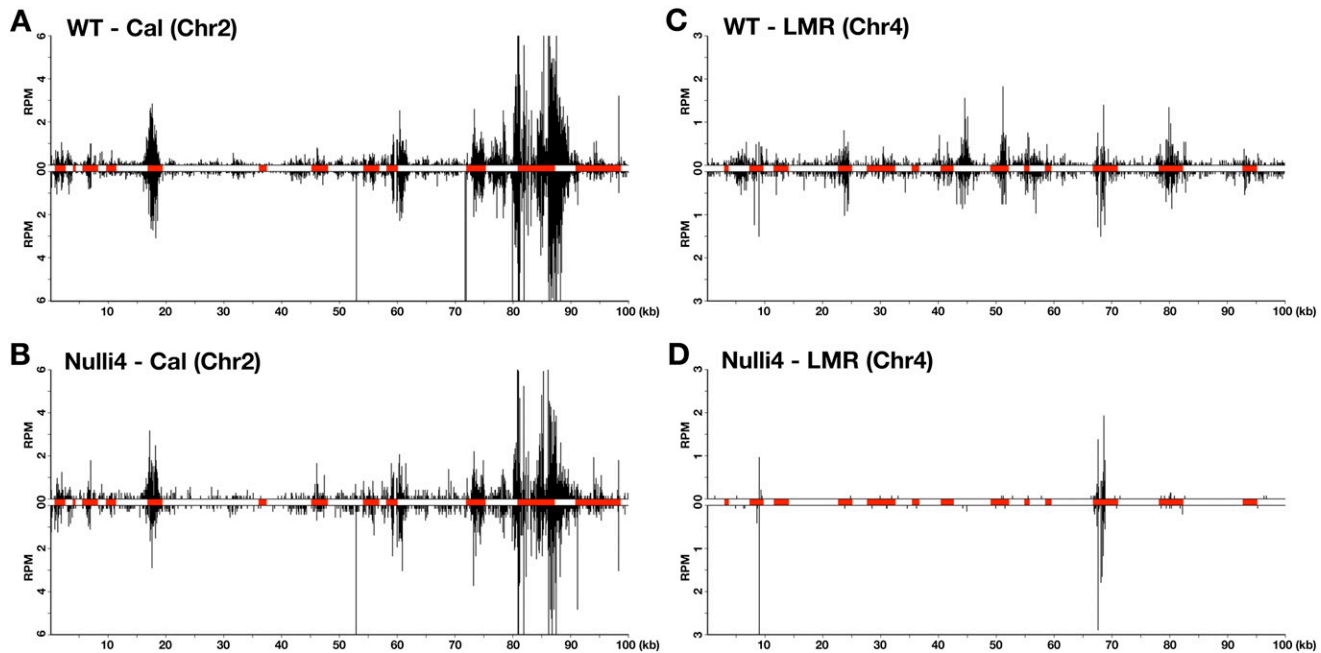


Figure 2. Analysis of scnRNAs from wild-type and Nulli4 strains. The Cal locus on the Mic Chr2 (A,B) or the LMR locus on the Mic Chr4 (C,D) was mapped by the scnRNA sequences obtained from wild-type (WT) (A,C) or Nulli4 (B,D) cells. The number (reads per million scnRNAs [RPM]) of scnRNAs uniquely mapping to the loci are shown as histograms with 100-nt bins. The bars toward the plus and minus directions indicate the numbers of scnRNAs matching the sense and antisense strands of the loci, respectively. The IESs are marked in red.

in this study. To analyze the original scnRNA population produced by Dcl1p in the Mic, we used *TWI1* knockout cells because the scnRNA selection process (Fig. 1K; see also below) that occurs after binding of scnRNAs to Twi1p should not occur in the absence of Twi1p.

The first 25-nt sequences of scnRNAs (26–32 nt) from *TWI1* knockout cells at 3 h post-mixing were mapped to the annotated Mic genome. Among the scnRNAs from *TWI1* knockout cells, 67.0% mapped to the annotated Mic genome without any mismatch (Supplemental Table S1). Of these matched scnRNAs, only 15.3% were exclusively complementary to MDSs, while 80.1% mapped exclusively to IESs (Fig. 3A, right; Supplemental Table S1). This distribution differed dramatically from the profile of the annotated genome sequence, which included 73.2% MDSs and 24.6% IESs (Fig. 3A, left). Therefore, in contrast to expectations, the production of scnRNAs is heavily biased toward IESs.

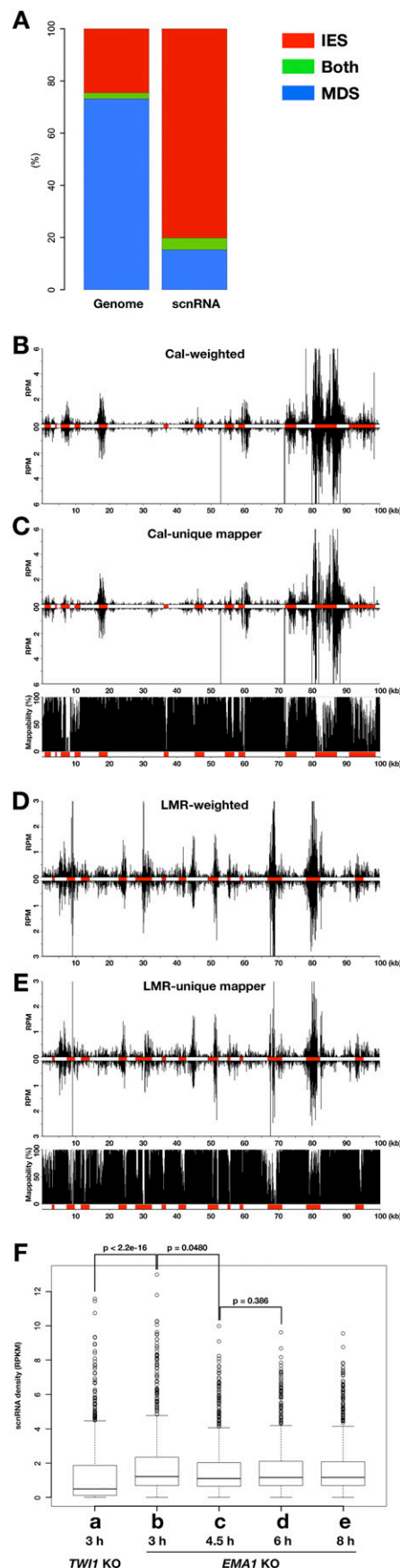
A similar conclusion was obtained by mapping scnRNAs to the Cal and LMR loci. The numbers of scnRNAs mapping to each position of these loci were calculated, and the repetitiveness of the sequences in the Mic genome was normalized by either dividing the numbers of mapped scnRNAs by their sequence frequencies in the draft Mic genome (Fig. 3B,D, “weighted”) or excluding scnRNAs that were complementary to more than one position in the draft Mic genome (Fig. 3C,E, “uniquely mapped”). Both of the analyses demonstrated that the majority of scnRNAs were complementary to IESs and that the other scnRNAs were complementary to MDSs (Fig. 3B–E, IESs are marked in red). These results further

indicate that scnRNAs are preferentially produced from IESs rather than MDSs.

scnRNAs are produced from a variety of Mic genome loci at various levels

It has been proposed that scnRNAs are produced not from specific chromosomal regions, but rather from the entire Mic genome (Mochizuki et al. 2002; Mochizuki and Gorovsky 2004c). As shown above, this is not the case at a local level because the scnRNA production is biased to IESs (Fig. 3A–E). Therefore, we next analyzed the origin of scnRNAs at a global level. We mapped scnRNA sequences to MDSs to analyze scnRNA production from different chromosomal locations without the effect of varied IES densities at different loci (Fass et al. 2011). All Mac genome scaffolds longer than 100 kb, which account for ~60% of the sequenced Mic genome in total, were divided into 100- to 200-kb fragments and used as representative MDS fragments.

We first analyzed scnRNAs from *TWI1* knockout cells at 3 h post-mixing. The density (reads per kilobase per million scnRNAs, RPKM) of scnRNAs that uniquely mapped to each of the MDS fragments was analyzed. Because the total length of the mappable sequences in the MDS fragments was 87.3 Mb, and 2.16 million of the 18.4 million scnRNA reads uniquely mapped to these fragments, the density was expected to be 1.34 RPKM if the scnRNA sequences mapped homogeneously to the genome. However, the densities of the mapped scnRNAs differed dramatically between the MDS fragments (Fig. 3F, column a), with a median density of 0.506 RPKM.



We hypothesized that the observed variability in the scnRNAs from *TWI1* knockout cells that mapped to the MDS fragments could account for the timing but not the levels of scnRNA production from different chromosomal locations. This is because scnRNAs are unstable in the absence of Twilp (Mochizuki and Gorovsky 2004a), and thus only scnRNAs produced immediately before harvesting the cells were likely sequenced from *TWI1* knockout cells. In wild-type cells, scnRNAs are stabilized in a complex with Twilp (Mochizuki and Gorovsky 2004a). On the other hand, in wild-type cells, Twilp transports scnRNAs to the parental Mac, where scnRNAs complementary to the Mac genome are degraded (Fig. 1K; Aronica et al. 2008). The Twilp-associated RNA helicase Emalp is required for this degradation in the parental Mac but not for the formation of Twilp-scnRNA complexes, which stabilize scnRNAs (Supplemental Fig. S3; Aronica et al. 2008).

Therefore, to circumvent scnRNA degradation and determine the level of scnRNA production from each locus more precisely, scnRNA sequences obtained from *EMA1* knockout cells at early to late stages of conjugation were mapped to the MDS fragments (Fig. 3F, columns b–e). Although many of the MDS fragments were sparsely (<1 RPKM) mapped by scnRNA sequences from *TWI1* knockout cells (Fig. 3F, column a), this fraction was greatly

Figure 3. Production of scnRNAs from the Mic genome. (A) The draft Mic genome was annotated as three sequence classes: sequences that appear only in MDSs (blue), sequences that appear in both MDSs and IESs (green), and sequences that appear only in IESs (red). The percent occupancy of genomic sequences in each class is shown in the *left* column (Genome). The fraction of the sequenced scnRNAs from *TWI1* knockout cells at 3 h post-mixing that is complementary to each of the classes of the annotated Mic genome is shown in the *right* column (scnRNA). (B–E) The Cal (B,C) and LMR (D,E) loci are mapped by the sequenced scnRNAs from *TWI1* knockout cells at 3 h post-mixing. For B and D, the number (reads per million scnRNAs [RPM]) of scnRNAs mapping to each position of the loci was divided by its sequence frequency in the draft Mic genome (“weighted”). For C and E, only the number of scnRNAs that mapped to a single location of the draft Mic genome is shown (“uniquely mapped”). In both cases, the numbers of scnRNAs matching the sense and antisense strands are shown in the plus and minus directions, respectively, as histograms with 100-nt bins. (C,E, bottom) The fraction of unique sequences that can be mapped by scnRNAs by the latter method in each 100-nt bin is shown as “Mappability”. The positions of IESs are marked in red. (F) The number of scnRNA sequences that uniquely matched each of the 100- to 200-kb MDS fragments was divided by the total length of the unique sequences of the MDS fragments (scnRNA density, RPKM [reads per kilobase of unique sequences per million] scnRNAs) and is shown as box plots. scnRNAs from *TWI1* knockout cells at 3 h post-mixing and from *EMA1* knockout cells at 3, 4.5, 6, or 8 h post-mixing were analyzed. The box indicates the interquartile range (IQR). The horizontal bar in the box indicates the median value. The bars on the *top* and *bottom* of the box indicate the minimum and maximum values, respectively, within $1.5 \times$ IQR. Circles indicate outliers. P-values were calculated using the Wilcoxon rank-sum test.

reduced when the MDS fragments were mapped by scnRNA sequences from *EMA1* knockout cells at 3 h post-mixing (Fig. 3F, column b). This result indicates that some population of scnRNAs that was derived from specific chromosomal locations existed in *EMA1* knockout cells but was missing in *TWI1* knockout cells. The variability of the mapped scnRNA density on different MDS fragments was smaller at 4.5 h post-mixing (Fig. 3F, column c) than at 3 h post-mixing (Fig. 3F, column b) in *EMA1* knockout cells but did not significantly change further at later stages (Fig. 3F, columns d,e). These results indicate that scnRNA production from different chromosomal locations indeed occurs at distinct conjugation stages before 3 h post-mixing and between 3 and 4.5 h post-mixing.

This different timing of scnRNA production from distinct chromosomal locations partially explains the variable densities of *TWI1* knockout scnRNAs mapped to different MDS fragments (Fig. 3F, column a). However, different MDS fragments were variably mapped even by the scnRNAs from the *EMA1* knockout cells at late stages of conjugation (Fig. 3F, columns d,e), when scnRNA production was expected to be completed, indicating that, although scnRNA production occurs from a variety of chromosomal locations in the Mic, it is not homogeneous at either the local (i.e., IES-biased) (Fig. 3A–E) or global (Fig. 3F) level.

Sequence repetitiveness in the genome does not fully explain IES-biased scnRNA production

The IES-biased production of scnRNAs may be caused by preferential production of scnRNAs from repeated sequences in the genome because some IESs are related to transposon-derived sequences and are repeated in the Mic genome (Yao and Gorovsky 1974; Wuitschick et al. 2002; Fillingham et al. 2004), and scnRNAs are processed from dsRNA precursors (Malone et al. 2005; Mochizuki and Gorovsky 2005; Noto et al. 2010), which can be more frequently formed by hybridization of transcripts from repeated sequences. To evaluate this possibility, we classified the annotated Mic genome sequences into three classes according to their repetitiveness. All possible 25-nt sequences from the draft Mic genome were sorted to three classes: unique (“Unique” in Fig. 4A), moderately repeated (two to 10 times in the genome) (“Mid-Rep” in Fig. 4A), and highly repeated (>10 times, in the genome) (“High-Rep” in Fig. 4A), which comprise 86.9%, 10.0%, and 3.1% of the annotated Mic genome, respectively (Fig. 4A). The first 25-nt sequences of scnRNAs were then mapped to the sequences in each class.

Among the scnRNAs from *TWI1* knockout cells at 3 h post-mixing, 52.2%, 29.1%, and 18.6% were complementary to Unique, Mid-Rep, and High-Rep sequences, respectively (Fig. 4E), indicating that scnRNAs are indeed produced from repeated sequences more frequently (Fig. 4, cf. A and E). However, because half of the scnRNAs were derived from unique sequences, repetitiveness in the genome seemed not necessary for scnRNA production. To determine whether the repeat bias explains the IES-biased

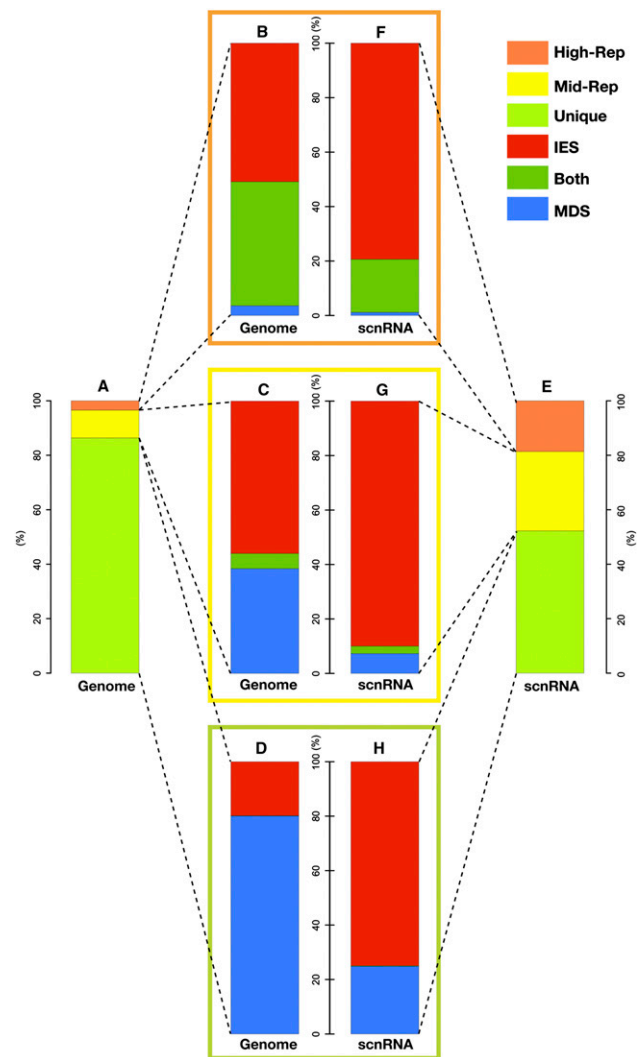


Figure 4. Production of scnRNAs from unique and repetitive sequences. (A) The draft Mic genome sequences were classified among three categories: unique sequences (Unique [light green]), moderately repeated sequences (repeated two to nine times in the draft genome, Mid-Rep [yellow]), and highly repeated sequences (repeated >10 times in the draft genome, High-Rep [orange]). The percent occupancy of the sequences in each category is shown. (B–D) The sequences in each category in A were subdivided into the sequences annotated as MDSs (blue), IESs (red), and both MDSs and IESs (dark green). (E–H) The fraction of scnRNA sequences from *TWI1* knockout cells at 3 h post-mixing that belongs to each sequence category is shown.

production of scnRNAs, we separately analyzed the IES and MDS biases of scnRNA production from the three repetitiveness classes (Unique, Mid-Rep, and High-Rep) (Fig. 4B–H). scnRNAs complementary to all three repetitiveness classes were strongly biased for IESs (Fig. 4F–H). Importantly, even scnRNAs complementary to Unique sequences were biased for IESs (Fig. 4, cf. D and H). Therefore, the IES-biased production of scnRNAs cannot be explained by the repeated nature of IESs.

IES-biased scnRNA production is attributed to biased transcription in the Mic

If the repetitiveness of the sequences in the genome is not the main contributor to IES-biased scnRNA production, what mechanism is then responsible for the bias? One obvious possibility is that Mic transcription, which produces the precursors of scnRNAs, preferentially occurs at IESs. To investigate this possibility, we analyzed Mic transcription by global run-on sequencing (GRO-seq), which provides a genome-wide profile of the position and orientation of all transcription-engaged RNA polymerases (Core et al. 2008).

A Mic-enriched (>99.5%) nuclear fraction was prepared by a stepwise differential centrifugation method (Sweet and Allis 2006) from wild-type cells at 2 h post-mixing for use in the GRO-seq assay. We used the cells at 2 h post-mixing because the Mic extensively elongates at meiotic prophase (~3–4 h post-mixing) (Sugai and Hiwatashi 1974), making it technically difficult to prepare an intact Mic at these later stages. Although scnRNA accumulation peaks at ~4 h post-mixing, scnRNA production has already initiated at 2 h post-mixing (see Fig. 7, below; Mochizuki et al. 2002). Therefore, Mic transcription should have started at this stage.

Nuclear run-on transcripts were purified and sheared to ~30- to 250-nt fragments, a cDNA library was constructed, single-read sequences were produced, and the first 25-nt sequences were used for mapping (see the Materials and Methods for details). Although we used a Mic-enriched nuclear fraction, ~9% of the reads matched rRNAs (Supplemental Table S2). Because rDNA is amplified to ~10,000 copies only in the Mac and exists as a single copy in the Mic genome, most of the reads matching rRNAs were likely derived from Mac contamination of the Mic-concentrated nuclear fraction.

The filtered GRO-seq reads were mapped to the Cal locus (Fig. 5A). Most of the transcripts occurred bidirectionally, and a large fraction was complementary to IESs. The pattern of these transcripts was highly similar to that of scnRNAs from *TWI1* knockout cells at 3 h post-mixing (Fig. 5B). Transcripts exhibiting mapping patterns that were highly similar to the scnRNA profiles were also detected from other loci (Supplemental Fig. S1). Therefore, we conclude that the bias of scnRNAs toward IESs is mainly attributed to the IES-biased transcription of the precursor RNAs in the Mic.

When the GRO-seq reads were mapped to the LMR locus, most of the transcripts were complementary to IESs, and their pattern was highly similar to the scnRNA

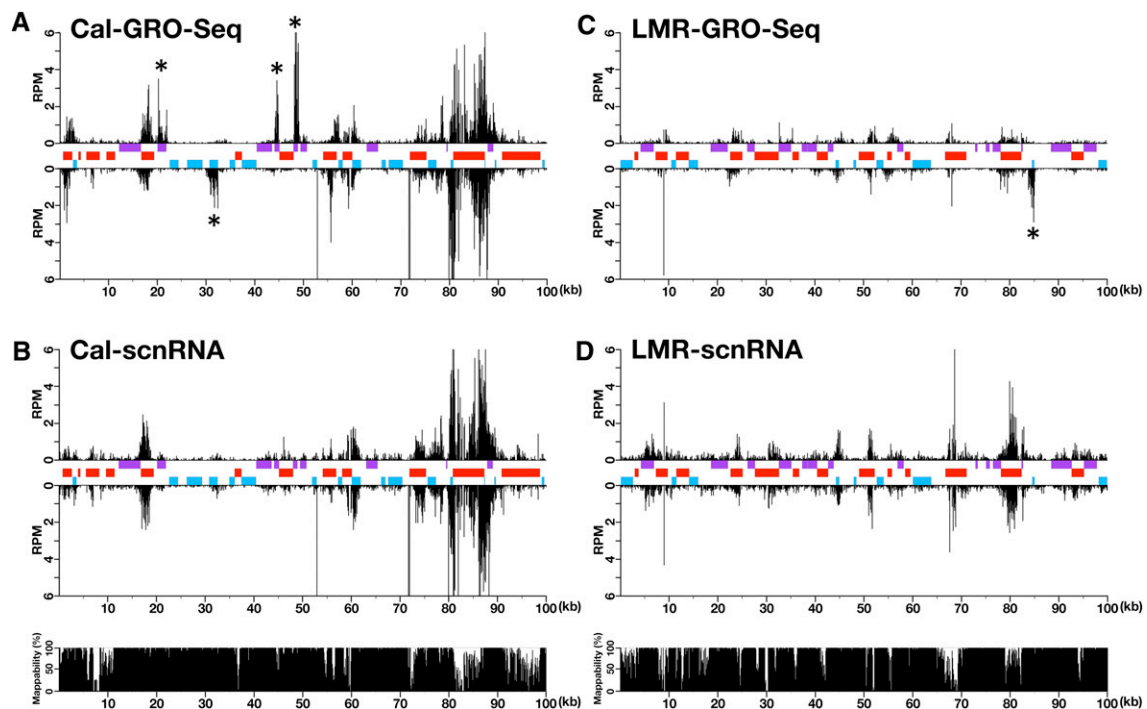


Figure 5. Nuclear run-on assay of the Mic-enriched nuclear fraction. (A,C) The sequences of nascent transcripts obtained by the global nuclear run-on assay using a Mic-enriched nuclear fraction of wild-type cells at 2 h post-mixing were mapped to the Cal (A) or LMR (C) locus. Numbers (reads per million RNAs [RPM]) of the sequences that uniquely mapped to the sense and antisense strands of the loci are shown in the plus and minus directions, respectively, as histograms with 100-nt bins. The IESs are marked in red. The regions of genes (including exons and introns) that are predicted to be transcribed from left to right and from right to left are labeled as purple and blue boxes, respectively. The reads that are predicted to be derived from unidirectional mRNA transcripts are marked with asterisks. (B,D) The profile of scnRNAs (*TWI1* knockout cells, 3 h post-mixing) mapped to the Cal (B) or LMR (D) locus. The same results are also shown in Figure 3, C and E, respectively. (B,D, bottom) The fraction of unique sequences that can be mapped by scnRNAs or GRO-seq reads in each 100-nt bin is shown as “Mappability”.

profiles at the same locus (Fig. 5C). However, overall levels of transcription from this locus were very low. The Mic LMR locus may not be highly transcribed at the stage at which the nuclei were prepared (at 2 h post-mixing) because different chromosomal locations probably produce scnRNAs at distinct times (Fig. 3F).

Some of the transcripts detected by GRO-seq occurred unidirectionally, and the number of these transcripts increased steeply at the 5' ends of predicted ORFs (signals marked with asterisks in Fig. 5A,C; Supplemental Fig. S1A). These likely represent mRNAs derived from contaminating polyploid Mac because although the Mic-enriched fraction contained <0.5% Macs, the Mac contains ~20 times more DNA than the Mic, and thus some transcripts from Mac DNA were expected to be detected in this study. Consistently, the number of reads mapped to the coding regions was correlated to the RNA expression levels of the corresponding genes detected by microarray hybridization with total RNA extracted from whole cells at the same stage (Miao et al. 2009). However, we cannot exclude the possibility that these mRNA-type transcripts were derived from the Mic. Nonetheless, they were transcribed unidirectionally and thus should not be processed to scnRNAs.

scnRNAs complementary to IESs are selectively enriched during conjugation

Despite the strong bias of scnRNA production for IESs, a significant fraction of scnRNAs at 3 h post-mixing was complementary to MDSs (Fig. 3A–E), including many protein-coding regions (see Fig. 5B,D; Supplemental Fig. S1B,D). Because scnRNAs are involved in the targeting of heterochromatin formation, which in turn induces DNA elimination (Liu et al. 2004, 2007), the presence of scnRNAs complementary to MDSs could cause the elimination of many essential genes. It has been proposed that this mistargeting of DNA elimination is prevented by the enrichment of scnRNAs complementary to IESs and that this enrichment is mediated by the interactions between scnRNAs and the Mac genome (Fig. 1K; Mochizuki et al. 2002; Mochizuki and Gorovsky 2004c; Aronica et al. 2008). This proposed process is called scnRNA selection.

To determine whether scnRNA selection occurs at a genome-wide level, we compared scnRNAs from wild-type cells at 3, 4.5, 6, and 8 h post-mixing. In this analysis, scnRNAs were oxidized prior to cDNA library construction. The oxidation inhibits linker ligation at the first step of cDNA library construction, and thus any oxidation-sensitive RNAs were excluded from the sequencing. Because scnRNAs loaded onto Twilp are 2'-O-methylated (Kurth and Mochizuki 2009), which prevents the periodate oxidation reaction, studying oxidation-resistant scnRNAs allowed us to specifically analyze scnRNAs that were loaded onto Twilp (HM Kurth and K Mochizuki, in prep.) and committed to the DNA elimination pathway.

First, the sequenced scnRNAs were mapped to the annotated Mic genome sequence described above. Among the scnRNAs from wild-type cells at 3 h post-mixing, when the selective degradation process should have just begun,

13.0% and 84.0% of scnRNAs mapped to the MDSs and IESs of the annotated genome, respectively (Fig. 6A; Supplemental Table S1). This ratio was similar to the one obtained from nonoxidized scnRNAs of *TWI1* knockout cells (15.3% to MDSs and 80.1% to IESs) (Fig. 3B; Supplemental Table S1), supporting our expectation that scnRNA selection had largely not begun at 3 h post-mixing.

The fraction of scnRNAs that was complementary to IESs was gradually enriched through conjugation (Fig. 6A; Supplemental Table S1). At 8 h post-mixing, when the heterochromatin formation that preceded DNA elimination had begun and thus the scnRNA selection was expected to be completed, only 5.6% of the scnRNAs were complementary to MDSs, while 93.1% were complementary to IESs (Fig. 6A, "WT_8 h"; Supplemental Table S1). These results are consistent with our previous Southern blot-based observations that total scnRNAs extracted from wild-type cells at 2 h post-mixing hybridized about three times more to Mic DNA than to Mac DNA, while scnRNAs at 8 h post-mixing hybridized ~22 times more to Mic DNA than to Mac DNA (Mochizuki and Gorovsky 2004a). Therefore, we conclude that scnRNAs complementary to IESs are enriched during conjugation at a genome-wide level.

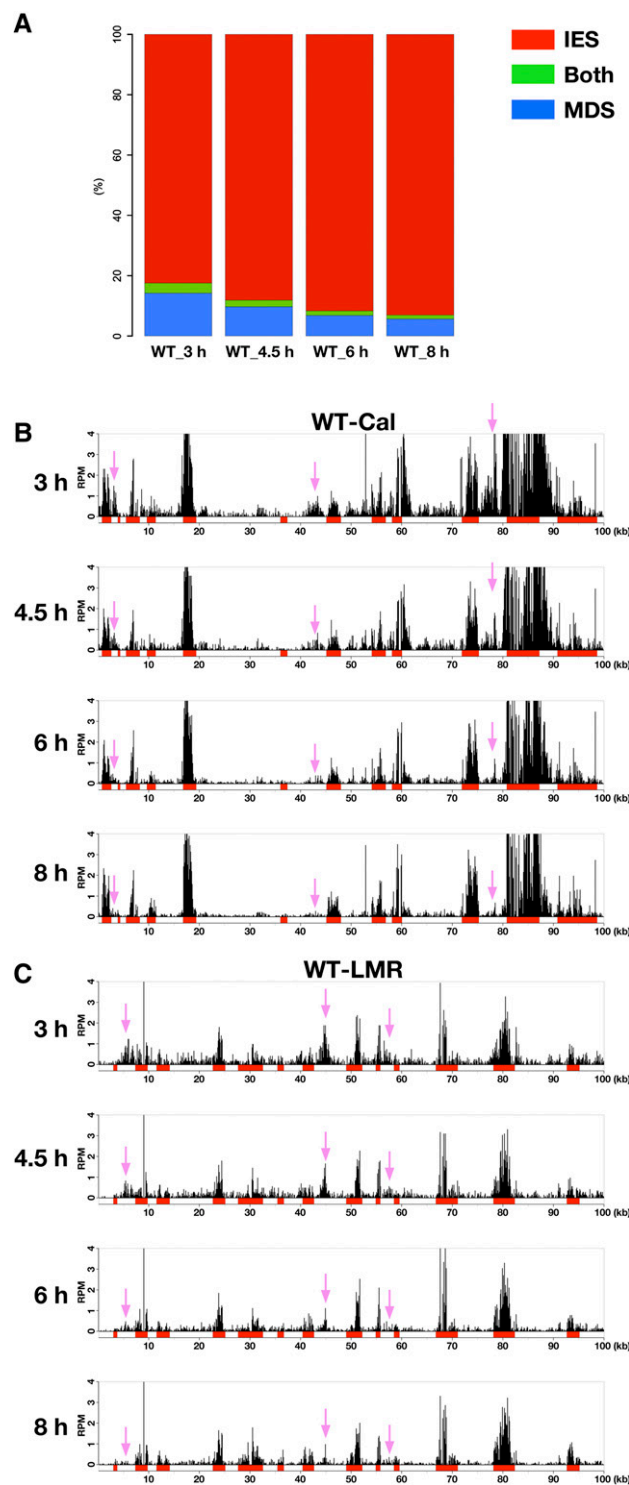
Enrichment of IES-complementary scnRNAs was also detected by analyzing scnRNAs from wild-type cells without oxidation (Supplemental Fig. S2). This is not surprising because almost all scnRNAs in cells at the mid to late stages of conjugation (after ~6 h post-mixing) are loaded onto Twilp (HM Kurth and K Mochizuki, in prep.). Therefore, oxidation-resistant scnRNAs and total scnRNAs during the later stages of conjugation should mostly overlap. This result also indicates that the periodate oxidation reaction does not perturb the IES/MDS ratio of the cDNA libraries.

A similar scnRNA selection process was detected by mapping scnRNAs to the Cal and LMR loci (Fig. 6B,C). Like the scnRNAs from *TWI1* knockout cells (Fig. 3B–E), the majority of oxidation-resistant scnRNAs from wild-type cells at 3 h post-mixing were complementary to IESs, and only a smaller fraction of scnRNAs mapped to MDSs (Fig. 6B,C, "3h"). While scnRNAs complementary to IESs persisted through conjugation, those complementary to MDSs were gradually diminished (Fig. 6B,C, peaks marked by pink arrows), and at 8 h post-mixing, most scnRNAs were complementary to IESs (Fig. 6B,C, "8h"). Therefore, scnRNA selection occurs at both the genome-wide and local levels.

We previously demonstrated that Ema1p, the Twilp-associated RNA helicase, is required for the degradation of scnRNAs complementary to some MDSs and for proper DNA elimination (Aronica et al. 2008). Therefore, we next analyzed scnRNAs from *EMA1* knockout cells. Although scnRNAs complementary to MDSs were reduced in wild-type cells during conjugation, as described above (Fig. 6A–C), scnRNAs complementary to MDSs remained at high levels throughout conjugation in *EMA1* knockout cells (Supplemental Fig. S3A–C). Therefore, we conclude that *EMA1* is required for the genome-wide process of scnRNA selection.

Although scnRNAs complementary to MDSs were dramatically reduced during conjugation in wild-type cells, such scnRNAs still remained at 8 h post-mixing (Fig. 6A–C), when scnRNA-mediated heterochromatin formation had already begun. We believe that this was caused in part by asynchronicity of conjugation; although asynchronous mating initiation was limited by adding nutrient medium

at ~3 h post-mixing, some cells at 8 h post-mixing might have mated for only 5 h. However, it is also possible that some scnRNAs complementary to MDSs remain in cells that are at late conjugation stages. There may be a threshold of scnRNAs required for efficient induction of DNA elimination, and MDS-complementary scnRNAs detected at 8 h post-mixing could be below this threshold.



Turnover of scnRNAs complementary to MDSs is responsible for scnRNA selection

Transcription in the Mic and the localization of Dcl1p, which are necessary for scnRNA biogenesis, both exclusively occur at the early stage of conjugation (Sugai and Hiwatashi 1974; Martindale et al. 1985; Mochizuki and Gorovsky 2004b, 2005; Malone et al. 2005). Therefore, the enrichment of scnRNAs complementary to IESs must be attributed to either selective degradation of scnRNAs complementary to MDSs or Dcl1p-independent selective secondary amplification of scnRNAs complementary to IESs by an RNA-dependent RNA polymerase activity.

If the observed relative increase of scnRNAs complementary to IESs (82.4% at 3 h to 93.1% at 8 h) (Fig. 6A) was solely achieved by their selective secondary amplification of scnRNAs complementary to IESs, the total scnRNAs would have to be amplified ~2.6-fold $[(100 - 82.4)/(100 - 93.1)]$. However, the amount of scnRNA, which was analyzed by gel electrophoresis followed by fluorescent dye staining of RNAs (Fig. 7A), peaked at ~4 h post-mixing and decreased slightly at later conjugation stages. Moreover, we recently found that secondary scnRNA production is limited after scnRNA processing by Dcl1p because scnRNAs from all analyzed stages exhibited the signature of Dicer processing (HM Kurth and K Mochizuki, in prep.). Therefore, selective amplification of IES-complementary scnRNAs, if any, cannot fully explain the enrichment of IES-complementary scnRNAs.

Our previous Northern hybridization studies have demonstrated that the levels of scnRNAs complementary to two different transposon-like IES elements remain constant during the course of conjugation, while the number of scnRNAs complementary to the MI repeat, a repetitive sequence that occurs in both IESs and MDSs, dramatically decreases at the mid-conjugation stage (Aronica et al. 2008). This is consistent with the observation in this study that the fraction of scnRNAs complementary to both MDSs and IESs of the annotated Mic genome

Figure 6. Comparison of scnRNAs from different conjugation stages. (A) The fractions of the oxidation-resistant scnRNAs from wild-type cells at 3, 4.5, 6, or 8 h post-mixing that are complementary only to MDSs (blue), to both MDSs and IESs (green), or only to IESs (red) are shown. (B,C) The numbers of sequences (reads per million scnRNAs [RPM]) of the oxidation-resistant scnRNAs from wild-type cells at 3, 4.5, 6, or 8 h post-mixing that uniquely mapped to the Cal (B) or LMR (C) locus are shown as histograms with 100-nt bins. The numbers of scnRNAs matching the sense and antisense strands were combined. The positions of the peaks that were markedly decreased during the course of conjugation are labeled by pink arrows.

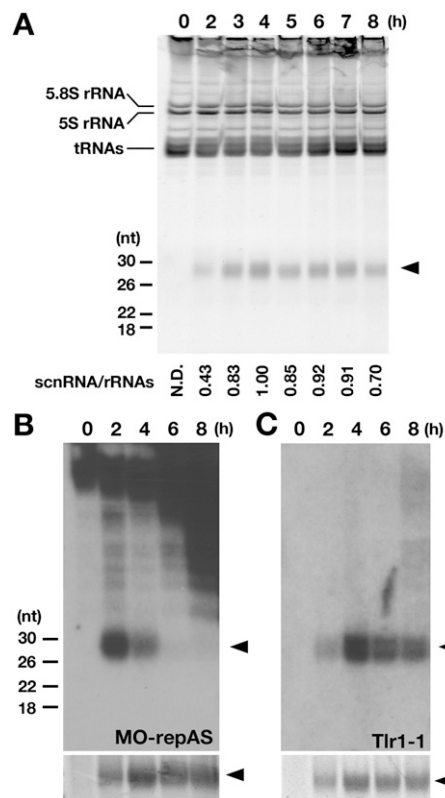


Figure 7. Analysis of the turnover of scnRNAs. (A) The total RNAs of wild-type cells during conjugation at the indicated time points (shown at the top) post-mixing were separated in a sequencing gel, and RNA was stained with GelRed, a fluorescent dye. The positions of the small RNA markers are shown on the left. The position of the scnRNA is marked by an arrowhead. The positions of the tRNAs, 5S rRNAs, and 5.8S rRNAs are shown on the left. The loading was normalized to the 5S and 5.8S rRNAs, the amount of scnRNAs at 4 h post-mixing was defined as 1.00, and the relative ratios of the scnRNAs at other time points are indicated at the bottom. (B,C) Total RNAs of conjugating wild-type cells at the indicated time points (shown at the top) post-mixing were analyzed by Northern hybridization (top panels) or GelRed staining (bottom panels). For Northern hybridizations, the radiolabeled probe MO-repAS (B) or Tlr1-1 (C) was used; these are complementary to the MO repeats that only appear in MDSs or a part of a repetitive Tlr1 IES element, respectively. The positions of the small RNA markers are shown at the left. The positions of the scnRNAs are marked by arrowheads.

decreased during conjugation (Fig. 6A, “Both,” shown in green, 3.6% at 3 h to 1.3% at 8 h). However, we previously failed to reliably detect scnRNAs complementary to any sequences that only exist in MDSs, most likely because our Northern hybridization conditions were not sensitive enough to detect scnRNAs derived from unique or low-copy genomic loci (Aronica et al. 2008).

By analyzing the genomic sequences, we identified a short repeated sequence, the MO (MDS-Only) repeat, that occurred 125 times exclusively in the MDSs of the annotated Mic genome. Total RNAs were extracted from wild-type cells in the early to late stages of conjugation and analyzed by Northern hybridization with a probe

complementary to the MO-repeat (MO-rpAS). As shown in Figure 7B, scnRNAs complementary to the MO repeat were detected at early stages of conjugation (2–4 h post-mixing) and became undetectable at mid to late conjugation stages (after 6 h post-mixing), indicating that these MDS-complementary scnRNAs were produced once and then degraded. In contrast, scnRNAs complementary to the Tlr1-1 probe, which is part of the Tlr1 transposon-related IES element, persisted into the late conjugation stages (Fig. 7C), as observed previously (Aronica et al. 2008). Taken together, our data indicate that the enrichment of IES-complementary scnRNAs is mainly achieved by selective degradation of scnRNAs complementary to MDSs.

Discussion

In this study, we addressed some of the key issues of the small RNA-directed epigenetic regulation of DNA elimination in *Tetrahymena*. As proposed in previous models (Fig. 1), we found that scnRNAs are produced exclusively from the Mic (Fig. 2) and a variety of Mic chromosomal locations (Fig. 3F) and that *EMA1*-dependent scnRNA selection occurs at a genome-wide level (Fig. 6; Supplemental Figs. S2, S3) by RNA turnover (Fig. 7). However, in contrast to expectations, this study demonstrates that scnRNA production is strongly biased toward IESs (Fig. 3A–E) and occurs nonhomogeneously from the Mic chromosomes at the global level (Fig. 3F). We further revealed that IES-biased scnRNA production is not caused mainly by the repetitiveness of IESs (Fig. 4), but is attributed to IES-biased transcription (Fig. 5; Supplemental Fig. S1).

The nonhomogeneous production of scnRNAs may explain position-dependent transgene elimination

The nonhomogeneous pattern of scnRNA production from the Mic genome (Fig. 3F) may address a previously unresolved question about the epigenetic regulation of DNA elimination: the position dependencies of transgene elimination. A neomycin resistance (neo) gene integrated into an MDS locus on a Mic chromosome is occasionally deleted from the genomes of new Macs during successive rounds of conjugation. The efficiency of neo elimination from different Mic loci varies significantly (Yao et al. 2003; Liu et al. 2005; Howard-Till and Yao 2007). Because scnRNA production from MDSs is generally lower than that from IESs (Fig. 3A–E) and the production of scnRNAs from the Mic genome is not homogeneous (Fig. 3), the variable efficiencies of neo elimination from different Mic loci may be due to the amount of scnRNAs produced from the loci.

It has been proposed that scnRNA-mediated DNA elimination acts as a genome surveillance mechanism against transposons (for review, see Schoeberl and Mochizuki 2011). However, the nonhomogeneous production of scnRNAs from the Mic genome indicates that scnRNAs cannot target transposons for DNA elimination if they are inserted into a Mic genome locus at which scnRNAs are not highly produced. The insertion of transposons may not immediately activate the DNA elimination pathway. Only after multiple insertions may

some of the transposons integrate into scnRNA-producing loci, enabling the cells to produce scnRNAs complementary to the transposon to eliminate it. In this way, the nonhomogeneous production of scnRNAs may set a threshold for DNA elimination such that only transposons that can efficiently multiply in the Mic genome are targeted for DNA elimination.

How does IES-biased scnRNA production occur?

Our unexpected observation that scnRNA production is heavily biased for IESs at the transcriptional level has raised a new question: How is the IES-biased transcription of scnRNA precursors achieved? Theoretically, the biased transcription could be mediated by specific DNA sequences that reside in or around all >6000 IESs. However, we believe this is unlikely for two reasons. First, different IESs share no obvious consensus sequences. Second, the transcriptional start sites of the Mic transcripts seem to be distributed across broad regions of the genome (Fig. 5; Supplemental Fig. S1; Chalker and Yao 1996). An alternative mechanism for the biased transcription is that scnRNAs are transported not only to the developing new Mac to target IESs for DNA elimination, but also to the new Mic to mark IES sites for future transcription. In this way, DNA elimination in the new Mac may be epigenetically and transgenerationally controlled not only by the genome contents of the parental Mac through the selective degradation of scnRNAs, but also by those of the grandparental Mac through transcriptional regulation of the Mic.

If the latter hypothesis is true, we predict that (1) a protein other than Twilp transports scnRNAs to the new Mic because Twilp does not localize to the Mic at any stage (Mochizuki et al. 2002); (2) IESs and MDSs in the Mic are differently marked by some chromatin modifications or chromatin-associated proteins, and this difference is maintained during vegetative growth (one such candidate is histone H3 Lys 27 methylation, which is maintained in the Mic during vegetative growth) (Liu et al. 2007); (3) an epigenetic “block” of DNA elimination through the ectopic introduction of an IES to the parental Mac (Chalker and Yao 1996) not only inhibits the elimination of the IES in the new Mac, but also inhibits Mic transcription, and thus scnRNA production, from the IES during the conjugation of the progeny. We await future research that would confirm or reject these predictions and reveal how ancestral genomes can epigenetically regulate the behavior of genomes of successive generations.

Materials and methods

Small RNA cDNA library preparation and sequencing

Wild-type *T. thermophila* strains B2086 and CU428, Chr4 nullisomic strains CU357 and CU367 (Bruns et al. 1983), *Twil1* complete (germ plus soma) knockout (Mochizuki and Gorovsky 2004a) strains 20-1 and 20-4, and *EMA1* knockout (Aronica et al. 2008) strains 7-17-1 and 8-6-1 were used for total RNA preparation. B2086, CU428, CU357, and CU367 were obtained from the *Tetrahymena* Stock Center at Cornell University. Cell mating

was induced by mixing equal numbers of prestarved cells of two strains possessing complementary mating types, and asynchronous mating initiation was limited by adding 1/3 vol of 4× SPP at ~3 h post-mixing. RNA was isolated from 2.5×10^6 mating cells at the indicated time points post-mixing with the mirVana miRNA isolation kit (Ambion) or TRIzol reagent (Invitrogen). For periodate oxidation of small RNAs (Akbergenov et al. 2006), the RNA pellet was dried and resuspended in 17.5 μ L of borax buffer (4.375 mM borax, 50 mM boric acid at pH 8.6). Then, 2.5 μ L of 0.2 M sodium periodate was added, and the reaction was incubated for 10 min at room temperature in the dark. Two microliters of glycerol was added, the solution was incubated for another 10 min at room temperature in the dark, and RNAs were purified by ethanol precipitation.

cDNA libraries of small RNAs were constructed as described (Hafner et al. 2008) with some modifications. In brief, RNA (with or without periodate oxidation) was mixed with radiolabeled 26- and 32-nt oligoribonucleotides (5'-GUC GUACGCGGAAUAGUUUAAACUGU-3'; 5'-AUCUUGGUCG UACGCGGAAUAGUUUAAACUGU-3'; PmeI restriction sites are underlined) and fractionated in a sequencing gel, and the gel regions containing the two radiolabeled markers (plus a few bases on both sides) were excised. RNA was recovered from the gel and ligated to a 5'-adenylated 3' adapter oligonucleotide (5'-App-TCGTATGCCGTCTTCTGCTTG-L-3', where L is the 3' OH-blocking group; IDT) by T4 RNA ligase 2-truncated or T4 RNA ligase 2-truncated K227Q (New England Biolabs). After gel fractionation and purification as above, the product was ligated to the 5' adapter oligoribonucleotide (5'-GUUCAGAGUUCUACAG UCCGACGAUC-3') by T4 RNA ligase (Ambion). The gel-purified final ligation product was then reverse-transcribed with the Sol-rv primer (5'-CAAGCAGAAGACGGCATACGA-3') and PCR-amplified with the Sol-rv primer and Sol-fw primer (5'-AATGATACCG CGACCACCGACAGGTTTCAGAGTTCTACAGTCCGA-3'). After phenol-chloroform extraction followed by ethanol precipitation, the PCR product was digested with PmeI (to cut the size markers) and fractionated by agarose gel electrophoresis, and the undigested PCR product was purified from the gel. The amplified cDNA libraries were mixed with 1/10 (by mass) of ϕ X174 phage DNA, and single-read 36-base sequences were generated on the Illumina GAII platform, resulting in ~17 million–27 million reads. The ϕ X174 phage DNA was included because most scnRNAs begin with 5'U (HM Kurth and K Mochizuki, in prep.); therefore, sequencing only scnRNAs disturbs the cluster recognition process during sequencing reactions. Raw sequencing data and processed data sets have been deposited at the NCBI Gene Expression Omnibus (<http://www.ncbi.nih.gov/geo>) as GSE38768.

Following the removal of the adapter sequences from the sequence reads, we removed reads containing any non-A/C/G/T bases; reads from rRNAs, tRNA, snRNAs, and the mitochondrial genome; and reads from mixed ϕ X174 phage DNA. Then, sequences from 26- to 32-nt RNAs were used as scnRNA sequences. The first 25 nt of each scnRNA sequence was mapped to the annotated Mic genome, the Cal locus, the LMR locus, or the selected 218 Mac genome scaffolds. After mapping scnRNAs to the Cal or LMR locus, the repetitiveness of the genome was normalized by either dividing the numbers of mapped scnRNAs by their sequence frequencies in the draft Mic genome (“weighted”) or excluding scnRNAs that were complementary to more than two positions in the Mic genome (“uniquely mapped”).

Annotation of the draft Mic genome

The draft Mic genome sequence (version 2, supercontigs) was obtained from the *Tetrahymena* Comparative Sequencing Project (Broad Institute of Harvard and Massachusetts Institute of

Technology) Web site (<http://www.broadinstitute.org>). The Mac genome sequence (the version released in October 2008 from the J. Craig Venter Institute) was obtained from the *Tetrahymena* Genome Database (<http://ciliate.org>). To predict IESs, all scaffolds of Mac genome sequences were concatenated. The concatenated sequence was then split into 100-nt sequences, each possessing 50-nt overlaps with the neighboring sequences. When a sequence did not map to the Mic draft genome sequence, the 100-nt sequence was further split into all possible pairs of pieces, both of which contained at least 24 nt. If both pieces mapped to the Mic draft genome sequence with a gap longer than 100 nt and shorter than 100,000 nt (because all known IESs are ~0.2–20 kb), the gap sequence was stored as an IES candidate. The prediction was further tested by comparing 200-nt sequences upstream of and downstream from the IES sites in the draft Mic genome and the Mac genome. Only when these 200-nt sequences were identical between the Mac and the Mic did we use the gap sequences as IESs. By this method, we predicted 8721 IESs (28.0 Mb in total). The positions of predicted IESs are listed in Supplemental Table S3. We likely underestimated the number of IESs in this prediction because we defined an IES only when it was assembled in the draft Mic genome with at least 200-bp flanking MDSs at both ends, while the median size of the supercontigs of the draft Mic genome was 2.3 kb and the median size of the predicted IESs was 2.7 kb. Moreover, many repetitive sequences, which were expected to be enriched in the IESs, may have failed to be assembled in the current Mic draft genome obtained by short-read pyrosequencing. Therefore, we also included unassembled DNA sequences that were obtained by an independent Mic genome Sanger sequencing and that did not map to the Mac genome sequence (Fass et al. 2011) as IESs. As MDSs, we used the Mac genome scaffolds that were confirmed to be enriched in the Mac by a microarray study (102.8 Mb in total) (Coyne et al. 2008). Then, all possible 25-nt sequences were extracted from the draft Mic genome, and any sequences containing non-A/C/G/T bases were removed and mapped to the IESs and MDSs described above to prepare the “annotated” Mic draft genome. Among the 147.3 Mb of the draft Mic genome, 25-nt sequences corresponding to 98.8 Mb, 33.1 Mb, and 3.0 Mb of genome were found only in MDSs, only in IESs, and in both MDSs and IESs, respectively. The remaining 12.8 Mb could not be annotated. The annotated Mic genome sequence was used for mapping scnRNAs (Figs. 3, 4, 6; Supplemental Figs. S2, S3; Supplemental Table S1). To analyze the repeat bias of the scnRNA origin (Fig. 4), all possible 25-nt sequences of the draft Mic genome were classified among three classes: Unique (25-nt sequences that appeared only once in the draft genome), Mid-Rep (appeared two to nine times), and High-Rep (appeared >10 times). Four 100-kb Mic regions containing, respectively, (1) the calmodulin gene and Cal-IES (Cal locus, corresponding to 127,977–277,725 bases of supercontig 2.218 in the second assembly of the draft Mic genome sequence); (2) the L-IES, M-IES, and R-IES (LMR locus, corresponding to 64,932–165,807 bases of supercontig 2.231); (3) the NgoA gene (NgoA locus, corresponding to 159,772–259,771 bases of supercontig 2.97); and (4) the THERM_01062860 gene (THERM_01062860 locus, corresponding to 120,184–220,187 bases of supercontig 2.116) were selected to locally map scnRNAs or GRO-seq reads (Figs. 2, 3, 5, 6; Supplemental Figs. S1, S3). By manually comparing them with the corresponding Mac genome regions, the positions of IESs in these loci were annotated, some gaps were filled, and potential misassemblies were corrected.

GRO-seq

Nuclear fractions were isolated from wild-type strains B2086 and CU428 at 2 h post-mixing as described previously (Sweet

and Allis 2006) with some modifications. All fractionation steps were performed at 4°C or on ice. Cells (1.2×10^8) were suspended in 100 mL of medium B and blended for 30 sec at the highest speed (Waring 7009G commercial blender). Samples were divided among two 50-mL conical tubes and centrifuged with a swing rotor at 1750g, 2750g, 3300g, and 4200g for 5 min each and twice at 4500g for 10 min. The supernatants of each step were blended as above for 20 sec and then centrifuged at the subsequent speed. Each pellet was washed once in modified nuclei resuspension buffer (NRB; 50 mM Tris-Cl at pH 8.0, 30% glycerol, 5 mM MgCl₂, 0.1 mM EDTA), resuspended in 200 μ L of NRB (per 50-mL tube), and stored at –80°C. A portion of the pellet was removed for DAPI staining, and the ratios between the Mac and Mic and the concentrations of nuclei were examined on a hemocytometer. The Mic-enriched nuclear fraction used for the experiments shown in Figure 5 and Supplemental Figure S1 consisted of ~99.5% Mic and ~0.5% Mac, as judged by the sizes of the nuclei.

Mic-enriched nuclear fractions were used to identify nascent transcription by applying the GRO-seq technique (Core et al. 2008) with some modifications. Briefly, $\sim 1.3 \times 10^7$ nuclei of the Mic-enriched fraction were used for the run-on reaction, which contained 500 μ M ATP, 500 μ M GTP, 500 μ M BrUTP, and 0.33 μ M α -³²PCTP. After DNase I treatment, the total RNA was isolated with TRIzol LS reagent (Invitrogen) and resuspended in 20 μ L of water. Base hydrolysis was performed for 20 min on ice to digest RNA products to ~30–250 nt. The RNAs were treated with DNase I again, and RNA was purified twice through P30 spin columns (Bio-Rad). BrdU-incorporated RNA was enriched on anti-BrdU beads (Santa Cruz Biotechnology) for 2 h at 4°C and resuspended in 21 μ L of water. A 20- μ L aliquot of the RNA was end-repaired and resuspended in 10.5 μ L of water.

cDNA was synthesized as described above in the section “Small RNA cDNA Library Preparation and Sequencing” with some modifications. For 3' adapter ligation, 10 μ L of the end-repaired RNAs was incubated with 1 μ L of 50 μ M 3' adapter, 0.5 μ L of Ribolock (Fermentas), 2.15 μ L of ATP-free T4 RNA ligase buffer, 6 μ L of 50% PEG-8000, and 2 μ L of T4 RNA ligase 2-truncated K227Q (New England Biolabs) overnight at 16°C. The RNAs were enriched on anti-BrdU beads as above and resuspended in 10 μ L of water. For 5' adapter ligation, 10 μ L of 3' adapter-ligated RNAs was incubated with 1 μ L of 50 μ M 5' adapter, 0.5 μ L of Ribolock, 2.2 μ L of T4 RNA ligase buffer, 1 μ L of 10 mM ATP, 6 μ L of 50% PEG-8000, and 2 μ L of T4 RNA ligase (Ambion) overnight at 16°C. The RNAs were purified again with anti-BrdU beads and resuspended in 10 μ L of water. The adapter-ligated RNA was incubated with 2 μ L of 10 μ M Sol-rv primer and 1 μ L of 10 mM dNTP mix for 5 min at 65°C and for 2 min on ice. Then, 4 μ L of 5 \times first strand buffer, 1 μ L of 0.1 M DTT, 1 μ L of Ribolock, and 1 μ L of SuperScript III reverse transcriptase (Invitrogen) were added. The mixture was incubated for 30 min at 50°C, for 30 min at 55°C, and for 15 min at 70°C. Then, 5 U of RNase H (Ambion) was added, followed by incubation for 10 min at 37°C. The cDNA was amplified by subjecting 9.5 μ L of the reaction to 17 cycles of PCR with KAPA Hifi Hot Start DNA polymerase (Kapa Biosystems) and Sol-fw and Sol-rv primers. The amplified cDNA library was fractionated in a 2% agarose gel; ~100- to 200-base pair (bp) products were extracted, and single-read 50-base sequences were generated on a HiSeq 2000 platform (Illumina) to obtain ~200 million reads. Raw sequencing data and processed data sets have been deposited at the NCBI Gene Expression Omnibus (<http://www.ncbi.nih.gov/geo>) as GSE38768. Reads containing any non-A/C/G/T bases and reads derived from rRNAs, tRNAs, snRNAs, and the mitochondrial genome were removed. The first 25 nt of each GRO-seq read was used to uniquely map the representative Mic loci.

RNA staining and Northern hybridization

Total RNA was isolated with TRIzol (Invitrogen). The total RNA from 1×10^5 cells was separated in a 15% sequencing gel, and RNAs were stained with GelRed (Gentaur). RNA was then blotted onto a nitrocellulose membrane and probed with ^{32}P -labeled oligo DNA probes as described previously (Aronica et al. 2008). The probes used in this study were Tlr1-1 (Aronica et al. 2008), which is complementary to the Tlr1 transposon-like IES, and MO-repAS (5'-TTTACCCAAGATAAAAGTGTCAAATTTGA GAGATTAGTGAATTTGCTAA-3'; the MO repeat is underlined), which is complementary to the MDS-limited MO repeat. The MO repeat (5'-ACTAATCTCTCAAATTTGACACTTT-3') was identified as the most repetitive 25-nt sequence that only existed in the annotated MDSs.

Acknowledgments

We thank Kathleen Collins for sharing Mic genome Sanger sequencing data, Eileen Hamilton and Ed Orias for sharing Mic chromosomal location data, Grzegorz Sienski for technical advice on GRO-seq, and Julius Brennecke and Josef Loidl and their laboratories for discussions and technical advice. The research leading to these results was supported by an ERC Starting Grant (204986) under the European Community's Seventh Framework Programme, by the Doktoratskolleg RNA Biology (W1207-B09) and the Special Research Program (SFB) "RNA Regulation of the Transcriptome" (F4307-B09) from the Austrian Science Fund (FWF), and by core funding from the Austrian Academy of Sciences to K.M.

References

- Akbergenov R, Si-Ammour A, Blevins T, Amin I, Kutter C, Vanderschuren H, Zhang P, Gruissem W, Meins Jr F, Hohn T, et al. 2006. Molecular characterization of geminivirus-derived small RNAs in different plant species. *Nucleic Acids Res* **34**: 462–471.
- Aronica L, Bednenko J, Noto T, Desouza LV, Siu KW, Loidl J, Pearlman RE, Gorovsky MA, Mochizuki K. 2008. Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev* **22**: 2228–2241.
- Austerberry CF, Allis CD, Yao MC. 1984. Specific DNA rearrangements in synchronously developing nuclei of *Tetrahymena*. *Proc Natl Acad Sci* **81**: 7383–7387.
- Austerberry CF, Snyder RO, Yao MC. 1989. Sequence microheterogeneity is generated at junctions of programmed DNA deletions in *Tetrahymena thermophila*. *Nucleic Acids Res* **17**: 7263–7272.
- Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**: 1387–1392.
- Bruns PJ, Brussard TB, Merriam EV. 1983. Nullisomic *Tetrahymena*. II. A set of nullisomics define the germinal chromosomes. *Genetics* **104**: 257–270.
- Chalker DL, Yao MC. 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.
- Chalker DL, Yao MC. 2001. Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. *Genes Dev* **15**: 1287–1298.
- Chalker DL, Fuller P, Yao MC. 2005. Communication between parental and developing genomes during *Tetrahymena* nuclear differentiation is likely mediated by homologous RNAs. *Genetics* **169**: 149–160.
- Cheng CY, Vogt A, Mochizuki K, Yao MC. 2010. A domesticated piggyBac transposase plays a key role in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in *Tetrahymena thermophila*. *Mol Biol Cell* **21**: 1753–1762.
- Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845–1848.
- Coyne RS, Nikiforov MA, Smothers JF, Allis CD, Yao MC. 1999. Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation. *Mol Cell* **4**: 865–872.
- Coyne RS, Thiagarajan M, Jones KM, Wortman JR, Tallon LJ, Haas BJ, Cassidy-Hanley DM, Wiley EA, Smith JJ, Collins K, et al. 2008. Refined annotation and assembly of the *Tetrahymena thermophila* genome sequence through EST analysis, comparative genomic hybridization, and targeted gap closure. *BMC Genomics* **9**: 562. doi: 10.1186/1471-2164-9-562.
- Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**: 157–161.
- Duharcourt S, Butler A, Meyer E. 1995. Epigenetic self-regulation of developmental excision of an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev* **9**: 2065–2077.
- Fass JN, Joshi NA, Couvillion MT, Bowen J, Gorovsky MA, Hamilton EP, Orias E, Hong K, Coyne RS, Eisen JA, et al. 2011. Genome-scale analysis of programmed DNA elimination sites in *Tetrahymena thermophila*. *G3 (Bethesda)* **1**: 515–522.
- Fillingham JS, Thing TA, Vythilingum N, Keuroghlian A, Bruno D, Golding GB, Pearlman RE. 2004. A non-long terminal repeat retrotransposon family is restricted to the germ line micronucleus of the ciliated protozoan *Tetrahymena thermophila*. *Eukaryot Cell* **3**: 157–169.
- Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holloch D, Lim C, Tuschl T. 2008. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44**: 3–12.
- Howard-Till RA, Yao MC. 2007. Tudor nuclease genes and programmed DNA rearrangements in *Tetrahymena thermophila*. *Eukaryot Cell* **6**: 1795–1804.
- Katoh M, Hirono M, Takemasa T, Kimura M, Watanabe Y. 1993. A micronucleus-specific sequence exists in the 5'-upstream region of calmodulin gene in *Tetrahymena thermophila*. *Nucleic Acids Res* **21**: 2409–2414.
- Koizumi S, Kobayashi S. 1989. Microinjection of plasmid DNA encoding the A surface antigen of *Paramecium tetraurelia* restores the ability to regenerate a wild-type macronucleus. *Mol Cell Biol* **9**: 4398–4401.
- Kurth HM, Mochizuki K. 2009. 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA* **15**: 675–685.
- Lee SR, Collins K. 2006. Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes Dev* **20**: 28–33.
- Liu Y, Mochizuki K, Gorovsky MA. 2004. Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in *Tetrahymena*. *Proc Natl Acad Sci* **101**: 1679–1684.
- Liu Y, Song X, Gorovsky MA, Karrer KM. 2005. Elimination of foreign DNA during somatic differentiation in *Tetrahymena thermophila* shows position effect and is dosage dependent. *Eukaryot Cell* **4**: 421–431.

- 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, 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.
- Martindale DW, Allis CD, Bruns PJ. 1985. RNA and protein synthesis during meiotic prophase in *Tetrahymena thermophila*. *J Protozool* **32**: 644–649.
- Miao W, Xiong J, Bowen J, Wang W, Liu Y, Braguinets O, Grigull J, Pearlman RE, Orias E, Gorovsky MA. 2009. Microarray analyses of gene expression during the *Tetrahymena thermophila* life cycle. *PLoS ONE* **4**: e4429. doi: 10.1371/journal.pone.0004429.
- Mochizuki K. 2010. RNA-directed epigenetic regulation of DNA rearrangements. *Essays Biochem* **48**: 89–100.
- Mochizuki K, Gorovsky MA. 2004a. Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev* **18**: 2068–2073.
- Mochizuki K, Gorovsky MA. 2004b. RNA polymerase II localizes in *Tetrahymena thermophila* meiotic micronuclei when micronuclear transcription associated with genome rearrangement occurs. *Eukaryot Cell* **3**: 1233–1240.
- Mochizuki K, Gorovsky MA. 2004c. Small RNAs in genome rearrangement in *Tetrahymena*. *Curr Opin Genet Dev* **14**: 181–187.
- 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.
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* **23**: 314–318.
- Noto T, Kurth HM, Kataoka K, Aronica L, Desouza LV, Siu KW, Pearlman RE, Gorovsky MA, Mochizuki K. 2010. The *Tetrahymena* Argonaute-binding protein Giw1p directs a mature Argonaute-siRNA complex to the nucleus. *Cell* **140**: 692–703.
- Schoeberl UE, Mochizuki K. 2011. Keeping the soma free of transposons: Programmed DNA elimination in ciliates. *J Biol Chem* **286**: 37045–37052.
- Sugai T, Hiwatashi K. 1974. Cytologic and autoradiographic studies of the micronucleus at meiotic prophase in *Tetrahymena pyriformis*. *J Protozool* **21**: 542–548.
- Sweet MT, Allis CD. 2006. Isolation and purification of *Tetrahymena* nuclei. *Cold Spring Harb Protoc* **2006**: doi: 10.1101/pdb.prot4500.
- Taverna SD, Coyne RS, Allis CD. 2002. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. *Cell* **110**: 701–711.
- Wuitschick JD, Gershan JA, Lochowicz AJ, Li S, Karrer KM. 2002. A novel family of mobile genetic elements is limited to the germline genome in *Tetrahymena thermophila*. *Nucleic Acids Res* **30**: 2524–2537.
- Yao MC, Gorovsky MA. 1974. Comparison of the sequences of macro- and micronuclear DNA of *Tetrahymena pyriformis*. *Chromosoma* **48**: 1–18.
- Yao MC, Duhaucourt S, Chalker DL. 2002. Genome-wide rearrangements of DNA in ciliates. In *Mobile DNA II* (ed. N Craig et al.), pp. 730–758. Academic Press, New York.
- Yao MC, Fuller P, Xi X. 2003. Programmed DNA deletion as an RNA-guided system of genome defense. *Science* **300**: 1581–1584.
- You Y, Aufderheide K, Morand J, Rodkey K, Forney J. 1991. Macronuclear transformation with specific DNA fragments controls the content of the new macronuclear genome in *Paramecium tetraurelia*. *Mol Cell Biol* **11**: 1133–1137.



Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in *Tetrahymena*

Ursula E. Schoeberl, Henriette M. Kurth, Tomoko Noto, et al.

Genes Dev. 2012, **26**:

Access the most recent version at doi:[10.1101/gad.196493.112](https://doi.org/10.1101/gad.196493.112)

Supplemental Material

<http://genesdev.cshlp.org/content/suppl/2012/08/01/26.15.1729.DC1>

Related Content

Intercepting noncoding messages between germline and soma

Shan Gao and Yifan Liu

Genes Dev. August , 2012 26: 1774-1779

References

This article cites 44 articles, 23 of which can be accessed free at:

<http://genesdev.cshlp.org/content/26/15/1729.full.html#ref-list-1>

Articles cited in:

<http://genesdev.cshlp.org/content/26/15/1729.full.html#related-urls>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Use CRISPRmod for targeted modulation of endogenous gene expression to validate siRNA data

horizon
a PerkinElmer company

The advertisement features a dark blue background. On the left, there is a colorful, abstract image of what appears to be a molecular structure or a cluster of cells in shades of red, orange, and purple. To the right of this image, the text 'Use CRISPRmod for targeted modulation of endogenous gene expression to validate siRNA data' is written in white. Further to the right, the 'horizon' logo is displayed in white, with 'a PerkinElmer company' written in smaller text below it.