Journal Pre-proof

Roles of non-coding RNAs in ciliate genome architecture

Sarah E. Allen, Mariusz Nowacki

PII: S0022-2836(20)30026-7

DOI: https://doi.org/10.1016/j.jmb.2019.12.042

Reference: YJMBI 66391

To appear in: *Journal of Molecular Biology*

Received Date: 14 October 2019

Revised Date: 19 December 2019

Accepted Date: 20 December 2019

Please cite this article as: S.E. Allen, M. Nowacki, Roles of non-coding RNAs in ciliate genome architecture, *Journal of Molecular Biology*, https://doi.org/10.1016/j.jmb.2019.12.042.

Title:

Roles of non-coding RNAs in ciliate genome architecture

Authors:

Sarah E. Allen, Mariusz Nowacki*

Institute of Cell Biology, University of Bern, Switzerland

*to whom correspondence should be addressed

Highlights:

Ciliate somatic genomes undergo massive rearrangement and reduction during sexual development.

Transposon elimination is guided by transposon-derived small RNAs and carried out by domesticated transposases

Novel pathways for small and long non-coding RNA production have recently been discovered in ciliates

Diverse ciliate species have different mechanisms and machinery for RNA-guided genome rearrangements, though common pathways indicate ancient origins of RNA-directed genome defence.

Abstract

Ciliates are an interesting model system for investigating diverse functions of non-coding RNAs, especially in genome defence pathways. During sexual development, the ciliate somatic genome undergoes massive rearrangement and reduction through removal of transposable elements and other repetitive DNA. This is guided by a multitude of non-coding RNAs of different sizes and functions, the extent of which is only recently becoming clear. The genome rearrangement pathways evolved as a defence against parasitic DNA, but interestingly also employ the transposable elements and transposases to execute their own removal. Thus, ciliates are also a good model for the co-evolution of host and transposable element, and the mutual dependence between the two. In this review we summarise the genome rearrangement pathways in three diverse species of ciliate, with focus on recent discoveries and the roles of non-coding RNAs.

Introduction

An example of the diversity of the functions of non-coding RNAs across evolutionary distance is their role in the fascinating genome rearrangement processes that take place in different species of ciliated protozoa (ciliate). In the last two decades non-coding RNAs have been shown to have crucial roles in these processes, roles that are divergent both within and across species and processes. Genome rearrangement is a process fundamental to the development of many ciliates and is thought to have emerged around one billion years ago, prior to the divergence of spirotrichea and oligohymenophorea, as a defence against transposable element invasion of the genome [1][2].

Nuclear dimorphism: One of the hallmarks of ciliates is their nuclear dimorphism: the presence of two distinct nuclei within the same unicellular organism. The larger nucleus is termed the macronucleus and contains all the DNA required for the day-to-day life of the ciliate, often at a high copy number (up to 2000n for certain nanochromosomes in Oxytricha), due to the large complex cells. The smaller nucleus (termed the micronucleus) comprises the germline and is kept mostly silent and quiescent until it is required to build a new soma via sexual reproduction. Ciliates reproduce sexually through conjugation, in which two ciliates of opposite mating types join and exchange one pronucleus produced from meiotic and mitotic divisions of the germline micronucleus. This haploid pronucleus fuses with a second pronucleus in the recipient cell and undergoes a number of mitotic divisions, the products of which go on to form the new micronuclei and macronuclei (the numbers of the various nuclei vary with the species of ciliate). The old macronucleus is degraded and a new one is formed from one of these nuclei. Thus, like in other organisms, the germline is immortal while the soma dies and is replaced every generation. In Paramecium, sexual development can also occur without a partner, in a process known as autogamy. Here, the two haploid pronuclei produced in a developing cell simply fuse with each other, and development continues similarly to in a conjugating cell.

The ciliate germline can be compared to the germlines of more well-known organisms in that it must remain totipotent, with the ability to differentiate into any type of somatic nucleus, but must also manage the risks inherent to totipotency, namely transposon replication and invasion. A ciliate micronucleus contains a large proportion of transposable elements (TEs) [3], many of which may still be active and could replicate and reinsert themselves if not strictly controlled. The way that the micronucleus prevents TE replication is to remain almost exclusively silent, with no detectable transcription taking place during the vegetative (growth) phase of the life cycle [4]. The macronucleus must, of course, be transcriptionally active and is protected from transposon invasion through an unusual mechanism – all the transposons and transposon-derived sequences are removed by excision during the development of the macronucleus from the micronucleus. This produces a streamlined somatic genome consisting almost entirely of genes and gene regulatory elements.

Genome rearrangement: The development of a new macronucleus is a staggeringly complex feat of DNA excision and repair. Apart from transposable elements, microsatellites and other and repetitive DNA, ciliate germlines also contain large numbers of non-coding sequences termed Internal Eliminated Sequences (IESs), which are remnants of ancient TEs

and which frequently interrupt coding genes. In Tetrahymena, 34% of the micronuclear genome is removed in around 12,000 sections, while at the same time the chromosomes are fragmented via de novo telomere addition and amplified to produce ~200 macronuclear chromosomes, at a ploidy of 45n, from the five diploid chromosomes present in the micronucleus [5][6][7]. This all happens accurately and reproducibly over a period of 12 hours. In Paramecium, the same process removes 25% of the germline genome in over 45,000 fragments, while amplifying the ploidy to ~800n [8]. Perhaps the most impressive genome rearrangement occurs in the spirotrichea such as Oxytricha, in which 95% of the micronuclear genome is removed and the remaining 5% is reassembled into 16,000 tiny nanochromosomes, most of which contain a single gene each. Additionally, around 20% of Oxytricha genes are scrambled in the micronucleus, meaning that they exist in several fragments that may not be adjacent or even in the correct order, and must be rearranged to form functional genes in the new macronucleus [9]. The sheer scale and complexity of the excision and repair that must occur during macronuclear development in ciliates is quite remarkable, and the questions of how the cell can identify which sequences to retain, the precise breakpoints, which broken ends to reanneal etc. is the subject of much study.

RNA-guided genome rearrangements: It has long been known that genome rearrangement configurations are inherited maternally, through a cytoplasmic factor that guides the removal of TEs and IESs based on their presence or absence in the parental macronucleus. This was first demonstrated with crosses performed with the d48 strain, a strain of Paramecium in which the germline is wild-type but the somatic macronucleus harbours a deletion in a region containing the surface antigen A gene. The cells were shown to transmit the deletion maternally through an unknown cytoplasmic factor [10,11]. Later, it was shown in both Paramecium and Tetrahymena that excision of germline-limited sequences could be prevented in the new macronucleus via microinjection of their retained forms into the maternal macronucleus, demonstrating that the maternal inheritance is sequence-dependent [12–14]. In the beginning of the 21st century it was demonstrated that non-coding RNA is the factor that mediates this cytoplasmic inheritance [15,16][17]. Since then, a range of different classes and functions of RNAs have been discovered and their functions in the genome rearrangement process elucidated.

Ciliate non-coding RNAs have pivotal roles in determining which micronuclear sequences to retain in the new macronucleus and in which order. Non-coding RNAs are thus used as a means of information transfer from the old macronucleus to the new developing one, directing the development of the daughter macronucleus to mimic the maternal macronucleus. In other words, information about the state of the maternal soma is passed to the daughter soma without affecting the germline, allowing true transgenerational epigenetic inheritance while maintaining the Weismann barrier. In this review we will discuss the various roles of non-coding RNA in genome rearrangements in ciliates, with focus on recent developments in the three most-studied organisms Paramecium, Tetrahymena and Oxytricha.

IESs, background:

The genome rearrangement pathways in ciliates provide an interesting example of the co-evolution of parasitic DNA and its host, and their interdependence. While the IES excision pathways evolved as a defence against TEs, they also heavily rely on TE and TE-derived sequences for their function. The clearest example of this is the excisases that perform the excision of IESs, they are transposases that have been co-opted by the cell to remove transposons and transposon-derived sequences. The requirements of the transposases impose limitations on the TEs and IESs that can remain in the germline – if an IES cannot be removed during development then it will not persist over evolutionary time. Therefore, IESs in different ciliates have different distinctive characteristics, reflecting the requirements of the excision machineries. The transposases and their requirements for excision will be discussed later. As IESs are the remnants of ancient TEs and the mechanisms for their removal are largely the same, for simplicity we will henceforth refer to all sequences that are excised and removed from the germline during ciliate development as IESs.

IESs characteristics: In Tetrahymena, there are around 12,000 IESs that must be removed during development. They are highly repetitive, with plenty of sequence overlap between different IESs, and range in size from 134 bp to 43.4 kbp (with 85% of IESs between 1 and 10 kbp). The IESs are largely situated in sub-telomeric and centromeric gene-poor regions, and therefore do not interrupt coding sequences [7]. There are also a number of chromosomes that form transiently during sexual development but are not maintained in the vegetative life cycle, these are termed non-maintained chromosomes (NMCs) and contain actively transcribed genes that are important during development [7,18]. In Tetrahymena, IES removal is imprecise, with microheterogeneity and sometimes large variability observed in IES borders [6,7,18]. This is tolerated by the cells due to the absence of IESs in coding regions. However, recently a small class of 12 IESs situated within coding genes was discovered, the excision of which is highly precise so as not to interrupt the coding sequence [19].

In Paramecium, in the range of 45,000 IESs are removed from the germline during development. These are much shorter than the Tetrahymena IESs, ranging from 26 bp to 5 kbp with a median length of 51 bp and a mode (most frequently-found length) of 28 bp [8,20]. 93% of IESs in Paramecium are under 150 bp long and over a third are between 26 and 31 bp. Paramecium IESs are spread all over the genome, including within coding sequences, meaning that their excision has to be extremely precise so as not to interrupt the reading frame of the genes they inhabit. They are flanked by TA dinucleotide repeats, one of which is removed along with the IES and one of which is retained in the macronuclear sequence. They also exhibit a loose end consensus inverted repeat sequence of TAYAG, which varies slightly with the size of the IES and its dependence on small RNAs for excision (see section on piwi-associated RNAs) [20,21]. Interestingly, Paramecium IESs show a sinusoidal length distribution which peaks at 28 bp and then has progressively smaller peaks every ~10.2 bp, except for 38 bp where hardly any IESs are found. While it is not clear exactly what causes this length distribution, it is hypothesised that it has something to do with bending of the DNA during the excision process, as 10.2 bp corresponds to one turn of the double helix in B-form DNA [20,22].

Oxytricha IESs make up 95% of the germline genome and the macronuclear-destined sequences (MDSs) are frequently scrambled and reversed. Due to this scrambling, the smallest Oxytricha IES is 0 bp, as it is simply two adjacent scrambled MDSs. The sizes of Oxytricha IESs vary with whether the MDSs they flank are scrambled or not, scrambled IESs are shorter (median 27 bp) than nonscrambled IESs (median 61 bp). MDSs align by way of flanking 'pointer' sequences, which are between 2 and 20 nt long [3,23,24]. The pointers behave similarly to the TA repeats in Paramecium in that they are located on either end of an MDS, and one pointer is retained in the macronuclear genome after rearrangement while one is removed along with the IES. Like Tetrahymena, Oxytricha has a number of non-maintained chromosomes that are formed and then removed during sexual development. These NMCs express 810 protein-coding genes that are almost exclusively conjugation-specific [3].

Piwi-associated small RNAs in genome rearrangements.

PIWI-associated RNAs, or piRNAs, are a well-studied class of small RNAs found in metazoan germlines and necessary for the protection against TE expansion [25]. In some ciliates, Piwi-bound small RNAs help to direct the elimination of germline-limited sequences including transposable elements, minisatellites and IESs during the development of the new macronucleus [15][26]. In this sense they can be seen as analogous to the piRNAs in germlines of animals, only the latter direct transcriptional silencing and heterochromatinisation of TEs while the former direct their complete removal from the genome. Interestingly, Oxytricha piRNAs act in the opposite way to the scnRNAs of Paramecium and Tetrahymena, targeting sequences for retention rather than elimination. In this section, the IES-targeting small RNAs of Paramecium and Tetrahymena will be discussed and compared in detail first, and Oxytricha piRNAs will be addressed later.

scanRNAs: The scanning model for IES and TE elimination in Paramecium and Tetrahymena was first proposed in 2002 in Tetrahymena [15], and elegantly explains how information about which micronuclear sequences are present in the old maternal macronucleus is transmitted to the developing macronucleus. In the early stages of conjugation, the entire micronucleus is bidirectionally transcribed and cleaved to form small RNAs, termed scanRNAs or scnRNAs. The scnRNAs are bound to PIWI proteins and are transported to the maternal macronucleus, where they are compared against the entire macronuclear genome. The scnRNAs that find a perfect match are removed and degraded. The scnRNAs that fail to find a match, by definition, correspond to sequences not present in the maternal macronucleus and therefore not desired in the new macronucleus. These scnRNAs are sent to the developing macronucleus where they target their matching sequences for elimination [15,16,27,28]. The scnRNA pathway is outlined in figure 1, panels 1 and 2 (Tetrahymena and Paramecium), with the similarities and differences outlined.

scnRNA production: More is known about scnRNA production in Tetrahymena than in Paramecium. In Tetrahymena, the premeiotic micronuclear genome is bidirectionally transcribed by RNA polymerase II (RNApolII) [11,13–15] and then cleaved by a Dicer-like enzyme, Dcl1, into 26-32 nt long RNAs [30][31]. These are then loaded onto the Piwi protein Twi1, 2' O-methylated, and transported to the maternal macronucleus [15,27][32]. Upon

loading of the scnRNA, the passenger strand is cleaved by the Slicer domain of Twi1. The passenger strand removal is necessary both for the stable accumulation of scnRNAs and for the transport of the Twi1-scnRNA complex to the parental macronucleus. The latter is mediated by a protein called Giw1, which binds to the mature Twi1-scnRNA complex and escorts it to the nucleus [33]. scnRNAs in Tetrahymena have certain sequence features including a strong 5' U bias and a weaker A bias 3 bases from the 3' end. These base preferences are indicative of Dicer cleavage, which cuts dsRNA with a 2 nt 3' overhang and a preference for 5' U [34]. The transcription of scnRNA precursors is a well-coordinated process that requires the rapid global activation of a normally completely silent nucleus and occurs shortly prior to chromosome condensation in meiosis I [35]. Recently, developmental-specific components of the transcriptional regulator complex Mediator have been characterised in Tetrahymena and shown to localise to the pre-meiotic micronucleus where they coordinate the burst of transcription associated with scnRNA production [36]. In Tetrahymena, transcription of the micronucleus is not uniform as initially expected, rather it is biased towards repetitive regions and a class of IESs called Type A IESs [21,22]. This preferential transcription of non-macronuclear-destined sequence appears to be directed by the Mediator-associated protein Rib1, which directs RNApolII to pericentromeric and subtelomeric regions where the repetitive regions are located [36].

In Paramecium, the transcription of the micronucleus has not been globally measured, so it is not clear whether transcription is uniform or whether there is a bias towards non-macronuclear destined sequences. However, it is assumed that the transcription is uniform based on the relative abundances of IES-matching and MDS (macronuclear-destined sequence)-matching scnRNAs during early development, along with the fact that in Paramecium, IESs and transposable elements are spread much more uniformly across the genome than in Tetrahymena [20,39]. The cleavage of scnRNA precursors in Paramecium is carried out by two Dicer-like enzymes, Dcl2 and 3, which have complementary functions in producing scnRNAs. Dcl2 is responsible for the scnRNA length, precisely 25 nt, while Dcl3 conveys a sequence cleavage preference of 5'UNG [39,40]. Following their production, Paramecium scnRNAs are loaded onto two Piwi proteins, Ptiwi01 and Ptiwi09, which have 98% identity at the amino acid level and are thought to have identical functions [26,41].

scnRNA selection:

The Piwi-complexed scnRNAs are transported to the maternal macronucleus where the genome-wide comparison event called 'scanning' takes place [27,28,39]. During scanning the scnRNA pool is enriched in IES- and TE-matching scnRNAs, presumed due to a loss of MDS-matching scnRNAs (although an amplification of IES- and TE-matching scnRNAs cannot be ruled out) [38,39]. The details of how this 'scanning' takes place are still mysterious, although it appears that the PIWI-complexed scnRNAs bind to genome-wide transcripts rather than to the genomic DNA itself. This notion is based on experiments in Tetrahymena, where the RNA helicase Ema1p has been shown to be necessary for binding of scnRNA-Twi1 to macronuclear chromatin, and nascent transcripts from the parental macronucleus were detected by RT-PCR [42]. Such transcripts have also been detected in Paramecium and are necessary for the maternal inheritance of excision and retention of certain IESs [28]. Whatever the precise mechanism for scnRNA selection in the old macronucleus, it must be agreed that it is stunningly complicated process, the logistics of which are almost unimaginable. Take Paramecium, arguably the simplest genome rearrangement system with

only ~25% of its 98 Mb germline sequences removed during development. If the entire micronuclear sequence is bidirectionally transcribed only once, it would give rise to 2*98Mb/25nt = 3.92 million unique scnRNAs, each of which must scan 72 Mb worth of macronuclear genome sequence to find its perfect match. Unless there is an unknown mechanism helping to guide scnRNAs to their matching macronuclear sequences, the number of scnRNA-genome interactions required is astronomical. This entire process takes place in a period of 2-4 hours.

scnRNA targeting of DNA for elimination:

The scnRNAs remaining after scanning has taken place are those mapping to regions destined for elimination in the new macronucleus. Still in complex with their Piwi proteins they are transported to the developing macronucleus where they again perform a genome-wide scanning event, this time to find matching sequences to target for elimination. In Tetrahymena, the targeting is fairly well understood and involves heterochromatinisation via methylation of histone H3 lysine 9 and 27 (H3K9 and H3K27) [43][44,45]. This is dependent on the scnRNA-Twi1 complex and on the histone methyltransferase Ezl1, a homologue of Drosophila E(z). The methylation of H3K9 and H3K27 leads to binding and heterochromatin assembly by the HP1 homologue Pdd1. Pdd1 is likely the effector molecule for heterochromatinisation, recruiting a number of downstream factors including the excision machinery which removes the heterochromatinised DNA [46,47]. The heterochromatinised IESs and TEs group into so-called heterochromatin bodies or elimination bodies, which can be seen as distinct foci in the nucleus when staining for heterochromatin or one of the proteins involved in DNA elimination [48][49]. Heterochromatin body formation is dependent on the dephosphorylation of Pdd1, which is hyperphosphorylated upon heterochromatin formation [50,51]. As yet, it is unclear whether the heterochromatin body formation occurs before or after excision of the IESs and TEs has begun, however it seems to be necessary for excision as knockout of proteins involved in heterochromatin body formation leads to retention of IESs in the daughter macronucleus [52–56]. Interestingly, one of these proteins is an RNA-binding protein that forms prion-like aggregates, suggesting that RNA may be involved in the aggregation of heterochromatin into foci [56].

In Paramecium, the histone methyltransferase Ezl1 is also active during sexual development and is necessary for the deposition of H3K9me3 and H3K27me3 marks in new macronuclei [57,58]. In the developing macronucleus, H3K9me3 and H3K27me3 form foci similarly to in Tetrahymena; these foci progressively become smaller and fewer in number as development progresses. Recently, H3K9 and H3K27 marks have been shown to be present on a number of transposable elements [58]. However, unlike in Tetrahymena, in Paramecium the heterochromatin marks have not been shown to directly bind the excision machinery, and so indirect effects such as effects on the expression of relevant genes cannot be ruled out. Notably, in Paramecium the majority of IESs are significantly shorter than the binding footprint of a nucleosome, and many of these short IESs are dependent on Ezl1 for their excision [57]. How histone marks could be responsible for targeting these IESs for elimination is not understood. A chromatin assembly factor called PtCAF-1 is also necessary for H3K9me3 and HeK27me3 deposition in the zygotic macronucleus and has a similar localisation pattern to Ezl1, suggesting that they may interact. Interestingly, PtCAF-1 appears to be necessary for the completion of scanning in the maternal macronucleus; when it is removed the MDS-matching scnRNAs are not eliminated [59].

Late-scnRNAs and iesRNAs: During the TE and IES elimination process, the genomes of Tetrahymena and Paramecium are amplified up to 45n and 800n respectively. In both species, a second class of IES-targeting small RNAs has evolved that is active during the DNA amplification stage of new macronuclear development [37,39]. It is hypothesised that this secondary small RNA pathway has evolved to ensure complete removal of all copies of IESs in a rapidly-amplifying but only partially-rearranged genome. In Tetrahymena, the secondary RNA pathway is named 'Late-scnRNAs' due to the RNAs' production in the later part of the developmental process. They are produced in the zygotic macronucleus from all IESs, and are around 29 nt in length [37]. Late-scnRNAs are thought to be produced by Dcl1 and they bind to Twi1, however, interestingly, they only bind to the zygotic Twi1 as opposed to early scnRNAs which bind maternal Twi1. The discovery of the late-scnRNAs led to the division of Tetrahymena IESs into two groups, termed type A IESs and type B IESs. Type A IESs produce early-scnRNAs while type B IESs produce no scnRNAs in early development. In late development, both type A and type B IESs produce late-scnRNAs in an early-scnRNA-dependent manner, so if the early-scnRNA pathway is silenced then no late-scnRNAs are produced. A loss of late-scnRNAs leads to the failure to excise type B IESs. This can be explained by sequence overlap between the two types of IESs: a subset of scnRNAs from type A IESs will bind to type B IESs in *trans* and initiate heterochromatinisation, which leads to production of late-scnRNAs by an unknown mechanism. Late-scnRNA production has been shown to take place prior to IES excision and is not excision-dependent [60]. In some ways the production of secondary scnRNAs through targeting of genomic regions by primary scnRNAs can be seen as analogous to the secondary piRNA production in flies and mice (for review see [61]), where transposon transcripts are targeted by primary piRNAs and cleaved to form secondary piRNAs. However, there are some key differences between the systems: for example, in ciliates the Late-scnRNA precursors are transcribed only in response to targeting by Early-scnRNAs, as opposed to being active transposon mRNA, and are cleaved by a Dicer-like enzyme rather than by the primary piRNA-bound Argonaute.

In Paramecium, the second class of small RNAs is named iesRNAs, as the RNAs exclusively bind to IESs and TEs. Their length varies from around 26-31 nt, and they have a weak end consensus sequence of 5'UAG. Like in Tetrahymena, iesRNAs are produced in the developing zygotic macronucleus and are necessary for the excision of a subset of IESs. In contrast to Tetrahymena, iesRNAs are produced from all IESs, are cleaved by a distinct Dicer-like enzyme, Dcl5, and bind to their own Piwi proteins, Ptiwi10 and Ptiwi11 [39–41]. Their production is rather interesting: following scnRNA targeting and excision of IESs, the excised IESs are ligated at their ends to provide a circular template for dsRNA transcription and cleavage by Dcl5. The majority of IESs are too short to circularise, these shorter IESs concatenate together end on end until they reach a length whereby they can circularise and provide a template for RNA polymerase [62]. Dcl5 is able to reliably produce iesRNAs that correspond to IESs from these randomly-assembled templates thanks to its sequence cleavage preference, which recognises and cleaves at IES-IES junctions [40].

Through the secondary small RNA pathway, both Tetrahymena and Paramecum establish a positive feedback loop which leads to large numbers of secondary IES-targeting RNAs being produced and ensuring complete removal of all IES sequences in the highly polyploid zygotic macronucleus.

piRNAs in Oxytricha: A class of small piwi-binding RNAs is produced during Oxytricha development and is necessary for correct development of the new macronucleus. In contrast to Tetrahymena and Paramecium, Oxytricha piRNAs are produced in the maternal macronucleus and identify sequences to be retained, rather than sequences to be excised. This makes sense given that the vast majority (~95%) of the germline is removed during macronuclear development, so identifying the relatively few macronuclear-destined sequences is energetically more efficient. Oxytricha piRNAs are 27 nt long and bind to the Piwi protein Otiwi1 [63,64]. How the piRNAs target sequences for retention is not known, however a model based on recent work in the related ciliate Stylonychia proposes a mechanism based on replication stalling caused by binding of the Piwi-RNA complex. In this model, the bound Piwi-RNA complex prevents replication at macronuclear-destined sequences and thus inhibits replication-dependent H3K27 methylation, which in turn leads to sequence elimination [65].

Non-Piwi RNAs in ciliate development:

Guide RNAs in Oxytricha: Due to the scrambled nature of the Oxytricha micronuclear genome, identifying the sequences that are to be removed versus retained is not sufficient information to produce a functioning macronucleus. As described earlier, Oxytricha IESs and MDSs contain pointer sequences at their ends which help to identify adjacent MDSs [3,24]. However, the pointer sequences are not unique and in some cases are very short, meaning that they do not in themselves provide enough information to reliably unscramble the germline. To guide the rearrangement process, Oxytricha generates long RNAs from the parental macronucleus that therefore correspond to the rearranged genome. These 'guide RNAs' are then transported to the new developing macronucleus where they are necessary for guiding the correct arrangement of MDSs [66,67]. Transcription of the guide RNAs appears to be rely on a specific subunit of RNA polymerase II that arose from a gene duplication in stichotrich ancestors, and has evolved to exclusively transcribe developmental guide RNAs [68]. It is possible to disrupt and alter the arrangement of the zygotic macronuclear genome through injection of alterative guide RNAs, demonstrating that changes in the parental macronucleus can be directly inherited via epigenetic processes. The experiments that showed this indicate that the guide RNAs act as a template or scaffold for host DNA rearrangement, rather than being involved in homologous recombination or similar. This is based on the observation that point substitutions included in injected alternative guide RNAs were generally not transmitted to the alternatively rearranged genomic DNA sequence [66].

Other Oxytricha non-coding RNAs: Interestingly, it was recently shown that Oxytricha TEs and non-repetitive micronucleus-limited sequences are circularised upon excision, and transcribed to form non-coding RNAs, similarly to Paramecium iesRNA production. While it is not yet clear what the Oxytricha RNAs are for, both circularisation and transcription are performed in a non-random fashion and peak during mid-development, suggesting that the RNAs may have a function in genome rearrangement [69].

Recently, a class of small RNAs in Oxytricha has been discovered that regulates DNA copy number during vegetative growth [70]. It was previously known that non-coding RNAs produced during conjugation guide copy number of nanochromosomes in the offspring [71].

Together, these studies illustrate the profound importance of non-coding RNAs for directing every level of genome architecture in Oxytricha.

Excisases and their requirements:

The main enzyme responsible for IES removal in Tetrahymena is the domesticated PiggyBac transposase TPB2, which removes the imprecisely excised majority of IESs [72,73]. TPB2 has been shown to bind heterochromatin, hence its recruitment to heterochromatinised IESs [65], but how the IES boundaries are identified was until recently mysterious. It has been shown that IESs are excised as complete units and circularised, rather than being sliced up into small pieces [74–76]. Thus, a mechanism for identifying IES boundaries and directing TPB2 cleavage must exist. It was known that cis-acting sequences helped to guide the excision from experiments in which shifting of certain IES-flanking sequences led to shifting of excision boundaries, but these sequences varied from IES to IES and no common consensus could be found [73,77–79]. Recently, a genome-wide approach to search for IES-flanking inverted repeat (IR) sequences was undertaken, and it was found that there are several conserved IR sequences that flank different subsets of IESs at similar distances on both sides of the IES [80]. The known IES boundary-defining protein LIA3 was found to aid in the excision of a subset of IESs, and these LIA3-dependent IESs share similar polypurine-rich IR sequences at their boundaries [80,81]. Meanwhile, a second boundary-defining protein, Ltl1, was found to be required for the excision of another subset of IESs. These Ltl1-dependent IESs also share an IR sequence at their boundaries that is distinct from the LIA3-dependent IESs' IR sequence [82]. Together, these studies suggest a model for IES boundary recognition that involves a number of boundary-defining proteins, each identifying a unique IR sequence flanking a subset of IESs. The boundary-defining proteins recruit or activate TPB2, allowing coordinated cleavage of each end of an IES.
In addition to TPB2, two more PiggyBac-like transposases necessary for genome rearrangement in Tetrahymena have been discovered. These are named TPB1 and TPB6 and together they are responsible for the excision of the 12 IESs located within protein-coding genes, the excision of which is by necessity highly precise [19,83]. Interestingly, TPB6 is found on one of the non-maintained chromosomes that appear transiently during development. The TPB1/TPB6-dependent IESs are flanked by the IR sequence TTAACHCTW, the TTAA from which is retained in the macronuclear genome. This indicates that the excision is similar to a canonical PiggyBac transposase event, although the IR is not alone sufficient for TPB6 and TPB1-mediated excision [83].

The only required end sequence for Paramecium IESs is the flanking TA repeat, however a loose IR consensus of TAYAG exists. This sequence is not necessarily identical at each end of the IES and many IESs exist with widely differing end sequences [8,20,21]. Recently, a protein required for excision of a subset of IESs, with specific end sequences, was discovered [84], raising the possibility that Paramecium may have a system similar to the Lia3-like proteins in Tetrahymena, whereby IESs depend on different proteins for excision depending on their specific end sequences. The enzyme that carries out the excision of IESs is another PiggyBac-like domesticated transposon called PiggyMac (PGM) [85]. PGM cleaves with a 4 nt 5' overhang centred over the flanking TA [86][87], and requires both ends of

each IES to be recognised and bound by both PGM and the non-homologous end-joining (NHEJ) machinery before it can cleave [22,88]. How this remote communication between IES ends occurs is not entirely clear, however it was recently shown that PGM does not perform the IES excision alone, rather it has a large number of “PiggyMac-Like” (PGML) co-factors that are necessary for the PGM-catalysed cleavage to occur [89,90]. One model for IES excision involves a huge complex of PGM and PGML proteins spanning both IES ends and thus coordinating cleavage [90].

It was thought that all Paramecium IESs are dependent on PGM for their excision, however recent progress in assembling the germline genome has called this into question, as a number of TEs and other non-coding elements were discovered in the germline genome assembly that are not present in genomic DNA isolated from PGM-silenced cells [8]. If there are non-PGM-dependent IESs, their excisase has yet to be found. So far, no NMCs such as the one carrying TPB6 have not been discovered in Paramecium, however their existence is a possibility that cannot yet be ruled out.

In Oxytricha, the DNA cleavage during genome rearrangement is carried out by transposases from the TBE (telomere-bearing element) families, which are Tc1/mariner transposons encoding three open reading frames including a 42 kD transposase [91]. Analysis of the micronuclear genome found 10,109 complete TBEs, clustering into four families [3,92]. The transposons are expressed during conjugation, and silencing through RNAi by feeding leads to failure to excise IESs and rearrange the genome [93]. This was shown by silencing the transposases using twelve different RNAi silencing constructs corresponding to a number of different identified versions of the TBE tranposase genes, meaning that due to high sequence similarity a large number of transposases would be silenced. It is not known what proportion of the transposases is required for genome rearrangement, but importantly silencing of all four families gave a much stronger effect than individual silencings [93]. It seems plausible that a large number if not all of the functional transposases are involved in the genome rearrangement process. This indicates that in Oxytricha, rather than domesticating a single or few transposases and upregulating them during development like Paramecium and Tetrahymena, the cells relies on the germline-limited expression of thousands of complete transposable elements, which mediate their own removal from the germline along with that of other IESs and unwanted DNA. This is a good example of a mutually beneficial relationship between parasitic DNA and its host [94].
The global expression of thousands of transposable elements raises the question of why this does not lead to reintegration of further copies of transposons into the germline at each sexual cycle. The answer to this is not clear, but may be related to the rapid circularisation of excised TEs and IESs.

Summary: As has been shown, the developmental processes of ciliates have many broad similarities, although they differ in the mechanistic details of how they remove non-coding and repetitive DNA from their somatic genomes. The similarities include transposon domestication and/or mutualism, the importance of Piwi-associated small RNAs, and the use of the old maternal macronucleus as a template for the formation of the new zygotic macronucleus. The latter makes ciliates a fascinating model for the study of epigenetic inheritance of acquired characteristics, as the maternal macronucleus is somewhat plastic over vegetative divisions [95]. The mechanistic differences in the pathways between

different ciliates often reflects their different structures and requirements. For example, the piRNAs in Oxytricha target sequences for retention rather than elimination, reflecting the relatively small proportion of sequences retained in Oxytricha (~5%) compared to Paramecium and Tetrahymena (~75% and 65% respectively). Even in the more closely related species Tetrahymena and Paramecium, mechanistic differences reflect differences in IES localisation and size. For example, Late-scnRNAs in Tetrahymena can be directly transcribed from unexcised IESs, whereas in Paramecium the short IES size means that to avoid excessive MDS transcription it is more efficient to concatenate already-excised short IESs and transcribe them into iesRNAs.

Importantly, at every step in the genome rearrangement pathways, non-coding RNAs are crucial, and often have homologous functions and binding proteins even in widely divergent ciliate species. This makes ciliates a useful model organism for discovering novel uses for RNA, and demonstrates the importance of non-coding RNA in ancient eukaryotes.

[1] G. Herrick, Developmental Genome Reorganization in Ciliated Protozoa: The Transposon Link, Progress in Nucleic Acid Research and Molecular Biology. 56 (1997) 1–62. https://doi.org/10.1016/S0079-6603(08)61001-6.

[2] X. Maurer-Alcalá, M. Nowacki, Evolutionary origins and impacts of genome architecture in ciliates, Annals of the New York Academy of Sciences. 1447 (2019). https://doi.org/10.1111/nyas.14108.

[3] X. Chen, J.R. Bracht, A.D. Goldman, E. Dolzhenko, D.M. Clay, E.C. Swart, D.H. Perlman, T.G. Doak, A. Stuart, C.T. Amemiya, R.P. Sebra, L.F. Landweber, The Architecture of a Scrambled Genome Reveals Massive Levels of Genomic Rearrangement during Development, Cell. 158 (2014) 1187–1198. https://doi.org/10.1016/j.cell.2014.07.034.

[4] M.A. Gorovsky, J. Woodard, STUDIES ON NUCLEAR STRUCTURE AND FUNCTION IN TETRAHYMENA PYRIFORMIS: I. RNA Synthesis in Macro- and Micronuclei, J Cell Biol. 42 (1969) 673–682. https://doi.org/10.1083/jcb.42.3.673.

[5] E. Hamilton, P. Bruns, C. Lin, V. Merriam, E. Orias, L. Vong, D. Cassidy-Hanley, Genome-Wide Characterization of Tetrahymena thermophila Chromosome Breakage Sites. I. Cloning and Identification of Functional Sites, Genetics. 170 (2005) 1611–1621. https://doi.org/10.1534/genetics.104.031401.

[6] J.N. Fass, N.A. Joshi, M.T. Couvillion, J. Bowen, M.A. Gorovsky, E.P. Hamilton, E. Orias, K. Hong, R.S. Coyne, J.A. Eisen, D.L. Chalker, D. Lin, K. Collins, Genome-Scale Analysis of Programmed DNA Elimination Sites in Tetrahymena thermophila, G3: Genes|Genomes|Genetics. 1 (2011) 515–522. https://doi.org/10.1534/g3.111.000927.

[7] E.P. Hamilton, A. Kapusta, P.E. Huvos, S.L. Bidwell, N. Zafar, H. Tang, M. Hadjithomas, V. Krishnakumar, J.H. Badger, E.V. Caler, C. Russ, Q. Zeng, L. Fan, J.Z. Levin, T. Shea, S.K. Young, R. Hegarty, R. Daza, S. Gujja, J.R. Wortman, B.W. Birren, C. Nusbaum, J. Thomas, C.M. Carey, E.J. Pritham, C. Feschotte, T. Noto, K. Mochizuki, R. Papazyan, S.D. Taverna, P.H. Dear, D.M. Cassidy-Hanley, J. Xiong, W. Miao, E. Orias, R.S. Coyne, Structure of the germline genome of Tetrahymena thermophila and relationship to the massively rearranged somatic genome, ELife. 5 (2016) e19090. https://doi.org/10.7554/eLife.19090.

[8] F. Guérin, O. Arnaiz, N. Boggetto, C. Denby Wilkes, E. Meyer, L. Sperling, S. Duharcourt, Flow cytometry sorting of nuclei enables the first global characterization of Paramecium germline DNA and transposable elements, BMC Genomics. 18 (2017) 327. https://doi.org/10.1186/s12864-017-3713-7.

[9] V.T. Yerlici, L.F. Landweber, Programmed genome rearrangements in the ciliate Oxytricha, Microbiology Spectrum. 2 (2014) 10.1128/microbiolspec.MDNA3-0025–2014. https://doi.org/10.1128/microbiolspec.MDNA3-0025-2014.

[10] L.M. Epstein, J.D. Forney, Mendelian and non-mendelian mutations affecting surface antigen expression in Paramecium tetraurelia., Mol. Cell. Biol. 4 (1984) 1583. https://doi.org/10.1128/MCB.4.8.1583.

[11] S. KOBAYASHI, S. KOIZUMI, Characterization of Mendelian and Non-Mendelian Mutant Strains by Micronuclear Transplantation in Paramecium tetraurelia, The Journal of Protozoology. 37 (1990) 489–492. https://doi.org/10.1111/j.1550-7408.1990.tb01253.x.

[12] D.L. Chalker, 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. 16 (1996) 3658. https://doi.org/10.1128/MCB.16.7.3658.

[13] S. Duharcourt, A.-M. Keller, E. Meyer, Homology-Dependent Maternal Inhibition of Developmental Excision of Internal Eliminated Sequences inParamecium tetraurelia, Molecular and Cellular Biology. 18 (1998) 7075–7085.

[14] E. Meyer, A. Butler, K. Dubrana, S. Duharcourt, F. Caron, Sequence-specific epigenetic effects of the maternal somatic genome on developmental rearrangements of the zygotic genome in Paramecium primaurelia, Mol Cell Biol. 17 (1997) 3589–3599. https://doi.org/10.1128/mcb.17.7.3589.

[15] K. Mochizuki, N.A. Fine, T. Fujisawa, M.A. Gorovsky, Analysis of a piwi-Related Gene Implicates Small RNAs in Genome Rearrangement in Tetrahymena, Cell. 110 (2002) 689–699. https://doi.org/10.1016/S0092-8674(02)00909-1.

[16] O. Garnier, V. Serrano, S. Duharcourt, E. Meyer, RNA-mediated programming of developmental genome rearrangements in Paramecium tetraurelia, Mol Cell Biol. 24 (2004) 7370–7379. https://doi.org/10.1128/MCB.24.17.7370-7379.2004.

[17] D.L. Chalker, P. Fuller, M.-C. Yao, Communication Between Parental and Developing Genomes During Tetrahymena Nuclear Differentiation Is Likely Mediated by Homologous RNAs, Genetics. 169 (2005) 149. https://doi.org/10.1534/genetics.104.032300.

[18] C.-Y.G. Lin, I.-T. Lin, M.-C. Yao, Programmed Minichromosome Elimination as a Mechanism for Somatic Genome Reduction in Tetrahymena thermophila, PLoS Genet. 12 (2016) e1006403–e1006403. https://doi.org/10.1371/journal.pgen.1006403.

[19] C.-Y. Cheng, J.M. Young, C.-Y.G. Lin, J.-L. Chao, H.S. Malik, M.-C. Yao, The piggyBac transposon-derived genes TPB1 and TPB6 mediate essential transposon-like excision during the developmental rearrangement of key genes in Tetrahymena thermophila, Genes Dev. 30 (2016) 2724–2736. https://doi.org/10.1101/gad.290460.116.

[20] O. Arnaiz, N. Mathy, C. Baudry, S. Malinsky, J.-M. Aury, C. Denby Wilkes, O. Garnier, K. Labadie, B.E. Lauderdale, A. Le Mouël, A. Marmignon, M. Nowacki, J. Poulain, M. Prajer, P. Wincker, E. Meyer, S. Duharcourt, L. Duret, M. Bétermier, L. Sperling, The Paramecium Germline Genome Provides a Niche for Intragenic Parasitic DNA: Evolutionary Dynamics of Internal Eliminated Sequences, PLoS Genetics. 8 (2012) e1002984. https://doi.org/10.1371/journal.pgen.1002984.

[21] E.C. Swart, C.D. Wilkes, P.Y. Sandoval, M. Arambasic, L. Sperling, M. Nowacki, Genome-wide analysis of genetic and epigenetic control of programmed DNA deletion, Nucleic Acids Research. 42 (2014) 8970–8983. https://doi.org/10.1093/nar/gku619.

[22] A. Marmignon, J. Bischerour, A. Silve, C. Fojcik, E. Dubois, O. Arnaiz, A. Kapusta, S. Malinsky, M. Bétermier, Ku-Mediated Coupling of DNA Cleavage and Repair during Programmed Genome Rearrangements in the Ciliate Paramecium tetraurelia, PLoS Genetics. 10 (2014) e1004552. https://doi.org/10.1371/journal.pgen.1004552.

[23] D.M. PRESCOTT, M.L. DuBOIS, Internal Eliminated Segments (IESs) of Oxytrichidae1, Journal of Eukaryotic Microbiology. 43 (1996) 432–441. https://doi.org/10.1111/j.1550-7408.1996.tb04502.x.

[24] L.F. Landweber, T.-C. Kuo, E.A. Curtis, Evolution and assembly of an extremely scrambled gene, Proceedings of the National Academy of Sciences. 97 (2000) 3298–3303. https://doi.org/10.1073/pnas.97.7.3298.

[25] E.-M. Weick, E.A. Miska, piRNAs: from biogenesis to function, Development. 141 (2014) 3458. https://doi.org/10.1242/dev.094037.
[26] K. Bouhouche, J.-F. Gout, A. Kapusta, M. Bétermier, E. Meyer, Functional specialization of Piwi proteins in Paramecium tetraurelia from post-transcriptional gene silencing to genome remodelling, Nucleic Acids Research. (2011). https://doi.org/10.1093/nar/gkq1283.
[27] K. Mochizuki, M.A. Gorovsky, Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement, Genes & Development. 18 (2004) 2068–2073. https://doi.org/10.1101/gad.1219904.
[28] G. Lepère, M. Bétermier, E. Meyer, S. Duharcourt, Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia, Genes & Development. 22 (2008) 1501–1512. https://doi.org/10.1101/gad.473008.
[29] D.L. Chalker, M.-C. Yao, Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila, Genes & Development. 15 (2001) 1287–1298. https://doi.org/10.1101/gad.884601.
[30] C.D. Malone, A.M. Anderson, J.A. Motl, C.H. Rexer, D.L. Chalker, 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. 25 (2005) 9151–9164. https://doi.org/10.1128/MCB.25.20.9151-9164.2005.
[31] K. Mochizuki, M.A. Gorovsky, A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase, Genes & Development. 19 (2005) 77–89. https://doi.org/10.1101/gad.1265105.
[32] H.M. Kurth, K. Mochizuki, 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena, RNA. 15 (2009) 675–685. https://doi.org/10.1261/rna.1455509.
[33] T. Noto, H.M. Kurth, K. Kataoka, L. Aronica, L.V. DeSouza, K.W.M. Siu, R.E. Pearlman, M.A. Gorovsky, K. Mochizuki, The Tetrahymena Argonaute-Binding Protein Giw1p Directs a Mature Argonaute-siRNA Complex to the Nucleus, Cell. 140 (2010) 692–703. https://doi.org/10.1016/j.cell.2010.02.010.
[34] K. Mochizuki, H.M. Kurth, Loading and pre-loading processes generate a distinct siRNA population in Tetrahymena, Biochemical and Biophysical Research Communications. 436 (2013) 497–502. https://doi.org/10.1016/j.bbrc.2013.05.133.
[35] T. SUGAI, K. HIWATASHI, Cytologic and Autoradiographic Studies of the Micronucleus at Meiotic Prophase in Tetrahymena pyriformis, The Journal of Protozoology. 21 (1974) 542–548. https://doi.org/10.1111/j.1550-7408.1974.tb03695.x.
[36] M. Tian, K. Mochizuki, J. Loidl, Non-coding RNA Transcription in Tetrahymena Meiotic Nuclei Requires Dedicated Mediator Complex-Associated Proteins, Current Biology. 29 (2019) 2359-2370.e5. https://doi.org/10.1016/j.cub.2019.05.038.
[37] T. Noto, K. Kataoka, J.H. Suhren, A. Hayashi, K.J. Woolcock, M.A. Gorovsky, K. Mochizuki, Small-RNA-Mediated Genome-wide trans-Recognition Network in Tetrahymena DNA Elimination, Molecular Cell. 59 (n.d.) 229–242. https://doi.org/10.1016/j.molcel.2015.05.024.
[38] U.E. Schoeberl, H.M. Kurth, T. Noto, K. Mochizuki, Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in Tetrahymena, Genes & Development. 26 (2012) 1729–1742. https://doi.org/10.1101/gad.196493.112.

[39] P.Y. Sandoval, E.C. Swart, M. Arambasic, M. Nowacki, Functional Diversification of Dicer-like Proteins and Small RNAs Required for Genome Sculpting, Developmental Cell. 28 (2014) 174–188. https://doi.org/10.1016/j.devcel.2013.12.010.

[40] C. Hoehener, I. Hug, M. Nowacki, Dicer-like Enzymes with Sequence Cleavage Preferences, Cell. 173 (2018) 234-247.e7. https://doi.org/10.1016/j.cell.2018.02.029.

[41] D.I. Furrer, E.C. Swart, M.F. Kraft, P.Y. Sandoval, M. Nowacki, Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA, Cell Rep. 20 (2017) 505–520. https://doi.org/10.1016/j.celrep.2017.06.050.

[42] L. Aronica, J. Bednenko, T. Noto, L.V. DeSouza, K.W.M. Siu, J. Loidl, R.E. Pearlman, M.A. Gorovsky, K. Mochizuki, Study of an RNA helicase implicates small RNA–noncoding RNA interactions in programmed DNA elimination in Tetrahymena, Genes & Development. 22 (2008) 2228–2241. https://doi.org/10.1101/gad.481908.

[43] Y. Liu, K. Mochizuki, 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. 101 (2004) 1679–1684. https://doi.org/10.1073/pnas.0305421101.

[44] S.D. Taverna, R.S. Coyne, C.D. Allis, Methylation of Histone H3 at Lysine 9 Targets Programmed DNA Elimination in Tetrahymena, Cell. 110 (n.d.) 701–711. https://doi.org/10.1016/S0092-8674(02)00941-8.

[45] Y. Liu, S.D. Taverna, T.L. Muratore, J. Shabanowitz, D.F. Hunt, C.D. Allis, RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena, Genes & Development. 21 (2007) 1530–1545. https://doi.org/10.1101/gad.1544207.

[46] M.T. Madireddi, R.S. Coyne, J.F. Smothers, K.M. Mickey, M.-C. Yao, C.D. Allis, Pdd1p, A Novel Chromodomain-Containing Protein, Links Heterochromatin Assembly and DNA Elimination in Tetrahymena, Cell. 87 (1996) 75–84. https://doi.org/10.1016/S0092-8674(00)81324-0.

[47] R.M. Schwope, D.L. Chalker, Mutations in Pdd1 reveal distinct requirements for its chromodomain and chromoshadow domain in directing histone methylation and heterochromatin elimination, Eukaryot Cell. 13 (2014) 190–201. https://doi.org/10.1128/EC.00219-13.

[48] M.T. Madireddi, M.C. Davis, C.D. Allis, Identification of a Novel Polypeptide Involved in the Formation of DNA-Containing Vesicles during Macronuclear Development in Tetrahymena, Developmental Biology. 165 (1994) 418–431. https://doi.org/10.1006/dbio.1994.1264.

[49] D.L. Chalker, Dynamic nuclear reorganization during genome remodeling of Tetrahymena, Biochim Biophys Acta. 1783 (2008) 2130–2136. https://doi.org/10.1016/j.bbamcr.2008.07.012.

[50] K. Kataoka, K. Mochizuki, Phosphorylation of an HP1-like Protein Regulates Heterochromatin Body Assembly for DNA Elimination, Dev Cell. 35 (2015) 775–788. https://doi.org/10.1016/j.devcel.2015.11.017.

[51] K. Kataoka, T. Noto, K. Mochizuki, Phosphorylation of an HP1-like protein is a prerequisite for heterochromatin body formation in Tetrahymena DNA elimination, Proc Natl Acad Sci U S A. 113 (2016) 9027–9032. https://doi.org/10.1073/pnas.1606012113.

[52] M.A. Nikiforov, J.F. Smothers, M.A. Gorovsky, C.D. Allis, Excision of micronuclear-specific DNA requires parental expression of Pdd2p andoccursindependentlyfromDNA replication in Tetrahymena thermophila, Genes & Development. 13 (1999) 2852–2862.

[53] C.H. Rexer, D.L. Chalker, Lia1p, a Novel Protein Required during Nuclear Differentiation for Genome-Wide DNA Rearrangements in Tetrahymena thermophila, Eukaryot Cell. 6 (2007) 1320. https://doi.org/10.1128/EC.00157-07.
[54] A.W.Y. Shieh, D.L. Chalker, LIA5 Is Required for Nuclear Reorganization and Programmed DNA Rearrangements Occurring during Tetrahymena Macronuclear Differentiation, PLoS ONE. 8 (2013) e75337. https://doi.org/10.1371/journal.pone.0075337.
[55] S.A. Horrell, D.L. Chalker, LIA4 Encodes a Chromoshadow Domain Protein Required for Genomewide DNA Rearrangements in Tetrahymena thermophila, Eukaryot Cell. 13 (2014) 1300. https://doi.org/10.1128/EC.00125-14.
[56] K. Kataoka, K. Mochizuki, Heterochromatin aggregation during DNA elimination in Tetrahymena is facilitated by a prion-like protein, J. Cell Sci. 130 (2017) 480. https://doi.org/10.1242/jcs.195503.
[57] M. Lhuillier-Akakpo, A. Frapporti, C. Denby Wilkes, M. Matelot, M. Vervoort, L. Sperling, S. Duharcourt, Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements, PLoS Genet. 10 (2014) e1004665–e1004665. https://doi.org/10.1371/journal.pgen.1004665.
[58] A. Frapporti, C. Miró Pina, O. Arnaiz, D. Holoch, T. Kawaguchi, A. Humbert, E. Eleftheriou, B. Lombard, D. Loew, L. Sperling, K. Guitot, R. Margueron, S. Duharcourt, The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in Paramecium, Nat Commun. 10 (2019) 2710–2710. https://doi.org/10.1038/s41467-019-10648-5.
[59] M. Ignarski, A. Singh, E.C. Swart, M. Arambasic, P.Y. Sandoval, M. Nowacki, Paramecium tetraurelia chromatin assembly factor-1-like protein PtCAF-1 is involved in RNA-mediated control of DNA elimination, Nucleic Acids Research. 42 (2014) 11952–11964. https://doi.org/10.1093/nar/gku874.
[60] M. Mutazono, T. Noto, K. Mochizuki, Diversification of small RNA amplification mechanisms for targeting transposon-related sequences in ciliates, Proc Natl Acad Sci USA. 116 (2019) 14639. https://doi.org/10.1073/pnas.1903491116.
[61] S.S. Parhad, W.E. Theurkauf, Rapid evolution and conserved function of the piRNA pathway, Open Biology. 9 (n.d.) 180181. https://doi.org/10.1098/rsob.180181.
[62] S.E. Allen, I. Hug, S. Pabian, I. Rzeszutek, C. Hoehener, M. Nowacki, Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs, Cell. 168 (2017) 990-999.e7. https://doi.org/10.1016/j.cell.2017.02.020.
[63] A.M. Zahler, Z.T. Neeb, A. Lin, S. Katzman, Mating of the Stichotrichous Ciliate Oxytricha trifallax Induces Production of a Class of 27 nt Small RNAs Derived from the Parental Macronucleus, PLoS ONE. 7 (2012) e42371. https://doi.org/10.1371/journal.pone.0042371.
[64] W. Fang, X. Wang, J.R. Bracht, M. Nowacki, L.F. Landweber, Piwi-Interacting RNAs Protect DNA Against Loss During Oxytricha Genome Rearrangement, Cell. 151 (2012) 1243–1255. https://doi.org/10.1016/j.cell.2012.10.045.
[65] J. Postberg, F. Jönsson, P.P. Weil, A. Bulic, S.A. Juranek, H.-J. Lipps, 27nt-RNAs guide histone variant deposition via “RNA-induced DNA replication interference” and thus transmit parental genome partitioning in Stylonychia, Epigenetics Chromatin. 11 (2018) 31–31. https://doi.org/10.1186/s13072-018-0201-5.

[66] M. Nowacki, V. Vijayan, Y. Zhou, K. Schotanus, T.G. Doak, L.F. Landweber, RNA-mediated epigenetic programming of a genome-rearrangement pathway, Nature. 451 (2008) 153–158. https://doi.org/10.1038/nature06452.

[67] K.A. Lindblad, J.R. Bracht, A.E. Williams, L.F. Landweber, Thousands of RNA-cached copies of whole chromosomes are present in the ciliate Oxytricha during development, RNA. 23 (2017) 1200–1208. https://doi.org/10.1261/rna.058511.116.

[68] J.S. Khurana, X. Wang, X. Chen, D.H. Perlman, L.F. Landweber, Transcription-independent functions of an RNA polymerase II subunit, Rpb2, during genome rearrangement in the ciliate, Oxytricha trifallax, Genetics. 197 (2014) 839–849. https://doi.org/10.1534/genetics.114.163279.

[69] V.T. Yerlici, M.W. Lu, C.R. Hoge, R.V. Miller, R. Neme, J.S. Khurana, J.R. Bracht, L.F. Landweber, Programmed genome rearrangements in Oxytricha produce transcriptionally active extrachromosomal circular DNA, Nucleic Acids Research. 47 (2019) 9741–9760. https://doi.org/10.1093/nar/gkz725.

[70] J.S. Khurana, D.M. Clay, S. Moreira, X. Wang, L.F. Landweber, Small RNA-mediated regulation of DNA dosage in the ciliate Oxytricha, RNA. 24 (2018) 18–29. https://doi.org/10.1261/rna.061333.117.

[71] M. Nowacki, J.E. Haye, W. Fang, V. Vijayan, L.F. Landweber, RNA-mediated epigenetic regulation of DNA copy number, Proc Natl Acad Sci U S A. 107 (2010) 22140–22144. https://doi.org/10.1073/pnas.1012236107.

[72] C.-Y. Cheng, A. Vogt, K. Mochizuki, M.-C. Yao, A Domesticated piggyBac Transposase Plays Key Roles in Heterochromatin Dynamics and DNA Cleavage during Programmed DNA Deletion in Tetrahymena thermophila, Molecular Biology of the Cell. 21 (2010) 1753–1762. https://doi.org/10.1091/mbc.E09-12-1079.

[73] A. Vogt, K. Mochizuki, A Domesticated PiggyBac Transposase Interacts with Heterochromatin and Catalyzes Reproducible DNA Elimination in Tetrahymena, PLoS Genetics. 9 (2013) e1004032. https://doi.org/10.1371/journal.pgen.1004032.

[74] S.V. Saveliev, M.M. Cox, The fate of deleted DNA produced during programmed genomic deletion events in Tetrahymena thermophila., Nucleic Acids Research. 22 (1994) 5695–5701.

[75] M.C. Yao, C.H. Yao, Detection of circular excised DNA deletion elements in Tetrahymena thermophila during development., Nucleic Acids Research. 22 (1994) 5702–5708.

[76] S.V. Saveliev, M.M. Cox, Product analysis illuminates the final steps of IES deletion in Tetrahymena thermophila, The EMBO Journal. 20 (2001) 3251–3261. https://doi.org/10.1093/emboj/20.12.3251.

[77] D.L. Chalker, A. La Terza, A. Wilson, C.D. Kroenke, M.-C. Yao, Flanking Regulatory Sequences of the Tetrahymena R Deletion Element Determine the Boundaries of DNA Rearrangement, Molecular and Cellular Biology. 19 (1999) 5631–5641.

[78] J.S. Fillingham, D. Bruno, 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. 29 (2001) 488–498. https://doi.org/10.1093/nar/29.2.488.

[79] R. Godiska, C. James, M.C. Yao, A distant 10-bp sequence specifies the boundaries of a programmed DNA deletion in Tetrahymena., Genes & Development. 7 (1993) 2357–2365. https://doi.org/10.1101/gad.7.12a.2357.

[80] C.-Y.G. Lin, J.-L. Chao, H.-K. Tsai, D. Chalker, M.-C. Yao, Setting boundaries for genome-wide heterochromatic DNA deletions through flanking inverted repeats in Tetrahymena thermophila, Nucleic Acids Res. 47 (2019) 5181–5192. https://doi.org/10.1093/nar/gkz209.

[81] C.M. Carle, H.S. Zaher, D.L. Chalker, A Parallel G Quadruplex-Binding Protein Regulates the Boundaries of DNA Elimination Events of Tetrahymena thermophila, PLoS Genet. 12 (2016) e1005842–e1005842. https://doi.org/10.1371/journal.pgen.1005842.

[82] V.N. Jaspan, M.E. Taye, C.M. Carle, J.J. Chung, D.L. Chalker, Boundaries of eliminated heterochromatin of Tetrahymena are positioned by the DNA-binding protein Ltl1, Nucleic Acids Res. 47 (2019) 7348–7362. https://doi.org/10.1093/nar/gkz504.

[83] L. Feng, G. Wang, E.P. Hamilton, J. Xiong, G. Yan, K. Chen, X. Chen, W. Dui, A. Plemens, L. Khadr, A. Dhanekula, M. Juma, H.Q. Dang, G.M. Kapler, E. Orias, W. Miao, Y. Liu, A germline-limited piggyBac transposase gene is required for precise excision in Tetrahymena genome rearrangement, Nucleic Acids Res. 45 (2017) 9481–9502. https://doi.org/10.1093/nar/gkx652.

[84] S. Bhullar, C. Denby Wilkes, O. Arnaiz, M. Nowacki, L. Sperling, E. Meyer, A mating-type mutagenesis screen identifies a zinc-finger protein required for specific DNA excision events in Paramecium, Nucleic Acids Res. 46 (2018) 9550–9562. https://doi.org/10.1093/nar/gky772.

[85] C. Baudry, S. Malinsky, M. Restituito, A. Kapusta, S. Rosa, E. Meyer, M. Bétermier, PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia, Genes & Development. 23 (2009) 2478–2483. https://doi.org/10.1101/gad.547309.

[86] A. Kapusta, A. Matsuda, A. Marmignon, M. Ku, A. Silve, E. Meyer, J.D. Forney, S. Malinsky, M. Bétermier, Highly Precise and Developmentally Programmed Genome Assembly in Paramecium Requires Ligase IV–Dependent End Joining, PLoS Genetics. 7 (2011) e1002049. https://doi.org/10.1371/journal.pgen.1002049.

[87] A. Gratias, M. Bétermier, Processing of Double-Strand Breaks Is Involved in the Precise Excision of Paramecium Internal Eliminated Sequences, Molecular and Cellular Biology. 23 (2003) 7152–7162. https://doi.org/10.1128/MCB.23.20.7152-7162.2003.

[88] A. Gratias, G. Lepère, O. Garnier, S. Rosa, S. Duharcourt, S. Malinsky, E. Meyer, M. Bétermier, Developmentally programmed DNA splicing in Paramecium reveals short-distance crosstalk between DNA cleavage sites, Nucleic Acids Research. 36 (2008) 3244–3251. https://doi.org/10.1093/nar/gkn154.

[89] E. Dubois, N. Mathy, V. Régnier, J. Bischerour, C. Baudry, R. Trouslard, M. Bétermier, Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements, Nucleic Acids Res. 45 (2017) 3204–3216. https://doi.org/10.1093/nar/gkw1359.

[90] J. Bischerour, S. Bhullar, C. Denby Wilkes, V. Régnier, N. Mathy, E. Dubois, A. Singh, E. Swart, O. Arnaiz, L. Sperling, M. Nowacki, M. Bétermier, Six domesticated PiggyBac transposases together carry out programmed DNA elimination in Paramecium, Elife. 7 (2018) e37927. https://doi.org/10.7554/eLife.37927.

[91] T.G. Doak, F.P. Doerder, C.L. Jahn, G. Herrick, A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common “D35E” motif, Proc Natl Acad Sci U S A. 91 (1994) 942–946. https://doi.org/10.1073/pnas.91.3.942.

[92] X. Chen, L.F. Landweber, Phylogenomic analysis reveals genome-wide purifying selection on TBE transposons in the ciliate Oxytricha, Mob DNA. 7 (2016) 2–2. https://doi.org/10.1186/s13100-016-0057-9.

[93] M. Nowacki, B.P. Higgins, G.M. Maquilan, E.C. Swart, T.G. Doak, L.F. Landweber, A Functional Role for Transposases in a Large Eukaryotic Genome, Science (New York, N.Y.). 324 (2009) 935–938. https://doi.org/10.1126/science.1170023.

[94] A. Vogt, A.D. Goldman, K. Mochizuki, L.F. Landweber, Transposon Domestication versus Mutualism in Ciliate Genome Rearrangements, PLoS Genetics. 9 (2013) e1003659. https://doi.org/10.1371/journal.pgen.1003659.

[95] S.E. Allen, M. Nowacki, Necessity Is the Mother of Invention: Ciliates, Transposons, and Transgenerational Inheritance, Trends in Genetics. 33 (2017) 197–207. https://doi.org/10.1016/j.tig.2017.01.005.

Figure 1: The RNA-guided genome rearrangements of three ciliate species. Thick horizontal lines depict chromosomes; IESs are orange, MDSs are black. Long non-coding RNA is represented by wavy lines, small Piwi-bound RNAs are depicted as short horizontal lines. Proteins are mentioned in the legend but not depicted for simplicity.

Left, Tetrahymena. A: dsRNA scnRNA precursors are transcribed bidirectionally from IES-rich A-regions in the meiotic micronucleus. They are cleaved by a Dicer-like enzyme into short, 26-32 nt long Early-scnRNAs which are imported into the maternal macronucleus. B: Here, scnRNAs that match to macronuclear sequence are removed and the remaining IES-matching scnRNAs are transported to the developing zygotic macronucleus. C: in the developing macronucleus, while amplification and chromosome fragmentation are ongoing, scnRNAs target both type A (orange) and type B (green) IESs for heterochromatinisation and elimination, which triggers transcription of Late-scnRNAs from both type A and type B IESs. Late-scnRNAs then further target the amplifying copies of IESs to ensure complete elimination.

Middle, Paramecium: A: scnRNA precursors are transcribed uniformly from the micronuclear genome and cleaved by Dicer-like enzymes Dcl2 and Dcl3 to 25 nt scnRNAs. The scnRNAs are transported by Piwi proteins into the maternal macronucleus, B, where scnRNAs that find matches are removed. The remaining IES-matching sequences are transported to the developing zygotic macronucleus, C. Here they target IESs for excision. Once excised, IESs concatenate end-on-end to form circles, which are transcribed bidirectionally and cleaved by a second Dicer-like enzyme, Dcl5, to form iesRNAs. iesRNAs then target amplifying copies of IESs to ensure complete excision. Meanwhile, chromosomes are fragmented and telomerised.

Right, Oxytricha: A: The Oxytricha macronucleus contains a huge number of IESs and its genes are scrambled, depicted as numbered/lettered gene fragments out of order. B: Both piRNA precursors and guide RNAs are produced from bidirectional transcription of the short nanochromosomes in the maternal macronucleus. They are transported to the zygotic macronucleus where the piRNAs target macronuclear-destined sequences for retention, and IESs are removed. The guide RNAs help to arrange the macronuclear-destined sequences in the correct order on gene-sized chromosomes.