universität wien

DISSERTATION

DNA elimination in the ciliate Tetrahymena – Excision of chromatin by the domesticated *piggyBac* like transposase Tpb2p

Verfasser

Diplom Biologe Alexander Vogt

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2012

Studienkennzahl lt. Studienblatt:	A 091 490
Dissertationsgebiet lt. Studienblatt:	Molekulare Biologie
Betreuer:	Dr. Kazufumi Mochizuki

Table of Contents

1. Abstract 3

2. Zusammenfassung 4

3. Introduction 5

3.1 *Tetrahymena thermophila* as a model organism 5

3.2 Life cycle of *Tetrahymena thermophila* 6

3.3 Programmed genome rearrangements during the macronuclear development 8

3.4 The IES excision process during programmed genome rearrangements 11

3.4.1 Origin of IESs in Tetrahymena 11

3.4.2 Regulation of IES excision and the excision machinery 11

3.5 Aim of the thesis 13

4. Publications 14

4.1 A Domesticated *piggyBac* Transposase plays Key Roles in Heterochromatin dynamics and DNA cleavage during Programmed DNA Deletion in *Tetrahymena thermophila* 14

4.2 Transposon Domestication versus Mutualism in Ciliate Genome Rearrangements 25

4.3 Local chromatin modifications and a sequence preference of the catalytic activity of the domesticated *piggyBac* transposase Tpb2p enable precise excision of germline limited sequences in Tetrahymena **43**

5. Discussion **80**

6. Literature **84**

7. Abbreviations **87**

8. Curriculum vitae **89**

9. Acknowledgements **90**

1. Abstract

During sexual reproduction the new developing macronucleus of *Tetrahymena* undergoes massive programmed DNA rearrangement, where over 30 % of the genome is eliminated. This process involves the sequence specific recognition of internal eliminated sequences by an Argonaute-scan RNA complex followed by heterochromatin formation including the accumulations of methylated histone H3 at lysine 9 and lysine 27 and the chromodomain protein Pdd1p. This heterochromatin formation eventually leads to the formation of distinct heterochromatin bodies in which DNA elimination is believed to occur.

In my thesis I demonstrated that the *piggyBac*-like transposase Tpb2p is recruited to the heterochromatin bodies and that it is essential for DNA elimination. Furthermore, the recombinantly expressed Tpb2p from bacteria can recognize and cut boundaries of different internal eliminated sequences in vitro when they are placed in the middle of an artificially designed and synthesized oligo. Thus Tpb2p probably introduces the initial double strand break during DNA excision. To get insight into how the precision of excision is achieved, I first analyzed the nuclease activity of recombinant Tpb2p in more detail. Using synthesized oligo DNAs where every position of the reported left boundary of the well studied R element (sequence AGTGAT) was individually mutated, I found that the third and fourth positions in the boundary sequence are important for efficient cleavage by Tpb2p. Furthermore, an in vivo study confirmed that these two positions were crucial for the precise elimination of the R IES element. Therefore, some DNA sequence preference of Tpb2p clearly contributes to the precise elimination of internal eliminated sequences.

On the other hand, because Tpb2p is a component of heterochromatin and is required for heterochromatin body formation, heterochromatin interaction with Tpb2p might also be involved in the precise DNA elimination. Tpb2p has an endonuclease domain and a zinc finger domain. I found that the zinc finger domain, but not the endonuclease domain, was essential for heterochromatin body formation. I could show in vitro that the zinc finger domain binds to histone H3 peptides that are tri-methylated at lysines 9 or 27 suggesting that these modifications specific to internal eliminated sequences are - together with the sequence preference of Tpb2p- specifying the precise IES excision.

2. Zusammenfassung

Während der sexuellen Fortpflanzung finden im neu entstehenden Makronukleus von Tetrahymena weitreichende Umgestaltungen des Genoms statt, bei denen mehr als 30 % des Genoms eliminiert werden. Dieser Prozess beinhaltet die spezifische Erkennung interner eliminierter Sequenzen durch einen Argonaut-scanRNA Komplex und die darauf folgende Heterochromatinbildung. Diese führt zur Akkumulierung der methylierten Lysine 9 und 27 des Histons H3 woraufhin das Chromodomänen Protein Pdd1p rekrutiert wird. Im Anschluss daran bilden sich sichtbar abgegrenzte Heterochromatinkörperchen von denen man davon ausgeht, dass in ihnen die DNS Eliminierung stattfindet.

In meiner Doktorarbeit konnte ich zeigen, dass die *piggyBac*-ähnliche Transposase Tpb2p in diese Heterochromatinkörperchen rekrutiert wird und dass sie für die DNS Eliminierung unerlässlich ist. Zudem erkennt und schneidet das rekombinant in Bakterien hergestellte Tpb2p die Grenzen verschiedener interner eliminierter Sequenzen in vitro, wenn diese in der Mitte einer künstlich ausgearbeiteten und synthetisierten oligo DNS vorhanden sind. Daher führt Tpb2p wahrscheinlich auch den initialen Doppelstrangbruch während des Ausschneidens der internen eliminierten Sequenzen aus. Um einen Einblick zu erhalten wie die Präzission während des Ausschneidens erreicht wird, habe ich die Nuklease Aktivität des rekombinanten Tpb2p genauer untersucht. Dafür verwendete ich synthetisierte oligo DNS in der jede Position der linken Grenze des gut untersuchten R-Elements (Sequenz AGTGAT) individuell mutiert wurde. Ich konnte zeigen, dass die Positionen drei und vier wichtig für ein effektives Schneiden der Sequenz durch Tpb2p sind. Des Weiteren konnte ich veranschaulichen, dass diese Positionen auch für die präzise Eliminierung des R-Elements wichtig sind. Daher trägt die Präferenz von Tpb2p für bestimmte DNS Sequenzen sicherlich zur genauen Eliminierung der internen eliminierten Sequenzen bei.

Andererseits lokalisiert Tpb2p in den Heterochromatinkörperchen und ist für deren Ausbildung essentiell. Daher ist es auch möglich, dass Tpb2p direkt mit dem Heterochromatin interagiert, und dass diese Interaktion die präzise Eliminierung ermöglicht. Tpb2p hat eine Endonuklease Domäne und eine Zink Finger Domäne. Ich habe herausgefunden, dass die Zink Finger Domäne, allerdings nicht die Endonuklease Domäne essentiell für die Ausbildung der Heterochromatinkörperchen ist. Außerdem konnte ich zeigen, dass die Zink Finger Domäne in vitro an Peptide des Histons H3 bindet, wenn diese an Lysin 9 oder 27 tri-methyliert sind. Dies lässt die Schlussfolgerung zu, dass diese Modifikationen, die spezifisch sind für interne eliminierte Sequenzen – zusammen mit der Sequenz Präferenz von Tpb2p – das präzise Ausschneiden der internen eliminierten Sequenzen vermittelt.

3. Introduction

3.1 *Tetrahymena thermophila* as a model organism

Tetrahymena thermophila is a fresh water swimming ciliated protist. As most ciliates it exhibits nuclear dimorphism (Fig 1B). The smaller diploid micronucleus (MIC) has five chromosomes and represents the germline. The larger polyploid macronucleus (MAC) harbors the somatic genome which has around 300 chromosomes [1]. During vegetative growth the micronuclear genome is silenced and the MAC is providing the cell with most if not all mRNA. The high ploidy levels of the MAC enable a massive messenger RNA production thus making a high division rate possible.

Tetrahymena is used as a model organism in cellular and molecular biology. In both fields ground breaking discoveries had been made using this protist. Examples include the discovery of the first motor protein (dynein) [2], the discovery of self splicing RNAs [3], as well as the first identification of the telomeric repeat sequence and the discovery of the telomerase [4, 5]. More recently it was discovered that an RNAi like pathway is involved in the programmed genome rearrangement in the MAC of Tetrahymena [6-9].

Tetrahymena has various advantages to be used as a model organism. First of all it is possible to synchronize cell division in liquid culture which makes it easy to get sufficient material for diverse biochemical and molecular biological applications. Additionally there are robust methods for cell culture and storage as either short term stocks or long term frozen stocks [10]. Furthermore Tetrahymena is amenable for genetics as it is possible to target specific loci by homologous recombination in both nuclei [11-15]. Another important aspect is that the genome is sequenced. Since 2005 the MAC genome is available [16] and just recently the MIC genome was released and is constantly refined (*Tetrahymena* Comparative Sequencing Project, Eli and Edythe L. Broad Institute of Harvard and MIT, www.broadinstitute.org).

3.2 Life cycle of *Tetrahymena thermophila*

When enough nutrients are present, Tetrahymena grows by binary fission where the diploid MIC divides and segregates by closed mitosis. In contrast, the polyploid MAC does not propagate through mitosis and the chromosomes are most likely distributed randomly to the daughter nuclei. Till now the details of this process are poorly understood. However the high chromosome copy number seems to be sufficient to ensure that both daughter nuclei receive at least one copy of each chromosome [1]. An advantage of this amitotic division may be that the chromosomes do not need to condense and therefore gene expression can be easily maintained during the whole cell cycle.

When nutrients are scarce, the cells switch from a vegetative into a sexually reproducing state. During sexual reproduction, which is called conjugation in ciliates, two cells of different mating type form a mating pair and their MICs undergo meiosis (Fig.1A). One of the resulting four haploid nuclei is selected and divides once by mitosis, the other three meiotic products are degraded. After the selected nucleus has finished mitosis one of the resulting pronuclei is exchanged with the mating partner (Fig. 1A). Afterwards the two pronuclei derived from different parental cells fuse and form the zygote which divides twice mitotically resulting in four post-zygotic nuclei. Two of the nuclei differentiate and become macronuclear anlagen, the precursors of mature MACs. In parallel the parental MAC is destroyed by an apoptosis like process. Around this stage the mating pairs separate. To reach the mature state the new MAC genome undergoes massive programmed genome rearrangements. One type of genome rearrangements leads to the cleavage of the five MIC chromosomes into smaller pieces at specific sequences called chromosome breakage sites (CBSs) (Fig 1B). This results in around 300 MAC chromosomes which are endoreplicated around 50 times. Furthermore a large portion of internal sequences (IESs = internal eliminated sequences) are precisely removed (Fig. 1B). This finally leads to a loss of around 30 percent of sequence complexity compared to the MIC. However DNA amount is still higher in the MAC due to the polyploidy. After the programmed genome rearrangements have occurred, one of the two MIC is degraded, resulting in two new

MACs and one MIC in each cell. When nutrients become available again the odd MIC divides by mitosis and the cells undergo cytokinesis. Each daughter cell receives one MIC and one MAC.

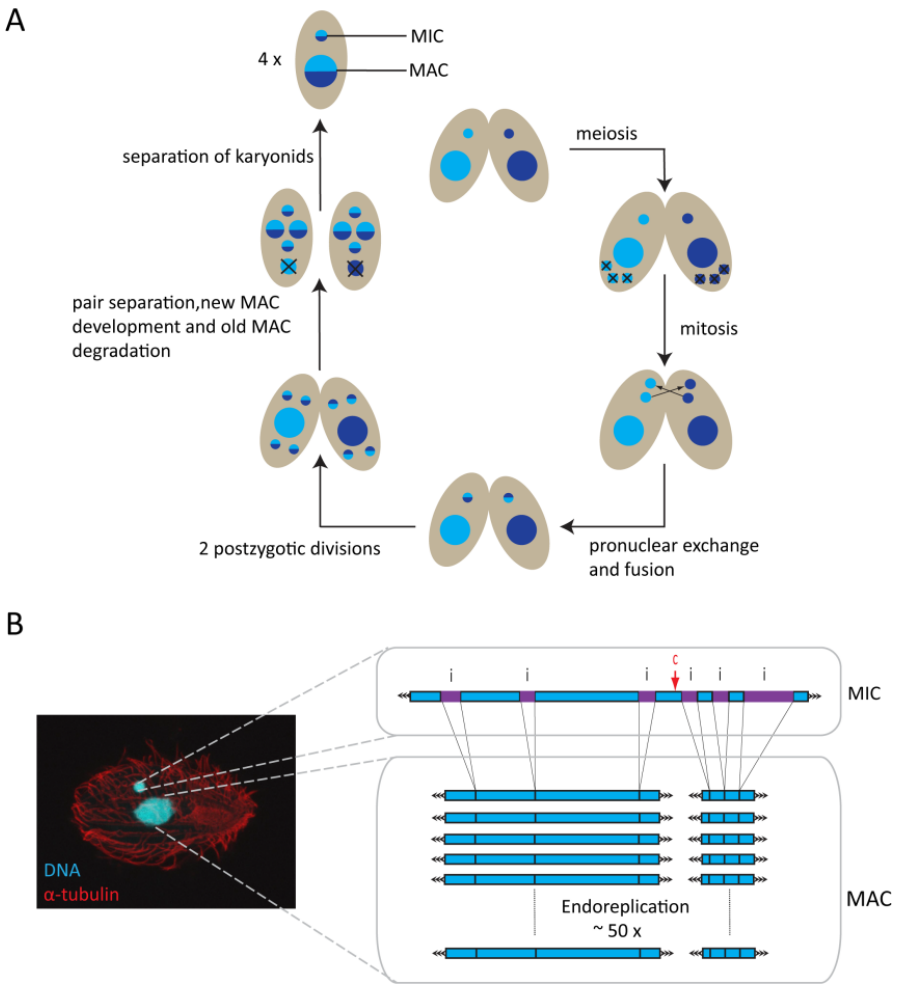

Figure 1: A) Sexual reproduction of Tetrahymena. Two mating partners form a pair and the MIC undergoes meiosis. One nucleus is selected and undergoes mitosis to form the pronuclei. One of them is exchanged with the mating partner where the two pronuclei from different parental cells fuse to form the zygote. This divides twice mitotically and two of the nuclei become MICs the other two become MACs. Afterwards the cells proceed through cytokinesis and the nuclei are equally distributed to the progenies. B) Tetrahymena cells show nuclear dimorphism. The smaller MIC represents the germline and the larger MAC is the soma. The MIC genome has internal eliminated sequences (IES = i) which are removed from the MAC genome. Additionally MIC chromosomes are cleaved into smaller fragments at chromosome breakage sites (CBS = c). These smaller chromosomes are endoreplicated around 50 x in the MAC)

3.3 Programmed genome rearrangements during the macronuclear development

Two types of programmed genome rearrangements can be observed in the MAC anlagen of Tetrahymena. The first type is chromosome breakage and happens at around 250 to 300 positions in the genome at specific sequences which are defined by the 15 bp consensus motif (5' TAAACCAACCTCTTT 3') called chromosome breakage sequence (CBS) [17]. The machinery responsible for this process is still unknown. Between 15 and 60 bp next to the CBS are missing in the resulting MAC chromosomes. This argues for the introduction of a double strand break by a so far unknown endonuclease and the unprotected DNA ends are resected by exonucleolytic digestion until the telomeres are established at the newly created chromosome ends by the de novo telomere synthesis [18].

The second type of programmed genome rearrangement observed in Tetrahymena is the DNA elimination of around 9000 different MIC limited sequences (IESs) (*Tetrahymena* Comparative Sequencing Project, Eli and Edythe L. Broad Institute of Harvard and MIT, www.broadinstitute.org). These sequences range from 600 bp to more than 40 kbp in size. Some of them are highly repetitive in the MIC genome while others are unique sequences [16, 19]. According to their sequences they are likely to be remnants of transposons. The origin of IESs will be discussed in the next section. Since some of IESs maintain complete open reading frames encoding proteins required for transposon activities, they are probably still containing functional transposons. Therefore it is assumed that the DNA elimination is a process removing potentially harmful transposons from the transcriptionally active somatic genome. The targeting of the eliminated DNA is mediated by an RNAi related pathway leading to sequence specific heterochromatin formation on IESs. This process resembles the RNAi-directed heterochromatin formation in other eukaryotes such as fission yeast [20]. Ciliates like Tetrahymena do not maintain heterochromatin but instead proceed to eliminate the heterochromatic regions from the somatic genome.

The DNA elimination pathway starts with the bidirectional transcription of the MIC genome [21]. The resulting long double stranded RNA is processed into 27-30 nt long small RNAs called

scanRNAs (scnRNAs) by the Dicer like enzyme Dcl1p [6, 9]. These small RNAs are transported into the cytoplasm where they associate with an Argonaute protein of the Piwi clade, Twi1p [22]. After removal of the passenger strand mature Argonaute-scnRNA complexes are transported into the parental MAC. The import is tightly regulated in a way that just the complex with a single stranded scnRNA can enter the nucleus. This is mediated by Giw1p, a gatekeeper to the nucleus [23]. In the parental MAC a scanning process is happening, where effector complexes carrying scnRNAs complementary to IESs are enriched compared to those having scnRNAs complementary to macronucleus destined sequences (MDSs) [21]. The scanning is believed to involve the interaction of the Argonaute-scnRNAs complex with nascent transcripts [24]. However mechanistic details of this scanning process are poorly understood. Probably a complex with a scnRNA complementary to a MDS is degraded [21]. The remaining Argonaute-scnRNA complexes are transported into the new developing MAC where they lead to sequence specific heterochromatin formation on IESs [25]. This includes the methylation of histone H3 at lysines 9 and 27 by the histone methyltransferase Ezl1p [26, 27]. These modifications are then bound by the chromodomain protein Pdd1p [28]. During anlagen maturation distinct heterochromatic foci appear in which many different proteins essential for DNA elimination (e.g. Pdd1p, Pdd3p, Lia1p) and the non dispensable histone marks H3K9/K27me are localized [26-30]. Although DNA elimination is believed to happen in the heterochromatic foci, it was not clear which enzyme mediate DNA elimination and how this enzyme is targeted to the heterochromatin. Together with a collaborating group, the domesticated *piggyBac* like transposase Tpb2p was found to be required for DNA elimination and characterized during this thesis.

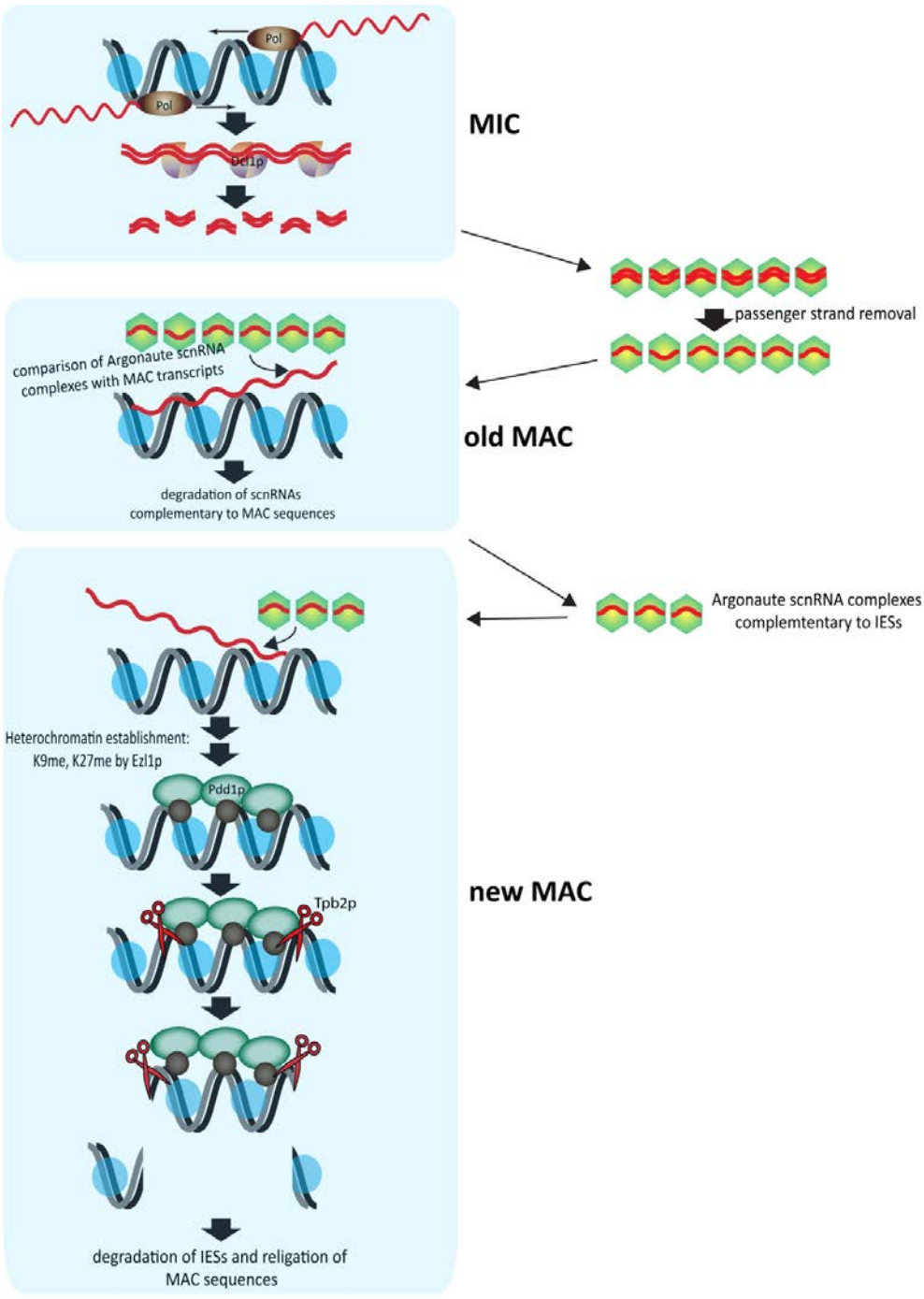

Figure 2: Schematic model for the DNA elimination pathway in Tetrahymena. First the MIC genome is transcribed bidirectionally by PoIII resulting in long dsRNA. This is processed by Dicer to scnRNAs (27-30 nt long). These are exported to the cytoplasm and associate with Twi1p (Argonaute). These complexes enter the parental MAC where scanning happens and complexes with RNA complementary to MDS regions are degraded. Complexes having RNA complementary to IESs are exported to the cytoplasm where the passenger strand is removed. Mature complexes can enter the new MAC where they lead to the sequence specific formation of heterochromatin on IESs. After the binding of the chromodomain protein Pdd1p the domesticated *piggyBac* transposase Tpb2p is recruited and excises IESs. Afterwards the flanking regions are relegated by an unknown mechanism.

3.4 The IES excision process during programmed genome rearrangements

3.4.1 Origin of IESs in Tetrahymena

The ~9000 different IESs in Tetrahymena most probably originated from transposable elements. There are many IESs showing sequence similarity to Tc1/mariner like transposons [16], or to non-LTR retrotransposons. The latter retrotransposons, called REP elements in Tetrahymena, are characterized by the presence of an open reading frame having endonuclease and reverse transcriptase domain [19]. There is another family of transposon-related IESs that is more moderately repeated in the Tetrahymena genome. Each element of this family of around 30 IESs encodes 15 open reading frames some of them are similar to transposon sequences or viruses. The internal region is harboring intrinsic telomeric sequences. The flanking regions are long inverted repeats which are composed of a conserved 47 bp core and additional variable regions that have multiple direct repeats being different in number and sequence between family members. The latter variable regions provide long and complex flanking sequences at both ends of each IES. Therefore this family is called Tlr (terminal long repeat) [31]. In contrast, many other IESs do not show obvious similarity to any transposon like sequence and thus their origin is unclear. However it is possible that these IESs could still be remnants of transposable elements.

3.4.2 Regulation of IES excision and the excision machinery

Most IESs are eliminated within a few to several base-pair precisions although some have a few alternative boundaries separated by a few hundred base-pairs or have exact base-pair precisions [32]. Therefore, the boundaries of IESs do not occur randomly but seem to be determined by some molecular feature. However, no specific consensus sequence at the boundaries has been found. This may not be surprising since different IESs are probably

different in their origin. It also seems unlikely that the boundary sequences of IESs are sufficient to determine their precise excision as for some elements cis-regulatory sequences were found which are necessary for their precise excision [33-36]. The best studied example of such a cis-regulatory sequence is the polypurine tract of the M-element which dictates the boundary to be always ~45 bp downstream of this sequence [36]. Another well studied IES, the R-element, needs cis-regulatory sequences on the left and right side to be removed precisely [35]. The cis-regulatory sequence on the left side of the R-element is around 70 bp long. How this sequence functions is still poorly understood as nucleotide and cluster substitutions do not show obvious effects on the R-element elimination. The cis-regulatory sequences identified so far do not share any obvious sequence property and thus it is difficult to imagine that common molecular machinery recognizes these diverse sequences to determine the boundaries of IESs. Instead, cis-regulatory sequences may provide some chromosomal structural constrain which limits regions where boundaries could occur.

To understand how IES boundaries are exactly targeted by the elimination machinery, the enzyme that directly catalyze the elimination of IESs has to be found and characterized. In the distantly related ciliate Paramecium IESs are flanked by a consensus sequence (5’ TAYAGYNR 3’) that is similar to the flanking sequence of Tc1/mariner transposons [37, 38]. A 5’ TA 3’ dinucleotide is kept at religated macronuclear destined sequence (MDS) boundaries, thus exhibiting the characteristic target site duplication from Tc1/mariner transposases. On the other hand Tc1/mariner transposases produce a double strand break with a 2 base 3’ protruding end whereas in Paramecium as well as in Tetrahymena it has been shown that IES elimination starts with a double strand break producing a 4 base 5’ overhang [37, 39]. This argues against the involvement of a Tc1/mariner transposase in IES excision in Paramecium and Tetrahymena. The geometry of the double strand break that occurs during IES excision in Paramecium and Tetrahymena rather resembles the double strand break product produced by PiggyBac transposases. During my thesis I demonstrated that the domesticated PiggyBac like transposase Tpb2p (Tetrahymena PiggyBac transposase like protein 2) plays an essential role during DNA elimination [40] and characterized its molecular functions both in vitro and in vivo. In a parallel study in Paramecium the PiggyBac like transposase PiggyMac (Pgm) was shown to

be involved in DNA elimination [41]. The study of these two enzymes can give new insights in the regulation of IES excision.

3.5 Aim of the thesis

In ciliates it was shown that an RNAi like pathway is leading to heterochromatin formation on specific DNA sequences that eventually are removed from the somatic genome. The key components of this pathway were described [6, 7, 9, 27] except the enzyme or the enzyme complex that excises the targeted sequences from the genome. The aim of this thesis was to identify the enzyme that excises IESs, to characterize its enzymatic activity, and to understand the mechanism by which the enzyme is recruited to the heterochromatinized IESs.

4. Publications

4.1 A Domesticated *piggyBac* Transposase plays Key Roles in Heterochromatin dynamics and DNA cleavage during Programmed DNA Deletion in *Tetrahymena thermophila*

Cheng CY*, Vogt A*, Mochizuki K, Yao MC.

*equally contributed to the work

Mol Biol Cell. 2010 May 15;21(10):1753-62. Epub 2010 Mar 31.

Molecular Biology of the Cell
Vol. 21, 1753–1762, May 15, 2010

A Domesticated *piggyBac* Transposase Plays Key Roles in Heterochromatin Dynamics and DNA Cleavage during Programmed DNA Deletion in *Tetrahymena thermophila*

Chao-Yin Cheng,*†‡ Alexander Vogt,‡§ Kazufumi Mochizuki,§ and Meng-Chao Yao*†

*Graduate Institute of Life Sciences, National Defense Medical Center, Taipei 114, Taiwan; †Institute of Molecular Biology, Academia Sinica, Nankang, Taipei 11529, Taiwan; and §Institute of Molecular Biotechnology of the Austrian Academy of Sciences, A-1030 Vienna, Austria

Submitted January 4, 2010; Revised March 5, 2010; Accepted March 22, 2010
Monitoring Editor: Kerry S. Bloom

Transposons comprise large fractions of eukaryotic genomes and provide genetic reservoirs for the evolution of new cellular functions. We identified *TPB2*, a homolog of the *piggyBac* transposase gene that is required for programmed DNA deletion in *Tetrahymena*. *TPB2* was expressed exclusively during the time of DNA excision, and its encoded protein Tpb2p was localized in DNA elimination heterochromatin structures. Notably, silencing of *TPB2* by RNAi disrupts the final assembly of these heterochromatin structures and prevents DNA deletion to occur. In vitro studies revealed that Tpb2p is an endonuclease that produces double-strand breaks with four-base 5′ protruding ends, similar to the ends generated during DNA deletion. These findings suggest that Tpb2p plays a key role in the assembly of specialized DNA elimination chromatin architectures and is likely responsible for the DNA cleavage step of programmed DNA deletion.

INTRODUCTION

Developmentally programmed, large-scaled DNA rearrangements have been known to occur in organisms ranging from protozoa to vertebrates (Borst and Greaves, 1987; Coyne *et al.*, 1996; Smith *et al.*, 2009). These processes drastically alter chromosome structures and genomic contents and have significant implications in gene regulation and genome integrity. However, their mechanisms and biological roles are not well understood. The ciliated protozoan *Tetrahymena* is among the best studied organisms to carry out such rearrangement processes and has offered considerable insights. *Tetrahymena thermophila* contains a silent germinal micronucleus and a transcription active somatic macronucleus in each cell. Both nuclei are derived from the same genetic source during conjugation. In this sexual process, the micronucleus goes through meiosis, postmeiotic mitosis, and cross-fertilization to generate zygotic nuclei, which further divide and differentiate to form the new macro- and micronuclei in the progeny cells while the old macronucleus is destroyed (see Figure 1A). New macronuclear development involves extensive genome-wide DNA rearrangements, deleting thousands of specific DNA segments, known as internal eliminated sequences (IESs) that comprise ~15% of the genome. The remaining DNA is fragmented at specific chromosome breakage sequences (Cbs) and endoreplicated ~23-fold to form the somatic genome (Yao and Chao, 2005).

Programmed DNA deletion is an RNA-guided process (Mochizuki *et al.*, 2002; Yao *et al.*, 2003). Double-stranded RNA is transcribed from IES during and after meiosis and processed into 27–30-nucleotide small RNA by the Dicer-like enzyme Dcl1p (Chalker and Yao, 2001; Malone *et al.*, 2005; Mochizuki and Gorovsky, 2005). Together with other proteins including the Argonaute protein Twi1p (Mochizuki *et al.*, 2002), the small RNA is believed to target corresponding IESs in the developing macronucleus, leading to histone modifications (H3K27me3 and H3K9me3; Taverna *et al.*, 2002; Liu *et al.*, 2007) and the recruitment of chromodomain proteins (Pdd1p and Pdd3p; Madireddi *et al.*, 1996; Nikiforov *et al.*, 1999; Nikiforov *et al.*, 2000) to form distinct heterochromatin structures that are eventually deleted. However, the enzymatic machinery that carries out the DNA cutting and joining reaction remains largely unknown.

Some transposase genes in eukaryotes have evolved to acquire functions that are beneficial to the host and become domesticated (Volff, 2006). Two well-known examples are *RAG1* from *Transib* transposons in the V(D)J recombination system of vertebrate immunoglobulin genes (Agrawal *et al.*, 1998; Kapitonov and Jurka, 2005) and *Cenp-B* from *pogo* DNA tranposons in the formation of chromosome centromere in some eukaryotes (Smit and Riggs, 1996; Casola *et al.*, 2008). Ciliate IES could have been evolved from transposons. In *Euplotes* and *Oxytricha*, the excision of some transposons resembled the deletion of IESs (Doak *et al.*, 1994; Klobutcher and Herrick, 1997). In *Tetrahymena*, most of the eliminated portion of the genome could be derived from transposons or other invading genetic elements during evolution. Studies of intermediates in the *Tetrahymena* DNA deletion process revealed distinct four-base 5′ protruding ends, resembling the end generated during the transposition

This article was published online ahead of print in *MBoC in Press* (http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E09–12–1079) on March 31, 2010.

‡ These authors contributed equally to this work.

Address correspondence to: Meng-Chao Yao (mcyao@imb.sinica.edu.tw).

of some transposons such as Tn7 (Saveliev and Cox, 1995, 1996).

The *piggyBac* transposon originally isolated from the genome of the moth *Trichoplusia ni* encodes a 594-amino acid transposase that mediates cut and paste excision and reinserts the transposon element into a TTAA target site in the genome (Elick *et al.*, 1996; Fraser *et al.*, 1996). Many *piggyBac*-like sequences were found in the genomes of different organisms including fungi, plants, insects, crustaceans, urochordates, amphibians, fishes, and mammals. The widespread distribution and phylogenetic analyses of *piggyBac*-like sequences in various organisms have led to the proposal that some of these ancient transposons might be domesticated by the host genome for cellular functions (Sarkar *et al.*, 2003). In this study, we explored the possibility that a transposase may have evolved to participate in programmed DNA deletion in *Tetrahymena*. In an independent parallel study, Baudry *et al.* (2009) have recently provided in vivo evidence that *PiggyMac* (*PGM*), a domesticated *piggyBac* transposase, is required for programmed genome rearrangements in *Paramecium tetraurelia*. Our results here show that a *piggyBac* transposase gene has also been domesticated in *Tetrahymena* to play essential roles in programmed DNA deletion. Further more, our results revealed an essential role for this gene in the formation of specialized DNA elimination heterochromatin structures and provided evidence that it is likely responsible for the DNA cutting activity in the DNA deletion process.

MATERIALS AND METHODS

Cell Culture

Wild-type *T. thermophila* strains B2086 II, CU428 (Mpr/Mpr [VII, mp-s]), and CU427 (Chx/Chx [VI, cy-s]) were obtained from Peter Bruns (Cornell University, Ithaca, NY). *Tetrahymena* strains were maintained as previously described. *Tetrahymena* cells were grown in SPP medium at 30°C and prepared for mating by washing cells with 10 mM Tris-HCl (pH 7.4) buffer and incubation overnight before mixing to initiate conjugation.

Construction of Hairpin RNA Expression Vectors and Green Fluorescent Protein Fusion Vector

The targeted region was amplified by PCR and the product (~500 base pairs) was cloned into the PCRII-I vector using two sets of primers with restriction enzyme cloning sites, one set with the ApaI-XhoI site and the other with the PmeI-SmaI site to generate the hairpin cassette. This hairpin cassette was then removed from the PCRII-I backbone by digestion with PmeI and ApaI and was ligated into the pIBF rDNA vector (Howard-Till and Yao, 2006). *TPB2* hairpin RNA was expressed under the control of a $CdCl_2$-inducible *MTT1* promoter. The full-length *TPB2* genomic DNA PCR product (4475 base pairs) was cloned into the PIGF-1 vector and fused to green fluorescent protein (GFP) at its N-terminus using the XhoI and ApaI cloning site. The sequences of the primers are listed in *Oligo DNAs* (see Supplemental Information).

Production of Tpb2p Antisera

Rabbits were immunized with a synthetic peptide "KQEHRSDQKKKNSY" corresponding to amino acid sequence from 1003 aa to 1016 aa of Tpb2p (commercial custom-made antibody from ProSci, Poway, CA).

RT-PCR

RNA samples were prepared using a RNA isolation kit (Roche, Indianapolis, IN). First strand cDNA was synthesized using Transcriptor reverse transcriptase with anchored-oligo$(dT)_{18}$ primer. It was then followed by either conventional PCR or quantitative PCR analysis. The sequences of the primers are listed in the "Oligo DNAs" section (see Supplemental Information). The quantitative-PCR analysis was performed by LightCycler Carousel-Based PCR System with the LightCycler FastStart DNA Master^{Plus} SYBR Green kit (Roche). Relative quantification was normalized with α-tubulin mRNA expression as an internal control.

Northern Blotting

Total RNA samples were prepared using TRIzol reagent (Invitrogen). RNA samples were mixed with RNA glyoxal reaction mixture loading dye and analyzed by electrophoresis in a 1.2% agarose gel. RNA was transferred to IMMOBILON-NY+ nylon membrane (Millipore), UV cross-linked and hybridized with probes overnight at 42°C in a hybridization buffer (Roche). Probes were made by random prime labeling reactions with PCR products amplified from genomic DNA. After hybridization, membrane was washed several times with 2–0.5× SSC and 0.1% SDS washing buffer and exposed to x-ray film. Bands were quantified and normalized by using Quantity One software (Bio-Rad, Richmond, CA).

Evaluation of TPB2 *Hairpin RNA Silencing Phenotype*

Two different mating types of *TPB2* hairpin transformants were starved and mated with each other in 10 mM Tris-HCl (pH 7.4) buffer without or with 0.05 μg/ml $CdCl_2$ to silence the *TPB2* gene during conjugation. To test for their abilities to produce viable progeny, 300 individual mating pairs were cloned to drops of growth medium for 48 h and tested for growth and drug resistance phenotypes that distinguish parents from progeny. To determine developmental stages, conjugating cells (14, 20, and 30 h after mating initiation) were stained with DAPI (4′,6′-diamidino-2-phenylindole; 1 μg/ml) to reveal their nuclear stages. To examine the IES elimination and chromosome breakage by using PCR assay, genomic DNAs were isolated from the late conjugating cell pools (30 h after mating initiation) and used to perform PCR analysis using primers that were specific for IES elimination (M element, R element, Cam element, and Tlr element; Coyne *et al.*, 1999; Aronica *et al.*, 2008) and chromosome breakage (Cbs819, 5-2, and 4L-2 and the telomere primer; Coyne *et al.*, 1999).

Immunofluorescence Analysis

Conjugating cells were fixed with 2% paraformadehyde, immobilized on slides, incubated with the primary antibodies, anti-Pdd1p (1:1000, Abcam, Cambridge, MA; ab5338) and anti-Tpb2p (1:500), respectively, and washed with 0.3% Triton X-100 in PBS buffer. The secondary antibodies were Cy5-conjugated AffiniPure F(ab′)2 fragment goat anti-rabbit IgG (1:500, Jackson ImmunoResearch, West Grove, PA; 111-176-003). Digital images were performed and analyzed using Zeiss Axio Imager system (Thornwood, NY) and Applied Precision Deltavision system (Issaquah, WA) fluorescence microscopy.

Chromatin Immunoprecipitation

Cells were cross-linked in 1% paraformadehyde, harvested, and washed as previously described (Dedon *et al.*, 1991). Cells were then resuspended at 5×10^4 cells/ml in 0.1% SDS lysis buffer (Chadee *et al.*, 1999) and sonicated for 4 min in 30-s bursts. Cell lysates were used to immunoprecipitate with antibodies, anti-Pdd1p (Abcam, ab5338), anti-H3K27me3 (Millipore, Bedford, MA; 07-449), and anti-H3K4me2 (Millipore, 07-030), respectively. The immunoprecipitated complexes were collected by protein A agarose (Millipore, 16-157) and then incubated in proteinase K buffer overnight. The cross-links were reversed at 65°C for 6 h, and DNA was extracted with phenol/chloroform and ethanol precipitation. DNA isolated from the lysate without antibody immunoprecipitation was used as an input control. The primer sets used are as previously described (Liu *et al.*, 2007). Quantitative-PCR analysis was performed as description above.

Production of Recombinant Proteins

To produce recombinant Tpb2p, DNA (*TPB2Ec*) encoding Tpb2p that had been optimized for codon usage in *Escherichia coli* was synthesized (GenScript, Piscataway, NJ). The sequence of *TPB2Ec* is shown in Supplemental Information. *TPB2Ec* was cloned into EcoRI and XhoI sites of pGEX4T1-TEV (gift from Tim Clausen, Institute of Molecular Pathology, Vienna) to make pGEX-TPB2. To generate pGEX-TPB2-CD, expressing a catalytic dead Tpb2p, D297 to L, D379 to L and D495 to L mutations were introduced into pGEX-TPB2 using the QuikChange II site-directed mutagenesis Kit (Stratagene, La Jolla, CA), and DNA oligos Tpb2pD297L fw/rv, Tpb2pD379L fw/rv, and Tpb2pD495L fw/rv. Sequences of the oligo DNAs used for the mutagenesis are listed in *Oligo DNAs* (see Supplemental Information). GST-Tpb2p and GST-Tpb2p-CD were expressed in *E. coli* strain BL21(DE3). After cultivation at 37°C to an A_{600} of ~0.8, cells were incubated with 0.5 mM IPTG for 8 h at 18°C. Cells were lysed in 500 mM NaCl, 80 mM Tris-HCl, pH 8.0, 0.5% Triton X-100, 0.2 mM PMSF, and 1× complete proteinase inhibitor cocktail (Roche). The cell lysate was incubated with glutathione Sepharose 4B (Amersham Biosciences, Piscataway, NJ) at 4°C. After washing with 500 mM NaCl, 80 mM Tris-HCl, pH 8.0, and 0.1% Triton X-100, the glutathione *S*-transferase (GST) fusion proteins were eluted in 160 mM reduced glutathione, 500 mM NaCl, and 80 mM Tris-HCl, pH 8.0, and subsequently the buffer was replaced with dialysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 4 mM $MnSO_4$, and 10% glycerol).

Tpb2p Endonuclease Assay

Sequences of the oligo DNAs used in this assay are listed in *Oligo DNAs* (see Supplemental Information). The oligo DNAs were 5′ end labeled by incubating 20 pmol single-stranded DNA with 10 U polynucleotide kinase (Roche) in the presence of 10 μCi [γ-^{32}P]ATP at 37°C for 1 h. Twenty picomoles of the complementary strand was added and purified with QIAquick Nucleotide removal Kit (QIAGEN, Chatsworth, CA). The samples were heat denatured at

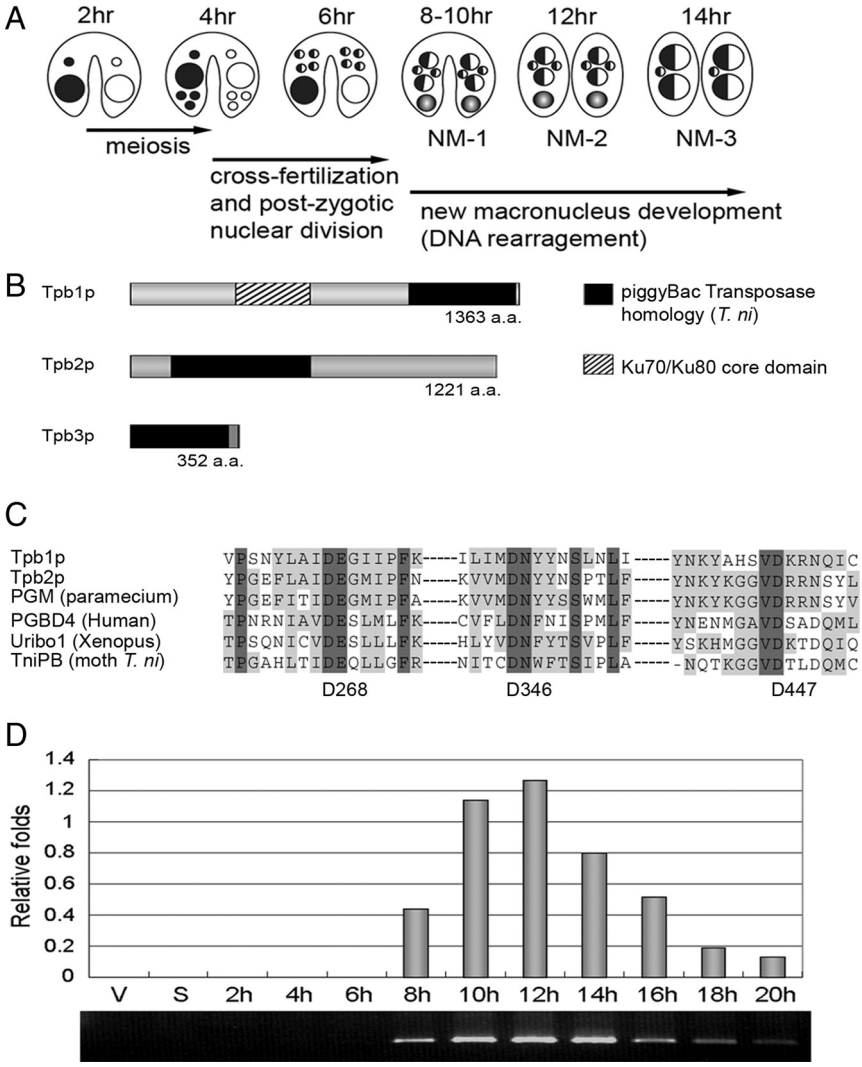

Figure 1. Characterization of TPB. (A) *Tetrahymena* nuclear developmental process during conjugation. (B) Schematic representation of Tpb proteins. (C) Comparison of the catalytic DDD-motif of Tpb2p and other piggyBac-like proteins. The DDD-motif of *T. ni piggyBac* consisting of D268, D346, and D447 is indicated. (D) Expression of *TPB2* by quantitative RT-PCR. Total RNA extracted from vegetative (V), starved (S), and conjugating cells (2, 4, 6, 8, 10, 12, 14, 16, 18, and 20 h after postmixing) were used as templates. Quantification was normalized with α-tubulin mRNA expression.

95°C for 5 min. Tubes were placed into 200 ml boiling water and allowed to cool to room temperature to allow the oligo DNAs to anneal. The double-stranded oligo DNA was incubated with 2.5 ng/μl recombinant GST-Tpb2p or GST-Tpb2p-CD at 30°C for 2.5 h in 10 mM $MnSO_4$, 10 mM Tris-HCl, pH 7.5. The reactions were treated with proteinase K, extracted with phenol-chloroform, and ethanol-precipitated. An aliquot of the samples with RArtfw and RArtrv were incubated with 10 U Klenow fragment (Fermentas, Hanover, MD) and 2 nmol dNTPs at 37°C for 30 min. The samples were separated on a 15% denaturing polyacrylamide gel. The dried gel was exposed to a storage phosphor screen (GE Healthcare, Piscataway, NJ), and images were captured using Typhoon Imager (GE Healthcare).

RESULTS

piggyBac *Transposase Homologues in* Tetrahymena

We searched the *Tetrahymena* genome database and found three *piggyBac* transposase homolog genes, *TPB1*, *TPB2*, and *TPB3* (*Tetrahymena piggyBac* transposase 1, 2, and 3) from the gene annotations of *Tetrahymena* genome database (TGD; http://www.ciliate.org/). The cDNA sequences of *TPB1* and *TPB2* have been submitted to DDBJ/EMBL/GenBank AB544065 and AB544066. Besides the transposase domain, Tpb1p contains a Ku70/Ku80 β-barrel domain in its N-terminal region. *TPB3* is present only in the micronuclear genome, and the available sequence is truncated at the C-terminal of the *piggyBac* transposase homology region (Figure 1B). Tpb1p and Tpb2p share transposase catalytic DDD motif (in D268, D346, and D447 of *T. ni piggyBac*; Sarkar *et al.*, 2003) with other *piggyBac-like* proteins (Figure 1C). In this study we focus on the analysis of *TPB2* due to its essential roles in programmed DNA deletion (see below).

RT-PCR analyses were carried out to determine the expression of *TPB2* mRNA at different stages of the life cycle. The results showed that the expression of *TPB2* was highly up-regulated at late stages of conjugation, especially at 12–14 h after cell mating begins (Figure 1D), which coincides with the stages of DNA rearrangements in new developing macronuclei (Austerberry *et al.*, 1984). Because the old macronuclei have largely degenerated at this stage, majority of *TPB2* mRNA is likely derived from new developing macronuclei.

Silencing of TPB2 *Gene by Hairpin RNA*

To study the function of *TPB2*, we generated genetic knockout strains lacking *TPB2* in their macronuclear genomes. The strains were able to grow and complete conjugation without notable phenotypic defects. Because *TPB2* expression is likely from the developing macronuclei, knocking out the germline copy of *TPB2* will be necessary to properly test its

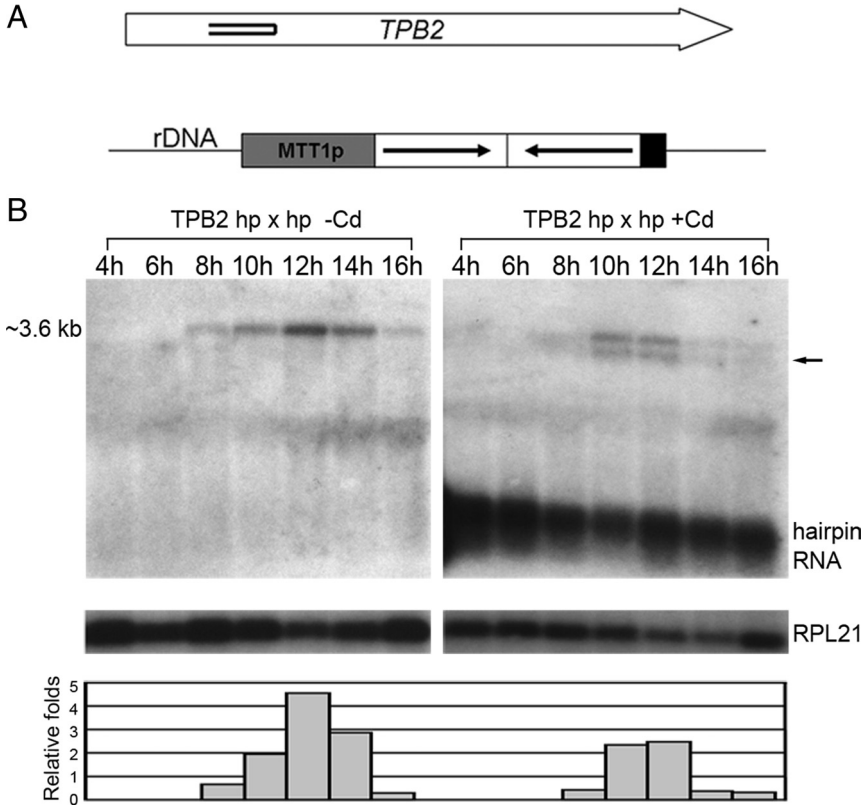

Figure 2. TPB2 gene silence by hairpin RNA expression. (A) Schematic map of hairpin RNA constructs. Open arrows represent *TPB2* gene with double lines inside indicating the region targeted by the hairpin RNA. (B) *TPB2* gene silencing by hairpin RNA expression. Northern blot of total RNA samples extracted from conjugating cells (4, 6, 8, 10, 12, 14, and 16 h after postmixing) transformed with the hairpin RNA construct with or without inducing with 0.05 μg/ml $CdCl_2$ in 10 mM Tris buffer. The approximate size of *TPB2* mRNA is indicated to the left of the panels, and the right arrow indicates a new band likely from the degraded RNA. The robust TPB2 hairpin RNA expression was also detected. The *RPL21* mRNA was used as an internal control to quantify relative folds of the full-length *TPB2* mRNA level.

effects. However, repeated attempts to generate germline knockout strains have not been successful.

Hairpin RNA-induced gene silencing (RNAi) in *Tetrahymena* had been developed as an efficient method to study gene functions (Howard-Till and Yao, 2006). We used this strategy to investigate the role of *TPB2* during conjugation. The region for hairpin construction in *TPB2* gene is shown in Figure 2A. *TPB2* hairpin RNA expression was driven from *MTT1* promoter, which is induced by the presence of Cd^{2+} ion. We examined the effect of gene silencing during conjugation by Northern blot analysis (Figure 2B). Hairpin RNA was present and maintained at all conjugating stages in the presence of $CdCl_2$. *TPB2* mRNA expression was reduced ~50% at the 12-h stage and more drastically (~90%) at the 14-h stage when DNA deletion takes place. A new band likely representing the degradation products was also detected (Figure 2B, arrow). These results indicate that *TPB2* mRNA expression can be efficiently down-regulated by the expression of *TPB2* hairpin RNA.

TPB2 *Is Essential for Nuclear Differentiation in Conjugation*

To determine whether *TPB2* is essential for conjugation process, we analyzed nuclear events at different conjugating stages by DAPI staining. *TPB2* silenced cells were arrested at the stage with two new Mic and two new Mac (Figure 3A). This phenotype is similar to several previously studied mutants (Δ*EMA1*, ΔEZL1, Δ*DCL1*, Δ*PDD1*, Δ*PDD2*, Δ*LIA1*, and Δ*TWI1*) that have defects in DNA rearrangements (Coyne *et al.*, 1999; Nikiforov *et al.*, 1999; Mochizuki *et al.*, 2002; Malone *et al.*, 2005; Mochizuki and Gorovsky, 2005; Liu *et al.*, 2007; Rexer and Chalker, 2007; Aronica *et al.*, 2008). To test for their abilities to produce progeny, single mating pairs were cloned, and their growth was analyzed. Significantly, more than 95% of the 300 individual *TPB2*-silenced mating cells failed to grow and eventually died (Figure 3B). Thus, we concluded that *TPB2* gene is essential for the completion of conjugation and for the formation of viable progeny.

IES Elimination and Chromosome Breakage Are Severely Affected in TPB2 *Silencing Cells*

To determine the importance of *TPB2* in DNA rearrangements, we examined IES elimination at four regions (M-, R-, Cam-, and Tlr element) and chromosome breakage at three sites (Cbs819, 5-2, and 4L-2) by PCR, using whole cell DNA isolated from pools of mating cells at the terminal stage of conjugation (30 h after mating initiation) as templates. In wild-type mating pools, the macronuclear form (Mac form) of PCR products was the majority compared with the micronuclear form (Mic form) at all regions studied. These ratios of the Mac/Mic form partly reflected the relative DNA amounts of the polyploid (45C) macronucleus and the diploid micronucleus. Significantly, in *TPB2* silenced mating pools, the Mic form was much more abundant than the Mac form (Figure 4B), indicating the failure of DNA rearrangements. The small amounts of Mac form were most likely derived from nonmating cells that were also present in the pool. These results demonstrate that IES elimination and chromosome breakage are severely affected in *TPB2*-silenced cells. We concluded that *TPB2* is essential for the programmed genome rearrangements in *Tetrahymena*.

GFP-Tpb2p Is Colocalized with Heterochromatin Markers and Pdd1p-containing Structures

To further understand the role of Tpb2p, its localization was analyzed by expressing Tpb2p fused with GFP at its N-terminus in an rDNA vector and then transformed into the macronuclear genome as high copy plasmids (Yu and Black-

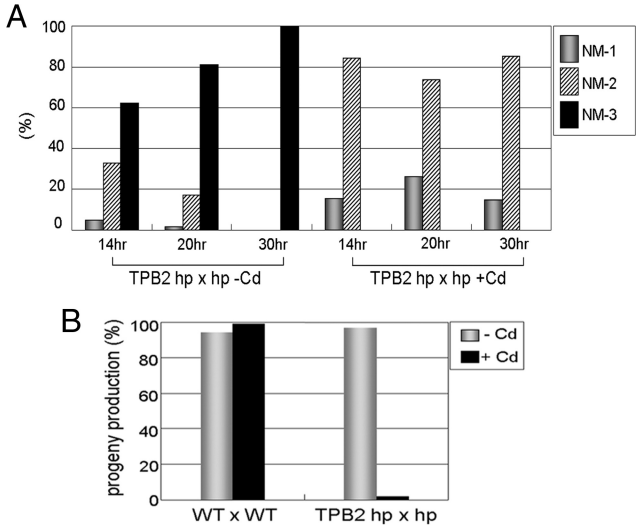

Figure 3. Phenotype of *TPB2* silencing in conjugation. (A) Late conjugating stages analysis of the mutant. Conjugating cells (14, 20, and 30 h after postmixing) of TPB2 hairpin RNA-containing strains with or without induction with 0.05 μg/ml $CdCl_2$ were stained with DAPI to reveal their nuclear stages (Figure 1A). NM (New Mac)-1, anlagen swelled in mating cells; NM-2, mating cells separated and two macronuclei and two micronuclei were present; NM-3, two macronuclei and one micronucleus were present. (B) Viability of progeny. Wild-type and TPB2 RNA hairpin cells were mated as indicated. Individual cell mating pairs were cloned into a drop of growth media at 10 h after postmixing. Cells were examined for progeny production by viability and drug resistance, which is specific for new developing macronucleus in progeny.

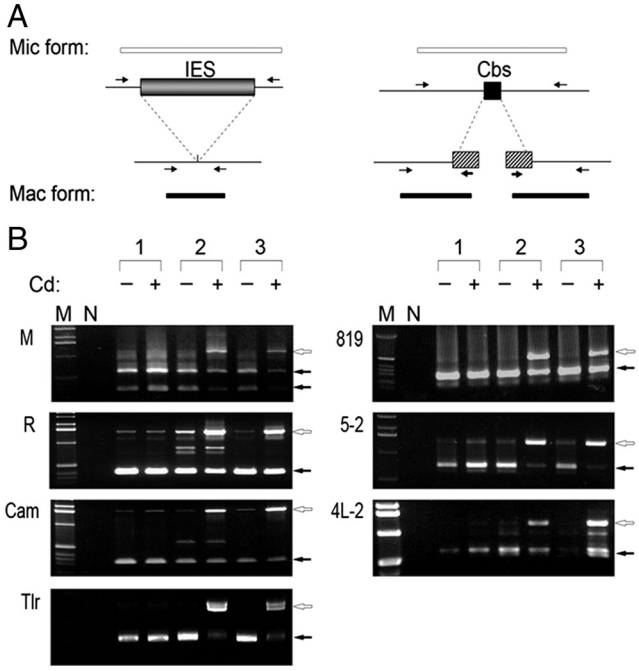

Figure 4. DNA elimination and chromosome breakage of *TPB2* silencing in conjugation. (A) Schematic drawing of the IES elimination and chromosome breakage PCR assay. The open box represented DNA region retained in the macronucleus, and the filled box represented IES, Cbs, or telomere regions. Arrows indicate the positions of PCR primers. The relative lengths of the expected products were shown at top (Mic form) and bottom (Mac form). (B) The result of IES elimination and chromosome breakage PCR assay. The genomic DNA were isolated from pools of conjugating cells (30 h after postmixing) of wild type (1) and TPB2 hairpin RNA-containing strains (2 and 3) with or without induction with 0.05 μg/ml $CdCl_2$. The open arrows indicate Mic form PCR product, and the filled arrows indicate Mac form PCR product.

burn, 1989; Godiska and Yao, 1990). The overexpressed GFP-Tpb2p was localized within the developing macronuclear anlagen in conjugating cells (Figure 5A). We performed immunofluorescence analysis of H3K27me3, H3K9me3, and Pdd1p in GFP-Tpb2p–overexpressed conjugating cells (~10 h). The results indicated that GFP-Tpb2p has similar localizations with H3K27me3 and H3K9me3, two markers known to be associated with heterochromatin, and Pdd1p, which is a chromodomain-containing protein involved in the establishment of DNA elimination heterochromatin structures (Figure 5B). GFP-Tpb2p colocalized with Pdd1p-containing nuclear structures at late stages of conjugation (~14 h; Figure 5C). These cytological studies indicate that the function of Tpb2p might be involved in chromatin modification process.

To support the GFP-Tpb2p localization results, we also generated antibodies against Tpb2p and performed immunofluorescence analysis. The results show that the endogenous Tpb2p was presented specifically within the developing macronuclear anlagen in conjugating cells (Figure 6, A–C) and especially colocalized with DNA elimination structures at the nuclear periphery similar to Pdd1p (see next section for more details; Madireddi *et al.*, 1994, 1996; Figure 6D). These results indicate that Tpb2p is targeted to the heterochromatic DNA elimination structure just like the overexpressed GFP-Tpb2p, further supporting its direct role in DNA deletion.

TPB2 *Is Not Required for the Physical Association of Pdd1p and H3K27me3 with IESs*

Genes involved in small RNA production, such as *DCL1* and *TWI1,* are expressed early in conjugation and are essential for the targeted heterochromatin formation on IESs (Liu *et al.*, 2007). Because of its late expression timing, *TPB2* might not be required for IES targeting. This issue was examined by chromatin immunoprecipitation (ChIP) studies. Wild-type and *TPB2* silenced cells at the 10-h stage were processed for ChIP, and two IESs, M-mic and PGM1-mic and two macronuclear genes, PGM1-mac and MTT1-mac, were analyzed by real-time PCR. In both wild-type and *TPB2*-silenced cells, Pdd1p and H3K27me3 were enriched in the IES regions, whereas H3K4me2, a euchromatin marker, was enriched in the macronuclear genes (Figure 7A). Thus, *TPB2* is dispensable for the IES-targeting process. These results place the action of *TPB2* downstream of the targeting to induce DNA elimination.

TPB2 *Is Essential for the Assembly of Pdd1p-containing Structures*

Previous studies have shown that H3K27me3 and H3K9me3 accumulate on IES loci shortly after new macronuclear formation begins (~7 h after mating) and before the assembly of large Pdd1p-containing structures. Pdd1p-containing structures first appeared homogeneously distributed in the developing macronuclei and later formed a few large spherical structures (~14 h after mating) near the nuclear periphery. Timing of DNA elimination (12–14 h) corresponds well with the aggregation of the Pdd1p-containing structures (Austerberry *et al.*, 1984; Madireddi *et al.*, 1994, 1996; Chalker,

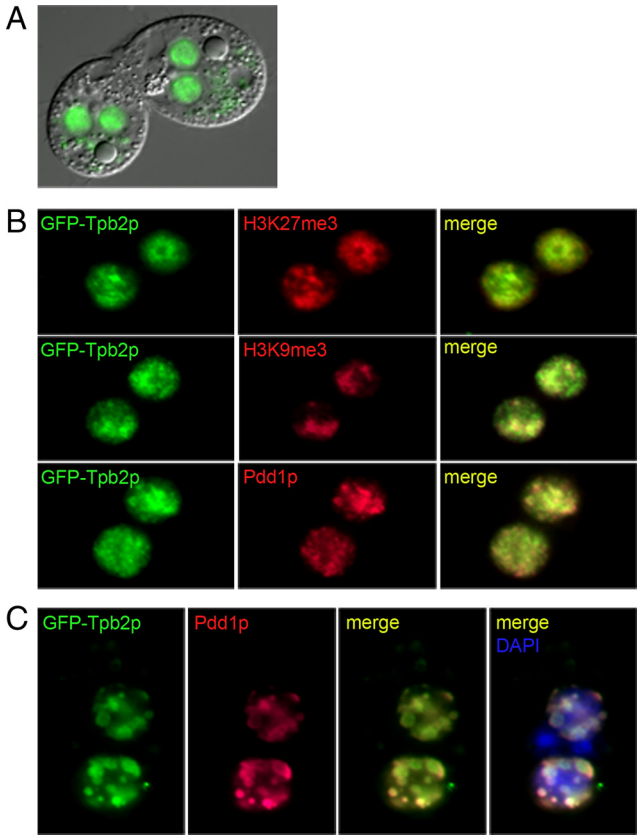

Figure 5. Localization of GFP-Tpb2p. (A) Overexpression of the N-terminal GFP-Tpb2p fusion protein. The GFP-Tpb2p fusion protein localized within the developing macronuclear anlagen in living conjugating cells. (B) Colocalization of GFP-Tpb2p and heterochromatin marker (H3K27me3, H3K9me3, and Pdd1p). Ten-hour conjugating cells with GFP-Tpb2p (green) were processed for immunofluorescence staining with the indicated antibodies (red). (C) Colocalization of GFP-Tpb2p and Pdd1p in the DNA elimination heterochromatin structure. Fourteen-hour conjugating cells expressing GFP-Tpb2p (green) were processed for immunofluorescence staining with an anti-Pdd1p antibody (red) and stained with DAPI for DNA (blue).

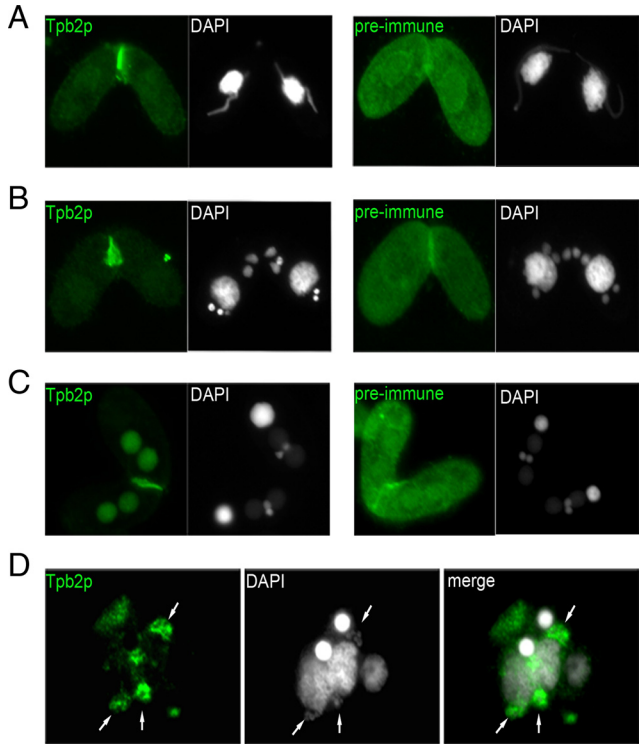

Figure 6. Localization of endogenous Tpb2p by antibody immunofluorescence staining. Cells were processed for immunofluorescence staining (green) and DAPI staining (white) using an anti-Tpb2p antibody and preimmune serum at different developmental stages in conjugation. (A) "Crescent" stage in micronuclei meiosis. (B) postmeiotic mitosis stage; (C) new developing macronuclei stage at early mating pair; (D) at the late conjugating stage after pair separation.

2008). We performed cytological analysis of the assembly of Pdd1p-containing structures by immunofluorescence staining using an anti-Pdd1p antibody. *TPB2*-silenced cells did not form large Pdd1p-containing structures, and some cells formed abnormal Pdd1p aggregates in the cytosol (Figure 7, B and C). Thus, *TPB2* plays an important role in the assembly of Pdd1p-containing structures.

Tpb2p Has an Endonuclease Activity in Vitro

Because Tpb2p shares the conserved transposase catalytic DDD motif (Figure 1C), we speculated that Tpb2p might posses a nuclease activity that carries out the final DNA-cutting step in programmed DNA deletion. To test this possibility, we examined possible nuclease activities of Tpb2p in vitro. Because of the difference in genetic codon usage, the *TPB2* gene sequence was modified and fused to GST to allow protein purification from *E. coli*.

PiggyBac recognizes 5′-TTAA-3′ sequence and induces double-strand break at the bond before the first T of 5′-TTAA-3′ (Cary *et al.*, 1989). Therefore we used a 50-base pair DNA that had 5′-TTAA-3′ sequence only once (the first T is at the 24th position). One strand of the DNA was end-labeled, incubated with recombinant GST-Tpb2p, and analyzed by denaturing gel-electrophoresis. A 23-base product was detected by autoradiography (Figure 8A, Tpb2p-TTAA), indicating that Tpb2p cuts the bond between 23rd and 24th (right before the first T of 5′-TTAA-3′) of the labeled DNA strand. Analysis of the same product in a native gel detected a ~25-base pair product indicating that Tpb2p produces double-strand break. No cleavage product was detected when we used a Tpb2p mutant in which all three aspartic acids in the DDD motif (Figure 1C) were replaced by leucines (Figure 8, A and B, Tpb2p-CD). These results suggest that Tpb2p has DDD motif-dependent endodeoxyribonuclease activity.

We then used 50-base pair DNA that has 5′-GTAG-3′ or 5′-GTTG-3′ instead of 5′-TTAA-3′ and found that Tpb2p could utilize the 50-base pair DNA with GTAG as a substrate and produce a 23-nucleotide (nt) product, albeit the amount of the product was lower than that from 50-base pair DNA with 5′-TTAA-3′ (Figure 8A, Tpb2p-GTAG). In contrast, 50-base pair DNA with 5′-GTTG-3′ could not produce any detectable 23-nt product (Figure 8A, Tpb2p-GTTG). These results suggest that Tpb2p has some sequence preference in its cutting site.

In *Tetrahymena*, DNA elimination machinery is thought to recognize 5′-ANNNNT-3′ sequence to induce DNA double-strand break (Saveliev and Cox, 1996). Therefore, we next tested the 50-base pair DNA substrate RArt and RArt30, which have the 5′-AGTGAT-3′ sequence. This sequence is found at the left boundary of the *Tetrahymena* R-element and

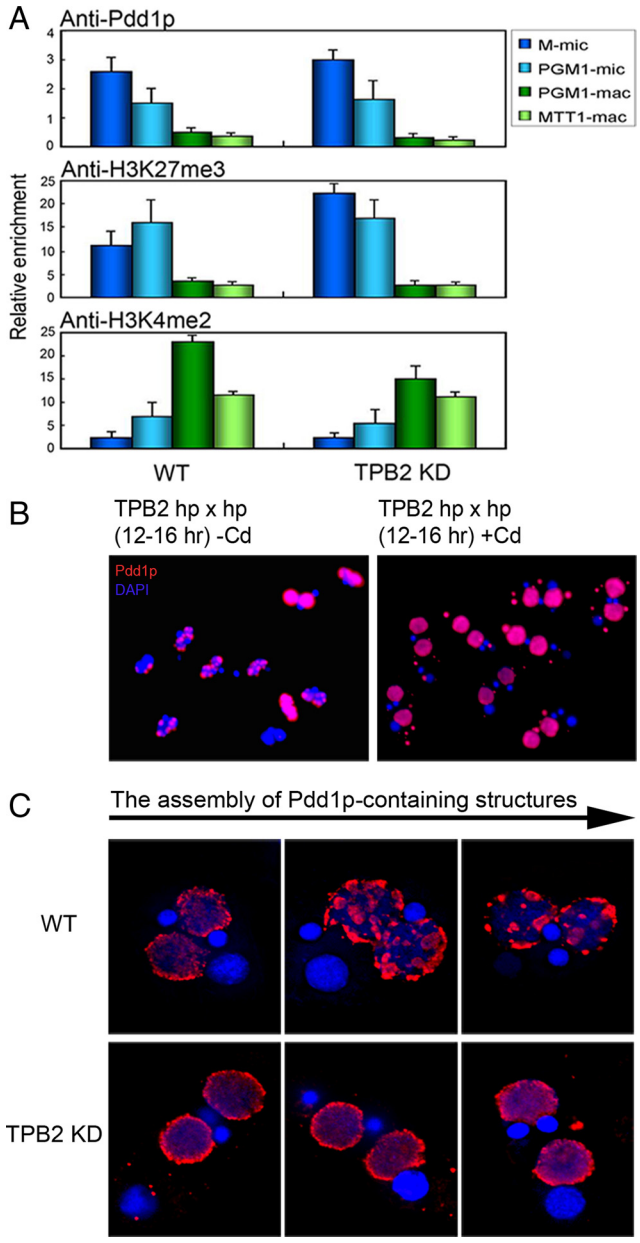

Figure 7. TPB2 is dispensable for recruiting heterochromatin markers to IES and is essential for the assembly of heterochromatin structures. (A) Conjugating cells of wild type (WT) and *TPB2* knockdown strains (TPB2 KD) were processed for ChIP using the antibodies indicated. Two micronuclear IESs (M-mic and PGM1-mic) and two macronuclear genes (PGM1-mac and MTT1-mac) were assayed by quantitative PCR, and the relative enrichment values are shown. (B) Conjugating cells (12–16 h) of TPB2 hairpin RNA strains with or without induction with 0.05 μg/ml $CdCl_2$ were processed for immunoflurescence staining with an anti-Pdd1p antibody (red) and DAPI staining (blue). (C) Magnified images of the assembly of Pdd1p-containing structures in WT and TPB2 KD cells.

the first A of the 5′-AGTGAT-3′ is located at the 25th and 30th position of the labeled strand of RArt and RArt 30, respectively. The observed cleavage site during DNA elimination of the R-element in vivo (Saveliev and Cox, 1996) corresponds to the bond between the first A and the first G of the 5′-AGTGAT-3′ sequence. The products were separated in a denaturing polyacrylamide gel. Twenty-five-nucleotide products from RArt (Figure 8C, RArt Tpb2p) were detected, indicating that Tpb2p has an endonuclease activity, which prefers to cut at the in vivo cleavage site of the R-element. A 30-nt product from RArt30 was also detected (Figure 8C, RArt30 Tpb2p), although the product was less abundant than that from RArt. This was possibly because Tpb2 needs certain length of DNA at both sides to make stable Tpb2p-DNA complex for inducing efficient cleavage. Tpb2p-CD could not produce any specific cleavage products from RArt (Figure 8C, RArt Tpb2p-CD), strongly support the specificity of the endonuclease activity detected for Tpb2p. Treatment of the product from RArt with Klenow fragment results in a migration shift of the prominent band from that of 25 nt to that of 29 nt (Figure 8C, RArt Tpb2p + Klenow), demonstrating that Tpb2p cleavage leaves a four-base 5′ protruding end (Figure 8D). This is consistent with the intermediate product of the previously reported DNA elimination pathway (Saveliev and Cox, 1995, 1996), suggesting that Tpb2p has the ability to produce proper double-strand break at the endogenous IES boundary sequence to initiate DNA elimination.

The 5′-TTAA-3′ sequence in the oligo used in Figure 8A was in the context of 5′-A"TTAA"C-3′. Therefore, we speculated that the double-strand break at the 5′-AGTGAT-3′ sequence in RArt oligo (Figure 8C) was possibly mediated not by the proposed consensus 5′-ANNNNT-3′ but by the internal 5′-GTGA-3′ sequence. To test this possibility, Tpb2p was incubated with the oligo possessing 5′-GTGA-3′, which was otherwise identical to the TTAA oligo used in Figure 8A. In contrast to our prediction, no detectable 23-nt product was produced (Figure 8E). This result indicates neither internal four bases nor the first and the last base of the 5′-AGTGAT-3′ sequence is sufficient to be cleaved by Tpb2p. We need further comprehensive analysis to find out what types of sequences Tpb2p recognizes and cuts.

DISCUSSION

In this study we presented evidence that a domesticated *piggyBac* transposase plays a key role in programmed DNA deletion in *Tetrahymena*. In an independent recent study, a *piggyBac* transposase gene was also found to have an essential role in DNA deletion in *Paramecium* (Baudry *et al.*, 2009), although its exact mode of action was unclear. Thus, the domestication of *piggyBac* transposase gene for DNA deletion appears to have preceded the split of *Tetrahymena* and *Paramecium* in evolution. In *Tetrahymena*, our studies have further demonstrated that *TPB2* participates in heterochromatin organization and likely carries out the DNA cleavage reaction, which was not known from the *Paramecium* study. The expression timing and localization of Tpb2p suggest that the actions of Tpb2p are with the DNA elimination heterochromatin structures at late stages of conjugation. The *TPB2* RNAi silencing studies further revealed its essential role in the assembly of these distinctive heterochromatin structures: their assembly was blocked without the normal expression of *TPB2*. In these cells the targeting of histone modifications and Pdd1p to the IES-containing chromatin occurred normally. Thus, *TPB2* acts downstream of the targeting and appears to play an essential role in inducing large heterochromatin structure formation. The timing of this cytological stage coincides with that of DNA cutting, placing the action of *TPB2* in this critical stage of DNA elimination. The evolutional conservation of the DDD catalytic sites in Tpb2p offered a hint for the endonuclease activity. The in vitro enzymatic assay results provided direct evidence that Tpb2p indeed has such endonuclease activities, which could

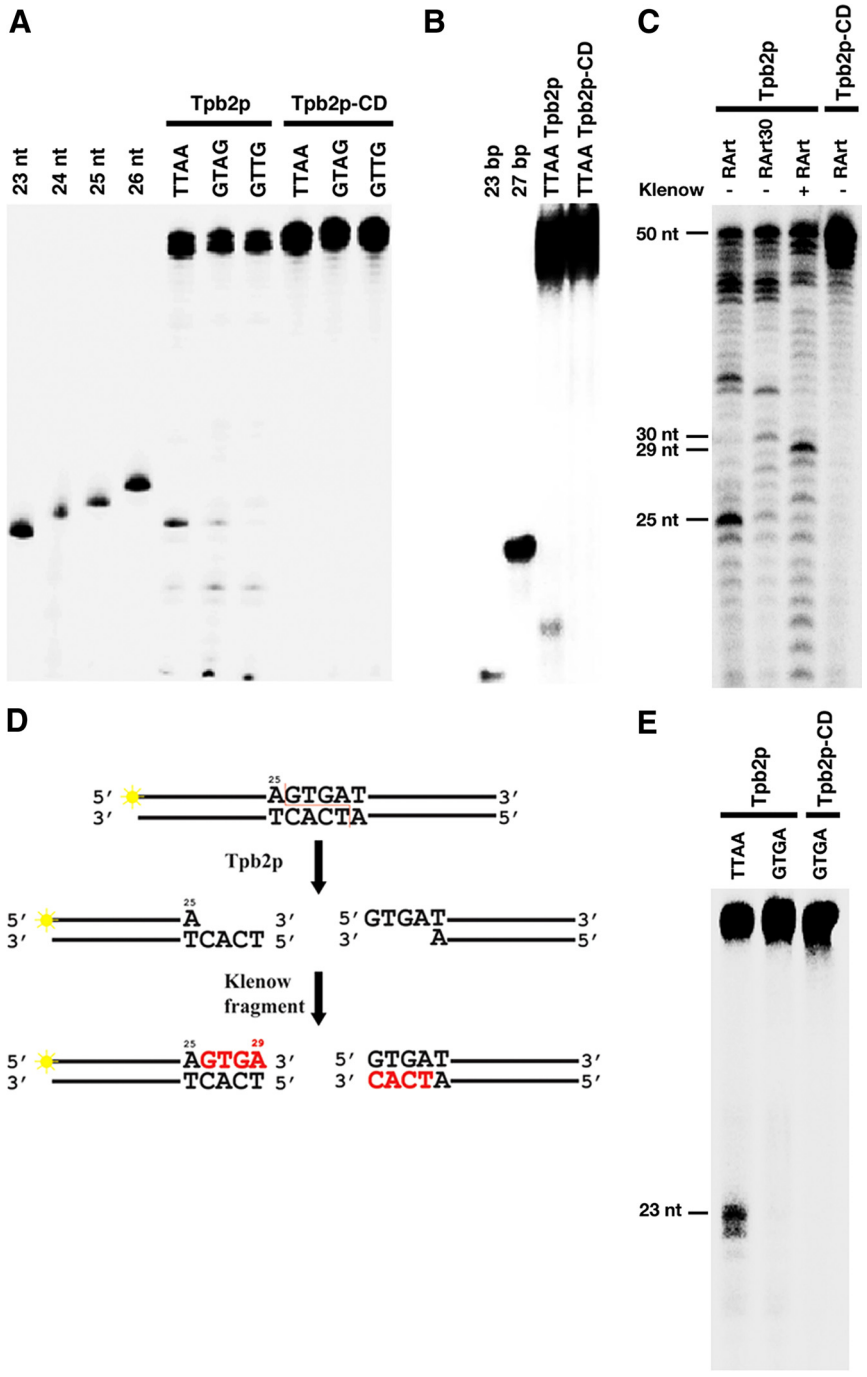

Figure 8. Tpb2p has DDD-motif dependent endonuclease activities. (A) Wild-type Tpb2p and a mutant Tpb2p in which all three aspartic acids in the DDD motif were replaced by leucines (Tpb2p-CD) were expressed in *E. coli* and incubated with 50-base pair DNA substrates that had TTAA, GTAG, or GTTG sequence once at their middle. One of two strands was 5′-end labeled. These products were separated in a denaturing polyacrylamide gel and detected by a phosphorimager. 23-, 24-, 25-, and 26-nt oligo DNAs were used as size markers. (B) The same product from the 50-base pair DNA substrate, which has TTAA shown in A, was separated in a native polyacrylamide gel and detected by a phosphorimager. Twenty-three- and 27-base pair oligo DNAs were used as size markers. (C) Tpb2p and Tpb2p-CD were expressed in *E. coli* and incubated with 50-base pair DNA substrates RArt or RArt30, which had AGTGAT sequence identified at a boundary of *Tetrahymena* R-element at their 25th and 30th position, respectively. One of two strands was 5′-end labeled. An aliquot of the product from RArt was also treated with Klenow fragment. These products were separated in a denaturing polyacrylamide gel and detected by a phosphorimager. Positions of 25-, 29-, 30-, and 50-nt oligo DNAs used as size markers are indicated in left. (D) Schematic representation of RArt cleavage by Tpb2p and followed by the fill-in reaction by Klenow fragment. ^{32}P-labeled end is marked with yellow stars, and nucleotides added by Klenow fragment are highlighted by red. (E) Tpb2p and Tpb2p-CD were expressed in *E. coli* and incubated with 50-base pair DNA substrates which had TTAA or GTGA sequence once at their middle. One of two strands was 5′-end labeled. These products were separated in a denaturing polyacrylamide gel and detected by a phosphorimager. Twenty-three-nucleotide oligo DNA was used as size markers.

have been evolved from the transposase activity. The fact that the same type of ends found in DNA deletion, a 4-nt 5′ protruding end, was generated by this protein further indicated the relevance of this endonuclease activity to DNA deletion.

In many transposons, their terminal repeat sequences are the targets of tranposases. However, IESs in *Tetrahymena* do not share common sequence feature at their termini, and their identities are apparently established through heterochromatin formation guided by small RNAs. Because Tpb2p colocalizes with the heterochromatin component Pdd1p (Figures 5C and 6D) and is required for the assembly of IES elimination heterochromatin structure, Tpb2p would have the ability to recognize some features of heterochromatin. We speculate that Tpb2p may be recruited to IES regions by heterochromatin structures but not by specific DNA sequences. This is reminiscent of the chromatin structure–dependent recruitment of some transposases. It is known that heterochromatin conformation enhances the recruitment of *Sleeping Beauty* transposases and facilitates transposition (Yusa *et al.*, 2004; Ikeda *et al.*, 2007), and RAG1/2 recombinases, which are domesticated transposases, can interact with H3K4me3-containing chromatin in the V(D)J recombination process (Matthews *et al.*, 2007). Understanding how Tpb2p interacts with heterochromatin could shed light on these processes.

On the other hand, our in vitro study showed that some preference for a DNA-cutting site sequence was preserved in Tpb2p (Figure 8). Previous studies have shown that, in the R and M elements, the deletion boundaries are determined by flanking *cis*-acting sequences located 40–50 base pairs away, and the precise nature of the deletion are dependent on the sequences at the boundaries (Chalker and Yao, 1996; Yao, 1996). Indeed, when new deletions were induced by RNA injections, no precise boundaries were observed, presumably due to the lack of these *cis*-acting sequences in regions not normally targeted for deletion (Yao *et al.*, 2003). The sequence preference of Tpb2 cutting activities we have observed in this study agrees well with those in vivo findings. Thus, deletion boundary is possibly established through an interaction between Tpb2p and the flanking and boundary sequences after it is targeted to the heterochromatin region.

Although piggyBac transposase exclusively targets TTAA sequence by generating a double-strand break at the 5′ end of TTAA, this sequence is not commonly found at the known deletion boundaries in *Tetrahymena*. Instead, ANNNNT was identified as a consensus sequence at the ends of a number of IESs (Saveliev and Cox, 1996). Our in vitro study indicated that Tpb2p has ability to induce a double-strand break not only at the TTAA site (Figure 8A) but also at GTAG (Figure 8A) and AGTGAT (Figure 8C) sites. Thus, Tpb2p has probably acquired an ability to target some non-TTAA sequence. Future detailed in vitro study of Tpb2p may elucidate how targeting specificity of Tpb2p determines deletion boundaries of IESs in vivo.

TPB2 appears to have gone through a domestication process to become a host gene and be maintained in the macronucleus. Because programmed DNA deletion is a common trait of most ciliates, we speculate that *TPB2* domestication occurred early in their evolution, agreeing well with the observation that *piggyBac* transposase homolog gene is also involved in programmed genome rearrangements in *Paramecium tetraurelia* (Baudry *et al.*, 2009). However, studies of a distantly related ciliate *Oxytricha trifallax* have suggested the involvement of transposases encoded by a group of abundant germline transposons (TBEs) in IES deletion (Nowacki *et al.*, 2009). Therefore, direct participation of transposons in their own removal in a programmed manner, which can be observed in *O. trifallax* today, could represent an ancient system in the evolution of DNA deletion, which could be followed by the domestication of one of the transposases to carry out most deletion processes like in *Tetrahymena*. Although several domesticated transposon proteins are known to be involved in host genome regulations in other eukaryotes, their evolutional histories are not clear. Ciliated protozoa may preserve traceable "fossil" records that help illuminate how a transposon has been domesticated to regulate eukaryotic genome.

ACKNOWLEDGMENTS

We thank all members of the Yao lab and the Mochizuki lab for helpful discussions and suggestions. This research received funding from European Research Council Starting Grant (204986) under the European Community's Seventh Framework Programme and from the Austrian Academy of Sciences to K.M., from the National Science Council (NSC97-2321-B-001-025) of Taiwan, ROC, and from Institute of Molecular Biology of Academia Sinica to M.C.Y.

REFERENCES

Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature *394*, 744–751.

Aronica, L., Bednenko, J., Noto, T., DeSouza, L. V., Siu, K. W., Loidl, J., Pearlman, R. E., Gorovsky, M. A., and 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, C. F., Allis, C. D., and Yao, M. C. (1984). Specific DNA rearrangements in synchronously developing nuclei of *Tetrahymena*. Proc. Natl. Acad. Sci. USA *81*, 7383–7387.

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., and Betermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. Genes Dev. *23*, 2478–2483.

Borst, P., and Greaves, D. R. (1987). Programmed gene rearrangements altering gene expression. Science *235*, 658–667.

Cary, L. C., Goebel, M., Corsaro, B. G., Wang, H. G., Rosen, E., and Fraser, M. J. (1989). Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni transposon* IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. Virology *172*, 156–169.

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

Chadee, D. N., Hendzel, M. J., Tylipski, C. P., Allis, C. D., Bazett-Jones, D. P., Wright, J. A., and Davie, J. R. (1999). Increased Ser-10 phosphorylation of histone H3 in mitogen-stimulated and oncogene-transformed mouse fibroblasts. J. Biol. Chem. *274*, 24914–24920.

Chalker, D. L. (2008). Dynamic nuclear reorganization during genome remodeling of *Tetrahymena*. Biochim. Biophys. Acta *1783*, 2130–2136.

Chalker, D. L., and Yao, M. C. (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, D. L., and Yao, M. C. (2001). Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. Genes Dev. *15*, 1287–1298.

Coyne, R. S., Chalker, D. L., and Yao, M. C. (1996). Genome downsizing during ciliate development: nuclear division of labor through chromosome restructuring. Annu. Rev. Genet. *30*, 557–578.

Coyne, R. S., Nikiforov, M. A., Smothers, J. F., Allis, C. D., and Yao, M. C. (1999). Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation. Mol. Cell *4*, 865–872.

Dedon, P. C., Soults, J. A., Allis, C. D., and Gorovsky, M. A. (1991). Formaldehyde cross-linking and immunoprecipitation demonstrate developmental changes in H1 association with transcriptionally active genes. Mol. Cell. Biol. *11*, 1729–1733.

Doak, T. G., Doerder, F. P., Jahn, C. L., and Herrick, G. (1994). A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc. Natl. Acad. Sci. USA *91*, 942–946.

Elick, T. A., Bauser, C. A., and Fraser, M. J. (1996). Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase. Genetica *98*, 33–41.

Fraser, M. J., Ciszczon, T., Elick, T., and Bauser, C. (1996). Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of *Lepidoptera*. Insect Mol. Biol. *5*, 141–151.

Godiska, R., and Yao, M. C. (1990). A programmed site-specific DNA rearrangement in *Tetrahymena thermophila* requires flanking polypurine tracts. Cell *61*, 1237–1246.

Howard-Till, R. A., and Yao, M. C. (2006). Induction of gene silencing by hairpin RNA expression in *Tetrahymena thermophila* reveals a second small RNA pathway. Mol. Cell. Biol. *26*, 8731–8742.

Ikeda, R., Kokubu, C., Yusa, K., Keng, V. W., Horie, K., and Takeda, J. (2007). Sleeping beauty transposase has an affinity for heterochromatin conformation. Mol. Cell. Biol. *27*, 1665–1676.

Kapitonov, V. V., and Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. *3*, e181.

Klobutcher, L. A., and Herrick, G. (1997). Developmental genome reorganization in ciliated protozoa: the transposon link. Prog. Nucleic Acid Res. Mol. Biol. *56*, 1–62.

Liu, Y., Taverna, S. D., Muratore, T. L., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in *Tetrahymena*. Genes Dev. *21*, 1530–1545.

Madireddi, M. T., Coyne, R. S., Smothers, J. F., Mickey, K. M., Yao, M. C., and Allis, C. D. (1996). Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in *Tetrahymena*. Cell *87*, 75–84.

Madireddi, M. T., Davis, M. C., and Allis, C. D. (1994). Identification of a novel polypeptide involved in the formation of DNA-containing vesicles during macronuclear development in *Tetrahymena*. Dev. Biol. *165*, 418–431.

Malone, C. D., Anderson, A. M., Motl, J. A., Rexer, C. H., and Chalker, D. L. (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.

Matthews, A. G., *et al.* (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature *450*, 1106–1110.

Mochizuki, K., Fine, N. A., Fujisawa, T., and Gorovsky, M. A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. Cell *110*, 689–699.

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

Nikiforov, M. A., Gorovsky, M. A., and Allis, C. D. (2000). A novel chromodomain protein, pdd3p, associates with internal eliminated sequences during macronuclear development in *Tetrahymena thermophila*. Mol. Cell. Biol. *20*, 4128–4134.

Nikiforov, M. A., Smothers, J. F., Gorovsky, M. A., and Allis, C. D. (1999). Excision of micronuclear-specific DNA requires parental expression of pdd2p and occurs independently from DNA replication in *Tetrahymena thermophila*. Genes Dev. *13*, 2852–2862.

Nowacki, M., Higgins, B. P., Maquilan, G. M., Swart, E. C., Doak, T. G., and Landweber, L. F. (2009). A functional role for transposases in a large eukaryotic genome. Science *324*, 935–938.

Rexer, C. H., and Chalker, D. L. (2007). Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in *Tetrahymena thermophila*. Eukaryot. Cell *6*, 1320–1329.

Sarkar, A., Sim, C., Hong, Y. S., Hogan, J. R., Fraser, M. J., Robertson, H. M., and Collins, F. H. (2003). Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. Mol. Genet. Genom. *270*, 173–180.

Saveliev, S. V., and Cox, M. M. (1995). Transient DNA breaks associated with programmed genomic deletion events in conjugating cells of *Tetrahymena thermophila*. Genes Dev. *9*, 248–255.

Saveliev, S. V., and Cox, M. M. (1996). Developmentally programmed DNA deletion in *Tetrahymena thermophila* by a transposition-like reaction pathway. EMBO J. *15*, 2858–2869.

Smit, A. F., and Riggs, A. D. (1996). Tiggers and DNA transposon fossils in the human genome. Proc. Natl. Acad. Sci. USA *93*, 1443–1448.

Smith, J. J., Antonacci, F., Eichler, E. E., and Amemiya, C. T. (2009). Programmed loss of millions of base pairs from a vertebrate genome. Proc. Natl. Acad. Sci. USA *106*, 11212–11217.

Taverna, S. D., Coyne, R. S., and Allis, C. D. (2002). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. Cell *110*, 701–711.

Volff, J. N. (2006). Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays *28*, 913–922.

Yao, M. C. (1996). Programmed DNA deletions in *Tetrahymena*: mechanisms and implications. Trends Genet. *12*, 26–30.

Yao, M. C., and Chao, J. L. (2005). RNA-guided DNA deletion in *Tetrahymena*: an RNAi-based mechanism for programmed genome rearrangements. Annu. Rev. Genet. *39*, 537–559.

Yao, M. C., Fuller, P., and Xi, X. (2003). Programmed DNA deletion as an RNA-guided system of genome defense. Science *300*, 1581–1584.

Yu, G. L., and Blackburn, E. H. (1989). Transformation of *Tetrahymena thermophila* with a mutated circular ribosomal DNA plasmid vector. Proc. Natl. Acad. Sci. USA *86*, 8487–8491.

Yusa, K., Takeda, J., and Horie, K. (2004). Enhancement of Sleeping Beauty transposition by CpG methylation: possible role of heterochromatin formation. Mol. Cell. Biol. *24*, 4004–4018.

4.2 Transposon Domestication versus Mutualism in Ciliate Genome Rearrangements

Vogt A, Goldman AD, Mochizuki K, Landweber LF.

PLoS Genetics (submitted)

Transposon Domestication versus Mutualism in Ciliate Genome Rearrangements

Vogt A[1], Goldman AD[2], Mochizuki K[1], Landweber LF[2*]

[1]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA) Vienna, Austria

[2]Department of Ecology and Evolutionary Biology, Princeton University, Guyot Hall, Princeton, NJ 08544, USA

*corresponding author

Abstract

Ciliated protists rearrange their genomes dramatically during a process of development that fragments chromosomes and deletes DNA sequences to produce a trimmer--and sometimes massively-reorganized--somatic genome. The deleted sequences include potentially active transposons or transposon-like sequences that reside in the germline genome. Three independent studies recently showed that transposase proteins of the DDE/DDD superfamily are indispensible for DNA processing in three different and evolutionarily distantly-related ciliates. In the spirotrich ciliate *Oxytricha*, high copy-number germline-limited TBE transposons mediate their own removal from the somatic genome but also contribute to programmed genome rearrangement through a type of transposon mutualism with the host. By contrast, the genomes of two oligohymenophoran ciliates, *Tetrahymena* and *Paramecium*, encode homologous PiggyBac-like transposases as single copy genes in both their germline and somatic genomes. These domesticated transposases are necessary for the removal of thousands of different internal sequences from the respective genomes during nuclear development. Here we compare the events underlying DNA deletion in three different ciliates and discuss the evolutionary origins and relations of different DNA deletion and rearrangement mechanisms.

Introduction

A transposon rearranges its host's genome when it moves from one genomic locus to another. Because they sometimes invade coding or regulatory genomic regions, transposons can therefore interfere with gene expression. Also, transposon-induced DNA double-strand

breaks can cause chromosomal rearrangement that leads to aneuploidy. Thus, for a long time transposons were considered to be harmful and selfish, part of so-called “junk DNA” [1]. On the other hand, because most eukaryotic genomes have maintained transposons and transposon-derived DNA throughout the course of evolution, the presence of those sequences may sometimes confer an adaptive benefit to the host [2]. Indeed, the activity of transposons can accelerate genome evolution by fabricating new DNA sequences and facilitating genome rearrangement.

In many cases the host manages to recruit or to “domesticate” transposon-encoded genes and repurpose them for new host functions [3,4]. A famous example in jawed vertebrates is the evolution of the gene encoding Rag1 from a Transib-like element. Now a key component in *V(D)J* recombination, it is responsible for cutting and rejoining segments from *V*, *D* and *J* genes [5,6]. As this process is indispensable for the maturation of B and T cells, the domestication of the *RAG1* gene enabled the evolution of adaptive immune systems [6]. Other examples of domesticated transposases include the yeast *Klyveromyces lactis* α3 MULE transposase-like protein, which is necessary for switching from mating type α to mating type a [7,8]. In addition, the *C. elegans* HIM-17 is a domesticated P-element–like transposase that is essential for double strand break and chiasma formation during meiosis, as well as for the accumulation of histone H3 methylation at lysine 9 on meiotic prophase chromosomes [8]. Therefore, transposon domestication is widespread, and transposons provide the toolkits for host cells to acquire new functions on an evolutionary time scale. However, the processes by which transposons become domesticated are not always clear.

Recently, it has been reported that transposase-related proteins play crucial roles in large scale genomic rearrangements in three different ciliated protists [9,10,11]. Curiously, while two ciliate species that are members of the Oligohymenophorea use single-copy domesticated transposase genes in genomic rearrangements, *Oxytricha*, a member of a different, deeply-diverging ciliate class (Spirotrichea), requires instead the expression of thousands of active transposase genes that are still present on intact transposons. Therefore, comparison of transposases involved in the genome rearrangements of different ciliates may provide us with a unique opportunity to understand the process of transposon recruitment for new functions, a molecular exaptation event [12]. This review will discuss the functions and evolution of those transposon-derived proteins involved in genome rearrangements in ciliates.

Programmed genome rearrangements in ciliates

Ciliates are mostly single-celled eukaryotes and all members of the Alveolata [13]. A common feature of most ciliates is nuclear dimorphism. The cells harbor two distinct types of nuclei in the same cytoplasm. The larger DNA-rich macronucleus [14] is polyploid and provides most, if not all, gene expression during vegetative growth and can be considered the somatic nucleus. The smaller germline micronucleus [15] is diploid and transcriptionally silent during most life stages except for conjugation. The exact numbers of macronuclei and micronuclei vary among ciliate species [16]. For example, *Oxytricha trifallax* and *Paramecium tetraurelia* have one macronucleus and two micronuclei in interphase vegetative cells, whereas *Tetrahymena thermophila* has one macronucleus and one micronucleus (Fig 1A) [17]. During vegetative growth, both the macronucleus and micronucleus divide and are independently segregated to daughter cells. In contrast, during sexual reproduction (conjugation or autogamy) of ciliates, the zygotic micronucleus gives rise to both the macronucleus and new micronucleus in the daughter cells, thereby supplying the next generation with all its genetic information. However, the macronucleus and micronucleus do not have identical genetic content because the somatic genome undergoes a dramatic rewiring during the developmental cascade that restructures a new macronucleus from the zygotic micronucleus, after cell mating [18,19].

Two major types of genome rearrangement occur during macronuclear development in all ciliates: chromosome fragmentation and DNA elimination [17] (Fig. 1B). Chromosome fragmentation produces larger numbers of considerably smaller macronuclear chromosomes than germline, micronuclear chromosomes. Macronuclear chromosomes in oligohymenophoran ciliates have an average size of 300 kbp in *Tetrahymena* and 800 kbp in *Paramecium* [19,20,21,22]. In contrast, spirotrich ciliates like *Oxytricha* typically have “gene-sized” nanochromosomes in the macronucleus, with an average length of just ~3 kbp including telomeres, and 90% of them do encode just a single gene [Swart et al., in review]. The second type of genome rearrangement, DNA elimination, discards between 20% (*Tetrahymena*) and ~95% (*Oxytricha*) of the entire germline genome during macronuclear development [18]. In some spirotrich ciliates, as well as phyllopharyngeans [23], DNA elimination events in a proportion of loci are accompanied by a third type of genome rearrangement, gene unscrambling (Fig. 1C). These events reorder gene pieces in the

micronucleus by translocation and sometimes inversion to assemble coding information in the macronucleus [24]. Despite the genome downsizing via DNA elimination, the macronucleus contains a greater quantity of DNA than the micronucleus. This is because macronuclear chromosomes undergo endoreplication to amplification levels that typically range from 50- (*Tetrahymena*) to 2000-fold (*Oxytricha*) [18]. In this review, we focus on DNA elimination, and for a review of genome unscrambling and the role of RNA-guided epigenetic effects in this process, we refer the reader to Nowacki et al. [2011] [25]. Although DNA elimination is common to most ciliates, recent studies have revealed a dependence on strikingly different groups of transposase-related proteins for DNA elimination in different classes of ciliates [9,10,11].

Deletion of germline-limited sequences in *Oxytricha*

Elimination of germline-restricted DNA sequences occurs at precise, nucleotide-level resolution in *Oxytricha trifallax*. One well-studied example of precisely-removed germline-limited sequences are the Tc1/mariner transposons of the TBE (telomere bearing element) class, which are present in thousands of copies in the micronucleus [26] and occupy roughly as much of the micronuclear genome as its coding content (Bracht et al., in preparation). TBE terminal regions possess inverted repeats, with the most distal 17 bp composed of telomeric repeats $((G_4T_4)_2G)_n$ and the elements are flanked by a 3 bp 5'-ANT-3' target site duplication (Fig 2). TBE excision precisely removes one target site repeat, thereby restoring functional open reading frames (ORFs) even when TBEs interrupt protein coding regions in the micronucleus. Mechanistically, it is likely that excision starts with the introduction of a double-stranded break (DSB) creating a 3 nt 5' protruding end on one end of the transposon. The other target site serves as an “integration site” so that TBEs are excised in a circular form, the TBE ring is degraded and the macronuclear DNA is re-ligated [27] (Fig. 2).

One of the genes encoded in TBEs encodes a protein belonging to the superfamily of DDE transposases, suggesting an involvement of this enzyme in the transposon’s own removal [27,28] (Fig 2). Furthermore, all three ORFs of TBE transposons appear to be under purifying selection, which initially hinted at an essential function of TBE encoded transposases [29,30]. *Oxytricha trifallax* has three different types of TBE transposons, TBE1, TBE2 and TBE3. The transposases encoded by these elements share ≥ 83 % similarity at the protein level, and all three types of transposases are specifically expressed during

macronuclear development when DNA rearrangements occur. Silencing of all three groups of TBE transposases in unison (but not of TBE1, 2, or 3 on their own) via RNAi, in an experiment that uniquely targeted a significant fraction of the germline genome, results in severe defects in DNA elimination of both TBE transposons and non-TBE micronucleus-limited elements, as well as an accumulation of high molecular weight DNA [10]. These results suggest that TBE transposases act redundantly in excising both the transposons that encode them and other non-TBE micronucleus-limited sequences (called "internal eliminated sequences" or IES). It also suggests that a massive quantity of transposase may be required for *Oxytricha* genome rearrangement [10].

DNA deletion in *Paramecium* and *Tetrahymena*

Paramecium tetraurelia also has two different types of eliminated sequences. Most repetitive micronucleus-limited sequences, which are similar to minisatellites or transposons, are eliminated imprecisely [19]. In contrast, removal of approximately 60,000 different non-repetitive micronucleus-limited sequences (IES) occurs precisely at nucleotide-level resolution during development of the macronuclear genome. Both classes of eliminated sequences are flanked by a 5'-TA-3' dinucleotide, which is a part of a weakly conserved 8 bp sequence that bears similarity to the recognition sequence of some Tc1/mariner transposases [31,32]. Furthermore a 5'-TA-3' dinucleotide is preserved after religation of the flanking regions (Fig. 3), a feature that is also similar to excision of Tc1/mariner transposases. This could suggest that some *Paramecium* IES are remnants of Tc1 mariner transposons. However, DNA elimination starts with a DNA double-strand break that produces a 4-base 5'-overhang [33] (Fig. 3). This is in sharp contrast with the type of DNA double-strand break produced by Tc1/mariner transposases that yield 2-base 3'-overhangs [34]. Therefore, the enzyme(s) responsible for DNA elimination in *Paramecium* were expected to be unrelated to Tc1/mariner transposases.

DNA elimination in *Tetrahymena thermophila* [35] also produces double-strand breaks with 4-base 5'-overhangs. *Tetrahymena* removes approximately 9,000 different micronucleus-limited sequences from its macronuclear genome (*Tetrahymena* Comparative Sequencing Project, Broad Institute of Harvard and MIT, www.broadinstitute.org), an order of magnitude fewer IES than *Paramecium*. Most of *Tetrahymena*'s germline-limited sequences studied so far are eliminated imprecisely, and hence the splice junctions

generally fall outside of coding regions to avoid disrupting functional sequences. However, recently Fass et al. provided evidence for a small number of IES within exons [36], and hence these must be removed with nucleotide precision, unless they are present in weakly constrained regions. IES length in *Tetrahymena* usually ranges from 600 bp to ~20 kb, with some sequences bearing similarity to transposons of the Tc1/mariner family or non-LTR retrotranposons [22,37]. Although no obvious consensus sequence has been identified at IES boundaries in *Tetrahymena*, the common form of DNA double-strand breaks (i.e. 4 base 5'-overhangs) at their boundaries in *Paramecium* and *Tetrahymena* suggests that both ciliates use similar enzymes for DNA elimination.

The fact that transposases of the PiggyBac family of transposons do produce DNA double-strand breaks with 4-base 5'-overhangs [38] led researchers to search for PiggyBac transposases in *Paramecium* and *Tetrahymena*. Indeed, the macronuclear genomes of both species contain genes derived from PiggyBac family transposases, and Baudry et al. and Cheng *et al*. independently showed that a transposase of the PiggyBac family plays a crucial role in DNA elimination during maturation of the macronuclear genomes in *Paramecium* and *Tetrahymena* [9,39]. These ciliate transposon-derived proteins are called Pgm (PiggyMac) in *Paramecium* and Tpb2p (*Tetrahymena* PiggyBac like transposase 2) in *Tetrahymena*. In both *Paramecium* and *Tetrahymena*, silencing of the respective PiggyBac transposase-like genes by RNAi inhibits the processes of DNA elimination and macronuclear development [9,11]. Both Pgm and Tpb2p have a predicted catalytic domain with conserved DDD residues, similar to PiggyBac transposases. Consistent with this observation, *in vitro* studies with Tpb2p that was recombinantly expressed in *E. coli* revealed that Tpb2p produces a double-strand break leaving a 4-base 5' protruding end, which correlates with the typical cleavage signature of canonical PiggyBac transposases [11,38] and the observed form of DSB during DNA elimination *in vivo* [35] (Fig. 3). Therefore, Pgm and Tpb2p are probably the enzymes responsible for catalyzing DNA excision during the process of DNA elimination in *Paramecium* and *Tetrahymena*, respectively.

Both Pgm and Tpb2p localize to the newly developing macronucleus at the time when DNA elimination occurs. In *Tetrahymena*, Tpb2p localizes to the subnuclear heterochromatin bodies where DNA elimination is thought to occur [9,11]. These heterochromatin structures contain heterochromatin-specific histone modifications, tri-methylated histone H3 lysine 9 (H3K9me3) and lysine 27 (H3K27me3), and the

chromodomain protein Pdd1p [15]. The localization of Tpb2p to these structures is possibly mediated by an interaction with some of these or other heterochromatin components. Pgm and Tpb2p share a predicted zinc finger domain and a hydrophobic coiled-coil domain at the C-terminus (Fig. 4), and these domains may directly interact with some of the heterochromatin components. Because heterochromatin is specifically established on deleted DNA sequences via an RNAi-related pathway, the heterochromatin interaction of PiggyBac transposase-like proteins may restrict their action to programmed deleted sequences.

In addition to the association of PiggyBac transposase-like proteins with heterochromatin, their enzymatic preference for certain DNA sequences may also facilitate DNA elimination at specific sequences. For instance, the minimalistic 5'-TA-3' dinucleotide consensus sequence that flanks *Paramecium* deleted regions is also the smallest sequence recognized by the canonical PiggyBac transposase, whose optimal recognition sequence is 5'-TTAA-3' [38]. Moreover, recombinant Tpb2p can specifically cleave *in vitro* a dsDNA oligonucleotide containing 5'-TTAA-3' before the first position of this motif [11]. In addition, recombinant Tpb2p preferentially cleaves the left boundary of a deleted region in *Tetrahymena* (5'-AGTGAT-3') between the first A and the first G, when the motif was placed in the middle of an otherwise randomly-designed 50 bp dsDNA oligonucleotide. Therefore, although the 5'-TA-3' motif is not necessary for the catalytic action of Tpb2p, the enzyme probably recognizes limited sequence context to induce a double strand break. However, because no obvious consensus sequence exists between different deleted regions in *Tetrahymena*, though DSBs are induced at limited regions during DNA elimination *in vivo*, it is unlikely that primary DNA sequences solely determine the sites of DSBs. Most likely, both the heterochromatin interaction of Tpb2p and Pgm and their preference for certain DNA sequences determine DSB sites for DNA elimination.

Transposon domestication versus mutualism: possible evolutionary origins

Although Pgm and Tpb2p are similar to PiggyBac transposases, they are not a part of modern ciliate transposons. The genes encoding these proteins are present in the macronuclear genome and not in proximity to any other transposon-related sequences. Therefore, these PiggyBac transposase-like genes are classic examples of transposon domestication by the host genome to mediate a new function, in this case DNA elimination.

Pgm and Tpb2p are homologous proteins that share 31% identity (Fig. 4), which suggests the possibility of a single domestication of a PiggyBac-like transposase in a common oligohymenophoran ancestor; however, the alternative scenario of recruitment of similar transposases in both genomes cannot be excluded at this point. After the *Paramecium* and *Tetrahymena* lineages separated, the domesticated transposases accumulated substitutions that could contribute to the slight differences in their DNA deletion pathways, as well as the apparent promiscuity of Tpb2p at IES boundaries in *Tetrahymena*.

The situation of a single domesticated transposase in Oligohymenophorea is in sharp contrast to *Oxytricha*'s distributed system that appears to enlist an army of thousands of TBE transposases while they are still components of potentially active Tc1/mariner transposons (Fig. 4) and occupy a significant fraction of the germline genome. *Oxytricha*'s process of macronuclear development is itself much more complex, however, requiring hundreds of thousands of rearrangement events, and this may explain its need for increased transposase participation. The greater complexity of genome rearrangements does not, however, explain why *Oxytricha* should recruit undomesticated transposases to facilitate genome rearrangement, but this strategy may be easier to evolve, as active transposons would multiply in number, up to the ceiling tolerated by its host. This strategy would thus ensure production of an ample quantity of transposase protein, in part because these enzymes must also facilitate elimination of their parent transposons, to produce a mature macronucleus. This evolutionary situation can be considered a mutualism between the host and its resident germline transposons [40]. It also wonderfully displays a functional and essential role for the otherwise dispensable portion of the micronuclear genome occupied by TBE transposons [10]. During the evolution of spirotrich ciliates, TBE transposases may have gained promiscuity and acquired the ability to excise off-target DNA sequences [27,28]. Additionally, only DNA insertions with the ability to be excised by TBE transposases or other active enzymes would have been tolerated in the germline genome over time, and thus able to accumulate. Such a mutualistic system would have allowed not only the accumulation of germline transposons but also the production of a sufficient quantity of transposase protein to facilitate *Oxytricha*'s elaborate process of genome remodeling. These requirements would have provided the selective pressure to maintain high copy numbers of TBE transposons to facilitate DNA elimination. On the other hand, the processes of DNA elimination that domesticated PiggyBac transposases facilitate in *Paramecium* and

Tetrahymena do not require maintenance of germline transposons. This striking difference in two evolutionary lineages separated by over a billion years may have been exaggerated over time by an evolving trend in the *Oxytricha* lineage to eliminate and rearrange considerably more of its micronuclear genome than *Paramecium* or *Tetrahymena*.

The ostensible similarities and likely homology (Fig. 4) between PiggyBac and TBE transposons belie their differences. How did different ciliate lineages acquire different types of transposases and co-evolve such different strategies between the transposons and their hosts to mediate different pathways of genome differentiation? Because oligohymenophoran and spirotrich ciliates are evolutionarily more distant from each other than plants and animals, a plausible explanation for the recruitment of different types of transposases in DNA elimination pathways in these distant ciliates is independent origins. However, it is also possible that the mutualistic system in *Oxytricha* may have predated DNA elimination by a domesticated transposase. A later transposon-domestication event or events that result in a high quantity of active transposase in the ancestral oligohymenophoran lineage could have lessened the dependency on feral transposons distributed throughout the genome. The modern piggyBac-like element in *Paramecium* and *Tetrahymena* might be a relic from a transposon that was initially maintained in the micronucleus by a mutualistic system more like *Oxytricha*'s, and then later a copy of its transposase gene could have accidently lost the signals for DNA deletion and become a resident of the macronucleus as well, where it accumulated additional substitutions. Then this PiggyBac transposase, if expressed at sufficiently high levels, could have taken over the former roles of TBE or other transposases, reducing the levels of purifying selection that act on the germline transposases until they became redundant with the function of the domesticated transposase. This relaxation of constraints on germline transposons would permit them to ameliorate to the background of micronuclear-limited DNA, scattering transposon remnants in the micronuclear genome, until most are eventually unrecognizable. Accordingly, sequences related to TBE transposases are present in the *Tetrahymena* micronuclear genome, and some of them appear to possess functional open reading frames that maintain the DDE catalytic triad (one of these sequences is labeled "42 kDa transposase" in Fig. 4). Therefore, these DNA sequences could even be remnants from a TBE mutualistic system, and the minimal conservation suggests the possibility that TBE transposases could still contribute some role to DNA elimination in oligohymenophoran

ciliates. In this context, it would be fruitful to study the function of these newly discovered TBE transposase genes in oligohymenophorans.

Conclusions

As discussed above, the roles of transposase proteins in ciliate programmed DNA elimination are just coming to light. It will be necessary to study how TBE and PiggyBac transposases interact with chromatin and how they induce DNA double-strand breaks. DNA elimination events, which are initiated by double-strand breaks, must be swiftly followed by DSB repair. The DNA elimination pathway in *Paramecium* and *Tetrahymena* does require non-homologous end joining (NHEJ) DSB repair machinery [41,42]. Therefore, it would also be important to understand how transposases and the NHEJ machinery interact with each other and how they cooperatively regulate DNA elimination. From an evolutionary point of view, it is unclear how two such distant groups of ciliates, oligohymenophorans (*Paramecium and Tetrahymena*) and spirotrichs (*Oxytricha*) evolved such different DNA deletion systems, dependent on PiggyBac and TBE transposases, respectively. Because these two lineages represent just a modest fraction of ciliate biological diversity, and because some level of DNA elimination may be ancestral to ciliates [43], it would be valuable to investigate the functional and evolutionary relationships between transposases and DNA elimination events in different, deeply divergent classes of ciliates. Such future studies will eventually provide a better natural history of transposase recruitment and the forces of mutualism versus domestication on an evolutionary time scale.

Acknowledgements

We want to thank Kensuke Kataoka, Janine Beisson and Wenwen Fang for providing us with ciliate images. We thank Mireille Betermier and several current and past members of the Mochizuki and Landweber labs, especially Brian Higgins and Tom Doak, for helpful discussion.

Figure legends

Figure 1: Nuclear dimorphism and genome rearrangements in ciliates. A) From left to right: *Tetrahymena thermophila*, *Paramecium tetraurelia*, *Oxytricha trifallax*. DNA is shown in

cyan, yellow represents tubulin staining. Images were kindly provided by Kensuke Kataoka (IMBA, Vienna), Janine Beisson (CNRS, Gif sur Yvette) and Wenwen Fang (Princeton University, Princeton). i = micronucleus, a = macronucleus. In *Oxytricha trifallax*, two lobes of a macronucleus are connected by a thin nuclear bridge (not visible in the image) B) Genome rearrangements in all ciliates shown include elimination of micronucleus [15]-limited sequences (i, purple) and chromosome breakage, which in *Tetrahymena* occurs at specific chromosome breakage sites (c). After re-ligation of the flanking macronuclear (MAC) sequences, *Tetrahymena* chromosomes undergo endo-replication to produce 50 identical copies. C) DNA unscrambling in *Oxytricha* involves the re-shuffling and occasional inversion of precursor germline DNA sequences (numbered blue boxes) to assemble them in the correct macronuclear order.

Figure 2: TBE transposases in *Oxytricha* are germline-limited sequences and they participate in their own removal. The encoded transposases of the Tc1/mariner family have a DDE catalytic motif. Cleavage of the germline limited sequences starts with a 3 nucleotide 5' overhang at a TNA recognition site; the second target site serves as the integration site.

Figure 3: Transposases in *Tetrahymena* and *Paramecium* belong to the PiggyBac family. As domesticated transposases they are present as single copy genes in the micronuclear and macronuclear genomes. After expression from the somatic genome they facilitate IES excision from the new macronuclear genome. IES removal occurs via a 4 base 5' protruding end. In *Paramecium*, all deleted sequences have a TA dinucleotide at both boundaries, whereas *Tetrahymena* displays no consensus sequence.

Figure 4: Phylogenetic analysis of representative transposases of the DDE/DDD superfamily. It supports the conclusion that TBE elements belong to the Tc1/mariner superfamily of transposons and also that there are TBE-like elements present in *Tetrahymena* (labeled “42 kDa transposase”). Additionally, this analysis supports the conclusion that the two PiggyBac-like transposases, Pgm and Tpb2p, in *Paramecium* and *Tetrahymena* are homologous to

each other. The tree was created with MRBayes pylogenetic inference software [44] using the alignment shown in supplementary figure S1, edited to remove regions with gaps in the consensus sequence. The phylogeny was created using a mixed amino acid substitution model and invariable gamma rate model over 200,000 iterations with a burn-in of 25%. Branch confidence values represent conditional probabilities generated by the Bayesian inference process.

1. Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. Nature 284: 604-607.
2. Kazazian HH, Jr. (2004) Mobile elements: drivers of genome evolution. Science 303: 1626-1632.
3. Volff JN (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28: 913-922.
4. Sinzelle L, Izsvak Z, Ivics Z (2009) Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci 66: 1073-1093.
5. Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol 3: e181.
6. Agrawal A, Eastman QM, Schatz DG (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394: 744-751.
7. Barsoum E, Martinez P, Astrom SU Alpha3, a transposable element that promotes host sexual reproduction. Genes Dev 24: 33-44.
8. Reddy KC, Villeneuve AM (2004) C. elegans HIM-17 links chromatin modification and competence for initiation of meiotic recombination. Cell 118: 439-452.
9. Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, et al. (2009) PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. Genes Dev 23: 2478-2483.
10. Nowacki M, Higgins BP, Maquilan GM, Swart EC, Doak TG, et al. (2009) A functional role for transposases in a large eukaryotic genome. Science 324: 935-938.
11. Cheng CY, Vogt A, Mochizuki K, Yao MC A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. Mol Biol Cell 21: 1753-1762.
12. Gould SJ, Vrba ES (1982) Exaptation - a Missing Term in the Science of Form. Paleobiology 8: 4-15.
13. Baroin-Tourancheau A, Delgado P, Perasso R, Adoutte A (1992) A broad molecular phylogeny of ciliates: identification of major evolutionary trends and radiations within the phylum. Proc Natl Acad Sci U S A 89: 9764-9768.
14. Thatcher TH, MacGaffey J, Bowen J, Horowitz S, Shapiro DL, et al. (1994) Independent evolutionary origin of histone H3.3-like variants of animals and Tetrahymena. Nucleic Acids Res 22: 180-186.

15. Madireddi MT, Coyne RS, Smothers JF, Mickey KM, Yao MC, et al. (1996) Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. Cell 87: 75-84.
16. Prescott DM (1994) The DNA of ciliated protozoa. Microbiol Rev 58: 233-267.
17. Mochizuki K DNA rearrangements directed by non-coding RNAs in ciliates. Wiley Interdiscip Rev RNA 1: 376-387.
18. Jahn CL, Klobutcher LA (2002) Genome remodeling in ciliated protozoa. Annu Rev Microbiol 56: 489-520.
19. Le Mouel A, Butler A, Caron F, Meyer E (2003) Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. Eukaryot Cell 2: 1076-1090.
20. Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, et al. (2002) The complete genome sequence of Chlorobium tepidum TLS, a photosynthetic, anaerobic, green-sulfur bacterium. Proc Natl Acad Sci U S A 99: 9509-9514.
21. Meyer E, Caron F, Baroin A (1985) Macronuclear structure of the G surface antigen gene of Paramecium primaurelia and direct expression of its repeated epitopes in Escherichia coli. Mol Cell Biol 5: 2414-2422.
22. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, et al. (2006) Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biol 4: e286.
23. Katz LA, Kovner AM (2010) Alternative processing of scrambled genes generates protein diversity in the ciliate Chilodonella uncinata. J Exp Zool B Mol Dev Evol 314: 480-488.
24. Mochizuki K RNA-directed epigenetic regulation of DNA rearrangements. Essays Biochem 48: 89-100.
25. Nowacki M, Shetty K, Landweber LF (2011) RNA-Mediated Epigenetic Programming of Genome Rearrangements. Annu Rev Genomics Hum Genet 12: 367-389.
26. Hunter DJ, Williams K, Cartinhour S, Herrick G (1989) Precise excision of telomere-bearing transposons during Oxytricha fallax macronuclear development. Genes Dev 3: 2101-2112.
27. Williams K, Doak TG, Herrick G (1993) Developmental precise excision of Oxytricha trifallax telomere-bearing elements and formation of circles closed by a copy of the flanking target duplication. EMBO J 12: 4593-4601.
28. Klobutcher LA, Herrick G (1997) Developmental genome reorganization in ciliated protozoa: the transposon link. Prog Nucleic Acid Res Mol Biol 56: 1-62.
29. Doak TG, Doerder FP, Jahn CL, Herrick G (1994) 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: 942-946.
30. Witherspoon DJ, Doak TG, Williams KR, Seegmiller A, Seger J, et al. (1997) Selection on the protein-coding genes of the TBE1 family of transposable elements in the ciliates Oxytricha fallax and O. trifallax. Mol Biol Evol 14: 696-706.
31. Gratias A, Betermier M (2001) Developmentally programmed excision of internal DNA sequences in Paramecium aurelia. Biochimie 83: 1009-1022.
32. Klobutcher LA, Herrick G (1995) Consensus inverted terminal repeat sequence of Paramecium IESs: resemblance to termini of Tc1-related and Euplotes Tec transposons. Nucleic Acids Res 23: 2006-2013.
33. Gratias A, Betermier M (2003) Processing of double-strand breaks is involved in the precise excision of paramecium internal eliminated sequences. Mol Cell Biol 23: 7152-7162.
34. van Luenen HG, Colloms SD, Plasterk RH (1994) The mechanism of transposition of Tc3 in C. elegans. Cell 79: 293-301.
35. Saveliev SV, Cox MM (1996) Developmentally programmed DNA deletion in Tetrahymena thermophila by a transposition-like reaction pathway. EMBO J 15: 2858-2869.
36. Fass JN, Joshi NA, Couvillion MT, Bowen J, Gorovsky MA, et al. (2011) Genome-Scale Analysis of Programmed DNA Elimination Sites in Tetrahymena thermophila. G3 (Bethesda) 1: 515-522.

37. Fillingham JS, Thing TA, Vythilingum N, Keuroghlian A, Bruno D, et al. (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.
38. Mitra R, Fain-Thornton J, Craig NL (2008) piggyBac can bypass DNA synthesis during cut and paste transposition. EMBO J 27: 1097-1109.
39. Cheng CY, Vogt A, Mochizuki K, Yao MC (2010) A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. Mol Biol Cell 21: 1753-1762.
40. Kidwell MG, Lisch DR (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55: 1-24.
41. Lin IT, Chao JL, Yao MC (2012) An essential role for the DNA breakage-repair protein Ku80 in programmed DNA rearrangements in Tetrahymena thermophila. Mol Biol Cell 23: 2213-2225.
42. Kapusta A, Matsuda A, Marmignon A, Ku M, Silve A, et al. (2011) Highly precise and developmentally programmed genome assembly in Paramecium requires ligase IV-dependent end joining. PLoS Genet 7: e1002049.
43. Riley JL, Katz LA (2001) Widespread distribution of extensive chromosomal fragmentation in ciliates. Mol Biol Evol 18: 1372-1377.
44. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

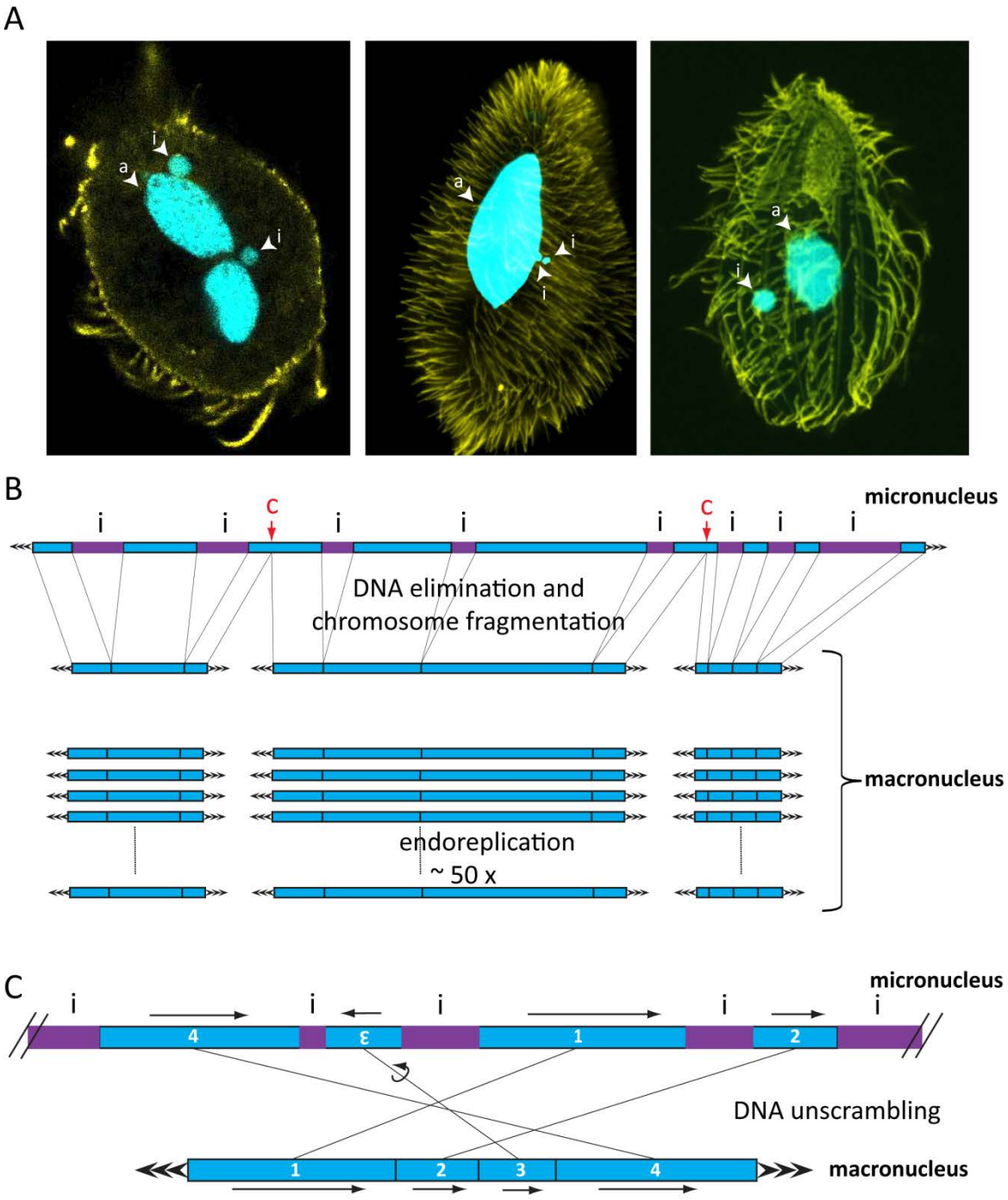
A
i
a
i
a
i
i
a
i
B
micronucleus
i
i
C
i
i
i
C
i
i
i
DNA elimination and
chromosome fragmentation
macronucleus
endoreplication
~ 50 x
C
micronucleus
i
i
i
i
i
4
ε
1
2
DNA unscrambling
1
2
3
4
macronucleus

Figure 1

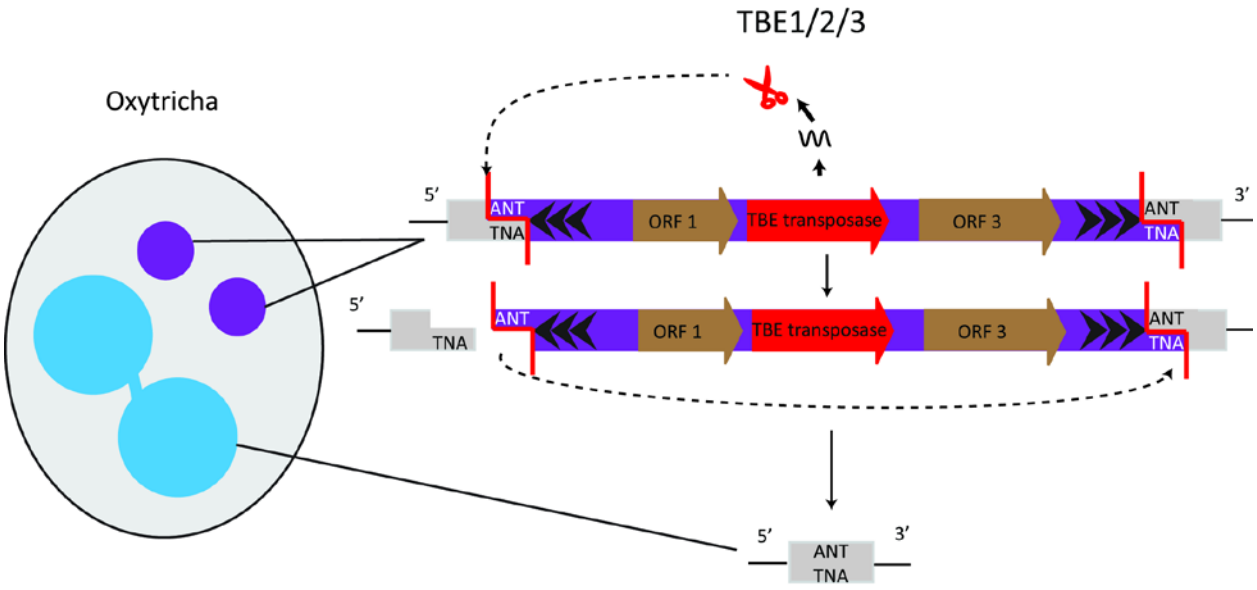

Figure 2

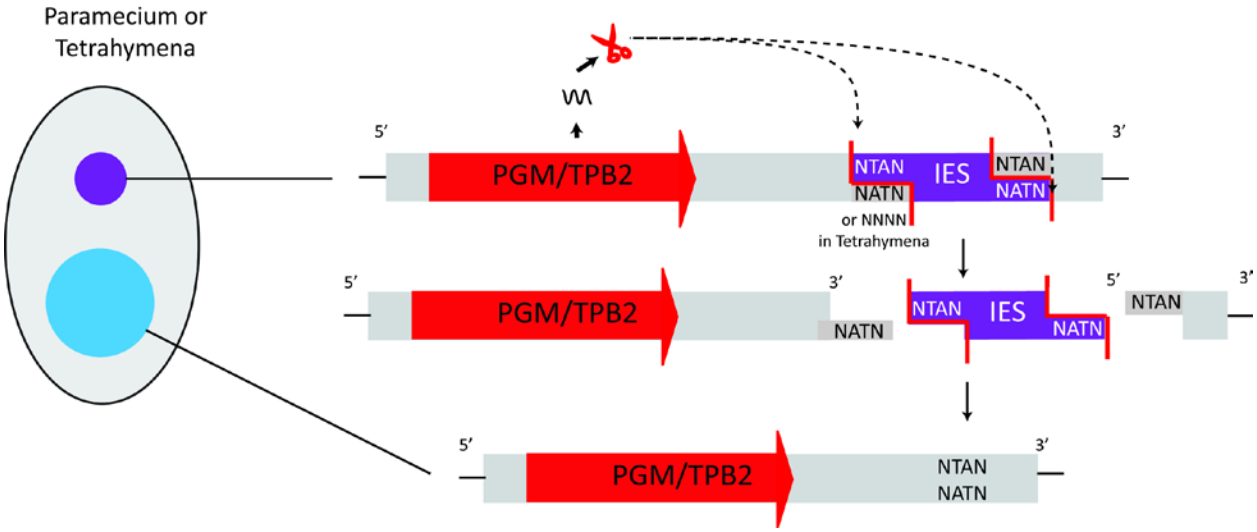

Figure 3

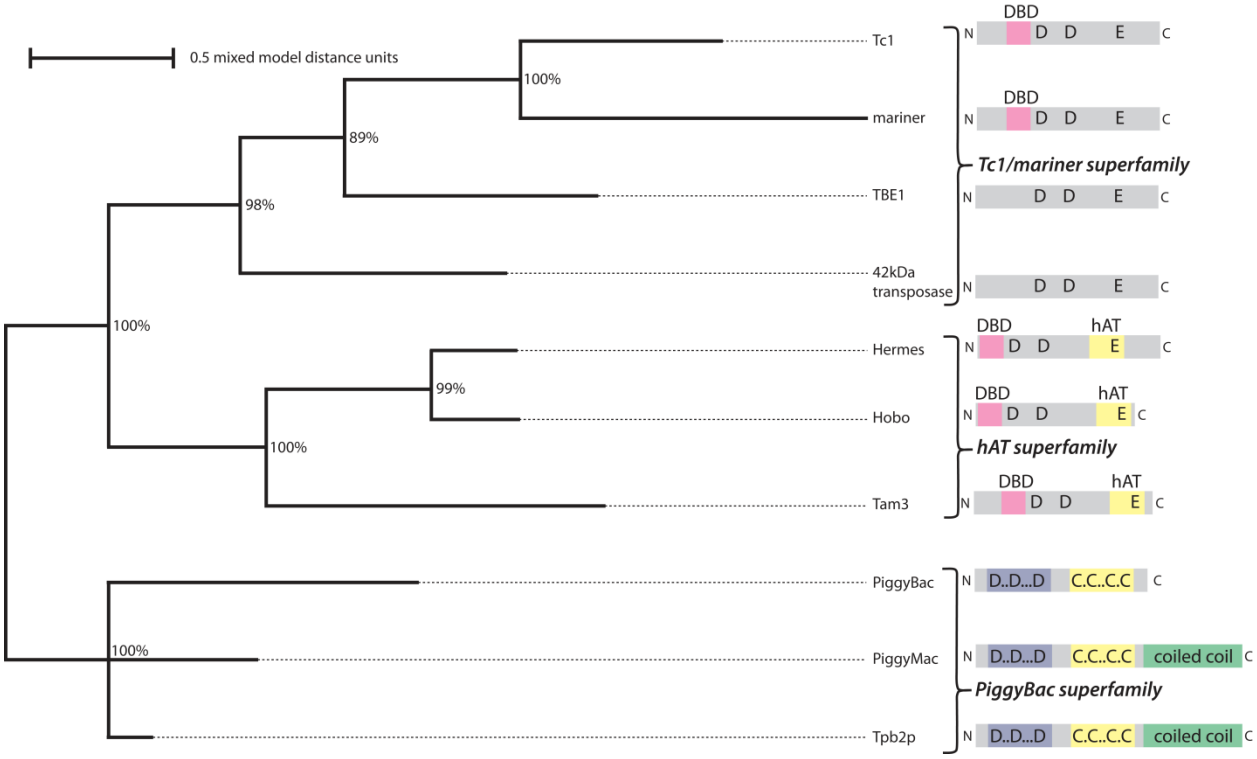
0.5 mixed model distance units
100%
89%
98%
100%
99%
100%
100%
Tc1
mariner
TBE1
42kDa transposase
Hermes
Hobo
Tam3
PiggyBac
PiggyMac
Tpb2p
DBD
D D E
Tc1/mariner superfamily
hAT
hAT superfamily
D..D...D
C.C..C.C
coiled coil
PiggyBac superfamily
N
C

Figure 4

4.3 Local chromatin modifications and a sequence preference of the catalytic activity of the domesticated *piggyBac* transposase Tpb2p enable precise excision of germline limited sequences in Tetrahymena

Vogt A, Mochizuki K.

(manuscript in preparation)

Local chromatin modifications and a sequence preference of the catalytic activity of the domesticated *piggyBac* transposase Tpb2p enable precise excision of germline limited sequences in Tetrahymena

Vogt A[1], Mochizuki K[1]*

[1]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA) Vienna, Austria

*corresponding author

Abstract

In ciliated protists the genome of the new developing macronucleus is massively rearranged. These programmed genome rearrangements involve the elimination of defined sequences called internal eliminated sequences. These sequences account for ~30 % of the genome. The process of DNA elimination involves the recognition of the eliminated sequences by an Argonaute – scan RNA complex which initiates the sequence specific heterochromatin formation. Characteristic heterochromatin marks are the tri-methylated lysines 9 and 27 of histone H3. After the chromodomain protein Pdd1p bound to these modifications distinct heterochromatin foci form in which DNA elimination is believed to occur.

We recently reported that the recruitment of the domesticated *piggyBac* like transposase Tpb2p is essential for DNA elimination and formation of heterochromatin foci in vivo. Furthermore recombinant Tpb2p recognizes and cleaves different boundaries from internal eliminated sequences in vitro. Yet it is still unclear if the catalytic activity of Tpb2p is essential for DNA elimination or if it is its role in heterochromatin dynamics that enables removal of internal eliminated sequences. Here we report that in deed catalytic activity of Tpb2p is necessary for DNA elimination. In contrast, the heterochromatin dynamics are linked to the zinc finger domain. This domain also binds histone H3 peptides that are tri-methylated at lysines 9 and 27 in vitro. We suggest a model in which Tpb2p is recruited to heterochromatin via its zinc finger and the sequence preference enables a more precise DNA elimination.

Introduction

Transposons represent harmful genetic elements because they rearrange host's genome and their integration into important coding or regulatory regions can cause deleterious effects on the host. Transposons are therefore considered to be "junk" DNAs [1] and hosts have invented many different ways to counteract these selfish elements [2, 3]. For example, different eukaryotes from yeast to mammals silence transposons by RNAi-related pathways [4]. On the other hand, transposons may just not be junk because transposons potentially contribute to the evolution of the host by genome rearrangements, by alternating gene expression networks, or by providing new genes from transposons to the hosts. One such transposon-driven evolution is V(D)J recombination in vertebrate immune system which is mediated by the domesticated Rag1 transposase [5]. Therefore, harmfulness and usefulness of transposons are two sides of a coin and all existing organisms must have evolved by balancing the two faces of transposons. An evolutional product likely created by such a balance is DNA elimination of the ciliated protist Tetrahymena in which transposon-related sequences are eliminated by a domesticated *piggyBac* transposase [6].

Most ciliates show nuclear dimorphism [7]. The small micronucleus (Mic) represents the germline and is transcriptionally inert during vegetative growth whereas the larger macronucleus (Mac) provides the cell with most if not all RNA and can be considered as the soma. When nutrients are scarce Tetrahymena does sexual reproduction (conjugation) where two mating partners form a pair and the Mic of each cell undergoes meiosis (Fig. 1). Three of the meiotic products are degraded and the remaining one divides mitotically. One of the products is exchanged with the mating partner and afterwards the two pronuclei fuse to form the zygote. The zygotic nucleus divides twice mitotically and out of the four mitotic products two become new MICs and the other two develop to become new Macs. The parental Mac is degraded during this process. Eventually one of the Mics is degraded and the other one undergoes mitosis. In parallel the parental cells undergo cytokinesis and one Mic and one Mac are distributed to each daughter cell.

Programmed genome rearrangements take place in the developing new Mac. Two major types of genome rearrangements occur in Tetrahymena. The first type is chromosome breakage that leads to the fragmentation of germline chromosomes. The chromosome breakages occur at specific DNA sequences called chromosome breakage sequences (Cbs), which exist ~250 copies in the Mic genome and include a conserved 15 nt sequence, 5'-TAAACCAACCTCTTT-3' [8, 9]. The sites of chromosome breakages are healed by de novo telomere formations. The second type of genome rearrangements in Tetrahymena is DNA elimination of internal eliminated sequences (IESs) followed by ligation of their flanking sequences. It has been estimated that there are over 9,000 different IESs, which represents around 30 % of the Mic genome [10]. Because many IESs contain transposon-related sequences, it is assumed that the DNA elimination is a process removing potentially harmful transposon-derived sequences from the transcriptionally active somatic genome. Although different IESs do not share any detectable common sequences in themselves and in their flanking regions, invariable sets of IESs are always eliminated from the Mac and the majority of their boundary occur within a few to several base pairs. This raises a question about how the identities and the boundaries of IESs are precisely determined.

An RNAi-directed heterochromatin formation mechanism targets IESs for DNA elimination in Tetrahymena like transposons are targeted for gene silencing by a similar mechanism in other eukaryotes. The ~28-29 nt siRNAs produced from the Mic genome are selected for IES-specificity by the way which the siRNAs complementary to the parental Mac genome are degraded. The selected IES-specific siRNAs eventually induce the establishment of heterochromatin specifically on IESs in the developing new Mac [11]. This heterochromatin comprises tri-methylated histone H3 at lysine 9 and lysine 27 (K9me3, K27me3) [12-14] and the chromodomain protein Pdd1p, which binds to these histone H3 modifications. The heterochromatinized IESs are then assembled into heterochromatin bodies located at the nuclear periphery [15]. Finally IESs are excised as a linier piece of DNA. There are indications that the identities of IESs are determined epigenetically by an RNAi-directed heterochromatin

formation mechanism. However, it is difficult to imagine that the same mechanism precisely determines the boundaries of IESs.

We previously reported that the domesticated *piggyBac* transposase-like protein Tpb2p is essential for DNA elimination [6]. Therefore, DNA elimination in Tetrahymena is a process in which transposon-related sequences are eliminated by a domesticated transposase. Furthermore, we reported that Tpb2p has ability to produce DNA double-strand break possessing 4-base 5' overhang at a boundary sequence of an IES in vitro. Because Tpb2p localizes to the heterochromatin bodies, an attractive hypothesis is that Tpb2p is recruited to the heterochromatin by directly interacting to a heterochromatin component and then inducing a DNA double-strand break at its preferential DNA sequence near by the heterochromatin. In this study, we analyzed the roles of individual domains of Tpb2p genetically and biochemically to understand how Tpb2p contributes to the precise DNA elimination in Tetrahymena.

Results

A conditional *TPB2* KO strain as a tool to study function of Tpb2p domains

It has been demonstrated that recombinantly-expressed Tpb2p has the ability to produce DNA double-strand break in vitro and that *TPB2* is essential for DNA elimination in vivo [6]. However, it is not clear whether the DNA double-strand break forming endonuclease activity of Tpb2p is necessary for DNA elimination because Tpb2p is also necessary for the formation of the heterochromatin bodies, where DNA elimination is believed to take place [6]. To test whether the endonuclease activity of Tpb2p is necessary for DNA elimination, we intended to establish a genetic system with which we can investigate different *TPB2* mutants. In the previous report, RNAi-knockdown was used to investigate the role of Tpb2p [6]. However, this strategy is very difficult to adopt genetic rescue assay using *TPB2* mutants because the current RNAi method requires to express hairpin RNA containing ~90 nt sequence complementary to a target mRNA [16], which makes difficult to selectively knockdown the endogenous TPB2 without targeting *TPB2* genes for rescue, and RNAi knockdown only partially down regulate TPB2 expression. Therefore we tried to obtain knockout (KO) strains *TPB2* but failed. This was probably because *TPB2* is a haplo-insufficient gene and heterozygous *TPB2* KO strain was not viable. To overcome this problem, we created a conditional *TPB2* knockout (cKO) strain.

To make *TPB2* expression conditional, we produced a *TPB2* cKO construct in which the endogenous *TPB2* promoter was replaced with the cadmium inducible MTT1 promoter (Fig. 1A). *TPB2* is expressed both from the parental and new Mac. Therefore, to eliminate *TPB2* expression, we need to KO the *TPB2* copies in the germline genome. *TPB2* cKO construct was introduced into the Mic genome by homologous recombination and then two heterozygous *TPB2* cKO strains were mated to obtain homozygous *TPB2* cKO strains. In these genetic crosses, *TPB2* expression was continuously induced during conjugation to obtain viable progeny. Tpb2p was undetectable in the absence of cadmium during conjugation of two homozygous *TPB2* cKO

strains (hereafter just called *TPB2* cKO strains), while Tpb2p expression was induced by addition of cadmium in these strains. Therefore, Tpb2p expression can be conditionally knocked out in the *TPB2* cKO strains.

It has been reported that the formation of heterochromatin bodies and DNA elimination were inhibited by the RNAi knockdown of *TPB2* [6]. Formation of heterochromatin bodies in the *TPB2* cKO strains was observed by localization of the heterochromatin component Pdd1p. In the absence of *TPB2* induction by cadmium, mature heterochromatin bodies were not detected but Pdd1p stained heterochromatin stayed as dispersed small foci in the new Mac (Fig. 2C, "not induced"). In contrast, in the presence of cadmium, large, matured heterochromatin bodies were formed in the new Macs (Fig. 2C, "induced"). Next, DNA elimination in the *TPB2* cKO strains was observed by DNA FISH against the moderately repeated (~30 copies per Mic genome) Tlr1 IESs. DNA elimination in the wild-type cells is normally completed at ~16 hr post-mixing [17]. Tlr1 IESs remained in the new Mac even at 36 hr post mixing when *TPB2* expression was not induced (Fig. 2C, "not induced"). In contrast, in the presence of cadmium, Tlr1 IESs were removed completely from the new MACs at 36 hr post mixing (Fig. 2C, "induced"). Thus the *TPB2* cKO strains recapitulate the phenotypic defects described for the previously reported TPB2 RNAi strains.

Using the *TPB2* cKO strains, we next intended to establish a phenotypic rescue system to assay functionalities of Tpb2p mutants. The non-essential *MTT1* locus of the parental Mac in the *TPB2* cKO strains was replaced with a rescue construct expressing a *TPB2* mutant under the control of the cupper inducible *MTT2* promoter (Fig. 3A, B). In this way, in the presence of cupper (but in the absence of cadmium), a *TPB2* mutant can be expressed without expressing the wild-type *TPB2* from the *TPB2* cKO locus (Fig. 3B). When the rescue construct containing the wild-type *TPB2* were introduced in the *TPB2* cKO strains, heterochromatin body maturation (Fig. 3C, WT) and DNA elimination (Fig. 3D, WT) were restored in the presence of cupper. This result

demonstrates that the phenotypic rescue system using the *TPB2* cKO strains and the rescue construct can be used to assay functionalities of Tpb2p mutants.

The endonuclease activity of Tpb2p is essential for DNA elimination and but is dispensable for heterochromatin body maturation

Tpb2p has the N-terminal endonuclease catalytic domain containing three aspartic acids forming the DDD catalytic triad (Fig 3A). We previously reported that the replacement of these three aspartic acids to leucines compromises the endonuclease activity of Tpb2p in vitro [6]. To understand the role of the endonuclease activity in vivo, a *TPB2* rescue construct in which the DDD catalytic triad was replaced with leucines (D297L; D379L; D495L, Catalytic-dead TPB2, "CD" in Fig. 2) was introduced into the TPB2 cKO strains. Expression of Catalytic-dead TPB2 did restore heterochromatin body maturation (Fig. 2C, CD) but did not support the elimination of Tlr1 IESs from the new Macs (Fig. 2D, CD). Therefore, endonuclease activity of Tpb2p is necessary for DNA elimination but not for the heterochromatin body formation. As far as we know this is the first demonstration that the heterochromatin body is not a product of DNA elimination but can be formed prior to the initiation of DNA elimination.

The zinc-finger domain of Tpb2p is essential for heterochromatin body maturation

Besides the N-terminal endonuclease catalytic domain, Tpb2p has the zinc-finger domain (Fig. 3A) that is conserved among piggyBac transposases. The role of this zinc-finger domain is unknown. Therefore, to understand the role of the zinc-finger domain, a *TPB2* rescue construct in which two of the four cysteines consisting of the metal-binding residues of the zinc-finger were replaced with alanines (C618A; C629A, Zinc-finger mutant TPB2, "ZM" in Fig. 2) was introduced into the TPB2 cKO strains. Expression of Zinc-finger mutant *TPB2* did not restore heterochromatin body maturation (Fig. 2C, ZM) and the elimination of Tlr1 IESs from the new MACs (Fig. 2D, ZM). Therefore, we conclude that the zinc-finger domain of Tpb2p is essential for heterochromatin body maturation. This result raises a possibility that Tpb2p is necessary to

interact with heterochromatin through its zinc-finger domain and this interaction is required for the heterochromatin body maturation.

Tpb2p zinc-finger domain binds histone H3 peptides with tri-methylated lysine 9 or 27 in vitro

To investigate the possible interaction of Tpb2p with chromatin we used the MODified histone peptide array (Actif motif) to test if Tpb2p interacts with specific histone modifications. On this array 384 unique combinations of histone modifications on different histone peptides from H3, H4, H2A and H2B are spotted. The full-length recombinant Tpb2p fused to the MBP (maltose binding protein)-tag was incubated with this array and peptide-bound Tpb2p was visualized by immunostaining using an anti-MBP antibody. We found that the full length Tpb2p preferentially binds histone H3 peptides. Among the 20 peptides to which >1.5 times more Tpb2p were bound than the control, 18 were peptide sequences from the N-terminal tail of histone H3 (see supplementary table T1). Importantly, 7 of them contained tri-methylated lysine 9 (H3K9me3) or tri-methylated lysine 27 (H3K27me3), indicating that Tpb2p may bind to histone H3 through H3K9/27me3 modifications.

We confirmed the interaction between Tpb2p and H3K9/27me3 by a peptide pull-down assay. To study H3K9me3 interaction a peptide representing the amino acids 1-19 of histone H3 with tri-methylation at the K9 position was used. To study H3K27me3 interaction with a peptide corresponding to the amino acids 16-35 of histone H3 with tri-methylation at the K27 position was used. For each experiment, unmodified and unmodified-scrambled peptides were used as controls. The peptides were immobilized on beads, incubated with recombinantly produced zinc-finger domain of Tpb2p fused to MBP-tag (MBP-Tpb2pZF) and peptide bound proteins were analyzed by western blot using anti-MBP antibody. Three to four times more MBP-Tpb2pZF was bound to the peptides having tri-methylated K9 or K27 (Fig. 3, H3 1-19 K9me3 and H3 16-35 K27me3) than to the unmodified-scrambled peptides (Fig. 3, H3 1-19 scrambled and H3 16-35 scrambled). We also found that slightly, but significantly, more interaction of MBP-

Tpb2pZF to the unmodified peptides (Fig. 3, H3 1-19 and H3 16-35) than to the unmodified-scrambled peptides. Two point mutations in the metal-bonding cysteins (C618A, C629A) completely abolished the interaction of MBP-Tpb2pZF to any of the peptides (Fig. 4 bottom). We conclude that the zinc-finger domain of Tpb2p directly interacts to the N-terminal tail of histone H3 and this interaction is highly enhanced by the presence of the tri-methylated K9 or K27. Because H3K9me3 and H3K27me3 were reported to accumulate specifically on IESs (ref), Tpb2p is likely recruited to IESs by the interaction between zinc-finger domain and H3K9/27me3.

Endonuclease activity of Tpb2p requires the sequence integrity of the R-IES boundary

The interaction of Tpb2p with the heterochromatin by Tpb2p-H3K9/K27me3 interaction may explain how Tpb2p is recruited to IESs. However chromatin structure is unlikely to determine the nucleotide precision of DNA elimination. The recruitment of Tpb2p to IESs by interaction with the heterochromatin may followed by the introduction of a double strand break whose position is determined by intrinsic sequence preference of the endonuclease activity of Tpb2p. We therefore decided to investigate the endonuclease activity of Tpb2p in more detail. Using an in vitro nuclease assay we have previously shown that Tpb2p can recognize and cleave the left boundary sequence of Tetrahymena R-IES (5'-AGTGAT-3') when present in an artificial oligo DNA environment [6]. To better understand how this sequence is cut preferentially by Tpb2p we performed in vitro cleavage assays using mutated versions of the left R-IES boundary. We prepared a set of synthetic 100-bp DNA duplex substrates containing the wild-type boundary sequence (5'-AGTGAT-3') (Fig. 5A) or mutant version of this boundary sequence in which every position of the AGTGAT sequence was individually replaced to every other possible nucleotide. When the substrate containing the wild-type boundary sequence (5'-AGTGAT-3') was incubated with MBP-fused Tpb2p that was recombinantly expressed in *E. coli*, the major radio labeled product at 50 nt was detected by a denaturing gel-electrophoresis (Fig. 5 most left lane) as observed previously. Next, each mutated substrate was incubated with MBP-fused Tpb2p. We found that most of the mutations of the AGTGAT sequence lead to a less efficient cleavage at

the expected site (Fig. 5). Especially, when positions +2 or +3 (the second and the third nucleotides downstream the expected cleavage site) are mutated the major band at 50 nt disappeared or became less intense and the cleavage site was shifted a few bases to the 3'. These results indicate that, at least in the model substrate we used, the integrity of the IES boundary sequence plays a role in the efficient catalysis of Tpb2p and the second and the third nucleotides downstream the cleavage site are crucial for Tpb2p to induce a double-strand break at the precise position.

The in vitro sequence preference of Tpb2p on substrate correlates with the choice of IES boundary in vivo

Next, to ask if the in vitro sequence preference of Tpb2p on its substrate DNA we observed above is involved in the reproducible occurrence of the IES boundaries in vivo, we used an episomal vector-based in vivo DNA elimination assay that was established previously [18, 19]. When the ribosomal DNA (rDNA) based episomal vector pD5H8 is introduced into the developing Mac during conjugation, rDNA is excised out from the vector, duplicated, the two copies of rDNAs are ligated in an inverted order, telomeres are added at the ends de novo, and is replicated independently. Virtually any piece of DNA can be inserted into near the end of rDNA without disturbing the function of rDNA and an IES inserted into this vector is removed like the corresponding endogenous IES.

To assess the importance of the boundary sequence of R-IES, the Mic genomic sequence including R-IES and 205 bp flanking regions upstream and 163 bp flanking regions downstream were cloned into pD5H8 and transformed into the new Mac of Tetrahymena by electroporation (Fig. 7A). Because the in vitro assay above indicated that the second and the third nucleotides downstream of the cleavage site are crucial for Tpb2p to induce a cleavage at the precise position, we produced three different constructs: the first had the wild-type boundaries (WT); the second had the T to G mutation at the +2 position of the left boundary (T2G); and the third

had the G to T mutation at the +3 position of the left boundary (G3T) (Fig. 7A), and compared the elimination of R-IESs from these constructs.

Twenty-four transformants for each construct were pooled for genomic DNA extraction and elimination of R-IES on the vector-derived rDNA was examined by PCR (Fig 7B). The PCR detects the presence or absence of the elimination of R-IES on the rDNA but not those on the endogenous Mac chromosome. From the construct having the wild-type boundary sequence, two major PCR products were detected (Fig. 7B, WT). A longer ~2.4 kb product corresponded to the R-IES region on the rDNA without DNA elimination and a shorter ~1.3 kb product to DNA in which the whole R-IES was eliminated. In addition, a minor ~1.8 kb product corresponding to DNA with a partial R-IES elimination (Fig. 7C filled arrowhead) was detected. A similar band was seen in previous studies using this system and assigned to an aberrant rDNA processing inheritant to the system [18, 20].To analyze the exact positions of boundaries produced by the DNA elimination, the ~1.3 kb products (= whole R-IES eliminated) were extracted from gel, cloned into a vector, and 20 clones each from different products were sequenced. In the all clones, their boundaries occurred at the same nucleotide position where the endogenous R-IES boundary was reported to occur. Therefore, as reported previously, the elimination of R-IES was precisely recapitulated in the rDNA vector based system, although the efficiency of DNA elimination was not 100 %.

Similar assays were performed with the constructs having T2G or G3T mutations. Like from the construct having the wild-type boundary sequence, two major PCR products: the ~2.4 kb band having no R-IES elimination and the ~1.3 kb band corresponding to the whole R-IES elimination product (Fig. 7B, T2G, G3T). In addition, the ~1.8 kb product became prominent in the presence of T2G mutation (Fig. 7C, T2G filled arrowhead) and another minor ~1.5 kb product was detected in the presence of G3T mutation (Fig. 7B, G3T, open arrowhead). Moreover, although the ~1.3 kb products from T2G and G3T constructs were indistinguishable to that from the wild-type (WT) construct, sequencing of the ~1.3 kb products showed that their boundaries were

diversified and shifted a few to several nucleotides downstream compared to the precise boundary in the wild-type (Fig. 7D, T2G, G3T). This pattern of shift observed in the mutated constructs in vivo was very similar to that what was observed by the same mutations in vitro (Fig. 6), indicating that Tpb2p probably directly recognizes these sequences to determine the site of DNA double-strand break for IES elimination in vivo. Altogether, these results indicate that the second and the third nucleotides after the cleavage site from the left wild-type boundary of the R-IES are critical to be recognized by Tpb2p and thus play a vital role in the determination of boundaries during IES elimination.

Interestingly, in the both mutation constructs (T2G and G3T), not only the left boundaries but also the right boundaries were frequently sifted (Fig. 7D, T2G, G3T). This dependency of the precise cleavage of the right boundary on the integrity of left boundary sequence suggests that there is a crosstalk between both IES ends. It is known that the *piggyBac* transposon, the element from which Tpb2p was predicted to be domesticated, assembles a DNA loop with a transpososome complex consisting of both ends of a transposon with one transposase at the each end when it jumps out from the host genome [21]. Furthermore a crosstalk between IES ends was also proposed for the distantly related ciliate *Paramecium* [22]. Therefore, Tpb2p may locate at both ends of an IES, form a dimer, and co-ordinate DNA-double strand break and the following DNA repair processes. We speculate that the underlying DNA sequences of IES boundaries play a role in this ends-dimer formation.

The endonuclease activity of Tpb2p does not require the zinc-finger domain in vitro

To test whether this sequence bias depends on the zinc finger or the C-terminal extension or if it is exclusively mediated by the catalytic domain we used vitro nuclease assays in which we incubated synthetic 100-bp DNA duplex containing the left boundary sequence of Tetrahymena R-IES (5'-AGTGAT-3') at the middle (Fig. 5A) with the recombinantly expressed full length Tpb2p (Fig. 7A, FL), N-terminal half of Tpb2p, which contains the endonuclease domain, (Fig. 7A, NT),

or C-terminal half of Tpb2p, which contains the zinc finger domain (Fig. 7A, CT). If double-strand break occurs as during the R-IES elimination in vivo, 50-nt radio labeled product should be detected by a denaturing gel-electrophoresis (Fig. 5A). Indeed, the full length Tpb2p produced the expected cleavage product (Fig. 7B, FL). The N-terminal half of Tpb2p also produced 50-nt radio labeled product with a comparable specificity to the full length Tpb2p (Fig. 7B, NT). In contrast, the C-terminal half of Tpb2p did not produce specific product (Fig. 7B, CT). These results indicate that N-terminal half, and thus probably the endonuclease domain alone, is enough to create a DNA double-strand break at the expected position and the zinc-finger domain contributes to neither the catalysis nor the substrate choice of the endonuclease in vitro.

Discussion

In this study we analyzed the roles of two conserved domains of the domesticated *piggyBac* transposase Tpb2p during DNA elimination in Tetrahymena genetically and biochemically. The first domain was the zinc-finger domain, which is necessary for the formation of the heterochromatin bodies (Fig. 3C, ZM) and for DNA elimination (Fig 3D, ZM). The interaction of Tpb2p with the chromatin is probably mediated by the direct binding of the zinc-finger domain to the heterochromatin-specific histone modifications H3K9me3 and H3K27me3, as the interaction between the zinc-finger domain of Tpb2p and H3 peptides were enhanced by these modifications in vitro (Fig. 4). The second domain we investigated was the endonuclease domain. We found that the endonuclease activity of Tpb2p is necessary for DNA elimination (Fig. 3D, CD) but not for the formation of the heterochromatin bodies in vivo (Fig. 3C, CD). This is the first demonstration that the heterochromatin bodies can be formed independently to DNA elimination. Furthermore we showed that the boundary sequence of an IES is a crucial determinant for the catalysis of Tpb2p in vitro (Fig. 6) and that the sequence preference of the Tpb2p endonuclease influences the choice of IES boundaries in vivo (Fig. 7). Because the N-terminal half of Tpb2p, which lacked the zinc-finger domain, showed similar sequence preference on substrate DNA as the full length Tpb2p, this preference is probably an intrinsic property of the endonuclease domain (Fig. 5). All together, we conclude that Tpb2p has two distinct functions in DNA elimination: the formation of the heterochromatin bodies through the interaction between heterochromatin and the zinc-finger domain; and the sequence-biased formation of DNA double-strand breaks mediated by the endonuclease domain.

Heterochromatin interaction and the sequence-biased catalysis of Tpb2p probably mediate the reproducible DNA elimination

The way in which Tpb2p interacts with the heterochromatin may provide us some hints about how this protein mediates the reproducible pattern of DNA elimination. In this study, we

showed that the zinc-finger domain of Tpb2p directly interacts with the heterochromatin specific histone modifications H3K9me3 and H3K27me3 in vitro. It is known that these modifications specifically occur on IESs (ref). Therefore Tpb2p is recruited to the IESs by its heterochromatin interaction and this recruitment may limit the occurrence of Tpb2p-endonuclease cleavage at or nearby IESs. Because an IES is removed as one piece of DNA, either in linear or circular form (ref), the endonucleolytic-cleavages of Tpb2p must be restricted not just on the heterochromatin but should only occur at the ends of heterochromatin. So how is Tpb2p action limited to the heterochromatin ends? The methylated histones H3K9me3 and H3K27me3 are also bound by one of the most abundant heterochromatin component Pdd1p [15]. Therefore, Tpb2p competes for chromatin-binding sites with Pdd1p. This competition may exclude Tpb2p from the body of heterochromatin and allows Tpb2p to bind edges of heterochromatin regions, where DNA double-strand breaks must be induced for DNA elimination. Alternatively, Tpb2p may localize throughout the heterochromatin segment while the heterochromatin structure or some heterochromatin proteins may inhibit the action of Tpb2p endonuclease at the body of heterochromatin. Future research should clarify the spatial as well as the temporal localization of Tpb2p and other heterochromatin components on chromatin by ChIP-seq experiments.

The heterochromatin localization of Tpb2p does not seem to be enough to explain the reproducible occurrence of the border of IESs because histone modifications are expected to be able to determine a chromatin segment only at the level of a size of nucleosome (~200 nt) while most of the boundaries of IESs occur within a few to several nucleotides. As we demonstrated in this study, the Tpb2p endonuclease has a sequence preference on its substrate. The combination of the heterochromatin localization and the sequence-biased action probably allow Tpb2p to precisely determine the boundaries of IESs at a sub-nucleosomal level. The intrinsic ability of Tpb2p-endonuclease to recognize some preferential DNA sequence at a sub-nucleosomal level is especially important as some IESs in Tetrahymena have size similar to a single nucleosome [23]. Although these small IESs could still be heterochromatinized, it seems

unlikely that the small IESs are specified by this small number of heterochromatin marks fitting on a single nucleosome. A possibility is that the endonucleolytic preference of Tpb2p alone is sufficient to determine the precise boundary of these small IESs. Alternatively, there might be some "protective" factor, that protect non-IES regions from Tpb2p. Interestingly most short IESs described to date are interrupting coding regions of genes [23]. The elimination of these elements can be nicely explained if some euchromatic marks acts as a protective factor against Tpb2p.

Possible roles of cis-regulatory element in Tpb2p-mediated DNA elimination

Previous studies described the significance of cis-regulatory elements adjacent to IESs. Also for the R-IES that we used for our experiments sequences on the left and right side were described to be indispensible [18]. In this study, we showed that the positions +2 and +3 of the left R-IES cleavage site are crucial for its precise digestion in vitro and in vivo. This is the first study describing that cis-regulatory sequences are not sufficient to promote precise excision of an IES and the integrity of the cleave site in the IES boundary plays a pivotal role in DNA elimination in Tetrahymena. The cis-regulatory sequence may contribute to create a structural environment which makes a boundary sequence recognizable by Tpb2p for IES excision. The bending of DNA could be such a structural feature because *piggyBac* insertions in Drosophila correlate with the bendability of target DNA [24]. Alternatively, the cis-regulatory sequence may be necessary for creating a nucleosome-free region where DNA is accessible for Tpb2p.

Evolutionary considerations

All organisms have to deal with genome invasion of transposons. Tetrahymena found a perfect way to cope with these mobile elements by removing them from the somatic genome. As long as they reside in the germline and are precisely removed from the soma they do not reduce the fitness of the host cell. We can detect many different types of transposon-related sequence in the Tetrahymena Mic genome. Moreover, because they are locked in the dormant genome,

many of their sequences are probably changing rapidly. A system that can eliminate such a variety of different invaders must not rely on a primary sequence to determine eliminated DNA but can use some alternative way to specify target sequences. In this study, we showed that the endonuclease activity of Tpb2p has broader target specificity than most transposases have. Although Tpb2p can efficiently induce DNA double-strand break at the TTAA sequence, which *piggyBac* transposases need as their substrate, Tpb2p is able to cut other sequences like the left R-IES boundary (Fig. 5). Yet, Tpb2p endonuclease has a sequence bias (Fig. 6) that is important for precision of DNA elimination (Fig. 7).

We believe this relaxed sequence specificity for Tpb2p endonuclease is complemented by the heterochromatin-binding ability of Tpb2p and IES-limited heterochromatin formation by a small RNA-directed mechanism. An epigenetic mechanism, in which a whole-genome comparison of the Mic and the Mac genome is involved, produces IES complementary small RNAs and these small RNAs induce accumulation of the heterochromatin-specific histone modifications H3K9me3 and H3K27me3 selectively on IESs (ref). We showed in this study that Tpb2p directly interacts with these heterochromatin-specific modifications through its zinc-finger domain (Fig. 4) and this domain is essential for DNA elimination (Fig. 3). Altogether, the small RNA-directed heterochromatin formation on IESs and the recruitment of Tpb2p to heterochromatin allow Tpb2p to localize on IESs independently of their primary sequences, and then sequence-biased action of Tpb2p endonuclease determines IES boundary to the limited sites. In this context, the biochemical features of Tpb2p we found in this study have been probably evolved to deal with many unrelated transposons by a single mechanism.

Figure legends

Figure 1: **Life cycle of Tetrahymena**. Under starvation conditions two Tetrahymena cells do sexual reproduction (conjugation) where two cells form a mating pair and the MIC undergoes meiosis. From the four meiotic products three are degraded and the remaining one divides mitotically. One of the two pronuclei is exchanged with the mating partner and fusion leads to formation of the zygotic nucleus. This divides twice by mitosis and two nuclei become MICs, the other two MACs. After cell division the nuclei are distributed to the daughter cells.

Figure 2: **Creation of a conditional TPB2 KO.** (A) Conditional KO construct (*MTT1-TPB2*) to replace endogenous TPB2 5' region by inducible MTT1 promoter. X indicates region where homologous recombination should occur. (B) Germline transformation of *MTT1-TPB2* resulted in one heterozygote clone which was crossed to WT to obtain different heterozygous clones. Transformants were phenotypically assorted until they lost macronuclear copies of *MTT1-TPB2* and afterwards crossed to obtain homozygous KO strains. (C) Western Blot analysis of induced cKO strains in which *TPB2* expression from *MTT1* promoter expression was induced with 0.1 ug/ml Cd^{2+} added 7 + 8 h after mixing mating partners. Tpb2p band is indicated on the right side. (D) Immunofluorescence staining of induced and non induced strains to observe heterochromatin body formation. Pdd1p serves as a marker for heterochromatin bodies. (Abbreviations: i = micronucleus, a = parental macronucleus, na = new macronucleus). (D) DNA FISH against the Tlr1 IES to assess the ability of conditional KO strains to do DNA elimination.

Figure 3: **Rescue system to investigate *TPB2* mutants in vivo.** (A) Rescue construct with flanking regions for homologous recombination (HR) in *MTT1* locus. X indicates region of HR. *TPB2* is under the control of the Cu^{2+}-inducible *MTT2* promoter. Three different constructs encoding for WT *TPB2*, catalytic dead *TPB2* (CD) or a zinc finger mutant (ZM) of *TPB2* were used. (B) Scheme of transformation into macronucleus of conditional KO strains. Genotypes are

written below the cells. The original allele is written first, the allele that replaced the original one is written after the :: (C) Immunofluorescence staining of induced rescue strains (+ 30 uM $CuSO_4$ added 7 + 8 h after mixing mating partners) to observe heterochromatin body formation. Pdd1p serves as a marker for heterochromatin bodies. (Abbreviations: i = micronucleus, a = parental macronucleus, na = new macronucleus). (D) DNA FISH against the Tlr1 IES to investigate DNA elimination phenotype.

Figure 4: **In vitro histone peptide binding assay.** Indicated biotinylated histone peptides were bound to streptavidin beads and incubated with recombinantly expressed zinc finger domain of Tpb2p or a mutated version (MBP-Tpb2pZF or MBP-Tpb2pZF C618A, C629A). The WB shows that only the non mutated zinc finger can bind to histone H3 peptides. The highest binding affinity is observed for peptides carrying tri-methylated lysines 9 or 27 (lanes 4 and 7). Quantification of the band intensity was done using IMAGEquant software. For statistical analysis students T-test was used (two tailed, unequal variances).

Figure 5: **In vitro nuclease assay using mutated "AGTGAT" substrate.** (A) 100 nt long artificially designed and synthesized substrate DNA having the 5'AGTGAT 3' sequences in the middle which corresponds to the predicted left boundary of the R-IES. (B) Substrate DNA oligos containing indicated nucleotide replacements in the left R-IES boundary were incubated with FL Tpb2p. Marker position is shown on the left. Expected cleavage position in the non mutated oligo is indicated by the arrowhead on the right side.

Figure 6: **In vitro nuclease assay with recombinant Tpb2p.** Top: Scheme of recombinant Tpb2p. Full length protein (FL), the N terminal part containing catalytic triade (NT) and the C terminal part including predicted zinc finger domain (CT) were used for nuclease assay. Bottom: Substrate shown in Fig. 5A was incubated with FL, NT or CT. The products were separated on a

denaturing Polyacrylamidgel (15 %). Marker position is shown on the left. Expected cleavage product size is indicated by arrowhead on the right side.

Figure 7: **In vivo elimination assay using mutated R-IES boundary.** (A) The rDNA vector containing the R-IES and important flanking regions was transformed into the new MAC of Tetrahymena. The left boundary was either WT or mutated at position 3 or 4 (T3G, G4T). Positive clones were selected and 24 clones each were pooled. PCR analysis with indicated primers (R205Lfw/pD5H8fw) was carried out to test for elimination of R-IES. (B) Agarose gel showing PCR result. WT, T3G and G4T can eliminate full length R-IES (full elimination, *1-3). Longer bands appear which are stronger when boundary is mutated (partial elimination, open and filled arrowheads). All bands were gel extracted, cloned and sequenced. (C) PCR product sequences were used to assign cleavage site. Lines indicate precise assignment of the cleavage site; open boxes could not be mapped precisely due to sequence redundancies before and after cleavage site.

Materials and Methods

Strains and culture conditions

Wild type strains B2086 or Cu428 were provided by Peter J. Bruns (Cornell University, Ithaca, NY). Tetrahymena cells were cultured in SPP medium with 2 % proteose peptone at 30 ^{o}C.

Mating was induced by mixing equal number of cells that were washed before in 10 mM Tris buffer (pH 7.5) and incubated 12-16 h at 30 ^{o}C.

Oligo DNAs

Oligo DNAs used in this study are listed below:

TN5MT 3'fw TCTAATAAAATAAATAATAATACTAAACTTAAAATAATGAAAAGAGATATTAACACGCC

TN5MT 3'rv GCGAGCACAGAATTAATACGACTCTCGAGCCAATAGGATCACTTCTGATTTTAGGG

TN5MT 5'fw GCTGATGGCGATGAATGAACACTGGGATCCAAGCTTATTATTTGAAATTTTGCCATATTG

TN5MT 5'rv CCATACTTTGAAGATATCAAGCTTATCGATACCTTTTAAAACGTAATAACAAATCAAAG

T2CPfw AGTGAATTCATGAAGCGTGACATCAACACCCCG

T2CPrv GACCTCGAGTTAGCTGTAACGGCAGGCCAGTTCC

T2ZPfw AGTGAATTCATGCCGAGCAAGAGCCAGCACGCC

TPB2_MBP_rv AGTCGACCTCGAGTTAGTCGCTGTCGATCTGGATGG

T2EC C618A fw GCCGTACCAAGAAAAAGAGCGCCATCG

T2EC C618A rv CGATGGCGCTCTTTTTCTTGGTACGGC

T2EC C629A fw CCCTGTTCAGCGCCAGCACCTGCAGC

T2EC C629A rv GCTGCAGGTGCTGGCGCTGAACAGGG

ArtlongFW

GGCCATCTGGCACAACGACCTGCTTGCAAGAGCCAAGCGCCCTGTCGCG**AGTGAT**GCGGAGAGACGGC

ATCCGGCTTGCGACGTCATGGCAACTGCGAAG

ArtlongRV

CTTCGCAGTTGCCATGACGTCGCAAGCCGGATGCCGTCTCTCCGC**ATCACT**CGCGACAGGGCGCTTGGC

TCTTGCAAGCAGGTCGTTGTGCCAGATGGCC

pD5H8fw ACTTTTTCATTCAAATCCACTTTCGC

pD5H8rv AGCATGAGAGTAGAATTTCCTACGATC

Bra1_OL_5RACErv GCTGATGGCGATGAATGAACACTGGTCGACGGCAAAAAAAAATATTATTATTGG

Bra1_OL_fw GCAAGGATTGCATTTACAGTTAGACATATGAGATATCTTCAAAGTATGG

MTT2_OL_fw CCATACTTTGAAGATATCTCATATGTCTAACTGTAAATGCAATCCTTGC

MTT2HA_AvrII_rv

GAGAGGTACCGAATTCACTAGTAGCATAATCAGGAACATCATAAGGATAGGATCCCATTTTCTTTATTGA

AATATTTATTGCTTTTTTAGCTTATTTTATAGTTGTTTTGTTTG

TPB2_ORF_AvrII_fw GCGCCTAGGATGAAAAGAGATATTAACACGCC

TPB2ORF_MluI_rv

GCGAGCACAGAATTAATACGACTACGCGTTCAGTCTGAGTCAATTTATATAGTAGTTTTTAGC

TPB2-C_MluI_rv GAGACGACGCGTTCATTTAGTTTTTACTTGTATTGGATTTTAATTGC

TPB2_D297L_fw CCTAGCTATTCTTGAAGGTATGATACC

TPB2_D297L_rv GGTATCATACCTTCAAGAATAGCTAGG

TPB2_D379L_fw GTAGTAATGCTTAATTATTATAACAGTCC

TPB2_D379L_rv GGACTGTTATAATAATTAAGCATTACTAC

TPB2_D495L_fw GGAGGTGTTCTTAGAAGAAACAG

TPB2_D495L_rv CTGTTTCTTCTAAGAACACCTCC

TPB2_C618A_fw CCCGTACTAAAAAAAAATCTGCTATTGAATGCAAATAATTAAC

TPB2_C618A_rv GTTAATTATTTGCATTCAATAGCAGATTTTTTTTTAGTACGGG

TPB2_C629A_fw GCAAATAATTAACTTTATTTTCTGCTAGTACATGCTC

TPB2_C629A_rv GAGCATGTACTAGCAGAAAATAAAGTTAATTATTTGC

TPB2_ZnF_fw GCGAGAATTCACCTACGAAGAAGTGGTGGAAGACATGTTCACCAAGCG

TPB2_ZnF_rv GAGCCTCGAGTTACTTGCTGGCGTGGAAGTCGTAGCACTGG

Creation of a conditional TPB2 KO strain

475 bp of TPB2 5' flanking region and the first 1132 bp of TPB2 genomic sequence were amplified from genomic DNA with primer pairs TN5MT 5' fw/rv and TN5MT3'fw/rv respectively. Neo5 resistance cassette fused to MTT1 promoter was amplified from pMNMM3 vector with primers TN5MTneo fw/rv. After PCR purification the fragments were combined using overlapping PCR and directly used for germline transformation of mating Tetrahymena UMPS strains at 3 h post mixing. After transformation 0.1 µg/ml cadmium chloride was added to the cells to induce TPB2 expression from MTT1 promoter for finishing conjugation. One clone resistant to pararomycin and 5-FOA was obtained and cultured until maturation. After mating to WT strain heterozygous conditional KO strains were genotyped by PCR and crossed to each other to obtain homozygous conditional KO strains. During all matings 0.1 µg/ml cadmium chloride was added.

Rescue system for conditional TPB2KO

Blasticidin resistance cassette was amplified by PCR from pBla1 vector using primers Bra 1_OL_5RACErv and Bra1_OL_fw. In parallel the MTT2 promoter was amplified from pDET2 vector with primers MTT2_OL_fw and MTT2_HA_AvrII_rv. The two PCR constructs were combined by overlapping PCR. Conditions were as described previously [25]. The overlapping PCR product was cloned in pBNMB1 vector with *Avr*II and *Sal*I restriction enzymes resulting in pBBM2B. The TPB2 open reading frame was amplified from genomic Tetrahymena DNA (strain B2086) using primers TPB2ORF_AvrII_fw and TPB2ORF_MluI_rv and subsequently cloned with enzymes *Mlu*I and *Avr*II in pBBM2B resulting in pBBM2B-TPB2. To get the C terminal deletion of TPB2, PCR was done on genomic DNA using primers TPB2ORF_AvrII_fw and TPB2-C_MluI_rv. The PCR product was cloned in pBBM2B using *Avr*II and *Mlu*I. By site direct mutagenesis (QuikChange® II Site-Directed Mutagenesis Kit) a catalytic dead version and a zinc finger mutant were created from the template pBBM2B-TPB2 using primers TPB2_D297L_fw/rv, TPB2_D379L_fw/rv, TPB2_D495L_fw/rv or TPB2_C618A_fw/rv, C629A_fw/rv respectively.

These rescue vectors were then transformed into somatic nucleus of two different mating types of conditional KO strains by ballistic transformation. Positive transformants were selected by blasticidin (up to 10 mg/ml).

Transformation of *Tetrahymena thermophila*

Germline transformation of Tetrahymena was done using particle gun as described earlier [26]. A hepta-adapter was used instead of a single adapter. Biolistic transformation of macronucleus was done as described previously [26]. The ribosomal vector was introduced into mating Tetrahymena cells by electroporation at 10 h post mixing as it was described earlier [27].

Immunofluorescence analysis

Fixation of Tetrahymena cells was done in 3.7% formaldehyde and 0.5% Triton-X 100 for 30 min at room temperature (RT). Cells were resuspended in 3.7% formaldehyde and 3.4% sucrose and dried on Superfrost Ultra Plus slides (Thermo Fisher scientific). The samples were blocked for 2 h with 3% BSA (Sigma), 10% normal goat serum (Invitrogen), and 0.1% Tween 20 in PBS followed by overnight incubation at 4°C in blocking solution containing a 1:1000 dilution of anti-HA (covance), anti-Pdd1p (abcam) or anti-Tpb2p antiserum. The anti-Tpb2p antibody was obtained by immunizing a rabbit with recombinant TPB2-CP. After washing with PBST, samples were incubated with a 1:2000 dilution of secondary antibody against mouse or rabbit conjugated to Alexa 488 or Alexa 568 (Invitrogen). The samples were washed, incubated with 10 ng/mL DAPI (Sigma) in PBST, mounted with ProLong Gold (Invitrogen), and observed by fluorescent microscopy.

Evaluation of DNA elimination by Fluorescence in situ hybridization (FISH)

A FISH probe against the Tlr1 IES was produced by mixing the plasmids Tlr1IntB, Tlr1 2 and Tlr1 4C1 [28]. Labeling of the DNA with Cy3 was achieved by nick translation. The genomic DNA and the probe were denatured by hot formamide and afterwards hybridized with the probe overnight at 37 ^{o}C. For detailed description see [29]

Production of Recombinant Proteins

DNA of recombinant proteins was amplified by PCR from pGEX-TPB2, pGEX-TPB2-CD [6] using primers T2CPfw/T2ECrv. N terminal (CP) and C terminal half (ZP) were amplified from pGEX-TPB2 using primers T2CPfw/T2CPrv and T2ZPfw/T2ECrv respectively. PCR products were cut with *Eco*RI and *Xho*I and cloned into pMalC2X vector cut with *Eco*RI and *Sal*I to obtain pMAL-TPB2, pMAL-TPB2-CD, pMAL-TPB2-CP and pMAL-TPB2-ZP. To generate pMAL-TPB2-ZM,

expressing a Zinc finger mutant, C618 to A and C629 to A mutations were introduced into pMAL-TPB2 using the QuikChange II site-directed mutagenesis Kit (Stratagene, La Jolla, CA), and DNA oligos T2EC C618A fw/rv and T2EC C629A fw/rv. To obtain pMAL-TPB2-ZP-ZM DNA was amplified from pMAL-TPB2-ZM using primers T2ZPfw/T2ECrv. To create the zinc finger fused to MBP protein the primers TPB2_ZnF_fw and rv were used with pMAL-TPB2. pMAL vectors containing TPB2 versions were expressed in E. coli strain BL21(DE3) which were cultivated to an A_{600} of ~ 0.8 and then incubated with 0.5 mM IPTG for 10 h at 16ºC. Cells were lysed in 500 mM NaCl, 80 mM Tris pH 8.0, 0.2 mM PMSF and 1x complete proteinase inhibitor cocktail [30]. The lysate was incubated with Amylose resin (New England Biolabs) at 4ºC and afterwards washed with 500 mM NaCl, 80 mM Tris pH 8.0 and finally eluted with 500 mM NaCl, 80 mM Tris pH 8.0, 20 mM maltose followed by dialysis in 20 mM Tris-HCl, pH 7.5, 100 mM KCl, 4mM $MgCl_2$ 4 mM $MnSO_4$, and 10% glycerol

Tpb2p Endonuclease Assay

The endonuclease assay was done as described earlier [6]. The oligomeric DNA used for the experiments was Artlong. In the oligo list the nucleotides mutated for the assay in figure 5 are labeled bold.

Peptide binding assays

The interactions of Tpb2p recombiant proteins to MODified^{TM} Histone Peptide array (Active motif) was analyzed according to the manual supplied with the peptide array. For pull down assays, peptides with the sequence of the N-terminal tail of histone H3 were synthesized with a C-terminal biotin separated from the most C-terminal amino acid by a PEG linker. 1 ug of peptide was incubated for 30 min at RT with streptavidin coupled dynabeads that were blocked with 2 % BSA in interaction buffer (recipe from MODified histone peptide array) for 1h at RT before. After washing with interaction buffer and blocking again with 2 % BSA 100 ug of the

recombinant Tpb2p proteins were added to the beads and incubated for 1h at 4 ^{o}C. After 5 washing steps with modified interaction buffer (300 mM KCl and 0.1 % Tween) the beads were resuspended in SDS PAGE loading buffer and boiled for 5 min. The samples were afterwards separated on a 10 % SDS- Polyacrylamid gel followed by Western Blot. Detection was done using anti-MBP antibody (NEB antiserum) followed by incubation with secondary antibody coupled with infrared dye. Visualization was done using Odyssey scanner.

Supplementary information

Table 1: Results from the MODified histone peptide array. Shown are the 25 highest hits

	peptide	Mod1	Mod2	Mod 3	Mod 4	mean intensity	intensity/ unmodified
1	H2B 1-19	K5ac	K12ac	S14P		1737.75	2.22
2	H3 1-19	R8me2a	K9me3			2048.00	2.10
3	H3 1-19	R2me2a	T3P	K4me3		1912.25	1.96
4	H3 1-19	R8me2s				1856.75	1.90
5	H3 1-19	R2me2a	K4me1	R8me2a	K9me3	1782.00	1.82
6	H3 1-19	R2me2s	K4me1	R8me2a	K9me3	1724.75	1.76
7	H2a 1-19	S1P	K5ac			1612.00	1.76
8	H3 1-19	K4ac	R8me2s	K9me3		1706.00	1.75
9	H3 1-19	R2me2a	K4me1	K9ac		1685.50	1.72
10	H3 1-19	R2me2a	T3P	K4me1		1680.00	1.72
11	H3 1-19	R2me2s	K4me2	R8me2a	K9me2	1672.00	1.71
12	H3 16-35	R26me2a	K27me3			2103.25	1.71
13	H3 1-19	R2me2a	T3P			1627.50	1.67
14	H3 1-19	R2me2s	K4ac	R8me2a	K9me1	1626.50	1.66
15	H3 1-19	R2Citr	K4me1			1612.50	1.65
16	H3 1-19	R2Citr	K4me3			1612.50	1.65
17	H3 1-19	R2me2a	T3P	K4ac		1561.75	1.60
18	H3 1-19	K9me3				1554.75	1.59
19	H3 1-19	R2me2s	K4ac	R8me2s	K9me2	1544.50	1.58
20	H3 1-19	R2me2s	K4ac	K9me3		1522.25	1.56
21	H3 1-19	R2me2s	K4me3	K9me3		1520.50	1.56
22	H3 1-19	K4me1	R8me2a	K9me3		1516.00	1.55
23	H3 1-19	K4ac	R8me2a	K9me3		1509.00	1.54
24	H3 1-19	R2me2s	K4me1	K9me3		1506.25	1.54
25	H3 1-19	K4me1	R8me2s	K9me3		1504.25	1.54

Literature

1. Orgel, L.E. and F.H. Crick, *Selfish DNA: the ultimate parasite.* Nature, 1980. 284(5757): p. 604-7.
2. Slotkin, R.K. and R. Martienssen, *Transposable elements and the epigenetic regulation of the genome.* Nat Rev Genet, 2007. 8(4): p. 272-85.
3. Yoder, J.A., C.P. Walsh, and T.H. Bestor, *Cytosine methylation and the ecology of intragenomic parasites.* Trends Genet, 1997. 13(8): p. 335-40.
4. Grewal, S.I., *RNAi-dependent formation of heterochromatin and its diverse functions.* Curr Opin Genet Dev, 2010. 20(2): p. 134-41.
5. Kapitonov, V.V. and J. Jurka, *RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons.* PLoS Biol, 2005. 3(6): p. e181.
6. Cheng, C.Y., et al., *A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila.* Mol Biol Cell, 2010. 21(10): p. 1753-62.
7. Prescott, D.M., *The DNA of ciliated protozoa.* Microbiol Rev, 1994. 58(2): p. 233-67.
8. Yao, M.C., *Ribosomal RNA gene amplification in Tetrahymena may be associated with chromosome breakage and DNA elimination.* Cell, 1981. 24(3): p. 765-74.
9. Yao, M.C., et al., *DNA elimination in Tetrahymena: a developmental process involving extensive breakage and rejoining of DNA at defined sites.* Cell, 1984. 36(2): p. 433-40.
10. Schoeberl, U.E., et al., *Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in Tetrahymena.* Genes Dev, 2012. 26(15): p. 1729-42.
11. Mochizuki, K. and M.A. Gorovsky, *Small RNAs in genome rearrangement in Tetrahymena.* Curr Opin Genet Dev, 2004. 14(2): p. 181-7.
12. Liu, Y., K. Mochizuki, and M.A. Gorovsky, *Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena.* Proc Natl Acad Sci U S A, 2004. 101(6): p. 1679-84.
13. Liu, Y., et al., *RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena.* Genes Dev, 2007. 21(12): p. 1530-45.
14. Taverna, S.D., R.S. Coyne, and C.D. Allis, *Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena.* Cell, 2002. 110(6): p. 701-11.
15. Madireddi, M.T., et al., *Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena.* Cell, 1996. 87(1): p. 75-84.

16. Howard-Till, R.A. and M.C. Yao, *Induction of gene silencing by hairpin RNA expression in Tetrahymena thermophila reveals a second small RNA pathway.* Mol Cell Biol, 2006. 26(23): p. 8731-42.
17. Austerberry, C.F., C.D. Allis, and M.C. Yao, *Specific DNA rearrangements in synchronously developing nuclei of Tetrahymena.* Proc Natl Acad Sci U S A, 1984. 81(23): p. 7383-7.
18. Chalker, D.L., et al., *Flanking regulatory sequences of the Tetrahymena R deletion element determine the boundaries of DNA rearrangement.* Mol Cell Biol, 1999. 19(8): p. 5631-41.
19. Godiska, R. and M.C. Yao, *A programmed site-specific DNA rearrangement in Tetrahymena thermophila requires flanking polypurine tracts.* Cell, 1990. 61(7): p. 1237-46.
20. Yao, M.C., C.H. Yao, and B. Monks, *The controlling sequence for site-specific chromosome breakage in Tetrahymena.* Cell, 1990. 63(4): p. 763-72.
21. Gueguen, E., et al., *The transpososome: control of transposition at the level of catalysis.* Trends Microbiol, 2005. 13(11): p. 543-9.
22. Gratias, A., et al., *Developmentally programmed DNA splicing in Paramecium reveals short-distance crosstalk between DNA cleavage sites.* Nucleic Acids Res, 2008. 36(10): p. 3244-51.
23. Fass, J.N., et al., *Genome-Scale Analysis of Programmed DNA Elimination Sites in Tetrahymena thermophila.* G3 (Bethesda), 2011. 1(6): p. 515-22.
24. Esnault, C., et al., *Intrinsic characteristics of neighboring DNA modulate transposable element activity in Drosophila melanogaster.* Genetics, 2011. 187(1): p. 319-31.
25. Kataoka, K., U.E. Schoeberl, and K. Mochizuki, *Modules for C-terminal epitope tagging of Tetrahymena genes.* J Microbiol Methods, 2010. 82(3): p. 342-6.
26. Bruns, P.J. and D. Cassidy-Hanley, *Biolistic transformation of macro- and micronuclei.* Methods Cell Biol, 2000. 62: p. 501-12.
27. Sweet, M.T. and C.D. Allis, *Transformation of Tetrahymena thermophila by Electroporation.* CSH Protoc, 2006. 2006(4).
28. Wuitschick, J.D., et al., *A novel family of mobile genetic elements is limited to the germline genome in Tetrahymena thermophila.* Nucleic Acids Res, 2002. 30(11): p. 2524-37.
29. Loidl, J. and H. Scherthan, *Organization and pairing of meiotic chromosomes in the ciliate Tetrahymena thermophila.* J Cell Sci, 2004. 117(Pt 24): p. 5791-801.
30. Keogh, M.C., et al., *A phosphatase complex that dephosphorylates gammaH2AX regulates DNA damage checkpoint recovery.* Nature, 2006. 439(7075): p. 497-501.

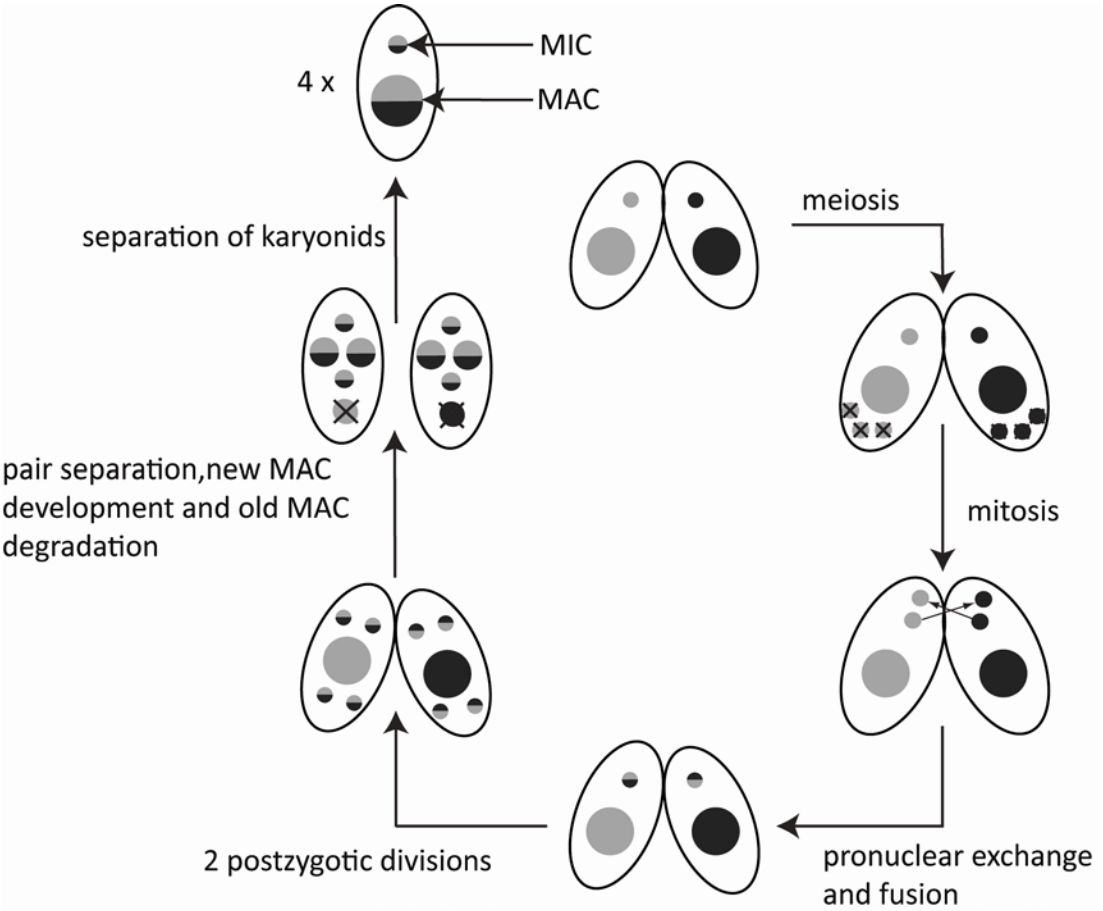
4 x
MIC
MAC
meiosis
separation of karyonids
pair separation,new MAC
development and old MAC
degradation
mitosis
2 postzygotic divisions
pronuclear exchange
and fusion

Figure 1

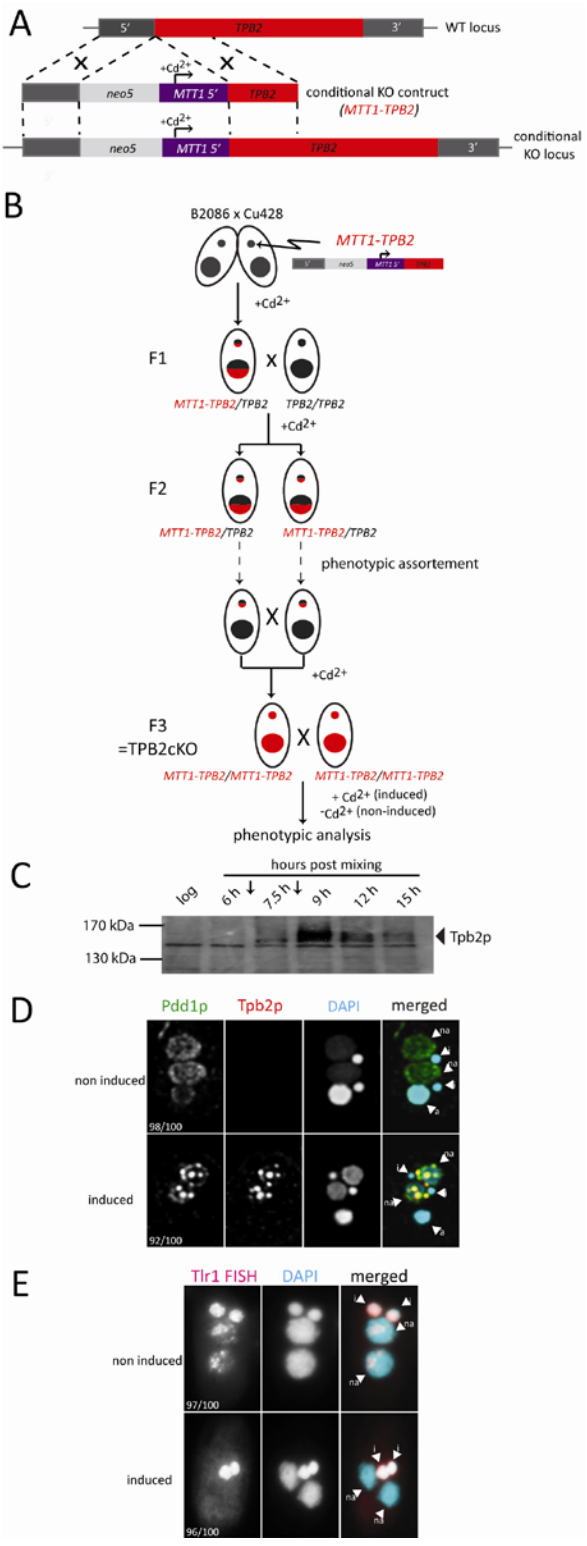
A
5'
TPB2
3'
WT locus
neo5
MTT1 5'
TPB2
conditional KO contruct
(MTT1-TPB2)
conditional
KO locus
B
B2086 x Cu428
MTT1-TPB2
F1
MTT1-TPB2/TPB2
TPB2/TPB2
F2
phenotypic assortement
F3
=TPB2cKO
MTT1-TPB2/MTT1-TPB2
+ Cd2+ (induced)
- Cd2+ (non-induced)
phenotypic analysis
C
hours post mixing
log
6 h
7.5 h
9 h
12 h
15 h
170 kDa
130 kDa
Tpb2p
D
Pdd1p
Tpb2p
DAPI
merged
non induced
induced
98/100
92/100
E
Tlr1 FISH
DAPI
merged
non induced
induced
97/100
96/100

Figure 2

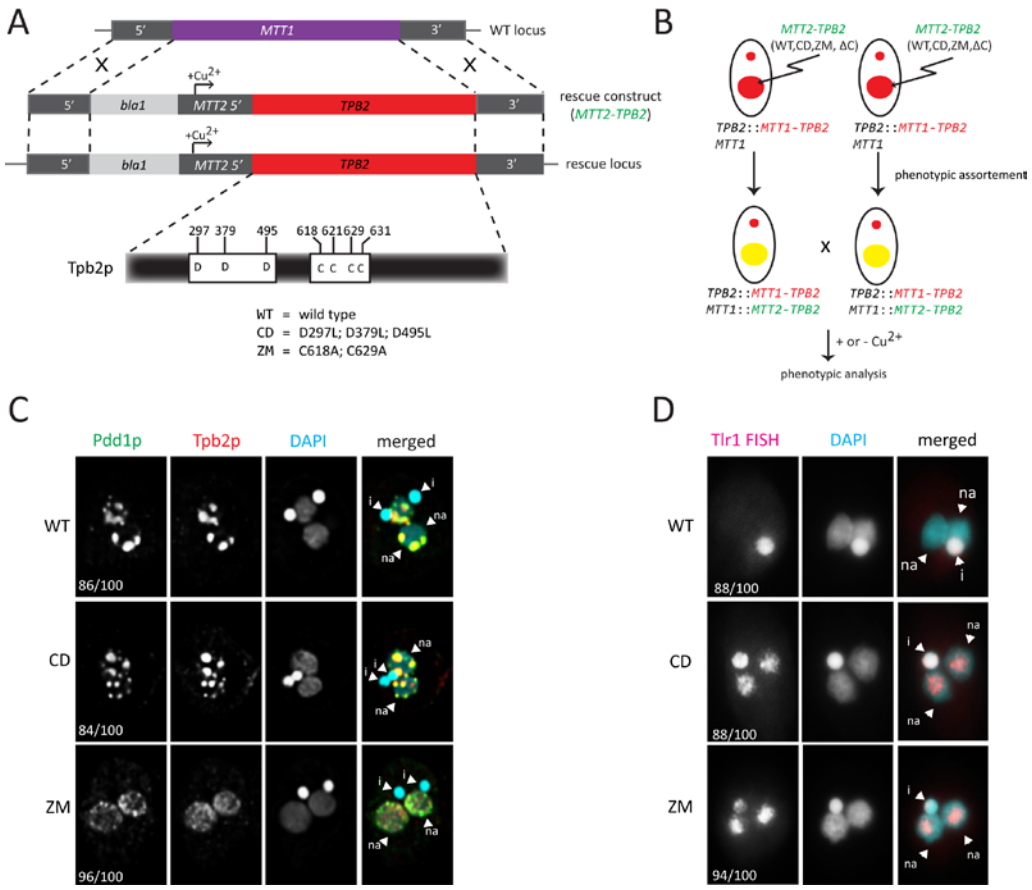
A
5′
MTT1
3′
WT locus
X
X
+Cu2+
5′
bla1
MTT2 5′
TPB2
3′
rescue construct
(MTT2-TPB2)
+Cu2+
5′
bla1
MTT2 5′
TPB2
3′
rescue locus
297 379 495 618 621629 631
Tpb2p
D D D C C CC
WT = wild type
CD = D297L; D379L; D495L
ZM = C618A; C629A
B
MTT2-TPB2
(WT,CD,ZM, ΔC)
MTT2-TPB2
(WT,CD,ZM,ΔC)
TPB2::MTT1-TPB2
MTT1
TPB2::MTT1-TPB2
MTT1
phenotypic assortement
X
TPB2::MTT1-TPB2
MTT1::MTT2-TPB2
TPB2::MTT1-TPB2
MTT1::MTT2-TPB2
+ or - Cu2+
phenotypic analysis
C
Pdd1p
Tpb2p
DAPI
merged
WT
86/100
CD
84/100
ZM
96/100
D
Tlr1 FISH
DAPI
merged
WT
88/100
CD
88/100
ZM
94/100

Figure 3

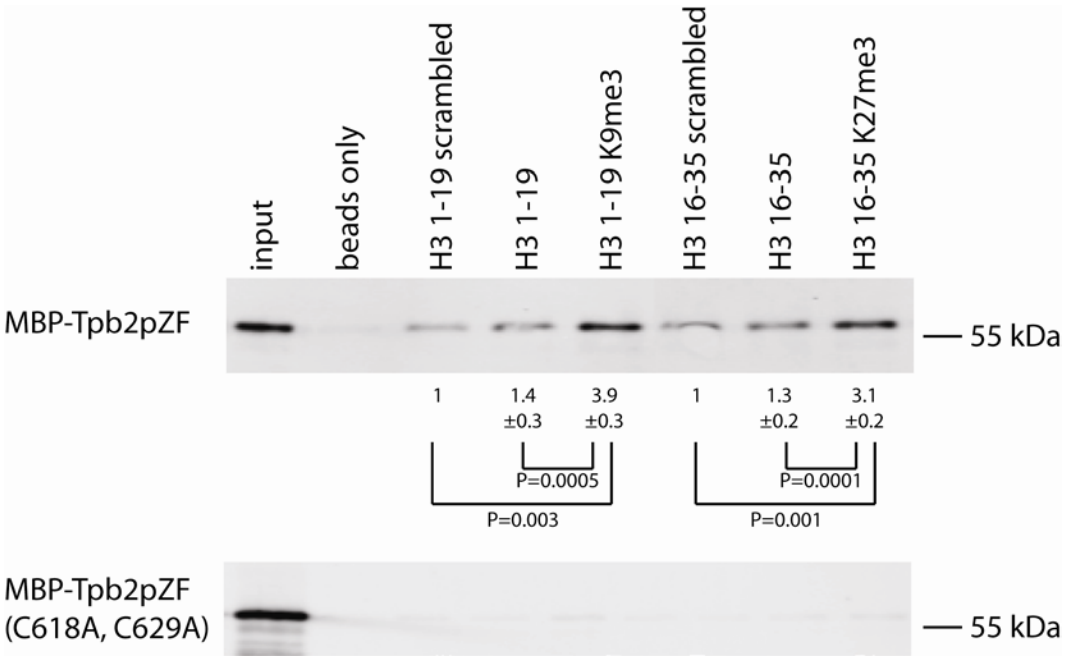
input
beads only
H3 1-19 scrambled
H3 1-19
H3 1-19 K9me3
H3 16-35 scrambled
H3 16-35
H3 16-35 K27me3
MBP-Tpb2pZF
55 kDa
1
1.4 ±0.3
3.9 ±0.3
1
1.3 ±0.2
3.1 ±0.2
P=0.0005
P=0.003
P=0.0001
P=0.001
MBP-Tpb2pZF (C618A, C629A)
55 kDa

Figure 4

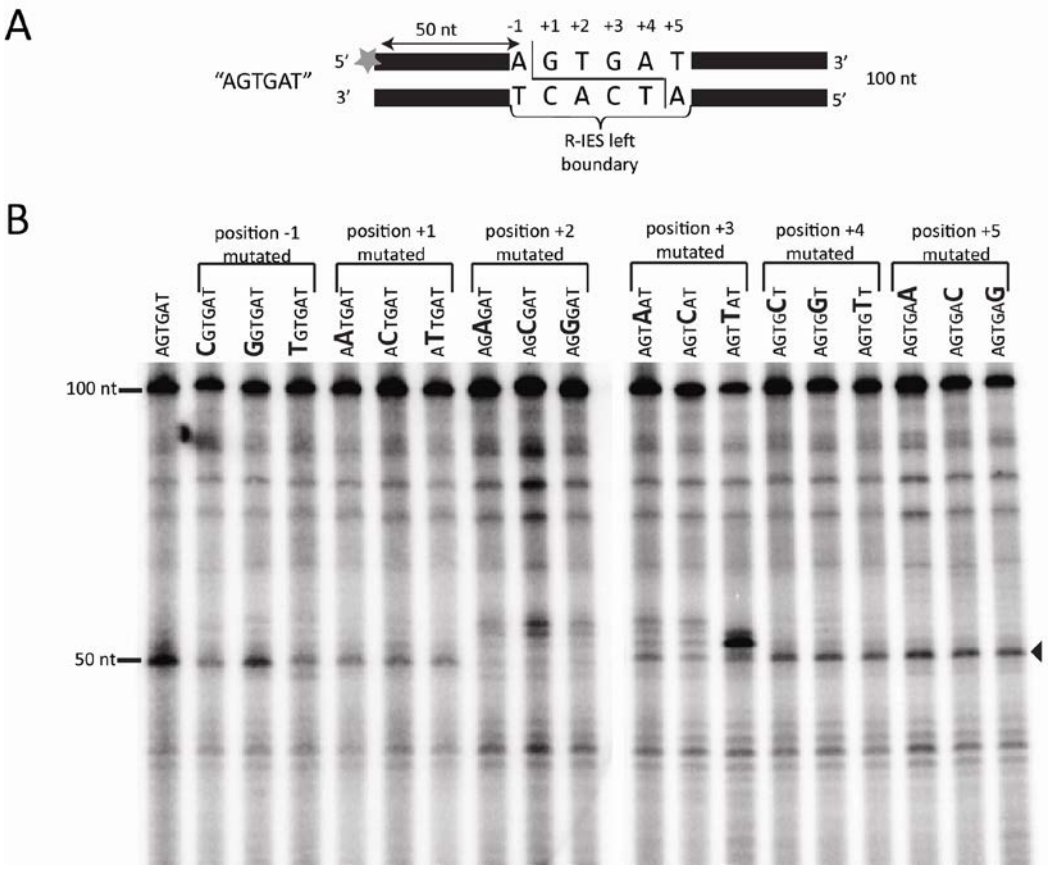
A
50 nt
-1 +1 +2 +3 +4 +5
"AGTGAT"
5′
3′
A G T G A T
T C A C T A
3′
5′
100 nt
R-IES left boundary
B
position -1 mutated
position +1 mutated
position +2 mutated
position +3 mutated
position +4 mutated
position +5 mutated
AGTGAT
CGTGAT
GGTGAT
TGTGAT
AATGAT
ACTGAT
ATTGAT
AGAGAT
AGCGAT
AGGGAT
AGTAAT
AGTCAT
AGTTAT
AGTGCT
AGTGGT
AGTGTT
AGTGAA
AGTGAC
AGTGAG
100 nt
50 nt

Figure 5

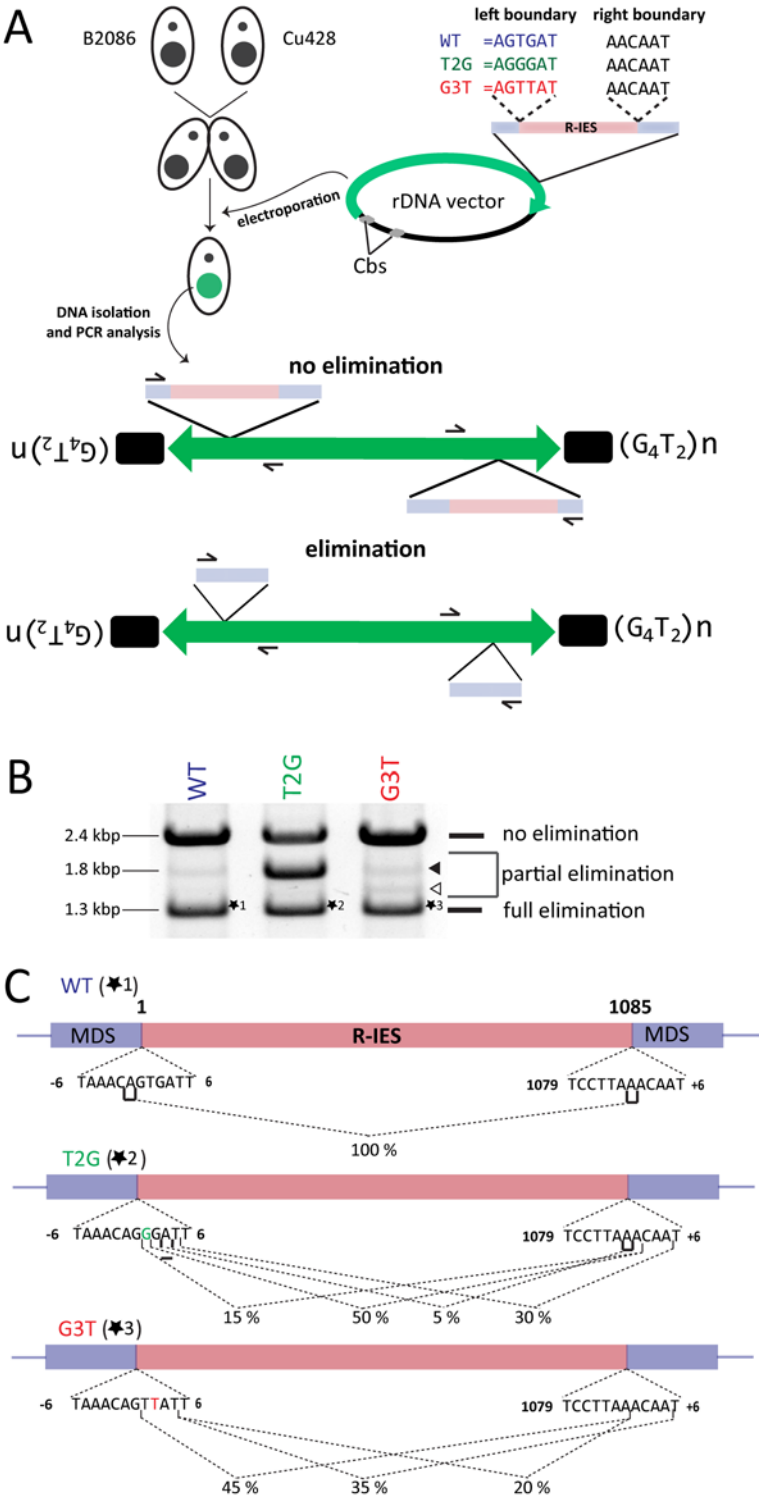
A
B2086
Cu428
left boundary
right boundary
WT =AGTGAT AACAAT
T2G =AGGGAT AACAAT
G3T =AGTTAT AACAAT
R-IES
electroporation
rDNA vector
Cbs
DNA isolation
and PCR analysis
no elimination
elimination
B
WT
T2G
G3T
2.4 kbp
1.8 kbp
1.3 kbp
no elimination
partial elimination
full elimination
C
WT (✱1)
1
1085
MDS
R-IES
MDS
-6 TAAACAGTGATT 6
1079 TCCTTAAACAAT +6
100 %
T2G (✱2)
-6 TAAACAGGGATT 6
1079 TCCTTAAACAAT +6
15 %
50 %
5 %
30 %
G3T (✱3)
-6 TAAACAGTTATT 6
1079 TCCTTAAACAAT +6
45 %
35 %
20 %

Figure 6

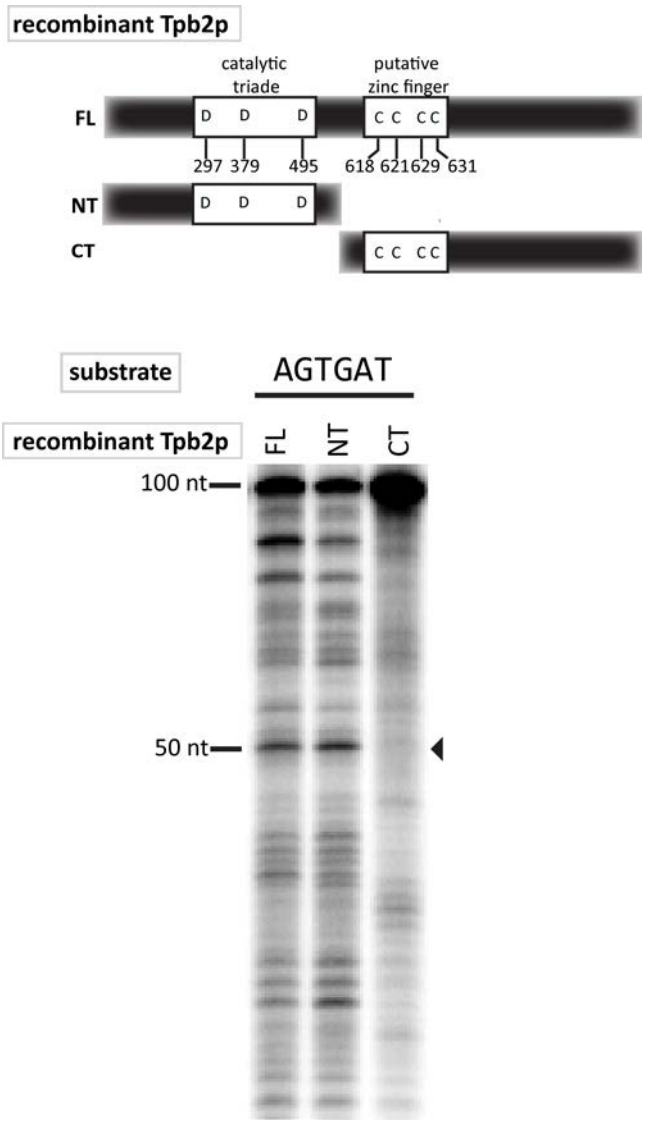
recombinant Tpb2p
catalytic triade
putative zinc finger
FL
D D D
C C C C
297 379 495
618 621629 631
NT
D D D
CT
C C C C
substrate
AGTGAT
recombinant Tpb2p
FL
NT
CT
100 nt
50 nt

Figure 7

5. Discussion

In my PhD thesis I investigated the final step of DNA excision during the programmed DNA elimination pathway in *Tetrahymena thermophila* during sexual reproduction. For many years this final step attracted the interest of researchers as the enzyme excising DNA was unknown. However genome sequencing projects [16] opened up new possibilities to find the "excisase". Because some IESs in Tetrahymena (as well as in other ciliates) share sequence similarities with transposons the idea already existed that transposases might be the enzymes responsible for DNA excision [42]. Furthermore it was known from a PCR based study in the mid 90's that excision of IESs resembles the signature some transposases show during transposition and that it happens with a 4 base 5' overhang [39]. This restricted the choice already to a few families within the DDE/DDD superfamily of transposases. One of them is the *piggyBac* family and we found two different genes in Tetrahymena that are similar to *piggyBac* transposase. Both of them have conserved catalytic aspartic acids with piggyBac transposase and a predicted zinc finger domain. Tpb1p (**T**etrahymena **p**iggy**B**ac like transposase **1**) has an additional Ku80 domain in the N terminal region. This was for us the strongest candidate to excise IESs because after the initial cut, the protein could directly bind this double strand break and initiate the religation of the flanking regions via non homologous end joining (NHEJ) pathway, from which Ku80 is a key component. However the TPB1 KO cells we created did not show defects in DNA elimination and chromosome breakage.

Together with a collaborating group I could show that the other *piggyBac* like transposase, Tpb2p (**T**etrahymena **p**iggy**B**ac like transposase **2**), is indispensable for DNA deletion and chromosome breakage [40]. An RNAi knockdown experiment revealed that different IESs are not excised during sexual reproduction and that chromosome breakage is impaired, resulting in lethality of sexual progeny in the absence of Tpb2p. Furthermore we have shown that Tpb2p localizes to the Pdd1p containing heterochromatin bodies in which DNA elimination is believed to happen, because many proteins necessary for DNA elimination have been shown to localize there [29, 30, 43]. Interestingly in the absence of Tpb2p the heterochromatin bodies did not

form suggesting a direct role for Tpb2p in heterochromatin dynamics. In the following study I could show that the maturation of heterochromatin bodies is depending on Tpb2p's zinc finger domain, but not on its catalytic activity. The fact that the zinc finger is important for chromatin dynamics fits together with the observation that the zinc finger of canonical *piggyBac* transposase is dispensable for its in vitro double strand formation activity on naked DNA while essential for in vivo transposition where chromatinized DNA is present [Nancy Craig, personal communication]. Furthermore zinc fingers in proteins were already shown to be associated with chromatin binding. One famous example is the 11 zinc finger protein CTCF that is involved in loop formation and chromatin dynamics during V(D)J recombination [44, 45].

This led us to speculate that Tpb2p's zinc finger domain is important for the direct interaction with chromatin and in an in vitro assay the zinc finger domain of Tpb2p binds to the N terminal tail of histone H3. The IES specific, tri-methylated lysines 9 and 27 (K9me3 and K27me3) enhance the affinity. However also unmodified H3 amino terminal tail is recognized. This raises the questions how Tpb2p is targeted specifically to IESs and how the rest of the genome is protected from DNA elimination by Tpb2p. An explanation could be that there are additional chromatin modifications on IESs that increase Tpb2p's specificity or that non-IES regions have specific chromatin modifications that reject Tpb2p. A good candidate for such a repelling modification is the acetylation of lysine 9 of histone H3 that was once shown to be enriched on non-IES regions [25]. This speculative theory should be investigated in future research.

My in vivo study has indicated that, in contrast to the zinc-finger domain, the catalytic domain is not required for the maturation of heterochromatin bodies but is necessary for the final DNA elimination process. Therefore, I analyzed the enzymatic activity of Tpb2p in detail to ask if Tpb2p has DNA double strand break activity and if it requires any sequence property on DNA substrate. I could show that Tpb2p recognizes the left R-IES boundary in vitro [40]. To reveal how Tpb2p does this and to get insight in how other IESs could be recognized we tested the effect of mutations in the left R-IES boundary sequence on Tpb2p's activity and we found that the mutation of position 3 and 4 impaired cleavage at the correct position. I could confirm the importance of these two positions also in vivo suggesting that the integrity of IES boundaries is

necessary for their precise excision. Together with the specific chromatin modifications this could ensure the precise excision of IESs by Tpb2p.

Altogether the results of my PhD study suggest Tpb2p is probably recruited to IESs by its interaction to the IES-restricted histone modifications and induces DNA double-strand break at its preferential sequence adjacent to heterochromatin. This mechanism may determine boundaries of IESs with a precision in the range of a few to several nucleotides. However, recently evidence was provided that there are also IESs in exons of Tetrahymena genes. This requires some mechanism that excises DNA with the precision of a single nucleotide at these exon invaded IESs. An interesting speculation is that the sequences at IES boundaries don't create a sequence consensus but rather lead to the formation of a certain DNA structure (bending) that can be recognized by Tpb2p. This would be especially interesting as a recent study showed that the neighboring sequence of *piggyBac* insertions in Drosophila is important for the integration and that this could correlate with the bendability of the DNA [46].

In an independent parallel study in the related ciliate *Paramecium tetraurelia* the homologous protein PiggyMac was identified [41]. Furthermore, another independent study in an evolutionary distinct ciliate, *Oxytricha trifallax*, showed that transposases of the Tc1/mariner family are involved in DNA elimination [47]. Thus in three different ciliates transposon derived genes are indispensable for programmed genome rearrangements. This provides a unique platform to study the evolution of DNA elimination systems in ciliates. Interestingly in the two hypotrichous ciliates Paramecium and Tetrahymena DNA elimination systems make use of a domesticated transposase as the both organisms encode the transposases responsible for IES excision as a single gene in their macronuclear, somatic genome. In contrast the spirotrichous ciliate Oxytricha has active transposons encoded in its germline genome that are involved in their own removal and in the deletion of other sequences. This contrasts domestication of a transposon gene versus the co – existence with an active transposon (domestication vs mutualism). This leaves room for the speculation that the mutualistic system in Oxytricha is the ancestral type and domestication of a transposase have occurred during the course of the

evolution of Paramecium and Tetrahymena. I discussed this circumstantially in the review article “Transposon domestication versus mutualism in ciliate genome rearrangements”.

The study of the Tpb2p piggyBac transposase not only provides knowledge about the process of DNA elimination in ciliates, but may also enable new usages of *piggyBac* transposons as genetic tools. *piggyBac* transposons have been used for mutagenesis, transgenesis and initial gene therapy experiments in various organisms. *Drosophila melanogaster* was succesfully mutagenized using *piggyBac* transposon and with this system researchers for example investigated the alcohol tolerance of *D. melanogaster* [48, 49]. The system was also used for transgenesis of mosquitos, chicken, pigs and human cells (reviewed in [50]). Another great achievement is the use of the piggyBac in gene therapy which already was demonstrated to work in mice [51]. However, it is unknown how piggyBac transposase exactly interacts with DNA and which domain is the DNA binding domain. Further investigation of Tpb2p in Tetrahymena is expected to provide a better understanding on the molecular action of general *piggyBac* transposons and help developing improved genetic tools based on *piggyBac* transposon.

6. Literature

1. Prescott, D.M., *The DNA of ciliated protozoa.* Microbiol Rev, 1994. **58**(2): p. 233-67.
2. Gibbons, I.R. and A.J. Rowe, *Dynein: A Protein with Adenosine Triphosphatase Activity from Cilia.* Science, 1965. **149**(3682): p. 424-6.
3. Kruger, K., et al., *Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.* Cell, 1982. **31**(1): p. 147-57.
4. Blackburn, E.H. and J.G. Gall, *A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena.* J Mol Biol, 1978. **120**(1): p. 33-53.
5. Greider, C.W. and E.H. Blackburn, *Identification of a specific telomere terminal transferase activity in Tetrahymena extracts.* Cell, 1985. **43**(2 Pt 1): p. 405-13.
6. Malone, C.D., et al., *Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila.* Mol Cell Biol, 2005. **25**(20): p. 9151-64.
7. Mochizuki, K., et al., *Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena.* Cell, 2002. **110**(6): p. 689-99.
8. Mochizuki, K. and M.A. Gorovsky, *RNA polymerase II localizes in Tetrahymena thermophila meiotic micronuclei when micronuclear transcription associated with genome rearrangement occurs.* Eukaryot Cell, 2004. **3**(5): p. 1233-40.
9. Mochizuki, K. and M.A. Gorovsky, *A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase.* Genes Dev, 2005. **19**(1): p. 77-89.
10. Collins, K. and M.A. Gorovsky, *Tetrahymena thermophila.* Curr Biol, 2005. **15**(9): p. R317-8.
11. Tondravi, M.M. and M.C. Yao, *Transformation of Tetrahymena thermophila by microinjection of ribosomal RNA genes.* Proc Natl Acad Sci U S A, 1986. **83**(12): p. 4369-73.
12. Gaertig, J. and M.A. Gorovsky, *Efficient mass transformation of Tetrahymena thermophila by electroporation of conjugants.* Proc Natl Acad Sci U S A, 1992. **89**(19): p. 9196-200.
13. Cassidy-Hanley, D., et al., *Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment.* Genetics, 1997. **146**(1): p. 135-47.
14. Gaertig, J., et al., *Electroporation-mediated replacement of a positively and negatively selectable beta-tubulin gene in Tetrahymena thermophila.* Proc Natl Acad Sci U S A, 1994. **91**(10): p. 4549-53.
15. Hai, B. and M.A. Gorovsky, *Germ-line knockout heterokaryons of an essential alpha-tubulin gene enable high-frequency gene replacement and a test of gene transfer from somatic to germ-line nuclei in Tetrahymena thermophila.* Proc Natl Acad Sci U S A, 1997. **94**(4): p. 1310-5.
16. Eisen, J.A., et al., *Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote.* PLoS Biol, 2006. **4**(9): p. e286.
17. Yao, M.C., K. Zheng, and C.H. Yao, *A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in Tetrahymena.* Cell, 1987. **48**(5): p. 779-88.
18. Fan, Q. and M. Yao, *New telomere formation coupled with site-specific chromosome breakage in Tetrahymena thermophila.* Mol Cell Biol, 1996. **16**(3): p. 1267-74.
19. Fillingham, J.S., et al., *A non-long terminal repeat retrotransposon family is restricted to the germ line micronucleus of the ciliated protozoan Tetrahymena thermophila.* Eukaryot Cell, 2004. **3**(1): p. 157-69.
20. Grewal, S.I., *RNAi-dependent formation of heterochromatin and its diverse functions.* Curr Opin Genet Dev, 2010. **20**(2): p. 134-41.

21. Schoeberl, U.E., et al., *Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in Tetrahymena.* Genes Dev, 2012. **26**(15): p. 1729-42.
22. Mochizuki, K. and M.A. Gorovsky, *Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement.* Genes Dev, 2004. **18**(17): p. 2068-73.
23. Noto, T., et al., *The Tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus.* Cell, 2010. **140**(5): p. 692-703.
24. Aronica, L., et al., *Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in Tetrahymena.* Genes Dev, 2008. **22**(16): p. 2228-41.
25. Taverna, S.D., R.S. Coyne, and C.D. Allis, *Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena.* Cell, 2002. **110**(6): p. 701-11.
26. Liu, Y., K. Mochizuki, and M.A. Gorovsky, *Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena.* Proc Natl Acad Sci U S A, 2004. **101**(6): p. 1679-84.
27. Liu, Y., et al., *RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena.* Genes Dev, 2007. **21**(12): p. 1530-45.
28. Madireddi, M.T., et al., *Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena.* Cell, 1996. **87**(1): p. 75-84.
29. Nikiforov, M.A., M.A. Gorovsky, and C.D. Allis, *A novel chromodomain protein, pdd3p, associates with internal eliminated sequences during macronuclear development in Tetrahymena thermophila.* Mol Cell Biol, 2000. **20**(11): p. 4128-34.
30. Rexer, C.H. and D.L. Chalker, *Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in Tetrahymena thermophila.* Eukaryot Cell, 2007. **6**(8): p. 1320-9.
31. Wuitschick, J.D., et al., *A novel family of mobile genetic elements is limited to the germline genome in Tetrahymena thermophila.* Nucleic Acids Res, 2002. **30**(11): p. 2524-37.
32. Fass, J.N., et al., *Genome-Scale Analysis of Programmed DNA Elimination Sites in Tetrahymena thermophila.* G3 (Bethesda), 2011. **1**(6): p. 515-22.
33. Fillingham, J.S., D. Bruno, and R.E. Pearlman, *Cis-acting requirements in flanking DNA for the programmed elimination of mse2.9: a common mechanism for deletion of internal eliminated sequences from the developing macronucleus of Tetrahymena thermophila.* Nucleic Acids Res, 2001. **29**(2): p. 488-98.
34. Patil, N.S. and K.M. Karrer, *A developmentally regulated deletion element with long terminal repeats has cis-acting sequences in the flanking DNA.* Nucleic Acids Res, 2000. **28**(6): p. 1465-72.
35. Chalker, D.L., et al., *Flanking regulatory sequences of the Tetrahymena R deletion element determine the boundaries of DNA rearrangement.* Mol Cell Biol, 1999. **19**(8): p. 5631-41.
36. Godiska, R. and M.C. Yao, *A programmed site-specific DNA rearrangement in Tetrahymena thermophila requires flanking polypurine tracts.* Cell, 1990. **61**(7): p. 1237-46.
37. Gratias, A. and M. Betermier, *Processing of double-strand breaks is involved in the precise excision of paramecium internal eliminated sequences.* Mol Cell Biol, 2003. **23**(20): p. 7152-62.
38. Klobutcher, L.A. and G. Herrick, *Consensus inverted terminal repeat sequence of Paramecium IESs: resemblance to termini of Tc1-related and Euplotes Tec transposons.* Nucleic Acids Res, 1995. **23**(11): p. 2006-13.
39. Saveliev, S.V. and M.M. Cox, *Developmentally programmed DNA deletion in Tetrahymena thermophila by a transposition-like reaction pathway.* EMBO J, 1996. **15**(11): p. 2858-69.
40. Cheng, C.Y., et al., *A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila.* Mol Biol Cell, 2010. **21**(10): p. 1753-62.

41. Baudry, C., et al., *PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia.* Genes Dev, 2009. **23**(21): p. 2478-83.
42. Klobutcher, L.A. and G. Herrick, *Developmental genome reorganization in ciliated protozoa: the transposon link.* Prog Nucleic Acid Res Mol Biol, 1997. **56**: p. 1-62.
43. Madireddi, M.T., M.C. Davis, and C.D. Allis, *Identification of a novel polypeptide involved in the formation of DNA-containing vesicles during macronuclear development in Tetrahymena.* Dev Biol, 1994. **165**(2): p. 418-31.
44. Sanchez, R. and M.M. Zhou, *The PHD finger: a versatile epigenome reader.* Trends Biochem Sci, 2011. **36**(7): p. 364-72.
45. Chaumeil, J. and J.A. Skok, *The role of CTCF in regulating V(D)J recombination.* Curr Opin Immunol, 2012. **24**(2): p. 153-9.
46. Esnault, C., et al., *Intrinsic characteristics of neighboring DNA modulate transposable element activity in Drosophila melanogaster.* Genetics, 2011. **187**(1): p. 319-31.
47. Nowacki, M., et al., *A functional role for transposases in a large eukaryotic genome.* Science, 2009. **324**(5929): p. 935-8.
48. Thibault, S.T., et al., *A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac.* Nat Genet, 2004. **36**(3): p. 283-7.
49. Eanes, W.F., et al., *Direct evidence that genetic variation in glycerol-3-phosphate and malate dehydrogenase genes (Gpdh and Mdh1) affects adult ethanol tolerance in Drosophila melanogaster.* Genetics, 2009. **181**(2): p. 607-14.
50. Kim, A. and I. Pyykko, *Size matters: versatile use of PiggyBac transposons as a genetic manipulation tool.* Mol Cell Biochem, 2011. **354**(1-2): p. 301-9.
51. Nakanishi, H., et al., *piggyBac transposon-mediated long-term gene expression in mice.* Mol Ther, 2010. **18**(4): p. 707-14.

7. Abbreviations

A	(deoxy)adenosine nucleotide
Asp	(D) aspartic acid
C	(deoxy)cytosine nucleotide
CBS	chromosome breakage sequence
Cys	(C) cysteine
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNS	Desoxyribonukleinsäure
dsDNA	double stranded deoxyribonucleic acid
dsRNA	double stranded RNA
DTT	dithiothreitol
G	(deoxy)guanosine nucleotide
Glu (E)	glutamic acid
GST	glutathione-S-transferase
hr	hour
IES	internal eliminated sequence
kDa	Kilodalton
Leu (L)	leucine
M	molar
MAC	macronucleus
MBP	Maltose binding protein
MCS	multiple cloning site
MIC	micronucleus
min	minute
mM	millimolar
mRNA	messenger ribonucleic acid

µg	microgram
µL	microlitre
N	unspecified (deoxy)ribonucleotide
nm	nanometre
nM	nanomolar
nt	nucleotide
OD600	absorption at 600 nm wavelength
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pH	*pondus Hydrogenii*
PMSF	phenylmethanesulfonyl fluoride
RNA	ribonucleic acid
RNS	Ribonukleinsäure
RNAi	ribonucleic acid interference
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
siRNA	small interfering ribonucleic acid
T	deoxythymidine nucleotide
Tris-HCl	2-Amino-2-hydroxymethyl-propane-1,3-diol hydrochloride
°C	degrees Celsius

8. Curriculum vitae

Personal dates

Name:	Alexander Vogt
Date of birth:	21/06/1982
Place of birth:	Ulm (Germany)
Gender:	male
Languages:	English, German, French

Education

1988 – 1992	Primary school in Illerrieden (Germany)
1992 – 2001	Secondary school in Ulm-Wiblingen (Germany)
20.06.2001	final secondary school examinations (grade: 2,3 (”good“)
10/2001 – 03/2003	Studies of French cultural sciences at the university of Saarbruecken (Germany)
23.09.2003-04.11.2008	Studies of biology at the University of Ulm. Graduated with the Diploma (grade 1,0 (“very good”)
11/2008-present	PhD thesis in molecular biology at the Institue of Molecular Biotechnolgy of the Austrian Academy of Sciences (IMBA) in Vienna (Austria)

9. Acknowledgements

I want to thank Kaz for the great supervision he gave me within the last four years. In the beginning, when I needed more support he always had an open door and time to talk about the poject. But he also gave me the opportunity to develop and learn how to drive projects on my own. He always found the right balance between supporting and giving freedom.

Furthermore I want to thank all current and previous members in the lab – Henri, Jan, Jana, Ken, Lucia, Sophie, Tomoko and Ursi – for creating a wonderful atmosphere to work in. You became more than just colleagues to me!

I also want to thank my family, especially my parents for always supporting me. I think we all remember the hard times I had in the beginning in Vienna!

Last but not least I want to thank Elisa for the great support she gave me and the wonderful time we are having together! Without you I wouldn't be where I am right now.