The taming of the shrew

Regulation of a catalytically active domesticated transposase

Alexander Vogt[1,*] and Kazufumi Mochizuki[2,*]

[1]European Molecular Biology Laboratory (EMBL); Heidelberg, Germany; [2]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA); Vienna, Austria

Transposons are mobile genetic elements that can be harmful for the host when mobilized. However, they are also genomic reservoirs for novel genes that can be evolutionarily beneficial. There are many examples of domesticated transposases, which play important roles in the hosts. In most cases domesticated transposases have lost their endonuclease activities and the hosts utilize their DNA-binding properties. However, some other domesticated transposases perform endonuclease activities for host biological processes. Because such a catalytically active transposase is potentially harmful for the integrity of the host genome, its activity should be tightly regulated. The catalytically active domesticated *piggyBac* transposase Tpb2p catalyzes programmed DNA elimination in the ciliate *Tetrahymena*. Here, we discuss the regulatory mechanism that prevents unintended DNA cleavage by Tpb2p and compare it to another well-studied catalytically active domesticated transposase, the RAG recombinase in V(D)J recombination. The regulatory mechanisms involve the temporarily regulated expression of the transposases, the target sequence preference of the endonuclease, and the recruitment of the transposases to locally restricted chromatin environments.

Transposases are enzymes encoded by DNA transposons. They catalyze the excision of the transposon at one genomic locus and its integration at a distant position. Both of these steps can be deleterious to the host. The excision step causes DNA double-strand breaks that are harmful if not repaired properly[1] and their integration may disrupt host genes. Nonetheless, the harmfulness of transposases is only one side of the coin, and it is becoming clear that many transposases have been domesticated for the benefit of their hosts.[2] In most cases, the host benefits from the DNA- or chromatin-interaction ability of the transposase. Examples include *Daysleeper* in *Arabidopsis*[3] and the widespread eukaryotic Cenp-b.[4] In contrast, there are only a few examples in which the domesticated transposase is still catalytically active. Because the endonuclease activity of transposases represents a great threat to the host, tight regulatory mechanisms are necessary to prevent unintended genome rearrangements. The most famous example of a domesticated transposase with catalytic activity is the RAG (recombination-activated gene) recombinase from jawed vertebrates.[5] The RAG enzyme complex containing RAG1 and RAG2 catalyzes the double-strand break formation in V(D)J recombination during lymphocyte development. Recently, we and others reported further examples of catalytically active transposases from the *piggyBac* family that play essential roles in programmed DNA elimination in the ciliated protists *Tetrahymena* and *Paramecium*.[6,7] Here, we discuss how the activity of the *Tetrahymena piggyBac* transposase 2 (Tpb2p) is regulated to catalyze programmed DNA elimination without causing harmful genome rearrangement events through a comparison of Tpb2p and the RAG recombinase.
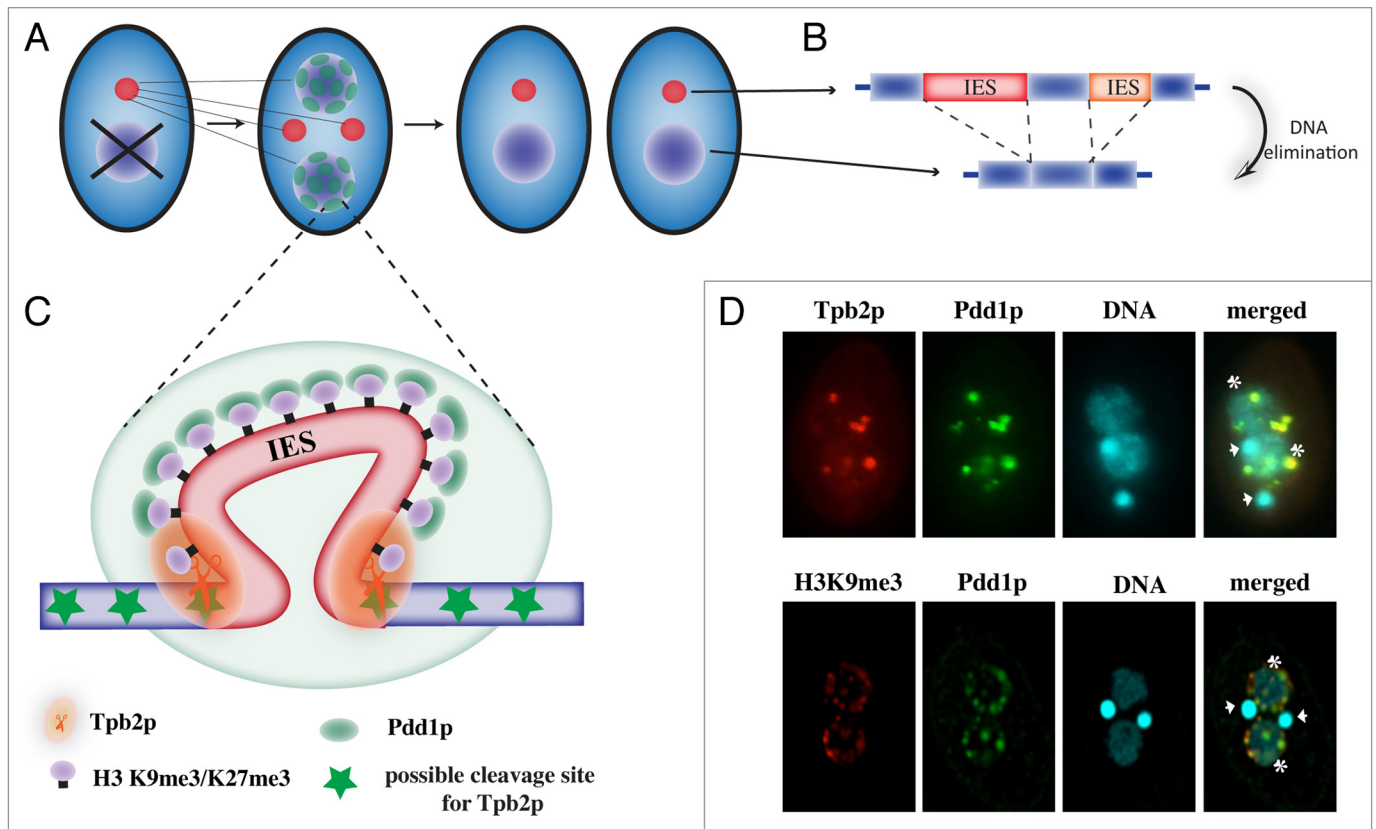
Figure 1. Conjugation and DNA elimination in *Tetrahymena*. (**A**) During conjugation, the Mic gives rise to the new Mic and the new Mac. DNA elimination occurs in the exconjugant, which has 2 Mics and 2 new Macs (middle); red = Mic, purple = Mac. (**B**) DNA elimination removes IESs from the developing Mac genome, and the flanking regions are ligated. (**C**) DNA elimination occurs in specific nuclear compartments, heterochromatin bodies, which contain histone H3 K9me3/K27me3, Pdd1p, and Tpb2p. We propose that this compartmentalization restricts Tpb2p's action (see text). (**D**) Localization of Tpb2p, Pdd1p, and H3K9me3 in heterochromatin bodies, as shown by immunofluorescence staining. Asterisk = new Mac, Arrow = Mic.

Most ciliates, such as *Tetrahymena thermophila,* have two structurally and functionally distinct nuclei in a single cell (**Fig. 1A**). The large, polyploid macronucleus (Mac) provides the cell with all RNAs necessary for vegetative growth, whereas the small, diploid micronucleus (Mic) is transcriptionally silenced during vegetative growth. During sexual reproduction, the Mic acts as the germline nucleus and produces both the Mac and the Mic of the progeny[8] (**Fig. 1A**). During new Mac development, two types of programmed genome rearrangement events occur in *Tetrahymena*. The first is chromosome breakage, by which five germline-derived chromosomes are fragmented to approximately 200 Mac chromosomes.[9] The breaks are repaired by de novo telomere addition. The second type of genome restructuring is the elimination of more than 8000 internal eliminated sequences (IESs), followed by ligation of their flanking sequences (**Fig. 1B**). The elimination of IESs occurs reproducibly: invariable sets of IESs are eliminated from the Mac, and the majority of their boundaries vary by only a few to several base pairs.[10] IESs are thought to be transposon remnants, with some of them still having the potential to be activated.[11] Therefore, IES excision is a process that removes harmful genetic elements from the transcriptionally active Mac. Moreover, it is essential for the streamlining of the somatic genome because it removes some IESs from gene-coding regions, creating functional genes.[12]

We recently showed that the endonuclease activity of the domesticated *piggyBac* transposase Tpb2p is required for IES elimination in *Tetrahymena*,[13] which strongly suggests that Tpb2p is the enzyme responsible for excising IESs. Indeed, our in vitro study showed a preferred activity of Tpb2p at the boundary sequences of well-studied R-IES[13] and other elements in vitro (unpublished). Different IESs in *Tetrahymena* do not share any detectable common sequences, neither within themselves nor in their flanking regions. Therefore, it is expected that, in sharp contrast to canonical DNA transposases, the enzyme that excises IESs should have promiscuous substrate specificity. Consistent with this idea, we found that only the 2nd and 3rd bases downstream of the cleavage site are crucial for the precision of the cleavage at the R-IES boundary both in vitro and in vivo.[13] However, this promiscuous substrate specificity of Tpb2p raises a question: how is the precise boundary of IESs specified? We believe one of the key components to solve this issue is the heterochromatin formation that occurs at IESs.

The IES elimination process can be divided into four distinct steps: i) small RNA-directed heterochromatin formation, ii) assembly of heterochromatin

	RAG	Tpb2p
Expression and stability	Only during lymphocyte development, proteasomal degradationafter V(D)J recombination	Only during new Mac development, specific degradation mechanism?
Chromatin environment	Euchromatin (H3K4me3)	Heterochromatin (H3K9me3, H3K27me3)
Endonuclease activity	activated upon H3K4me3 binding	Activated upon K9me3 or K27me3 binding? Inhibited by euchromatin?
Sequence preference	RSS	IES boundaries (including 5'TTAA3')
Higher order nuclear compartment	Recombination centers	Heterochromatin bodies

Figure 2. Comparison of the regulatory mechanisms of RAG recombinase and Tpb2p.

bodies (**Fig. 1C**), iii) DNA excision, and iv) ligation of the flanking DNA by a non-homologous end-joining (NHEJ) pathway.[14] Heterochromatin formation includes the de novo establishment of tri-methylation on lysine 9 and lysine 27 of histone H3 (H3K9me3 and K27me3), specifically at IESs[15,16] (**Fig. 1C**). The chromodomain protein Pdd1p subsequently binds one or both of the histone modifications and promotes the assembly of IESs into sub-nuclear foci (heterochromatin bodies)[17](**Fig. 1D**). We have shown that Tpb2p is required for the assembly of heterochromatin bodies and DNA excision in *Tetrahymena*.[7,13] The fact that endonucleolytically inactive Tpb2p can only promote the former step supports the longstanding idea that DNA excision occurs in heterochromatin bodies. The cysteine-rich domain (CRD) of Tpb2p, which might fold into a PHD finger, was found to preferentially interact with peptides having the histone H3 N-terminal tail sequence when they were tri-methylated at either position lysine 9 (K9me3) or lysine 27 (K27me3) in vitro.[13] Furthermore, the mutation of two cysteines in the CRD led to the inability of Tpb2p to promote heterochromatin body formation in vivo.[13] Therefore, Tpb2p is most probably recruited to IESs via interactions between H3K9/K27me3 and CRD, and this interaction is necessary for the assembly of heterochromatin bodies. We speculate that heterochromatin formation at IESs restricts the endonucleolytic action of Tpb2p in two ways: by specifically recruiting Tpb2p to the IESs and by preventing Tpb2p from accessing the middle region of IESs. Tpb2p may only be able to cleave euchromatic regions adjacent to heterochromatinized IESs. A potential regulation could also be mediated by cis-regulatory sequences that are found close to some IESs.[18-22] They might be able to position Tpb2p in a reproducible manner by localizing nucleosomes at precise positions and/or by arranging IESs in distinct chromatin compartments.

Although the interaction of Tpb2p with heterochromatin may restrict its action at the edges of IESs, the promiscuous substrate specificity of Tpb2p could still potentially cause undesirable DNA cleavage at non-target genomic locations. We currently do not know how such a deleterious action of Tpb2p is inhibited in *Tetrahymena*. To consider how Tpb2p might be regulated, it is worth recalling what is known about another catalytically active domesticated transposase, RAG recombinase. Here, we propose a potential regulatory mechanism of Tpb2p that was deduced from a comparison of Tpb2p and RAG recombinase (**Fig. 2**).

First, these potentially harmful enzymes are only expressed in a particular developmental process with a short time window. Tpb2p expression is limited to the stage of new Mac development when DNA elimination occurs.[7] The RAG proteins are only expressed during lymphocyte development and are quickly degraded after productive V(D)J recombination by the proteasome.[23,24] Therefore, it would be interesting to study whether there is a similar active protein degradation mechanism for Tpb2p in *Tetrahymena*.

Second, in both DNA elimination and V(D)J recombination, the chromatin environment of the target sequences regulates the interaction of the recombinase with the target sequences. As discussed above, Tpb2p is likely tethered to IESs through heterochromatin-specific histone modifications. In contrast, the heterochromatic environment inhibits the access of RAG recombinase to recombination signal sequences (RSS),[25] a consensus sequence at V, D and J genes. The recombinase can access an RSS only when the locus is activated for transcription, accompanied by euchromatic modifications, including H3K4me3.[26] Moreover, RAG's endonuclease is activated upon binding H3K4me3.[26] It has been proposed that the binding of RAG2 via its PHD finger to H3K4me3 releases an auto-inhibition of the RAG1 protein.[27] Analogous to

this mechanism, the endonuclease activity of Tpb2p may be auto-inhibited and specifically activated upon its interaction with heterochromatin. Alternatively, a protection mechanism that repels Tpb2p or inhibits its enzymatic activity in non-IES (and thus euchromatic) regions could be present. Further study is necessary to reveal whether there is such an additional layer of regulation of Tpb2p's endonuclease activity.

Lastly, both DNA elimination and V(D)J recombination likely occur in specialized nuclear compartments. In *Tetrahymena*, DNA elimination is believed to occur in the heterochromatin bodies[13] and in mammals, V(D)J recombination is thought to occur in recombination centers.[28] In addition, DNA looping helps to bring separated gene segments in closer proximity for recombination.[29] This compartmentalization may inhibit the spontaneous action of domesticated transposases outside of their targets. As it is not known how IESs are assembled into heterochromatin bodies in *Tetrahymena*, it would

be interesting to investigate the physical configuration of the chromosomes in the developing Mac using a genome-wide chromosome conformation capture (3C) technique.

Although our knowledge of the regulation of Tpb2p is still far behind our knowledge of the RAG recombinase, the study of transposases in DNA elimination in ciliates may shed light on the evolutionary process of transposon domestication, which is one of the black boxes of the biology of domesticated transposases. *Tetrahymena* and its closely related ciliate *Paramecium* use *piggyBac* transposases that have undergone the domestication process for DNA elimination. These transposases are no longer encoded within the context of a transposon but are single-copy genes in the macronuclear genome.[30] In contrast, the more distantly related ciliate *Oxytricha* eliminates DNA with the help of Tc1/mariner-like transposases encoded by the germline-limited TBE transposons.[31] Therefore, it is possible that ciliates preserve traceable "fossil" records

of the transposase domestication process. The system in *Oxytricha*, in which transposons participate in their own excision from the somatic genome, could represent a very ancient system of DNA elimination. During the evolution of the *Tetrahymena/Paramecium* clade, a common ancestor could have domesticated a *piggyBac* transposase that was then modified in such a way that it could excise all IESs. Because ciliates are an evolutionarily very diverged group of eukaryotes, and their mode of DNA elimination varies in different species, it would be interesting to search for "missing link" species that might have both systems in parallel. We believe that further biochemical studies of Tpb2p in the model ciliate *Tetrahymena* and the comparison of DNA elimination pathways in different ciliate species will help in illuminating the "taming of the shrew" process of transposase domestication.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. Nature 1980; 284:604-7; PMID:7366731; http://dx.doi.org/10.1038/284604a0

2. Volff J-N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 2006; 28:913-22; PMID:16937363; http://dx.doi.org/10.1002/bies.20452

3. Bundock P, Hooykaas P. An Arabidopsis hAT-like transposase is essential for plant development. Nature 2005; 436:282-4; PMID:16015335; http://dx.doi.org/10.1038/nature03667

4. Cam HP, Noma K, Ebina H, Levin HL, Grewal SIS. Host genome surveillance for retrotransposons by transposon-derived proteins. Nature 2008; 451:431-6; PMID:18094683; http://dx.doi.org/10.1038/nature06499

5. Kapitonov VV, Jurka J. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol 2005; 3:e181; PMID:15898832; http://dx.doi.org/10.1371/journal.pbio.0030181

6. Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Bétermier M. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. Genes Dev 2009; 23:2478-83; PMID:19884254; http://dx.doi.org/10.1101/gad.547309

7. Cheng C-Y, Vogt A, Mochizuki K, Yao M-C. A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. Mol Biol Cell 2010; 21:1753-62; PMID:20357003; http://dx.doi.org/10.1091/mbc.E09-12-1079

8. Prescott DM. The DNA of ciliated protozoa. Microbiol Rev 1994; 58:233-67; PMID:8078435

9. Yao MC, Zheng K, Yao CH. A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in Tetrahymena. Cell 1987; 48:779-88; PMID:3815524; http://dx.doi.org/10.1016/0092-8674(87)90075-4

10. Austerberry CF, Snyder RO, Yao MC. Sequence microheterogeneity is generated at junctions of programmed DNA deletions in Tetrahymena thermophila. Nucleic Acids Res 1989; 17:7263-72; PMID:2798093; http://dx.doi.org/10.1093/nar/17.18.7263

11. Chalker DL, Yao M-C. DNA elimination in ciliates: transposon domestication and genome surveillance. Annu Rev Genet 2011; 45:227-46; PMID:21910632; http://dx.doi.org/10.1146/annurev-genet-110410-132432

12. Fass JN, Joshi NA, Couvillion MT, Bowen J, Gorovsky MA, Hamilton EP, Orias E, Hong K, Coyne RS, Eisen JA, et al. Genome-Scale Analysis of Programmed DNA Elimination Sites in Tetrahymena thermophila. G3 (Bethesda) 2011; 1:515-22; PMID:22384362; http://dx.doi.org/10.1534/g3.111.000927

13. Vogt A, Mochizuki K. A domesticated PiggyBac transposase interacts with heterochromatin and catalyzes reproducible DNA elimination in Tetrahymena. PLoS Genet 2013; 9:e1004032; PMID:24348275; http://dx.doi.org/10.1371/journal.pgen.1004032

14. Lin I-T, Chao J-L, Yao M-C. An essential role for the DNA breakage-repair protein Ku80 in programmed DNA rearrangements in Tetrahymena thermophila. Mol Biol Cell 2012; 23:2213-25; PMID:22513090; http://dx.doi.org/10.1091/mbc.E11-11-0952

15. Liu Y, Mochizuki K, Gorovsky MA. Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena. Proc Natl Acad Sci U S A 2004; 101:1679-84; PMID:14755052; http://dx.doi.org/10.1073/pnas.0305421101

16. Liu Y, Taverna SD, Muratore TL, Shabanowitz J, Hunt DF, Allis CD. RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. Genes Dev 2007; 21:1530-45; PMID:17575054; http://dx.doi.org/10.1101/gad.1544207

17. Madireddi MT, Coyne RS, Smothers JF, Mickey KM, Yao MC, Allis CD. Pdd1p, a novel chromo-domain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. Cell 1996; 87:75-84; PMID:8858150; http://dx.doi.org/10.1016/S0092-8674(00)81324-0

18. Patil NS, Karrer KM. A developmentally regulated deletion element with long terminal repeats has cis-acting sequences in the flanking DNA. Nucleic Acids Res 2000; 28:1465-72; PMID:10684943; http://dx.doi.org/10.1093/nar/28.6.1465

19. Chalker DL, La Terza A, Wilson A, Kroenke CD, Yao MC. Flanking regulatory sequences of the Tetrahymena R deletion element determine the boundaries of DNA rearrangement. Mol Cell Biol 1999; 19:5631-41; PMID:10409752

20. Godiska R, James C, Yao MC. A distant 10-bp sequence specifies the boundaries of a programmed DNA deletion in Tetrahymena. Genes Dev 1993; 7(12A):2357-65; PMID:8253382; http://dx.doi.org/10.1101/gad.7.12a.2357

21. Godiska R, Yao MC. A programmed site-specific DNA rearrangement in Tetrahymena thermophila requires flanking polypurine tracts. Cell 1990; 61:1237-46; PMID:2364428; http://dx.doi.org/10.1016/0092-8674(90)90688-B

22. Fillingham JS, Bruno D, Pearlman RE. 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; 29:488-98; PMID:11139619; http://dx.doi.org/10.1093/nar/29.2.488

23. Kuo TC, Schlissel MS. Mechanisms controlling expression of the RAG locus during lymphocyte development. Curr Opin Immunol 2009; 21:173-8; PMID:19359154; http://dx.doi.org/10.1016/j.coi.2009.03.008

24. Jiang H, Chang FC, Ross AE, Lee J, Nakayama K, Nakayama K, Desiderio S. Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle. Mol Cell 2005; 18:699-709; PMID:15949444; http://dx.doi.org/10.1016/j.molcel.2005.05.011

25. Jones JM, Gellert M. Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. EMBO J 2002; 21:4162-71; PMID:12145216; http://dx.doi.org/10.1093/emboj/cdf394

26. Shimazaki N, Tsai AG, Lieber MR. H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. Mol Cell 2009; 34:535-44; PMID:19524534; http://dx.doi.org/10.1016/j.molcel.2009.05.011

27. Grundy GJ, Yang W, Gellert M. Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. Proc Natl Acad Sci U S A 2010; 107:22487-92; PMID:21149691; http://dx.doi.org/10.1073/pnas.1014958107

28. Ji Y, Resch W, Corbett E, Yamane A, Casellas R, Schatz DG. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. Cell 2010; 141:419-31; PMID:20398922; http://dx.doi.org/10.1016/j.cell.2010.03.010

29. Chaumeil J, Micsinai M, Ntziachristos P, Deriano L, Wang JM, Ji Y, Nora EP, Rodesch MJ, Jeddeloh JA, Aifantis I, et al. Higher-order looping and nuclear organization of Tcra facilitate targeted rag cleavage and regulated rearrangement in recombination centers. Cell Rep 2013; 3:359-70; PMID:23416051; http://dx.doi.org/10.1016/j.celrep.2013.01.024

30. Vogt A, Goldman AD, Mochizuki K, Landweber LF. Transposon domestication versus mutualism in ciliate genome rearrangements. PLoS Genet 2013; 9:e1003659; PMID:23935529; http://dx.doi.org/10.1371/journal.pgen.1003659

31. Nowacki M, Higgins BP, Maquilan GM, Swart EC, Doak TG, Landweber LF. A functional role for transposases in a large eukaryotic genome. Science 2009; 324:935-8; PMID:19372392; http://dx.doi.org/10.1126/science.1170023