

# Evolutionary dynamics of transposable elements in a small RNA world

Justin P. Blumenstiel

Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045, USA

**Transposable elements (TEs) are selfish elements that cause harmful mutations, contribute to the structure of regulatory networks and shape the architecture of genomes. Natural selection against their harmful effects has long been considered the dominant force limiting their spread. It is now clear that a genome defense system of RNA-mediated silencing also plays a crucial role in limiting TE proliferation. A full understanding of TE evolutionary dynamics must consider how these forces jointly determine their proliferation within genomes. Here I consider these forces from two perspectives – dynamics within populations and evolutionary games within the germline. The analysis of TE dynamics from these two perspectives promises to provide new insight into their role in evolution.**

## Genomes and selfish elements

In sexually reproducing species, iterated rounds of meiosis and fertilization establish evolutionary conflict because genetic mixing allows harmful selfish elements to spread to uncolonized genomes [1]. The classic example of a selfish element is the transposable element (TE) – an element that replicates itself in the germline and proliferates in the genome across generations. TEs are key contributors to variation in genome architecture [2]. Indeed, the genomes of some sexually reproducing species are famously encumbered by TEs. Approximately 40% of the human genome is made up of TEs, many of which are inactive and ancient [3]. In the *Hydromantes* salamander genome, one TE family alone contributes 5 Gb of DNA sequence [4]. By contrast, less than 3% of the pufferfish genome is made up of TEs [5]. Why is there such variation in TE content across species?

A tremendous body of work has been devoted to this question. This work is largely influenced by population genetic theory – in particular, our understanding of the balance between mutation and selection in populations. However, the recent discovery of an ancient RNA-based immune system that protects the genome from TE proliferation has added complexity to our understanding of TE evolutionary dynamics. Here I place our understanding of these dynamics in the light of recent discoveries in host TE control by small RNA.

## Modes of TE control

The replicative nature of TEs allows their rapid exponential growth within populations. This raises the question: what limits the spread of TEs? In response to this question, a framework has been developed [6] to tease apart

the various forces that determine genomic TE abundance – transposition acting to increase copy number, and excision and natural selection acting to decrease copy number. Considering these three forces, and also genetic drift (Glossary), a TE copy-number equilibrium can be established provided that certain conditions are met. Stabiliza-

## Glossary

**Argonaute:** protein mediators of RNA silencing. Argonaute proteins form complexes with small RNAs that are used as guides for the detection of RNA with shared sequence complementarity. Argonaute endonuclease activity mediates target mRNA destruction. This is known as Slicer activity. Argonaute proteins also mediate other forms of silencing through their small-RNA-mediated interactions with complementary target RNA.

**Dicer:** an RNase III endoribonuclease. Dicer recognizes double-stranded RNA and cleaves it into 20 to 25 nt small-interfering RNAs (siRNAs).

**Ectopic recombination:** non-allelic recombination between dispersed chromosomal sites. Dispersed repeat sequences such as TEs can mediate ectopic recombination through shared sequence complementarity. Ectopic recombination is a known source of chromosomal damage and is considered an important mode of selection acting against the fixation of TE insertions.

**Effective population size:** the size of an idealized population with random mating that provides an equivalent measure of genetic drift in the population of interest. The effective population size is a measure of the importance of genetic drift in determining the evolutionary dynamics of alleles. A population with a small effective population size is more strongly influenced by the random effects of genetic drift than one with a large effective population size. Thus, a mildly advantageous allele (such as a repressor allele) has a greater chance of fixation relative to a neutral allele in a population with a larger effective size. Populations with large census sizes (such as the current human population) could have small effective population sizes if they frequently experience population bottlenecks.

**Genetic drift:** changes in allele frequencies that can be attributed to random sampling of alleles across generations in finite populations. Genetic drift reduces the deterministic nature of a population. Because of genetic drift, advantageous alleles do not always fix and occasionally slightly harmful alleles do.

**Haldane's rule:** the observation by J.B.S. Haldane that if a hybridization event produces one sterile or inviable sex, it will be the sex with two different sex chromosomes (the heterogametic sex). The broad generality of this rule indicates its great significance to understanding mechanisms of post-zygotic reproductive isolation.

**Heterochromatin:** differentially staining chromatin with reduced transcriptional activity. Heterochromatin also has a reduced amount of recombination. Thus, TEs in heterochromatin are believed to be less likely to participate in ectopic recombination.

**piRNA:** piwi-interacting RNA. Piwi proteins represent a class of Argonaute proteins that have so far been found only in ciliates and animals. piRNAs are a class of small silencing RNAs, typically 24 to 30 nt in length, that are found in complex with Piwi proteins and are produced by Dicer-independent mechanisms. TE control is a known function of piRNAs, but sequence analysis suggests there are other functions that have not been elucidated. The Dicer-independent mechanisms of their production are poorly understood, but they apparently arise from distinct loci (piRNA clusters) at different sites in the genome.

**Repressor allele/silencing allele:** a TE insertion allele that produces TE-derived small silencing RNA and mediates TE silencing in *trans* by either transcriptional or posttranscriptional gene silencing. Generally, TE insertions that produce double-stranded RNA mediate silencing through the production of siRNAs and TE insertions that land in the proper orientation in piRNA clusters to yield piRNA.

Corresponding author: Blumenstiel, J.P. (jblumens@ku.edu)

**Box 1. Modes of selection against TEs**

Three major forms of selection have been proposed to limit the fixation of TEs [7,8], selection acting against (i) the mutagenic effects of insertion, (ii) the metabolic costs of TE transcription and translation, and (iii) chromosomal damage arising from ectopic recombination among dispersed copies, especially heterozygous copies that lack pairing partners on homologous chromosomes [9]. Natural selection acting against TE fixation has long been considered to be a dominant force constraining their proliferation [7]. Two major predictions can be made if natural selection limits the spread of TEs. First, TE insertions should segregate at low frequencies in species with large population sizes. This is because genetic drift is weaker in large populations and natural selection is more effective in limiting the increase of harmful alleles. Second, TE fixation events should be more likely in regions of the genome that have reduced recombination rates. This is due to the reduced efficacy of selection in low-recombining regions, known as the Hill–Robertson effect [10], as well as the reduced likelihood for these insertions to participate in ectopic recombination.

Results from extensive research in *Drosophila melanogaster*, which has a large population size, confirm the two predictions outlined above. In *D. melanogaster* most TE alleles segregate at low frequencies and fixation events occur predominantly in the low-recombining pericentric heterochromatin [7,11]. In vertebrates with larger genomes and smaller population sizes TEs tend to segregate at higher frequencies, indicating that selection is less effective [12,13]. Inbreeding species, where the evolutionary dynamics of regulation and selection are expected to differ [14,15], show disagreement with some of the above predictions. As in *D. melanogaster*, the low-recombining pericentric regions of the selfing *Arabidopsis thaliana* are highly enriched for TEs. However, there is no negative correlation between TE abundance and recombination rate across the remainder of the genome [16]. *A. thaliana* TEs also segregate at higher population frequencies compared to TEs in its close outcrossing relative, *A. lyrata* [17]. In the selfing nematode, *Caenorhabditis elegans*, TEs are not enriched in regions of low recombination and they segregate in populations at higher frequencies compared to an outcrossing relative [18,19]. These observations from inbreeding populations have been attributed to the fact that inbreeding leads to increased homozygosity. With greater homozygosity, ectopic recombination among dispersed TEs will be less frequent because even low-frequency insertions will commonly be paired with an insert at the homologous chromosomal position [20]. With decreased ectopic recombination, selection against TEs is expected to decrease.

tion of TE copy number may be achieved if net transposition rates – the difference between transposition and excision – equal zero. This could occur under transposition repression if the rate of transposition per TE copy decreases with increasing copy number. Stabilization of TE copy number may also be achieved under certain selection regimes, for example, if each new copy in the genome reduces the logarithm of fitness by an increased amount. Selection against chromosome damage by TE-mediated ectopic recombination (Box 1) satisfies this condition [7].

Natural selection is widely credited as the dominant force constraining TE proliferation in eukaryotes (Box 1), and these forces have been discussed extensively elsewhere [7,8]. However, hybrid dysgenesis in *Drosophila melanogaster* has long indicated that additional factors are at play. Hybrid dysgenesis is a phenomenon of hybrid sterility [21,22] that occurs when strains possessing particular TEs are mated with those that lack them [23–25]. The salient feature of hybrid dysgenesis is that repression of transposition across generations is mediated in the female germline by the TEs themselves [26,27]. Thus,

TEs inherited from the father but absent from the mother can become derepressed in progeny. This difference in TE regulation between the male and female germline indicates that transposition repression is, in addition to natural selection, an important factor controlling TE accumulation.

Repression of transposition can be mediated either by TE lineages themselves or by the host. Self-repression is exhibited by TE lineages that autoregulate and reduce transposition to a rate below the maximum. Host-repression refers to a property of the host that reduces the transposition rate. A host allele that reduces TE transposition rate is defined as a host repressor allele. The theoretical conditions necessary for the evolution of both self- and host-repressive mechanisms have been determined [28], and in both cases the rate of recombination is a crucial factor. In asexual lineages, lack of genetic exchange permanently binds proliferating TEs to the harmful effects of TE copies elsewhere in the genome [1,29]. Because asexual lineages with highly active TEs will be outcompeted by lineages with less active TEs, natural selection acting on TE lineages in asexuals will favor self-repression. For similar reasons, inbreeding can also favor TE lineages exhibiting self-repression [20,28].

In outbreeding species, one might also assume that selection favors self-repression. However, conditions for the evolution of self-repression are restrictive in outbreeding species with high recombination rates [28]. This is because recombination allows TE lineages with high transposition rates to become unburdened from their deleterious consequences. A TE with a high transposition rate can copy itself, resulting in a harmful insertion. However, recombination will break up the association between the parent copy and harmful daughter copy. Therefore, the fate of the parent copy is minimally influenced by the fate of the daughter copy, which could itself be removed quickly from the population by natural selection due to its harmful effects. Under this scenario, natural selection on the TE lineage will favor high transposition rates, but not self-repression. The exception to this are cases in which TE insertions cause dominant lethality or sterility [28]. In these cases, there is no opportunity for the parent copy to escape its detrimental consequences. A daughter copy that causes dominant lethality or sterility will eliminate both itself and the parental copy from the population.

Conditions for the evolution of host-mediated regulation of TE proliferation are more easily met. In particular, selection will always act in favor of host mechanisms that lead to TE repression, although the rate of recombination still influences the magnitude of selection acting on a host repressor. Importantly, the strength of selection acting in favor of a host repressor allele will be higher when there is greater linkage between the repressor allele and TE insertion sites [28]. This is because a host repressor allele gains an advantage when it ameliorates the harmful effects of TE insertions. Within populations, this will occur more frequently when a host repressor allele is physically linked to harmful TE insertions.

A variety of host repressor alleles that reduce transposition rates have been identified in different species. In maize, *Mu killer* initiates heritable dominant silencing of

### Box 2. Mechanisms of TE control by small RNA

Mechanisms of TE control by small RNA have been extensively reviewed elsewhere [36–39]. The two best understood classes of small RNA that regulate TEs are small-interfering RNA (siRNA) and piwi-interacting RNA (piRNA). siRNAs are produced from double-stranded RNA (dsRNA) processed by the Dicer endoribonuclease, whereas piRNAs are Dicer-independent. siRNAs and piRNAs repress TEs through post-transcriptional gene silencing (PTGS) and transcriptional silencing by both DNA methylation and heterochromatin formation. In all cases, silencing by small RNA is mediated in conjunction with members of the Argonaute (AGO) family of proteins.

siRNAs play a crucial role in TE control in plants where Dicer-dependent small RNAs (miRNAs and siRNAs) are the only known class of silencing RNA. In plants, siRNAs repress TEs primarily through RNA-directed DNA methylation [40]. Plant siRNAs are known to be generated by two primary mechanisms [41,42]. Many TEs possess inverted repeats and frequently insert in orientations that yield sense and anti-sense RNA, producing dsRNA that is processed into siRNAs. In addition, TEs residing within the heterochromatin contribute siRNAs simply as a consequence of their location in silent chromatin. Transcripts derived from these regions are recognized as ‘aberrant’ and are targets of RNA-dependent RNA polymerase activity that generates additional TE-derived dsRNA [41–43].

Unlike plants, many animals have a designated germline, and in the gonad Dicer-independent [44] piRNAs have assumed a dominant role for TE control [36]. piRNAs are so named because they mediate silencing through their interaction with Piwi proteins, a distinct clade of the Ago gene family. Mechanisms of piRNA biogenesis are best understood in *D. melanogaster* where they arise from TEs that have landed within certain genomic loci designated as piRNA clusters [45]. These clusters drive the production of primary piRNAs that prime a larger pool of piRNAs through iterated cycles of destruction of sense and antisense TE transcripts. Although it has been established that this cycle destroys target RNA by PTGS [46], the mechanism of piRNA biogenesis from piRNA clusters is poorly understood. In mammals, piRNAs have a crucial role for TE control in the male but not the female germline [39], the latter probably relying on Dicer-dependent siRNAs [47,48]. Most of the evidence to date suggests that piRNAs repress TEs through PTGS. However, mammalian piRNAs, like plant siRNAs, have been implicated in DNA methylation of TE sequences [49] and there is genetic and biochemical evidence for piRNAs contributing to TE transcriptional silencing in *D. melanogaster* [50].

the *MuDR* element [30]. In *D. melanogaster*, a dominant allele of *flamenco* is known to silence the *gypsy* endogenous retrovirus [31]. Interestingly, *P*-element insertions in *D. melanogaster* telomeres also exhibit a general ability to act as host repressor alleles of *P* elements across the genome [32,33]. It is now clear that these well-characterized alleles all mediate TE repression through small-RNA-mediated mechanisms.

### TE silencing by small RNA

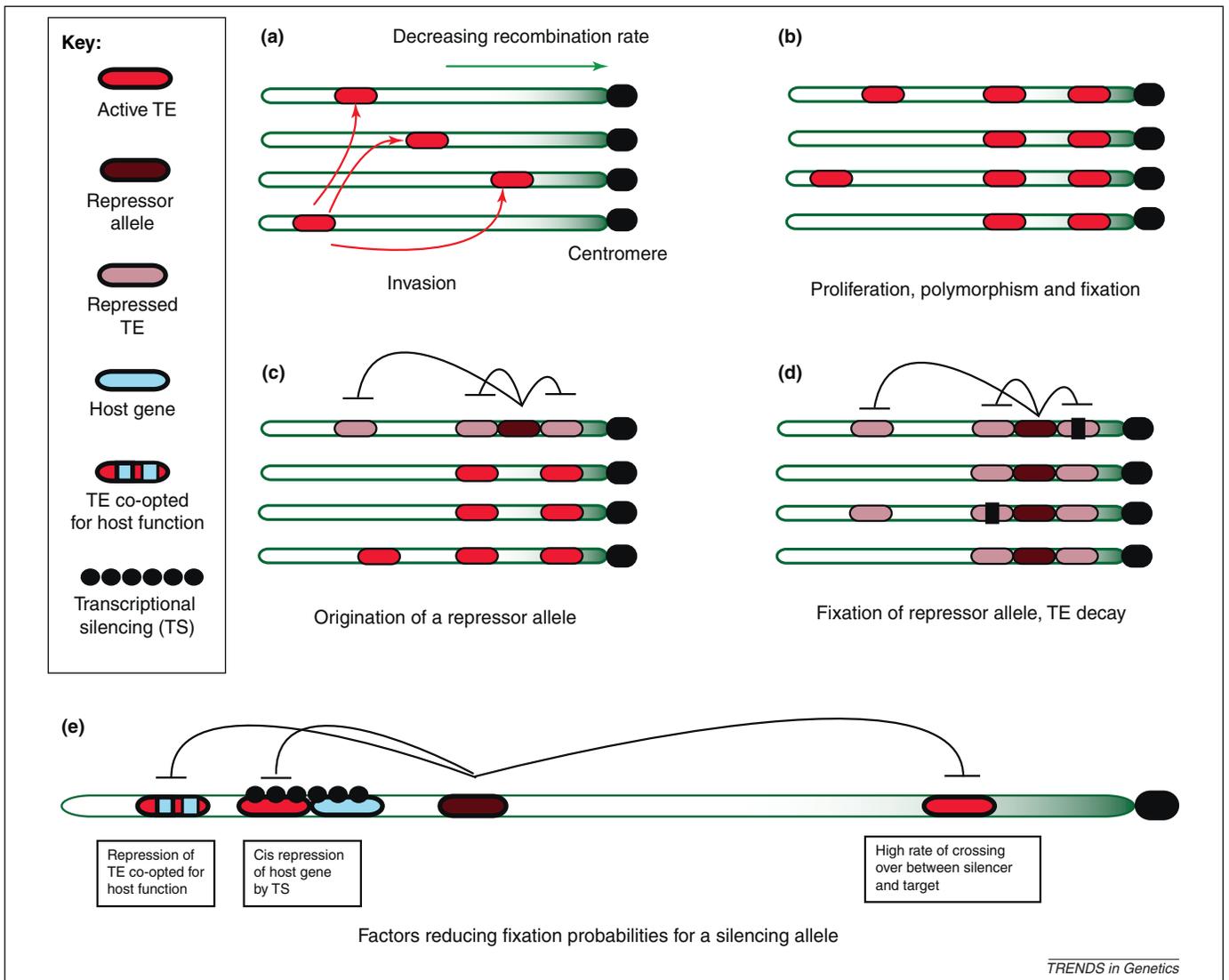
The prediction that host-mediated regulation would evolve more easily than TE self-regulation in outbreeding species is supported by the discovery that host-mediated RNA-based genome defense mechanisms are shared across eukaryotes. Soon after the discovery of RNA interference (RNAi), defined as post-transcriptional gene silencing (PTGS) by double-stranded RNA, a role for RNA-mediated silencing in host-mediated TE regulation was revealed when nematodes defective in RNAi were found to have greater TE activity [34,35]. Since then, diverse modes of RNA-mediated genome defense have been revealed in plants and animals (Box 2). Several core features of

small-RNA-mediated TE silencing are shared across eukaryotes. First, the proliferative nature of TEs makes them prone to insert into the genome in such a way that both sense and antisense transcripts are produced; these can then generate aberrant double-stranded RNA. Second, the RNA from these insertion alleles is recognized by several host mechanisms and is processed into small 21 to 30 nt RNAs and assembled into a complex that includes Argonaute proteins. Third, using these small RNAs as a guide, Argonaute proteins either mediate destruction, translational repression or transcriptional silencing of corresponding complementary RNAs [51] derived from other TE copies elsewhere in the genome. A crucial feature of TE silencing by small RNA is that silencing is mediated by TE insertion alleles themselves – either insertions that spontaneously produce aberrant double-stranded RNA, yielding siRNA, or those that land within distinct loci (Box 2). An example of the former is the *Mu killer* allele in maize, which itself is an altered *MuDR* TE insertion that produces a hairpin transcript and small RNAs that target the silencing of other *MuDR* TEs [52]. Examples of the latter include *gypsy* insertions into the *D. melanogaster flamenco* locus and *P* element insertions into *D. melanogaster* telomeres [45,53]. In both cases, TE insertions into distinct piRNA-generating loci (Box 2) yield a source of piRNAs that silence corresponding TEs throughout the genome.

### Evolutionary dynamics in populations

#### Phases of TE invasion

The discovery of RNA-based genome defense against TE proliferation places our understanding of TE evolutionary dynamics in a new light [54]. We now understand that TE invasions can occur in several distinct phases and this is supported by recent simulation studies of the population dynamics of repressor allele evolution [55] (Figure 1). The stereotypical scenario is likely to proceed in four distinct phases. First, there is an initial phase of TE invasion within a population [56–58]. Second, after invasion, there will be a period of time in which the TE family proliferates within the population. Third, TE insertion alleles that initiate the production of small silencing RNA (Box 2) could appear, increase in frequency, and fix within the population. Fixation takes place due to the fratricidal strategy of small-RNA production, which results in repression of other members of the TE family. Formally, this silencing should be considered host-mediated rather than self-mediated because repression is the property of a TE insertion allele, not the property of a TE lineage. Fourth, a more quiescent state will persist after the silencing allele has fixed within the population. In this time, unless the silencing allele is strong enough to drive transposition rates towards zero with increasing copy number, transposition will continue at a reduced rate and selection against insertions will still be required to stabilize copy number. During this final phase, TE extinction could occur through the accumulation of mutations within TE copies. With transposition rates on the order of  $10^{-5}$  to  $10^{-4}$  per copy per generation [59], and a genomic mutation rate on the order of  $10^{-8}$  per generation [60], a 5 kb TE is expected to experience on the order of one



**Figure 1.** Phases of TE invasion. TE proliferation and fixation of alleles that silence TEs by small RNA are likely to take place in discrete phases. Soon after (a) invasion, a persistent state of (b) proliferation, polymorphism and fixation will be maintained by a balance between transposition, natural selection removing TE copies from the population, and drift that allows mildly deleterious elements to either fix or segregate within the population. Dark green indicates decreasing recombination rate. Selection against ectopic recombination, gene disruption and the Hill–Robertson effect will favor fixation in these regions. (c) A repressor allele can arise either by a structured insertion that drives the formation of dsRNA or by insertion in the proper orientation into an RNA silencing locus. (d) Fixation of the repressor allele will be more favored when closely linked to the targets of repression, but this fixation will also lead to silencing of other unlinked targets. Fixation in a region of low recombination is also favored by the reduced likelihood of participation in ectopic recombination. If transposition rates are reduced to a sufficiently low value, TE insertions will accumulate damaging mutations faster than the transposition rate, leading to TE decay. (e) A silencing allele will not always fix. Some factors that act against fixation are shown.

mutation between transpositions, excluding the mutations that occur during transposition itself [8]. Thus, at low enough transposition rates, mutations can accumulate in a TE family, leading to its extinction within a population. The fixation of alleles that reduce transposition rates is expected to accelerate this process. Importantly, TE families are not expected to always pass through these four phases. The emergence of a host repressor allele is a serendipitous event that might never happen. Even in the circumstance that such an allele arises, it is not guaranteed to fix. In the absence of host repressor alleles, TEs could proliferate at a higher rate until extinction of the TE family occurs.

#### Evolutionary dynamics of repression by small RNA

An important consequence of host TE control is that transposition rates and selection coefficients acting against TE

insertions prior to the fixation of repressor alleles will be different from those in populations that have achieved RNA-mediated TE control. This has important implications for the evolution of TE repression. Previous estimates of transposition rates in *D. melanogaster* are of the order  $10^{-5}$  to  $10^{-4}$  per copy per generation [59]. These low transposition rates were previously used to argue that the selection pressure for host-control is modest (see discussion of these matters in [8,61]). However, these estimates were derived from TEs such as *roo*, which are now known to be under the influence of RNA-mediated silencing [44,62]. In the early invasion phase prior to the origination of *roo* silencing alleles, transposition rates could have been much higher. If true, selection for host control in this phase might be more plausible than previously considered.

Several important factors will influence the dynamics of repressor allele fixation. One of these factors is genetic drift

[6,7,63]. Reduced effective population size has been credited as a crucial factor leading to expansion of TE content in the human lineage due to reduced efficacy of natural selection in purging TEs [64]. Reduced effective population size will also lead to lower probabilities of fixation by natural selection for repressor alleles. A second factor influencing the dynamics of repressor allele fixation is recombination. Increased linkage between repressor alleles and target insertions increases the advantage of repressor alleles [28]. Simulation studies have confirmed that repressor alleles will increase in frequency when the targets of silencing reside on the same chromosome [65]. This role of linkage could contribute to the observation in *D. melanogaster* that piRNA clusters are frequently found in pericentric heterochromatin, where recombination rates are low. TE insertions in these regions could be important targets for driving piRNA repressor alleles to fixation. Finally, like any TE insertion, insertions that yield piRNA in these regions are more likely to fix because they are shielded from participating in ectopic recombination [54,65].

#### *Selection against TE insertions during different phases of invasion*

The magnitude of TE selective effects will change before and after the fixation of a silencing allele. This is suggested by the fact that TE insertions that are the targets of RNA-mediated silencing drift to higher frequencies, as a result of relaxed selective pressure, than those insertions that are not targets of RNA-mediated silencing [65]. Following the fixation of RNA-mediated silencing alleles, destruction of TE coding RNA by PTGS will eliminate the energetic costs of producing TE-encoded proteins [66]. Furthermore, transcriptional silencing of TEs will reduce the energetic costs of producing both TE mRNA and protein. Moreover, if a silencing allele initiates or maintains heterochromatin formation at TE insertion sites, this is expected to reduce ectopic recombination among dispersed homologs. Heterochromatin has a reduced rate of recombination [67] and a host repressor allele that initiates or maintains a heterochromatic state at TE insertions could reduce ectopic recombination within TE families [68]. Ectopic recombination is considered one of the primary forces of selection acting against TE insertions and the fixation of such *trans* acting factors could reduce the harmful effects of individual TE insertions [54,55] of a TE family with a given copy number.

#### *Limits on the evolution of host-repression*

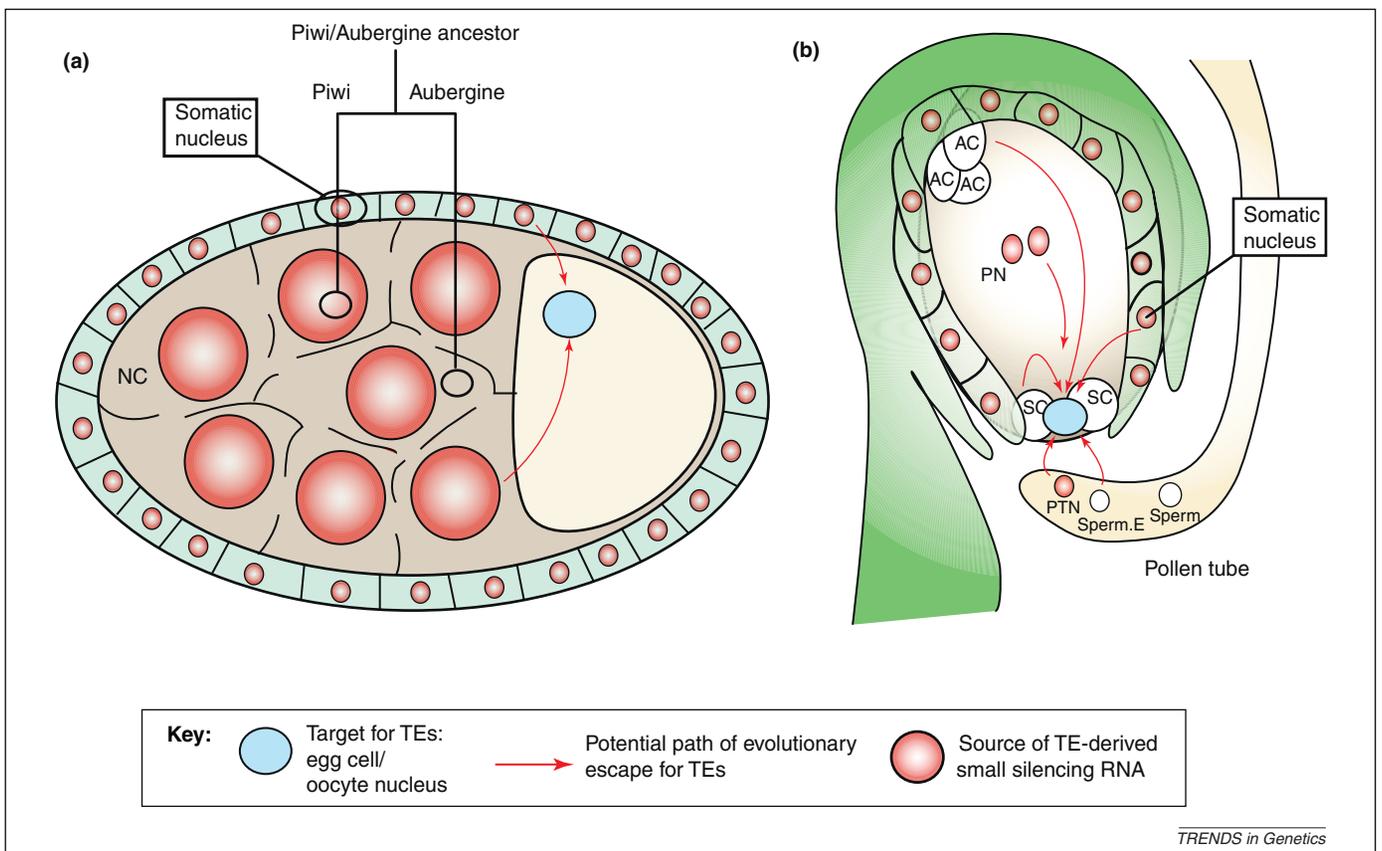
From this perspective, it might appear that natural selection on the part of the host would favor transcriptional silencing over PTGS. However, transcriptional silencing has some inherent negative effects. In particular, the nucleation of silent chromatin formation by small RNA can spread in *cis* [69] and have the harmful effect of silencing neighboring genes. A signature of natural selection against these *cis* effects has been observed in plants – methylated TEs are found further away from genes than unmethylated TEs [70]. In light of these potentially detrimental consequences, there are specialized mechanisms that limit small-RNA-mediated transcriptional silencing

in *trans* and *cis*. For example, the ERI-1 exonuclease in the fission yeast limits transcriptional silencing by siRNA in *trans* [71,72]. In *D. melanogaster*, the piRNA machinery contributes to *cis* transcriptional silencing by the 1360 element, but silencing is observed only in pericentric heterochromatin and not in the euchromatin [73]. This indicates the presence of mechanisms restricting *cis* silencing in gene rich regions of the genome.

Barring damaging *cis* effects, one might expect that natural selection on the part of the host would always favor full TE repression by small RNAs. However, there are some circumstances in which a repressor allele will be selected against. Although TE insertions are generally parasitic and harmful, they can also play a role in host adaptation through shaping the evolution of regulatory networks [74,75]. One example is the imprinted retrotransposon-derived gene *Peg10* [76]. *Peg10* is maternally methylated and paternal expression is necessary for proper placental growth in mice. Such a gene is unlikely to evolve in the presence of alleles that produce piRNA and target DNA methylation of this TE family in the male germline [49,77]. After fixation of *Peg10*, the fixation of a male germline host repressor allele could thus be precluded, allowing the continued proliferation of other members of this retrotransposon family until sufficient divergence accumulates between the TE family and the co-opted TE. This form of ‘hostage taking’ on the part of TE lineages could contribute to their success in some species. Because selection coefficients associated with host-mediated TE repression are not especially high [7], selection favoring a TE-derived regulatory module could outweigh selection for repression of an associated TE lineage.

#### *TEs, RNA-mediated silencing and hybrid sterility*

A final consideration is the role of RNA-mediated silencing and TE activity in hybrid sterility and reproductive isolation. Intraspecific syndromes of hybrid dysgenesis in *D. melanogaster* and *D. virilis* indicate that TE-induced sterility can occur when males that possess a particular TE family fertilize mothers lacking the corresponding silencing RNAs [53,78]. Similarly to hybrid dysgenesis, interspecific crosses between *Arabidopsis thaliana* and *A. arenosa* can lead to activation of paternally inherited TEs [79]. Reproductive barriers might thus be reinforced in populations with reciprocally different TE profiles and silencing alleles. However, despite strong evidence that TEs contribute to a variety of cases of hybrid sterility, it is not clear whether these mechanisms contribute substantially to the formation of long-term reproductive isolation between species. This is because soon after interbreeding both TEs [80] and repressor alleles could quickly invade across populations. In addition, if these dynamics were important during early speciation, we should expect maternal effects on sterility to evolve early, because protective small RNAs are primarily transmitted through the female germline in animals. The genus *Drosophila* has been extensively analyzed with respect to the genetics of speciation and it appears that TE mobilization [81] and maternal effects [82] do not contribute as substantially to isolation as do the hybrid incompatibilities that cause Haldane’s rule [83]. A promising avenue of future research will be



TRENDS in Genetics

**Figure 2.** Evolutionary games in the germline. **(a)** In *D. melanogaster*, fifteen nurse cells contribute maternal resources to the developing egg, but are fated for apoptosis. **(b)** In *A. thaliana*, multiple haploid nuclei contribute to the female gametophyte, but only the egg cell passes genetic material to the next generation. In both plants and animals, flanking somatic tissues also promote gametogenesis but are excluded from subsequent generations. In *D. melanogaster* **(a)** piRNAs are derived from somatic cells and in nurse cells within a structure called the nuage. An ancestral gene duplication has yielded specialization of the piRNA machinery within the developing egg chamber. piRNAs produced by Aubergine function solely within the germline where as piRNAs produced by Piwi function in somatic follicle cells as well. These somatic piRNAs are specialized to defend against retrovirus-like TEs that proliferate in the soma and spread into the germline. In flowering plants **(b)** multiple companion nuclei including somatic nuclei, nuclei fated to the endosperm and nuclei within the pollen tube also contribute TE-derived siRNAs. Red arrows indicate the potential paths of TE evolutionary escape from tissues that will not enter the next generation toward those tissues that do. AC, antipodal cells; NC, nurse cells; PN, polar nuclei; PTN, pollen tube nucleus; SC, synergid cells; SN, somatic nucleus; Sperm.E, sperm fertilizing the central cell, thus establishing the endoderm.

to determine the extent to which TE mobilization drives reproductive isolation in both outbreeding and inbreeding species. High levels of inbreeding could limit the spread of both TEs and silencing alleles so that modes of speciation involving regulation of TE proliferation might be more likely to evolve within inbreeding species.

### Evolutionary games in the germline

Cycles of immune system evasion and immune system adaptation can result in an evolutionary arms race that drives a high rate of adaptive evolution in parasites and host. The interactions between TEs and hosts are analogous and a signature of this evolutionary arms race has been identified in the piRNA machinery of *Drosophila* [84,85]. A second framework for understanding the evolutionary dynamics of TEs in a small RNA world is to consider these evolutionary games within the germline. Of particular interest are the strategies that TE lineages have evolved to proliferate within the germline and the counter-strategies that have evolved on the part of the host to limit this proliferation.

Gametogenesis can establish substantial conflict with TEs due to the varying fates of different germline and somatic components. In animals this is exemplified by

*D. melanogaster* oogenesis [86,87]. In the *D. melanogaster* egg chamber the oocyte nucleus enters the gamete whereas the fifteen nurse cell nuclei undergo rounds of endoreduplication and transcription, provisioning the developing egg chamber with maternal resources. Eventually these nurse cells undergo apoptosis. This mode of oogenesis sets up substantial conflict with TE copies residing in the nurse cells. Natural selection on the part of TE lineages will act strongly for these inserts to copy themselves or transpose from the nurse cells destined for apoptosis and into the oocyte nucleus (Figure 2). For example, I element mRNA transcripts leave the nurse cells [88] and directly target the oocyte [89]. Similar targeting of the oocyte nucleus is also seen by the *Jockey* and *G2* elements [90]. TE proliferation strategies are also not restricted to germline tissues. For example, the retrovirus-like TEs *ZAM* and *gypsy* [91,92] invade the germline from somatic follicle cells flanking the developing egg chamber.

In response to these evolutionary strategies on the part of TE lineages, *D. melanogaster* has evolved counter-strategies to protect the genome. During oogenesis, biogenesis of many piRNAs depends on the nuage, a structure that forms a coating over the nurse cells and is enriched for many components of the piRNA machinery [45,93–98]. Any

TE mRNA to escape the confines of the nurse cell must first pass through the nuage, probably getting 'sliced' by Argonaute-piRNA complexes along the way. Another example of host counter-strategy includes a gene duplication that has produced niche specialization on the part of the piRNA machinery to jointly combat TE proliferation from within the germline and the flanking somatic follicle cells. Three Argonaute paralogs, *aubergine*, *argonaute3* and *piwi*, have distinct functions in the ovary. Aubergine and Argonaute3 proteins function solely in the germline whereas Piwi contributes to somatic silencing of retroviruses such as *gypsy* and *ZAM* [94,99] in the follicle cells. *Piwi* and *aubergine* are most closely related and arose through a more recent gene duplication. This suggests that, as TEs evolve specialized strategies of proliferation, the TE silencing machinery likewise evolves specialized strategies of defense.

As with *D. melanogaster* nurse cells, companion nuclei play crucial roles in germline function in plants (Figure 2). The pollen tube nucleus propagates the pollen tube, and the two sperm nuclei separately fertilize either the central cell to form the endosperm or the egg gamete. During oogenesis the megaspore undergoes three rounds of mitosis to produce an assemblage of nuclei that include the egg cell, the polar nuclei, the synergid cells (which assist fertilization) and the antipodal cells. Similar to *D. melanogaster* nurse cells, these companion cells have an apparent role in genome defense. In the female gametophyte somatic companion cells [100] and endosperm cells [101] function as a source of TE-derived siRNA, although in the case of the endosperm it has not been established that this siRNA functions to repress TEs. In pollen, the pollen tube nucleus reactivates TEs at the transcriptional level through deactivation of DECREASE IN DNA METHYLATION, driving the production of a pool of TE-derived siRNA [102]. These siRNAs appear to limit TE proliferation in the sperm nuclei that contribute to fertilization.

### Concluding remarks

The discovery of genome defense by small RNA has revised our understanding of TE dynamics in populations. Natural selection was once considered the primary force controlling TE proliferation. It now appears that these two modes act jointly. Importantly, unless genome defense by small RNA reduces transposition rates with increasing copy number, natural selection is still required to stabilize copy number within populations. A promising line of further inquiry will be to determine the relative importance of these two factors by performing a joint evolutionary analysis of silencing alleles and TE proliferation. Population genetic analysis of host repressor alleles in wild populations could allow estimation of their selection coefficients. Coupled with a genomic analysis of TE dynamics, one can disentangle the evolutionary history of different phases of TE invasion. In particular, it is conceivable that one could identify correlations between TE proliferation events and the fixation of host repressor alleles. An evolutionary perspective will also provide insight into the many mechanistic questions that remain. How have evolutionary arms races shaped the machinery of RNA-mediated silencing? How have genome defense

strategies in the germline evolved in the context of TE proliferation strategies? The essential function of the germline is to protect genetic information across generations. In sexually reproducing species, TEs pose a tremendous challenge to this function. By considering the complex evolutionary game between host and TEs we will gain many insights in our understanding of genome structure, epigenetic inheritance and the mechanisms of gametogenesis.

### Acknowledgments

This work was funded by the National Science Foundation (grant MCB-1022165) and the University of Kansas. I am very grateful for the excellent comments and suggestions by five anonymous reviewers.

### References

- Hickey, D.A. (1982) Selfish DNA – a sexually-transmitted nuclear parasite. *Genetics* 101, 519–531
- Kazazian, H.H. (2004) Mobile elements: drivers of genome evolution. *Science* 303, 1626–1632
- Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
- Marracci, S. *et al.* (1996) *Gypsy*/*Ty3*-like elements in the genome of the terrestrial salamander *Hydromantes* (Amphibia, Urodela). *J. Mol. Evol.* 43, 584–593
- Aparicio, S. *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297, 1301–1310
- Charlesworth, B. and Charlesworth, D. (1983) The population dynamics of transposable elements. *Genet. Res.* 42, 1–27
- Charlesworth, B. and Langley, C.H. (1989) The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* 23, 251–287
- Nuzhdin, S.V. (1999) Sure facts, speculations, and open questions about the evolution of transposable element copy number. *Genetica* 107, 129–137
- Langley, C.H. *et al.* (1988) On the role of unequal exchange in the containment of transposable element copy number. *Genet. Res.* 52, 223–235
- Hill, W.G. and Robertson, A. (1966) Effect of linkage on limits to artificial selection. *Genet. Res.* 8, 269–294
- Petrov, D.A. *et al.* (2003) Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol. Biol. Evol.* 20, 880–892
- Batzer, M.A. and Deininger, P.L. (2002) *Alu* repeats and human genomic diversity. *Nat. Rev. Genet.* 3, 370–379
- Neafsey, D.E. *et al.* (2004) Different regulatory mechanisms underlie similar transposable element profiles in pufferfish and fruitflies. *Mol. Biol. Evol.* 21, 2310–2318
- Wright, S.I. and Schoen, D.J. (1999) Transposon dynamics and the breeding system. *Genetica* 107, 139–148
- Morgan, M.T. (2001) Transposable element number in mixed mating populations. *Genet. Res.* 77, 261–275
- Wright, S.I. *et al.* (2003) Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Res.* 13, 1897–1903
- Wright, S.I. *et al.* (2001) Population dynamics of an *Ac*-like transposable element in self- and cross-pollinating *Arabidopsis*. *Genetics* 158, 1279–1288
- Dolgin, E.S. *et al.* (2008) Population frequencies of transposable elements in selfing and outcrossing *Caenorhabditis* nematodes. *Genet. Res.* 90, 317–329
- Cutter, A.D. *et al.* (2009) Evolution of the *Caenorhabditis elegans* genome. *Mol. Biol. Evol.* 26, 1199–1234
- Charlesworth, D. and Charlesworth, B. (1995) Transposable elements in inbreeding and outbreeding populations. *Genetics* 140, 415–417
- Kidwell, M.G. and Novy, J.B. (1979) Hybrid dysgenesis in *Drosophila melanogaster*: Sterility resulting from gonadal dysgenesis in the P-M system. *Genetics* 92, 1127–1140
- Engels, W.R. and Preston, C.R. (1979) Hybrid dysgenesis in *Drosophila melanogaster*: the biology of female and male sterility. *Genetics* 92, 161–174
- Yannopoulos, G. *et al.* (1987) *hobo* Is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor *23.5MRF*. *Cell* 49, 487–495

- 24 Bucheton, A. *et al.* (1984) The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I-factor. *Cell* 38, 153–163
- 25 Bingham, P.M. *et al.* (1982) The molecular basis of P-M hybrid dysgenesis: The role of the P element, a P-strain-specific transposon family. *Cell* 29, 995–1004
- 26 Engels, W.R. (1979) Hybrid dysgenesis in *Drosophila melanogaster*: rules of inheritance of female sterility. *Genet. Res.* 33, 219–236
- 27 Marin, L. *et al.* (2000) P-element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric P copy. *Genetics* 155, 1841–1854
- 28 Charlesworth, B. and Langley, C.H. (1986) The evolution of self-regulated transposition of transposable elements. *Genetics* 112, 359–383
- 29 Dolgin, E.S. and Charlesworth, B. (2006) The fate of transposable elements in asexual populations. *Genetics* 174, 817–827
- 30 Slotkin, R.K. *et al.* (2003) *Mu* killer causes the heritable inactivation of the *Mutator* family of transposable elements in *Zea mays*. *Genetics* 165, 781–797
- 31 Prudhomme, N. *et al.* (1995) *Flamenco*, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* 139, 697–711
- 32 Stuart, J.R. *et al.* (2002) Telomeric P elements associated with cytotype regulation of the P transposon family in *Drosophila melanogaster*. *Genetics* 162, 1641–1654
- 33 Ronsseray, S. *et al.* (1991) The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by 2 P copies at cytological site 1a on the X chromosome. *Genetics* 129, 501–512
- 34 Tabara, H. *et al.* (1999) The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99, 123–132
- 35 Ketting, R.F. *et al.* (1999) *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 99, 133–141
- 36 Aravin, A.A. *et al.* (2007) The *Piwi*-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318, 761–764
- 37 Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108
- 38 Malone, C.D. and Hannon, G.J. (2009) Small RNAs as guardians of the genome. *Cell* 136, 656–668
- 39 Zamudio, N. and Bourc'his, D. (2010) Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? *Heredity* 105, 92–104
- 40 Matzke, M. *et al.* (2009) RNA-mediated chromatin-based silencing in plants. *Curr. Opin. Cell Biol.* 21, 367–376
- 41 Lisch, D. (2009) Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* 60, 43–66
- 42 Pikaard, C.S. *et al.* (2008) Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci.* 13, 390–397
- 43 Daxinger, L. *et al.* (2009) A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J.* 28, 48–57
- 44 Vagin, V.V. *et al.* (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313, 320–324
- 45 Brennecke, J. *et al.* (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089–1103
- 46 Sigova, A. *et al.* (2006) Measuring the rates of transcriptional elongation in the female *Drosophila melanogaster* germ line by nuclear run-on. *Cold Spring Harb. Symp. Quant. Biol.* 71, 335–341
- 47 Ma, J. *et al.* (2010) MicroRNA activity is suppressed in mouse oocytes. *Curr. Biol.* 20, 265–270
- 48 Suh, N. *et al.* (2010) MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Curr. Biol.* 20, 271–277
- 49 Aravin, A.A. *et al.* (2008) A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol. Cell* 31, 785–799
- 50 Klenov, M.S. *et al.* (2007) Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. *Nucleic Acids Res.* 35, 5430–5438
- 51 Hock, J. and Meister, G. (2008) The Argonaute protein family. *Genome Biol.* 9
- 52 Slotkin, R.K. *et al.* (2005) Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat. Genet.* 37, 641–644
- 53 Brennecke, J. *et al.* (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322, 1387–1392
- 54 Lee, Y.C.G. and Langley, C.H. (2010) Transposable elements in natural populations of *Drosophila melanogaster*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1219–1228
- 55 Lu, J. and Clark, A.G. (2010) Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Res.* 20, 212–227
- 56 Kidwell, M.G. and Lisch, D.R. (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* 55, 1–24
- 57 Le Rouzic, A. and Capy, P. (2005) The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. *Genetics* 169, 1033–1043
- 58 Brookfield, J.F.Y. (2005) The ecology of the genome – mobile DNA elements and their hosts. *Nat. Rev. Genet.* 6, 128–136
- 59 Nuzhdin, S.V. and Mackay, T.F.C. (1995) The genomic rate of transposable element movement in *Drosophila melanogaster*. *Mol. Biol. Evol.* 12, 180–181
- 60 Haag-Liautard, C. *et al.* (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445, 82–85
- 61 Nuzhdin, S.V. *et al.* (1998) Quantitative genetic analysis of  *copia* retrotransposon activity in inbred *Drosophila melanogaster* lines. *Genetics* 150, 755–766
- 62 Aravin, A.A. *et al.* (2003) The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* 5, 337–350
- 63 Langley, C.H. *et al.* (1983) Transposable elements in Mendelian populations. I. A theory. *Genetics* 104, 457–471
- 64 Gherman, A. *et al.* (2007) Population bottlenecks as a potential major shaping force of human genome architecture. *PLoS Genet.* 3, 1223–1231
- 65 Lu, J. and Clark, A.G. (2010) Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Res.* 20, 212–227
- 66 Nuzhdin, S.V. *et al.* (1996) Positive association between  *copia* transposition rate and copy number in *Drosophila melanogaster*. *Proc. R. Soc. Lond. B Biol. Sci.* 263, 823–831
- 67 Ellermeier, C. *et al.* (2010) RNAi and heterochromatin repress centromeric meiotic recombination. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8701–8705
- 68 Maside, X. *et al.* (2005) Fixation of transposable elements in the *Drosophila melanogaster* genome. *Genet. Res.* 85, 195–203
- 69 Noma, K. *et al.* (2004) RITS acts in *cis* to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat. Genet.* 36, 1174–1180
- 70 Hollister, J.D. and Gaut, B.S. (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 19, 1419–1428
- 71 Iida, T. *et al.* (2006) Conserved ribonuclease, Eri1, negatively regulates heterochromatin assembly in fission yeast. *Curr. Biol.* 16, 1459–1464
- 72 Buhler, M. *et al.* (2006) Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* 125, 873–886
- 73 Haynes, K.A. *et al.* (2006) Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Curr. Biol.* 16, 2222–2227
- 74 Slotkin, R.K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285
- 75 Feschotte, C. (2008) Opinion – transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405
- 76 Ono, R. *et al.* (2006) Deletion of *Peg10*, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* 38, 101–106
- 77 Aravin, A.A. *et al.* (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 316, 744–747
- 78 Blumenstiel, J.P. and Hartl, D.L. (2005) Evidence for maternally transmitted small interfering RNA in the repression of transposition in *Drosophila virilis*. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15965–15970
- 79 Josefsson, C. *et al.* (2006) Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* 16, 1322–1328

- 80 Hurst, G.D.D. and Schilthuizen, M. (1998) Selfish genetic elements and speciation. *Heredity* 80, 2–8
- 81 Coyne, J.A. (1986) Meiotic segregation and male recombination in interspecific hybrids of *Drosophila*. *Genetics* 114, 485–494
- 82 Sawamura, K. (1996) Maternal effect as a cause of exceptions for Haldane's rule. *Genetics* 143, 609–611
- 83 Laurie, C.C. (1997) The weaker sex is heterogametic: 75 years of Haldane's rule. *Genetics* 147, 937–951
- 84 Obbard, D.J. *et al.* (2009) The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 99–115
- 85 Vermaak, D. *et al.* (2005) Positive selection drives the evolution of *rhino*, a member of the Heterochromatin Protein 1 family in *Drosophila*. *PLoS Genet.* 1, 96–108
- 86 Buszczak, M. and Cooley, L. (2000) Eggs to die for: cell death during *Drosophila* oogenesis. *Cell Death Differ.* 7, 1071–1074
- 87 Roth, S. and Lynch, J.A. (2009) Symmetry breaking during *Drosophila* oogenesis. *Cold Spring Harb. Perspect. Biol.* 1, a001891
- 88 Chambeyron, S. *et al.* (2008) piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc. Natl. Acad. Sci. U. S. A.* 105, 14964–14969
- 89 Van De Bor, V. *et al.* (2005) *gurken* and the *I* factor retrotransposon RNAs share common localization signals and machinery. *Dev. Cell* 9, 51–62
- 90 Hamilton, R.S. *et al.* (2009) A bioinformatics search pipeline, RNA2DSearch, identifies RNA localization elements in *Drosophila* retrotransposons. *RNA* 15, 200–207
- 91 Brasset, E. *et al.* (2006) Viral particles of the endogenous retrovirus ZAM from *Drosophila melanogaster* use a pre-existing endosome/exosome pathway for transfer to the oocyte. *Retrovirology* 3, 25
- 92 Chalvet, F. *et al.* (1999) Proviral amplification of the *Gypsy* endogenous retrovirus of *Drosophila melanogaster* involves *env*-independent invasion of the female germline. *EMBO J.* 18, 2659–2669
- 93 Klattenhoff, C. *et al.* (2009) The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138, 1137–1149
- 94 Li, C.J. *et al.* (2009) Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in Flies. *Cell* 137, 509–521
- 95 Findley, S.D. *et al.* (2003) *Maelstrom*, a *Drosophila* spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. *Development* 130, 859–871
- 96 Harris, A.N. and Macdonald, P.M. (2001) *aubergine* encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* 128, 2823–2832
- 97 Lim, A.K. and Kai, T. (2007) Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6714–6719
- 98 Pane, A. *et al.* (2007) *zucchini* and *squash* encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev. Cell* 12, 851–862
- 99 Malone, C.D. *et al.* (2009) Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137, 522–535
- 100 Olmedo-Monfil, V. *et al.* (2010) Control of female gamete formation by a small RNA pathway in *Arabidopsis*. *Nature* 464, 628–632
- 101 Mosher, R.A. *et al.* (2009) Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. *Nature* 460, 283–286
- 102 Slotkin, R.K. *et al.* (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136, 461–472