Long-Term and Short-Term Evolutionary Impacts of Transposable Elements on *Drosophila*

Yuh Chwen G. Lee[1] and Charles H. Langley
Center for Population Biology and Department of Evolution and Ecology, University of California, Davis, California 95616

ABSTRACT Transposable elements (TEs) are considered to be genomic parasites and their interactions with their hosts have been likened to the coevolution between host and other nongenomic, horizontally transferred pathogens. TE families, however, are vertically inherited as integral segments of the nuclear genome. This transmission strategy has been suggested to weaken the selective benefits of host alleles repressing the transposition of specific TE variants. On the other hand, the elevated rates of TE transposition and high incidences of deleterious mutations observed during the rare cases of horizontal transfers of TE families between species could create at least a transient process analogous to the influence of horizontally transmitted pathogens. Here, we formally address this analogy, using empirical and theoretical analysis to specify the mechanism of how host–TE interactions may drive the evolution of host genes. We found that host TE-interacting genes actually have more pervasive evidence of adaptive evolution than immunity genes that interact with nongenomic pathogens in *Drosophila*. Yet, both our theoretical modeling and empirical observations comparing *Drosophila melanogaster* populations before and after the horizontal transfer of *P elements*, which invaded *D. melanogaster* early last century, demonstrated that horizontally transferred TEs have only a limited influence on host TE-interacting genes. We propose that the more prevalent and constant interaction with *multiple* vertically transmitted TE families may instead be the main force driving the fast evolution of TE-interacting genes, which is fundamentally different from the gene-for-gene interaction of host–pathogen coevolution.

HOST–PATHOGEN interactions affect the population dynamics and the evolutionary trajectories of both species. In particular, coevolutionary dynamics will affect the pattern of polymorphism and divergence of genes underlying host–parasite interactions either through an arms race (Van Valen 1973; Dawkins and Krebs 1979) or through balancing selection (Haldane 1949; Hughes *et al*. 1990; Takahata *et al*. 1992; Hughes and Yeager 1998; Rose *et al*. 2004). In either case, accelerated rates of protein evolution and/or recurrent adaptive substitutions are expected in genes engaged in these interactions, which has been observed in both immunity genes (Schlenke and Begun 2003; Jiggins and Kim 2007; Sackton *et al.* 2007; Obbard *et al.* 2009b) and antivirus *siRNA* genes (Obbard *et al*. 2006, 2009a,b, 2011) in *Drosophila*.

doi: 10.1534/genetics.112.145714
Manuscript received May 6, 2012; accepted for publication September 12, 2012
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.145714/-/DC1.
Sequences generated from this article have deposited with the EMBL/GenBank Data Libraries under accession nos. JX974565–JX974572 and KC115855–KC116217.
[1]Corresponding author: Department of Evolution and Ecology, One Shields Ave., University of California, Davis, CA 95616. E-mail: grylee@ucdavis.edu

Transposable elements (TEs) are ubiquitous genomic constituents that increase their copy number (the number of TEs in a host genome) by replicative transposition (copying to new genomic locations). Like *Drosophila*, most host genomes are occupied by multiple TE families, which are defined by sequence similarity (homology) and by replication and transposition mechanisms. Even though incidences of potentially adaptive individual TE insertions with high population frequencies have been reported (Daborn *et al.* 2002; Aminetzach *et al.* 2005; González *et al.* 2008; Schmidt *et al.* 2010), the mutagenic effects of TE insertions are typically deleterious because they disrupt gene structure and function (Finnegan 1992) and can lead to deleterious chromosomal rearrangement (Montgomery *et al.* 1987, 1991; Langley *et al.* 1988). Supporting this, TEs in natural populations of *Drosophila* are generally found in intergenic regions (Aquadro *et al.* 1986; Kaminker *et al.* 2002; Bergman *et al.* 2006) and present at low population frequencies (reviewed in Charlesworth and Langley 1989; Le Rouzic and Deceliere 2005; Lee and Langley 2010). *Drosophila melanogaster* strains with a larger copy number of several TE families surveyed also have lower fitness (Mackay 1989; Pasyukova *et al.* 2004).

Because of these deleterious fitness impacts of TEs on their hosts, the interaction between host and TEs has been suggested to be analogous to an arms race between host and other more familiar pathogens (Kidwell and Lisch 2001; Aravin *et al.* 2007; Siomi *et al.* 2008; Obbard *et al.* 2009a; Blumenstiel 2011). Surprisingly, there has been no systematic comparison of the evolution of host genes interacting with TEs and those interacting with pathogens to examine this analogy. Furthermore, no specific models for the evolutionary impact of antagonistic interactions between host and TEs on protein evolution have been analyzed.

Due to the horizontal transmission of pathogens, there can be strong associations between specific host alleles and the level of pathogen load, leading to large selective benefits for host alleles that can effectively suppress pathogen infections. On the contrary, TE families are inherited vertically as part of the host parental genome. Vertical inheritance of TE insertions, random mating, and recombination lead to weak associations between selectively favored, TE-suppressive host variants and any net reduction in TE replication. Indeed, in organisms with random mating and weak linkage disequilibrium (*e.g.*, *Drosophila*), a host variant suppressing a single TE family will enjoy only weak adaptive advantage (Charlesworth and Langley 1986). If the suppression of TE transposition involves specific associations between host alleles and TE variants, this is unlikely to drive the fast evolution of host genes involved in TE suppression.

In rare cases, TEs are observed to have been horizontally transferred between host species (reviewed by Silva *et al.* 2004; Loreto *et al.* 2008; Schaack *et al.* 2010). In these cases, the evolutionary impacts of TEs on the host may be more analogous to those of nongenomic pathogens. During the spread of a newly invaded TE family, the transposition rate can be exceptionally high (Kidwell *et al.* 1988; Good *et al.* 1989; reviewed in Silva *et al.* 2004). This is usually associated with a high incidence of deleterious insertions and sterilities, imposing a strong selective pressure on the host. However, the spread of horizontal transfer of TE families to all genomes of the new host species is found to happen on a short timescale, generally within thousands of generations (Daniels *et al.* 1990; Simmons 1992). After the initial horizontal transfer, copies of the new TE family appear to follow the typical TE mode of vertical transmission. A quantitative analysis of the hypothesis that horizontal transferred TE families are able to impose strong enough selection to elicit specific adaptive evolution in host genes remains unaddressed.

One of the best-studied examples of a TE horizontal transfer is the *P element* (reviewed in Engels 1997; Rio 2002), which invaded *D. melanogaster* from the distantly related *D. willistoni* <100 years ago (Brookfield *et al.* 1984; Anxolabéhère *et al.* 1988; Daniels *et al.* 1990; Clark *et al.* 1994). While virtually all more recently surveyed *D. melanogaster* populations have *P elements*, *D. melanogaster* strains collected early last century, and maintained in the laboratory since that time, are still free of *P elements* (Kidwell *et al.* 1983; Anxolabéhère *et al.* 1988). Matings between females lacking *P elements* (the M strains, said to have *M cytotype*) and males with *P elements* (the P strains, said to have *P cytotype*) result in progeny with "hybrid dysgenesis" syndrome, which consists of high rates of male recombination, sterility, mutation, and chromosomal rearrangement (Hiraizumi 1971; Kidwell *et al.* 1973, 1977; Engels 1979). These dysgenic effects are attributed to the high rate of *P-element* transposition (Bingham *et al.* 1982; Rubin *et al.* 1982). *P elements* of *D. melanogaster* thus provide a rare opportunity to investigate the evolutionary impacts of horizontally transferred TE families on the host genes interacting with them.

Our systematic comparisons demonstrate that the molecular evolution of TE-interacting genes exhibits comparable evidence of recurrent adaptive fixations to that of genes mediating the interactions between *Drosophila* and horizontally transferred pathogens. We took two approaches to investigate whether horizontally transferred TEs have discernable selective impacts and can contribute to the observed long-term adaptive evolution of TE-interacting genes. We developed and analyzed a model of TE invasion after horizontal transfer and we used the recently invaded *P element* as a system to empirically contrast the genetic differentiation of candidate host genes between *D. melanogaster* populations before and after *P-element* invasions. Both our analytical modeling and empirical observations suggest that the selective pressure imposed by horizontally transferred TE families is limited. We proposed a hypothesis other than the gene-for-gene host–TE coevolutionary model to explain our observations.

Materials and Methods

D. melanogaster variation of post–P-element invasion population and Drosophila divergence data

We used the release 1.0 assembly of 44 *D. melanogaster* genome sequences generated by the *Drosophila* Population Genomic Project [DPGP, www.dpgp.org (Langley *et al.* 2012)]. The DPGP data consist of 7 strains from Malawi, Africa, and 37 strains from North Carolina, which were all collected after *P-element* invasions and have *P elements*. Coding regions of each candidate gene were defined according to the *D. melanogaster* reference genome annotation (version 5.16) and parsed out from the above genomic sequences, using Perl scripts. Bases with quality scores <30 and regions that appeared as identical by descent (IBD) or exhibited residual heterozygosity (Langley *et al.* 2012) were treated as missing data. We also removed alleles when >50% of the bases were missing data. To compare the evolution of candidate genes with immunity genes, we used 236 immunity genes included in Sackton *et al.* (2007) and follow the categorization of the immunity genes of this previous study. Retrieval of coding sequences and population genetics analysis (see below) of these immunity genes were the same as those of candidate genes.

D. simulans (Begun *et al.* 2007) and *D. yakuba* (Clark *et al.* 2007) alleles were retrieved according to *D. melanogaster* coordinates from the DPGP multispecies alignment,

which includes *D. melanogaster*, *D. simulans*, *D. yakuba*, and *D. erecta* genomes (Langley *et al.* 2012). When a *D. simulans* allele was used as an outgroup in statistical inferences (see below), we chose the allele with the smallest proportion of missing data and alignment gaps (or highest base coverage) among the mosaic *D. simulans* genome (Clark *et al.* 2007) and six *D. simulans* genomes (Begun *et al.* 2007). For *Ago3*, all *D. simulans* alleles from the above seven genomes had low coverage. We used *D. melanogaster* exon sequences of *Ago3* to blast against the trace reads generated from the *D. simulans* population genomics project (Begun *et al.* 2007) and assembled retrieved reads, using the codoncode aligner (http://www.codoncode.com/aligner/). We removed reads whose alignment outside exons was incongruent with the majority of other reads. Consensus sequence was called if there were at least three reads covering the region.

Population genetics and molecular evolution analysis of candidate genes

π was estimated as average pairwise differences (Nei 1987). Lineage-specific divergences were estimated by maximum likelihood, using PAML version 4 (Yang 2007) on the branch leading to *D. melanogaster* and *D. simulans*, with *D. yakuba* as the outgroup. Genes with <100 sites included in the PAML analysis or with a d_S value <0.0001 were excluded. We used both *D. melanogaster* and *D. simulans* within-species polymorphism to carry out McDonald–Kreitman tests [two-species McDonald–Kreitman (MK) test (McDonald and Kreitman 1991)]. Codons having more than two states within species were removed. Codons that are both polymorphic within species and divergent between species were counted as both polymorphism and divergence. We used the mutational path minimizing the number of nonsynonymous differences. *P*-values of MK tests were determined by Fisher's exact tests (FET). For genes without *D. simulans* variation (*Ago3* and *mael*), we carried out one-species MK tests, using *D. melanogaster* polymorphism and the *D. simulans* allele with highest base coverage to count the number of fixations. For candidate genes with significant MK test results, we used Pfam (Finn *et al.* 2009) with an *E*-value cutoff of 10^{-5} to annotate known domains and perform a MK test on each annotated domain. We estimated average α (the proportion of amino acid fixations driven by positive selection) for different classes of genes, using Welch (2006) with default parameters. We also used the likelihood-ratio test to investigate whether a single-α or a two-α model better fits the data when we included both candidate and immunity/all genes in the analysis, testing whether there are differences in α between classes of genes.

When comparing population genetic estimates or statistics of candidate genes with genome-wide distribution, we used a conservative gene set used by DPGP (Langley *et al.* 2012), which consists of genes whose *D. melanogaster* alleles of DPGP data and outgroup alleles all have the same gene model as the reference annotations (canonical initiation codon, splice junction, and termination codon).

D. melanogaster variation data before P-element invasions

Variation data from *pre–P-element D. melanogaster* populations were collected by resequencing the coding regions of candidate genes from laboratory-maintained strains collected before the 1960s and previously identified as M strains (Kidwell *et al.* 1983). PCR with primers amplifying the second exon of *P-element* transposase (O'Hare and Rubin 1983; Clark *et al.* 1998) was used to confirm the absence of *P elements*. To have comparable sampling locations to those of the DPGP data, we first used four African strains (CA1, KSA2, KSA3, and KSA4) and four North and South Carolina strains (Wild 10E, Wild 11A, Wild 11C, and Wild 11D) in the initial survey for unusual temporal differentiation. Five candidate genes (*Irbp*, *squ*, *Spn-E*, *Krimp*, and *Hen1*) showed significant differentiation between alleles from the above eight strains and post–*P-element* alleles from DPGP data. Additional alleles were then collected on these five genes, using 4 Asian, 5 European, 1 South American, and 11 North American strains. Details of *D. melanogaster* strains used in this study can be found in supporting information, File S1, Table S1. For control genes near *Hen1*, we sequenced only the 15 North American M strains.

Despite exhaustive efforts to locate M strains collected before *P-element* horizontal transfer, the available pre–*P-element* strains are far from ideal. Within North America, where there is the largest set of pre–*P-element* strains, spatial locations are disperse: strains were collected on the West and the East Coast as well as in the northern and southern latitudes. Latitudinal clines for various loci have been observed in *D. melanogaster* (reviewed in Schmidt *et al.* 2005; Hoffmann and Weeks 2007). Unfortunately, this may increase the possibility of falsely concluding that there is temporal genetic differentiation while the actual difference would be a result of the geographic heterogeneity of between-time samples. Accordingly, we also examined other aspects of the data (heterozygosity and haplotypes) in addition to temporal differentiation and included control genes near candidate genes showing strong temporal differentiation before drawing conclusions (see below).

DNA samples were prepared from 30 flies from each *D. melanogaster* M strain. PCR and sequencing primers for coding regions of candidate genes were designed using the Primer3 program (Rozen and Skaletsky 2000) and the *D. melanogaster* reference genome. PCR products were purified and sequenced directly. Most of the *D. melanogaster* strains used in this study have been maintained in the laboratory for >50 years and are highly inbred. For targeted regions with residual heterozygosity within lines, PCR products were cloned with TOPO-TA cloning (Invitrogen, Carlsbad, CA) and one clone of each PCR product was sequenced.

Analysis of temporal differentiation between pre-P and post–P-element populations

Sequences of *pre–P-* and *post–P-element* populations were aligned using ClustalW (Chenna *et al.* 2003), followed by

manual curation. We estimated F_{st} according to Weir and Cockerham (1984) and used permutations to determine the *P*-values (Hudson *et al.* 1992). To further test for unusual haplotypic structures, we used methods based on the frequency of major haplotypes (Hudson *et al.* 1994), the number of haplotypes, and the heterozygosity of haplotypes (Depaulis and Veuille 1998) and used coalescent simulation without recombination to determine the *P*-values (Hudson 2002). Although these three haplotype-based tests are related conceptually, their power to detect deviations from the same null hypothesis varies with the alternatives and thus is not fully redundant (Depaulis and Veuille 1998). It is worth noting that the significance of haplotype tests was based on coalescent simulations *without* recombination, which is especially conservative, and our observation of strong evidence for a 20-kb haplotypic structure around *Hen1* is highly unusual (see *Results*).

For analysis of the upstream and downstream regions of *Hen1*, we used a sliding window of size 5 kb incremented every 100 bp to depict the divergence between *D. melanogaster* and *D. simulans*, the polymorphism within *D. simulans*, and the polymorphism of *D. melanogaster* African and North American populations separately.

Analytical model for the dynamics of host alleles that can reduce transposition during the spread of an invading transposable element such as the P element

We considered a panmictic population of diploid hosts with infinite population size and initially devoid of the invading TE. After invasion of the TE, each host genome carries a number ($n \geq 0$) of TEs and zero, one, or two suppressive alleles at the host locus of interest. We assumed that there is complete linkage equilibrium among the invading TEs, and the TEs and the host resistance locus. The low frequency of virtually all TE insertions in natural populations of *D. melanogaster* (Aquadro *et al.* 1986; Montgomery *et al.* 1987; Charlesworth and Langley 1989; Lee and Langley 2010) coupled with the small scale of linkage disequilibrium between SNPs with more intermediate frequency in *D. melanogaster* (Miyashita and Langley 1988; Long *et al.* 1998; Langley *et al.* 2000, 2012) ensures that the magnitude of linkage disequilibrium among elements is small and our assumption is reasonable. The assumption of no linkage disequilibrium and low TE frequencies motivates the further modeling of distribution of TE copy number as Poisson. The use of the Poisson distribution of TE copy number among individuals of a population has been developed as an approximation and successfully applied in theoretical analyses of TEs (Charlesworth and Charlesworth 1983; Langley *et al.* 1983). As mentioned above, it has an empirical basis in studies of specific TE families and surveys of genomic variation (reviewed in Charlesworth and Langley 1989). The transposition of TEs was also modeled following a Poisson process.

We considered two aspects of the deleterious effects of TEs on host fitness with some details of the model based on the specific biology of *P elements*. According to previous theoretical analysis (Charlesworth and Charlesworth 1983), to have stable containment of transposable elements, the logarithm of fitness must decline more rapidly than linearly with average copy number. We considered a synergistic epistasis for the deleterious effects of TE insertions described previously (Dolgin and Charlesworth 2006, 2008) and the fitness of an individual with n copies of a *P element* is

$$w(n) = e^{-an-bn^2/2}.$$

a and b were chosen as 10^{-5} and 10^{-6}, respectively (see *Appendix*). The other deleterious fitness effect is caused when the transposition of *P elements* generates more double-stranded breaks than the host recombination repair machinery can efficiently repair, leading to the reported reduced fertility (reviewed in Rio 2002). The *P element* transposes through a cut-and-paste mechanism and the process starts with a double-stranded break generated at the original *P-element* insertion (donor site). Approximately 85% of the double-stranded breaks at the donor sites are repaired using the sister chromatid as the template (Engels *et al.* 1990), resulting in regeneration of a *P element* at the donor site and an increase in copy number by one. With the assumption that every *P-element* transposition leads to a net gain of one *P-element* copy, we used a truncation selection model: an offspring with more than n_{HD} (the maximum number of new TE transpositions a host can tolerate before having hybrid dysgenic syndrome) new *P-element* insertions is sterile. The mean fitness of offspring of parents with an average of m *P-element* copies is

$$\bar{w}(m,u) = \sum_{n=0}^{\infty} \frac{e^{-m}m^n}{n!} \left(\sum_{i=0}^{n_{HD}} \frac{e^{-nu}(nu)^i}{i!} e^{-a(n+i)-b(n+i)^2/2} \right).$$

u, the transposition rate per copy per generation, changes according to parental cytotypes, with the *P-element* transposition rate in the $M \times P$ dysgenic cross (u_0) being much higher than that in other crosses (u_1). Individuals with *P elements* are set to have a *P cytotype* and others without are set to have an *M cytotype*.

One of the two segregating alleles of the host locus is able to reduce the *P-element* transposition rate by a proportion d in homozygotes [*i.e.*, the transposition rate is then $u(1 - d)$ in homozygotes]. This allele is referred to as "beneficial allele." The heterozygotic effect of this allele is h of that of homozygotes and the transposition rate in heterozygotes is thus $u(1 - hd)$. We were interested in three aspects of the host population that changed over generations: (1) the proportion of *P cytotype*, r; (2) the allele frequency of the host allele reducing *P-element* transposition (the beneficial allele), l; and (3) the average copy number of *P elements* among individuals with *P cytotype*, μ. With the assumption that there is linkage equilibrium between *P-element* insertions and the host locus, the reduction of the *P-element* transposition rate due to the host locus can be considered

Table 1 Information of candidate genes

Category	Flybase ID	Symbol	Gene name	Chr[a]	CDs length[b]	Gene functions
piRNA genes	FBgn0004872	*piwi*	*piwi*	2L	2532	*piRNA* generation
	FBgn0000146	*aub*	*aubergine*	2L	2601	*piRNA* generation
	FBgn0250816	*AGO3*	*Argonaute 3*	3L	2604	*piRNA* generation
	FBgn0041164	*armi*	*armitage*	3L	3825	*RNA* helicase
	FBgn0003483	*spn-E*	*spindle E*	3R	4305	*RNA* helicase
	FBgn0002652	*squ*	*squash*	2L	726	Nuclease
	FBgn0261266	*zuc*	*zucchini*	2L	762	Nuclease
	FBgn0033686	*Hen1*	*PIMET/DmHen1*	2R	1176	*piRNA* methyltransferase
	FBgn0016034	*mael*	*maelstrom*	3L	1389	Nuage component
	FBgn0034098	*krimp*	*krimper*	2R	2241	Nuage component
	FBgn0262526	*vas*	*vasa*	2L	661	Nuage component
	FBgn0004400	*rhi*	*rhino*	2R	1257	HP1 paralog, heterochromatin binding
P-element–specific genes	FBgn0014870	*Psi*	*P-element* somatic inhibitor	2R	2394	*P-element* somatic inhibitor; *mRNA* splicing factor
	FBgn0004838	*Hrb27C*	Heterogeneous nuclear ribonucleoprotein at 27C	2L	1266	*mRNA* splicing factor
	FBgn0011774	*Irbp*	Inverted repeat-binding protein	3R	1897	Recombination repair protein
	FBgn0041627	*Ku80*	*Ku80*	2L	2100	Recombination repair protein

[a] chromosome
[b] length of coding regions (bp)

as an independent event from the type of cross. The mean fitness of offspring of a specific cross with m parental *P-element* copies is

$$\bar{w}_{\text{cross}}(m,u,l) = l^2\bar{w}(m,u(1-d)) + 2l(1-l)\bar{w}(m,u(1-hd)) + (1-l)^2\bar{w}(m,u).$$

Based on this formula, we derived equations for r, l, and μ, which can be found in the *Appendix*. We calculated r, l, and μ for 1000–10,000 generations. For most cases, we reported the result for 1000 generations, the approximate number of *D. melanogaster* generations since the M strains were first collected. At generation zero, we set r_0 and $l_0 = 10^{-3}$. Parameters without a significant impact on the dynamics of l are set as constant values ($d = 0.5$, $h = 0.5$, $u_1 = 10^{-4}$, and $\mu_0 = 10$; see *Appendix* for results of all parameters tested). u_0 was tested for 10^{-1} and 1. n_{HD} was tested for 2, 3, 5, 7, and 10.

Results

Candidate genes

We took a candidate gene approach and focused on two groups of genes (Table 1). The first group consists of genes known to be involved in the *piwi-RNA* (*piRNA*) biogenesis [hereafter termed *piRNA* genes (reviewed in Klattenhoff and Theurkauf 2008)]. *piRNA* is a class of small RNAs that has been implicated in TE transposition rate regulation. Generation of *piRNAs* is disrupted in *piRNA* gene mutants (Aravin *et al.* 2004; Lim and Kai 2007; Pane *et al.* 2007 Klattenhoff *et al.* 2009; Li *et al.* 2009), leading to elevated expression levels of >10 TE families (Aravin *et al.* 2004; Vagin *et al.* 2004; Lim and Kai 2007; Pane *et al.* 2007; Klattenhoff *et al.* 2009; Li *et al.* 2009; Lu and Clark 2010). We also included *vasa* (*vas*), whose mutant phenotypes also include *piRNA* generation disruptions and elevated transcription of several TE families (Vagin *et al.* 2004; Lim and Kai 2007). *piRNAs* corresponding to *P-element* sequences have been observed in *P* strains but not in *M* strains, suggesting *P elements* are a common target of the *piRNA* pathway (Brennecke *et al.* 2008).

The second group of candidates contains genes known to interact with *P elements* via other pathways. The double-stranded breaks left after *P-element* transpositions in the germline are repaired by host recombination–repair machinery of heterodimers formed by *Irbp* and *Ku80* (Rio and Rubin 1988; Beall *et al.* 1994). Double-stranded breaks generated by other DNA-based TEs may be repaired using a similar mechanism. Splicing factor *Psi* is shown to specifically bind to the 5′-splice site of the *P-element* third intron, suppressing the proper splicing of *mRNA* of *P-element* transposase and thereby repressing the somatic transposition of *P elements* (Siebel and Rio 1990; Siebel *et al.* 1992, 1994, 1995; Adams *et al.* 1997). *Hrb27C* is a nuclear protein that forms an *mRNA* splicing complex with *Psi* (Siebel *et al.* 1992, 1994).

Genes interacting with transposable elements show evidence of positive selection

Recurrent directional selection can lead to an accelerated rate of protein divergence relative to synonymous site divergence. We used maximum-likelihood methods (Yang 2007) to estimate the d_N/d_S ratios on both the *D. melanogaster* and the *D. simulans* branches with *D. yakuba* as an outgroup. We then rank d_N/d_S estimates of candidate

Table 2 Linage-specific relative rates of protein evolution

	mel		*sim*	
Gene	d_N/d_S	Percentile[a]	d_N/d_S	Percentile[a]
AGO3	0.261	12.67	0.265	18.79
armi	0.297	10.22	0.391	10.88
aub	**0.382**	**6.60**	**0.530**	**6.31**
Hen1	0.246	13.64	0.266	18.66
Hrb27C	0.176	21.99	0.019	82.02
Irbp	0.101	38.42	0.164	32.83
krimp	**0.471**	**4.71**	**0.964**	**2.09**
Ku80	0.181	21.28	0.245	20.80
mael	**0.902**	**1.05**	**0.491**	**7.17**
piwi	0.081	45.87	0.194	27.71
Psi	0.091	42.12	0.027	78.55
rhi	**1.415**	**0.31**	**0.508**	**6.77**
spn-E	0.219	16.28	0.281	17.42
squ	0.240	14.25	0.338	13.72
vas	0.243	14.05	0.306	15.63
zuc	**0.359**	**7.48**	0.207	25.73

[a] d_N/d_S for each candidate gene was ranked among 9172 (*D. melanogaster*) and 9051 (*D. simulans*) genes that have PAML results. Candidate genes that ranked among the top 10% among all the genes with PAML results are in boldface type.

Table 3 McDonald–Kreitman test on candidate genes

	Two-species MK test[a]		*mel* MK test[b]	
Gene	No. codons	*P*-value	No. codons	*P*-value
AGO3	NA	NA[c]	689	0.351
armi	1186	<**0.001**	1231	<**0.001**
aub	835	<**0.001**	855	<**0.001**
Hen1	389	0.837	391	1.000
Hrb27C	421	0.095	421	0.532
Irbp	628	0.189	628	0.069
krimp	712	**0.020**	690	**0.001**
Ku80	694	**0.012**	699	0.156
mael	NA	NA	459	0.390
piwi	837	0.732	843	0.267
Psi	796	0.189	797	1.000
rhi	413	0.396	413	0.535
spn-E	1427	**0.021**	1433	0.731
squ	130	0.493	233	1.000
vas	634	**0.022**	639	0.151
zuc	253	1.000	253	1.000

MK tests with significant *p-values* and positive α are in bold type.
[a] MK tests using both *D. melanogaster* and *D. simulans* polymorphism (see *Materials and Methods*).
[b] MK tests using only *D. melanogaster* polymorphism.
[c] Not available due to lack of *D. simulans* polymorphism data.

genes among other *D. melanogaster* annotated genes (see Table 2 for d_N/d_S ratio and Table S2 for separate d_N and d_S values). Consistent with a previous report (Vermaak *et al.* 2005), *rhi* has a d_N/d_S ratio that is larger than one and is among the fastest-evolving genes in *D. melanogaster* (d_N/d_S = 1.415, rank = 0.31% genome-wide). Two nuage component genes, *krimp* and *mael*, both rank in the top 5% genome-wide while *aub* and *zuc* are among top 10%. Estimates of the d_N/d_S ratios on the *D. simulans* branch, except for that of *zuc*, gave similar results. Our inferences are generally in agreement with previous reports that used a branch-site model to detect recurrent amino acid substitutions on the phylogeny of 12 *Drosophila* species (Heger and Ponting 2007; Kolaczkowski *et al.* 2011). Even though *Spn-E* was identified as the *RNA interference* gene showing the most extensive signal of recurrent adaptation on the phylogenetic tree (Heger and Ponting 2007; Kolaczkowski *et al.* 2011), the branch leading to *D. melanogaster* was not significant (Kolaczkowski *et al.* 2011). The differences between studies are not surprising given the fundamental differences in methodology (comparing relative rates of amino acid substitutions of entire coding sequences among all genes *vs.* identification of a subset of sites or branches that recurrently substituted across the phylogeny).

We used the MK test (McDonald and Kreitman 1991) to detect genes whose evolution does not follow the neutral model of evolution. Rejection of the null hypothesis due to the presence of more than the expected number of amino acid fixations has been interpreted as evidence supporting adaptive protein evolution. To have greater statistical power, we considered polymorphisms from both *D. melanogaster* (Langley *et al.* 2012) and *D. simulans* (Begun *et al.* 2007). We identified *aub* and *armi* as significant while *spn-E*, *krimp*, *vas*, and *Ku80* were marginally significant (Table 3). All of these genes rejected the null hypothesis of neutral evolution in the direction of an excess of amino acid divergence (Table 3), consistent with a history of positive selection acting on these genes. To further localize protein domains showing signals of positive selection, we performed MK tests on Pfam-annotated domains. We found the PAZ domain of *aub* (*P*-value = 0.011) and the DEAD domain of *vas* (*P*-value = 0.0007) showed significant enrichment of amino acid substitutions. When considering only within *D. melanogaster* polymorphism, we did not find evidence of adaptive evolution for *spn-E*, *vas*, and *Ku80* (Table 3), perhaps due to a generally lower level of variation in *D. melanogaster* than in *D. simulans* (Aquadro *et al.* 1988; Andolfatto 2001; Andolfatto *et al.* 2011) and thus lower statistical power.

Genes interacting with transposable elements show more prevalent evidence of positive selection than immunity genes

The fundamental differences in the mechanism of transmission between TE families and other nongenomic pathogens raised the question regarding the relative intensity of evolutionary impacts they imposed on hosts. To address this question, we compared the proportion of genes exhibiting evidence of positive selection between our candidate genes and immunity genes, using the same population genetic and molecular evolution analysis.

We found 5 of 12 *piRNA* genes (41.67%) have *D. melanogaster* d_N/d_S estimates among the top 10% genome-wide, which is significantly greater than that of immunity genes (24 of 214 genes, 11.21%, FET P = 0.01, Figure 1A and Table S2). Studies have found that the rates of adaptive evolution vary with respect to the function of immunity genes (Sackton *et al.* 2007; Obbard *et al.* 2009b; reviewed in Lazzaro 2008). We categorized immunity genes into "recognition," "signaling," and "effector" categories and still found

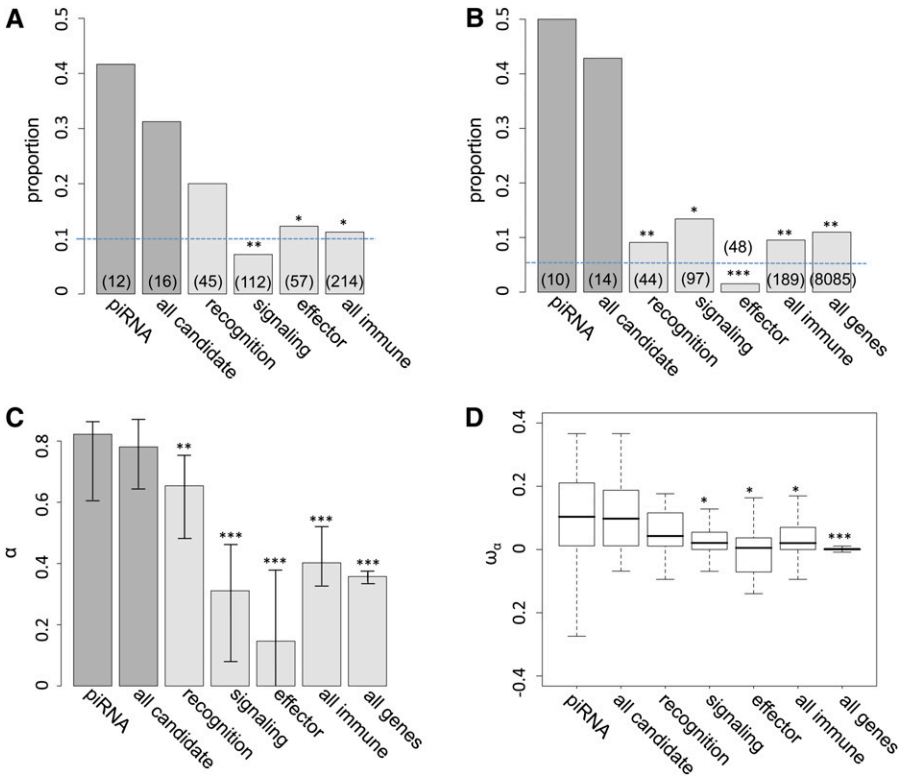

Figure 1 Proportion of candidate and immunity genes showing evidence of positive selection. (A) Proportions of candidate and immunity genes having *D. melanogaster* d_N/d_S among the top 10% genome-wide. (B) Proportions of candidate genes, immunity genes, and all genes having significant two-species MK tests (*P*-value < 0.05) and positive α. Dashed lines are the expectations assuming uniformity. The Number of genes with MK test and PAML results in each category is shown in parentheses. (C and D) Maximum-likelihood estimates of averaged α (C) and boxplots (25th, 50th, and 75th percentiles) of estimated ω_α (D) for different classes of genes. Error bars represent the 95% bootstrapping intervals around each estimate. Significant comparisons between *piRNA* genes and other classes of genes are denoted by * (*P*-value <0.05), ** (*P*-value <0.01), and *** (*P*-value <0.001). Comparisons of proportions (A and B) were based on Fisher's exact test, comparisons of maximum-likelihood estimated α (C) were based on permutations, and comparisons of ω_α (D) were based on a Mann–Whitney *U*-test.

piRNA genes have a higher proportion of fast-evolving genes than all categories of immunity genes. However, we found statistical significance only when comparing *piRNA* genes to either signaling or effector categories (Figure 1A). Comparisons considering all candidate genes (both *piRNA* genes and genes known to interact with *P elements*; Figure 1A) or focusing on relative rates of amino acid evolution on the *D. simulans* lineage (data not shown) are consistent with our findings using *piRNA* pathway gene evolution along the *D. melanogaster* lineage.

We observed even more dramatic enrichment in the proportion of genes showing evidence of recurrent adaptive protein evolution (rejection of MK tests with overabundant amino acid fixation) of *piRNA* genes (5 of 10 genes, 50%) than that of immunity genes (18 of 180 genes, 9.52%, FET *P*-values = 0.002; Figure 1B) and the genome-wide proportion (888 of 8085 genes, 10.98%, FET *P*-values = 0.003). Signaling genes have the largest proportion of genes showing adaptive evolution among three immunity gene categories (13 of 97 genes, 10.92%), which is still significantly lower than that of *piRNA* genes or all candidate genes (FET *P*-values = 0.012). We found a consistent pattern when including all candidate genes in the comparisons.

We used the maximum-likelihood method proposed by Welch (2006) to formally test whether averaged α (the proportion of amino acid substitution fixed by positive selection) is different between classes of genes. When comparing either "*piRNA vs.* immunity genes" or "*piRNA vs.* all genes," we found a two-α model consistently fitted the data better [$\Lambda = 446.46$ (*vs.* immunity genes) and 613.82 (*vs.* all genes); $P < 0.001$ for both comparisons]. Comparisons between *piRNA* genes and a specific subset of immunity genes were also highly significant ($P < 0.001$). Permutation analysis also found the maximum-likelihood estimated α of *piRNA* genes ($\alpha = 0.82$) is significantly higher than that of immunity genes ($\alpha = 0.40$ for all immunity genes) and all genes ($\alpha = 0.36$), except for recognition immunity genes (Figure 1C). Again, comparisons considering all candidate genes gave consistent results. Another estimate of adaptive protein evolution ω_α, the rate of adaptive substitution relative to the rate of neutral substitutions (Gossmann *et al.* 2010), supports the same conclusion based on α (Figure 1D). Either *piRNA* genes or all candidate genes have larger ω_α than immunity genes except for recognition immunity genes (Mann–Whitney *U*-test, $P < 0.05$) and all genes (Mann–Whitney *U*-test, $P < 0.001$), suggesting that the larger α of TE-interacting genes is not due to differences in proportion of effectively neutral mutations.

Horizontal transfer of TE families does not impose enduringly strong selection on host beneficial variants

We used a deterministic model to analyze the dynamics of host alleles that can reduce the transposition rate of a newly horizontally transferred TE family by a fixed proportion (referred to as "beneficial allele") during the spread of that TE family in a panmictic host population (see *Materials and Methods* and *Appendix* for model details). We specifically considered the well-documented horizontal transfer of *P elements*, which provided the biological context needed to specify details of the model. The model considered both epistatic selections against increases in *P-element* copy

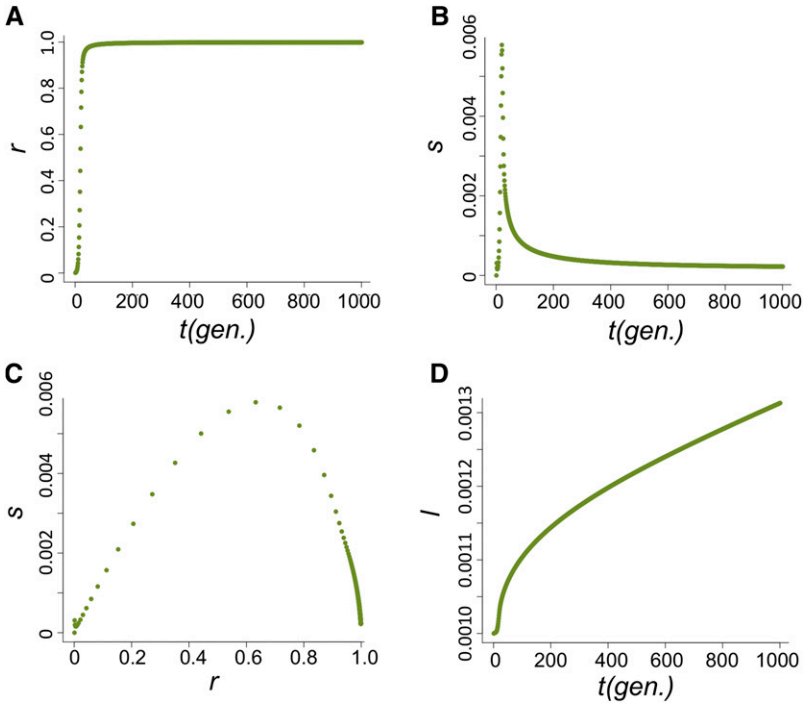

Figure 2 The dynamics of the host population during the spread of *P elements*. (A–D) The change of proportion of *P-cytotype* individuals, *r* (A); the selection coefficient against the nonbeneficial host allele, *s* (B); the relationship between *s* and *r* (C); and the allele frequency of the host beneficial allele, *l* (D) when u_0=1 and n_{HD}=5.

number (Charlesworth and Charlesworth 1983; Dolgin and Charlesworth 2006, 2008) and host sterility caused by too many double-stranded chromosomal breaks generated through *P-element* transposition. We set individuals with more than n_{HD} new *P-element* transpositions (and thus double-stranded breaks) to be completely sterile. This sterility effect would most likely occur in the *M*(female) × *P*(male) hybrid dysgenic cross, as the transposition rate of *P elements* in the dysgenic cross (u_0) is several orders of magnitude higher than in the nondysgenic cross (u_1) (Eggleston *et al.* 1988). The host beneficial allele will especially enjoy strong fitness advantages in the dysgenic crosses because hosts with this allele will be more likely to have fewer than n_{HD} double-stranded breaks and therefore higher expected fertility. Of course, the degree of sterility elicited by *P-element* transposition can be a continuous phenotype. The usage of the truncation selection model will maximize the selective benefit of a host allele that can reduce the transposition rate of *P element*, making our overall conclusion conservative.

The increase in frequency of the host beneficial allele (l) is dependent on how fast *P elements* spread through the population and, during their spread, how likely hybrid dysgenesis is to occur. We found that the spread of *P elements* is fast in most cases (Figure 2A for $u_0 = 1$ and $n_{HD} = 5$; see *Appendix* for discussions of other cases tested), a finding that is consistent with several caged experiments introducing *P elements* into M strain populations (Kidwell *et al.* 1988; Good *et al.* 1989). This quick spread leads to the host beneficial allele having selective advantage only during a narrow period (Figure 2B). We found that the largest selective advantage occurs when the proportion of *P cytotype* individuals (r) is intermediate and the probability of a hybrid dysgenic cross is high (Figure 2C). This is also reflected in the dynamics of l, which has a phase of rapid increase followed by a longer phase of slow increase (Figure 2D).

The combined effects of u_0 and n_{HD} had the most notable influence on the dynamics of the host beneficial allele. With increased n_{HD}, it takes fewer generations until the *P element* is found in nearly all genomes ($r > 0.99$) in the population ($t_{0.99}$) (Figure 3A) and host beneficial allele frequency at generation 1000 (l_{1000}) because of the decreased duration during which the probability of hybrid dysgenesis is high (Figure 3B). Generally, *P elements* spread faster when u_0 is larger (except for $n_{HD} = 2$, see below). Yet, the probability of hybrid dysgenesis is also higher, leading to larger l_{1000} (Figure 3, A and B). For all cases examined except one (see *Appendix* for all parameters tested), the difference between l_0 and l_{1000} (<2%) would hardly be detected with the regular size of samples.

The only exception is when $u_0 = 1$ and n_{HD}=2 in which the *temporal* F_{st} can be as large as 0.15. However, it takes >10,000 generations for the *P element* to become nearly fixed in the population due to the still high probability of hybrid dysgenesis even when the majority of the hosts have *P cytotype*. This is much longer than the timescale of actual *P-element* spread. Furthermore, the tolerance of nonprogrammed double-stranded breaks in the germline can be much higher than this threshold ($n_{HD} = 2$; Orsi *et al.* 2010).

It is worth noting that we modeled the distribution of *P-element* copy number as approximately Poisson. The derivation and previous application of the Poisson approximation were focused on situations where the TE population is near equilibrium and the mean copy number of TEs is much larger than one (Charlesworth and Charlesworth 1983;

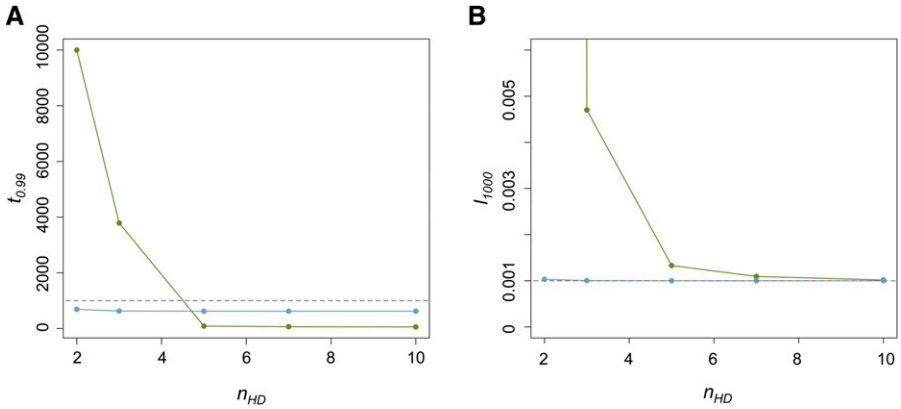

Figure 3 The influences of u_0 and n_{HD} on the time for near fixation of *P elements* and host beneficial allele frequency. (A) The generations until the *P element* is nearly fixed in the population ($t_{0.99}$) for different u_0 and n_{HD}. Green dots are when $u_0 = 1$ and blue dots are when $u_0 = 10^{-1}$. The dashed line denotes generation 1000. (B) The allele frequencies of the host beneficial allele (I_{1000}) at generation 1000 for different u_0 and n_{HD}. The dashed line denotes 0.001, which equals I_0.

Langley *et al.* 1983). However, the critical aspect of our analytical modeling here is the initial phase of the invasion of a new TE family (the *P element*), which may lead to a copy number distribution different from Poisson. We thus used Monte Carlo simulations (see File S1 for details of simulations) to investigate how the distribution of TE copy number reaches a Poisson distribution. We found that soon after the *P cytotype* is common in the population, the TE copy number distribution is close to Poisson (Figure S5 in File S1). More importantly, these simulations show that the estimated change of host beneficial allele frequency from the analytical approximation is within 2% of the simulated results (Figure S3 and Figure S4 in File S1), supporting our overall conclusions.

Recent horizontal transfer of the P element does not have widespread evolutionary impacts on candidate genes

To empirically investigate the short-term evolutionary impacts of TE horizontal transfer on hosts, we compared the genetic differentiation between *pre–P-element* invasion (*pre-P*) and current (*post-P*) populations (temporal differentiation). For five candidate genes that showed strong temporal differentiation in our initial survey with a smaller number of M strains (*Irbp*, *krimp*, *Hen1*, *spn-E*, and *squ*; see *Materials and Methods* and Table S3), only *Hen1* and *squ* still showed highly significant temporal differentiation with increased size of *pre–P-element* samples (Table 4). However, we observed strong genetic differentiation between North American and African contemporary *post-P* populations for our candidate genes (geographic differentiation, Table 4). If the geographic differentiation of the current population was also present in the *pre-P* population, the wide geographic distribution of M strains used may lead to false conclusions about the temporal differentiation (see *Materials and Methods* and Table S1). We further restricted our analysis to the North American population, which has the largest number of alleles for both *pre-P* (15) and *post-P* (37) samples and still found strong temporal differentiation of *Hen1* ($F_{st} = 0.173$, $P = 0.003$; Table 4). Such differentiation may reflect divergence in protein functions, as F_{st} estimated using amino acid sequences also showed significant temporal differentiation ($F_{st} = 0.261$, $P < 0.001$; Table 4).

The strong temporal differentiation of Hen1 is likely the result of genetic hitchhiking from a nearby, strongly selected gene

Given that the spread of *P elements* is fairly recent and fast, any associated directional selection on suppressive host variants is expected to result in reduced genomic polymorphism around the selected SNPs and increased linkage

Table 4 Temporal differentiation of a subset of candidate genes

	Geographic differentiation[a]		Temporal differentiation[b]					
			All samples		Only North American samples[c]			
	nucleotide differentiation[d]		nucleotide differentiation		nucleotide differentiation		amino acid differentiation[e]	
	F_{st}	*P*-value	F_{st}	*P*-value	F_{st}	*P*-value	F_{st}	*P*-value
Irbp	0.144	0.111	**0.170**	**0.038**	0.000	0.334	−0.010	0.455
krimp	0.143	**0.004**	0.011	0.262	−0.009	0.635	0.006	0.345
Hen1	0.632	<**0.001**	**0.303**	**0.001**	0.173	**0.003**	0.261	<**0.001**
Spn-E	0.459	<**0.001**	0.053	0.08	0.031	0.097	0.071	**0.033**
squ	0.028	0.25	**0.202**	**0.001**	0.071	**0.036**	0.039	0.103

All the significant results are in boldface type.
[a] Genetic differentiation between current (*post–P-element*) African and North American populations.
[b] Genetic differentiation between pre–*P-* and *post–P-element* invasion populations.
[c] The temporal differentiation was estimated considering only North American samples of both populations (before and after *P-element* invasions).
[d] Genetic differentiation estimated using nucleotide sequences.
[e] Genetic differentiation estimated using amino acid sequences.

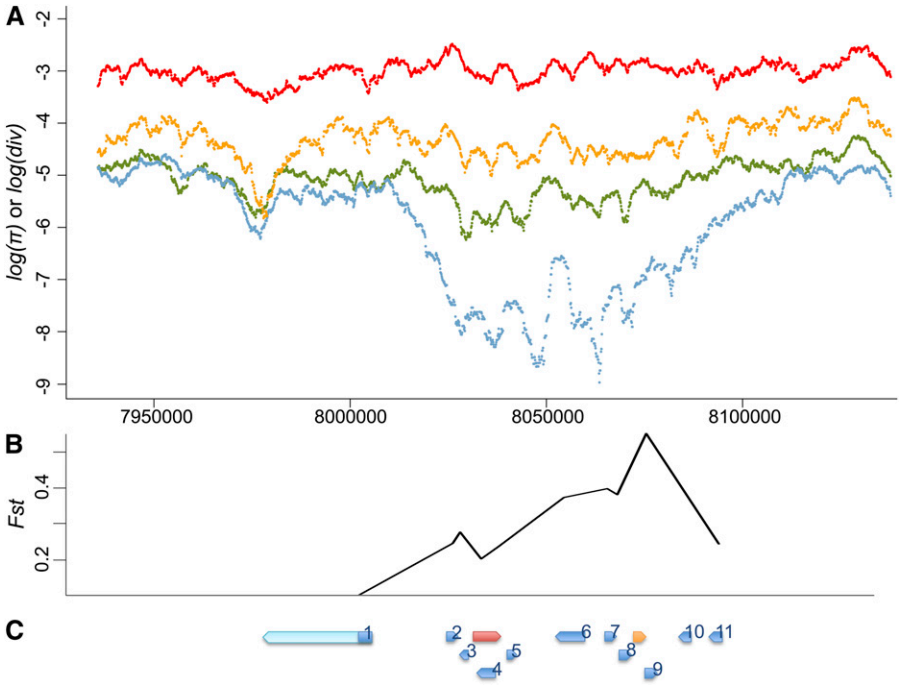

Figure 4 Polymorphism, divergence, and temporal differentiation around *Hen1*. (A) Divergence between *D. melanogaster* and *D. simulans* (red), polymorphism of *D. simulans* (orange), polymorphism of the post–*P-element* African *D. melanogaster* population (green), and polymorphism of the post–*P-element* North American *D. melanogaster* population (blue) of 100 kb upstream and downstream of *Hen1* (from 8,033,215 to 8,040,257) are shown on a log scale. There is a dramatic drop of polymorphism in the North American *D. melanogaster* populations around *Hen1*. (B and C) Temporal differentiation between pre–*P*- and post–*P-element* North American populations of control genes (B), and their relative position (C). Genes in (C) are *Hen1* (red), *Cyp6g1* (orange), *jeb* (1), *CG8378* (2), *CG13178* (3), *CG8878* (4), *CG8407* (5), *Oda* (6), *wash* (7), *CG33964* (8), *Cyp6t3* (9), *RpS11* (10), and *SrCII* (11). Sequenced regions of each control gene are shown in blue while unsequenced regions are in light blue. The coordinates of three figures are aligned.

disequilibrium due to genetic hitchhiking (Maynard Smith and Haigh 1974; Kim and Nielsen 2004; Stephan *et al.* 2006). Consistent with the analysis of temporal differentiation, we observed marginally significant haplotypic structures for the *Hen1* gene region for *post-P* North American populations [frequency of major haplotypes = 15, $P = 0.042$ (Hudson *et al.* 1994) and haplotype heterozygosity = 0.759, $P = 0.031$ (Depaulis and Veuille 1998)]. This strong haplotypic structure was extended at least 10 kb upstream and downstream of *Hen1* (frequency of major haplotypes = 10, $P = 0.003$ and haplotype heterozygosity = 0.290, $P = 0.001$). We also observed a dramatic reduction of post–*P-element* North American variation for an almost 100-kb genomic segment around *Hen1* when compared with polymorphism in post–*P-element* African *D. melanogaster*, polymorphism in closely related *D. simulans*, and divergence between *D. melanogaster* and *D. simulans* (Figure 4A).

However, the genomic region around *Hen1* is highly gene rich, with >20 genes, including *Cyp6g1*. Studies have identified a recent selective sweep associated with the *Cyp6g1* allele that has an *Accord* transposable element inserted upstream (Daborn *et al.* 2002; Catania *et al.* 2004; Chung *et al.* 2007; Schmidt *et al.* 2010). Functional analysis confirmed that this *Accord* insertion confers insecticide resistance (Daborn *et al.* 2002; Chung *et al.* 2007), which is the most likely force driving the strong, recent selective sweep on *Cyp6g1*. The *Accord* inserted allele was found fixed in non-African populations, yet it was intermediate in African populations (Catania *et al.* 2004), which may also explain why the reduction of heterozygosity around *Hen1* was most apparent in the North American populations while the *P element* is virtually fixed worldwide.

To further investigate whether the strong temporal differentiation observed on *Hen1* is the result of genetic hitchhiking from strongly selected *Cyp6g1* or any other genes in the nearby region, we estimated the temporal differentiation of coding regions for another 11 genes that are within 30 kb to either *Hen1* or *Cyp6g1* and have functions unrelated to either TE suppression or insecticide resistance ("control genes," Table S4 and Figure 4C). Compared with the observed strong temporal differentiation of these control genes, *Hen1* ($F_{st} = 0.173$) is no longer exceptional (Table 5 and Figure 4B). We also found that the closer a gene is to *Cyp6g1*, the stronger the temporal differentiation was (Table 5 and Figure 4B). The geographic differentiation was also strong for these 11 genes (Table 5), which is consistent with the scenario that application of insecticide in the non-African regions is leading to strong genetic hitchhiking on genes around *Cyp6g1* in the post-*P* North American population we studied.

Discussion

TEs are selfish genetic elements in the genome. Their interactions with their hosts are often analogized to the molecular arms race between hosts and other nongenomic pathogens, such as bacteria, viruses, fungi, and protozoa. This analogy deserves further mechanistic specification and analysis because of the fundamental differences in transmission mode between TEs and horizontally transmitted pathogens.

To address this analogy, we first systematically compared the long-term evolution of host TE-interacting genes to that of immunity genes. We found the proportion of TE-interacting genes with evidence of positive selection is at least as large as, if not greater than, that of genes in pathways conferring immunity to pathogens. *aub*, which showed strong evidence of adaptive protein evolution, is a key

Table 5 Temporal and geographic differentiation of control genes

	Distance to *Hen1*[c]	Distance to *Cyp6g1*[d]	M strain π		Temporal differentiation[a]		Geographic differentiation[b]	
			Nonsyn	Syn	F_{st}	*P*-value	F_{st}	*P*-value
jeb[e]	−31,911	−69,718	0.0021	0.0145	0.089	0.037	0.084	0.085
CG8378	−7,273	−45,080	0.0000	0.0013	0.239	0.001	0.372	0.011
CG13178	−5,378	−43,185	0.0007	0.0044	0.271	0.002	0.689	<0.001
CG8878	0	−37,807	0.0001	0.0013	0.195	0.001	0.365	<0.001
CG8407	4,426	−33,381	0.0000	0.0143	0.230	0.005	0.057	0.226
Oda	21,177	−16,630	0.0012	0.0007	0.369	0.001	0.669	<0.001
wash	32,266	−5,541	0.0010	0.0029	0.393	<0.001	0.583	<0.001
CG33964	34,832	−2,975	0.0009	0.0010	0.377	<0.001	0.654	<0.001
Cyp6t3	42,227	4,420	0.0021	0.0059	0.549	<0.001	0.329	0.007
RpS11	50,683	12,876	0.0000	0.0027	NA	NA	0.648	<0.001
Sr-CII	60,799	22,992	0.0025	0.0151	0.238	0.001	0.368	<0.001

NA, not available due to no SNP differences between pre–*P*- and post–*P-element* North American populations. The differentiations were estimated using nucleotide sequences.

[a] Comparisons between pre– and post–*P-element* invasion populations with samples from North American populations only.

[b] Comparisons between North American and African post–*P-element* invasion populations.

[c] Distance in base pairs between the midpoints of the focused gene and *Hen1*. *CG8878* has zero distance because it is nested within *Hen1*.

[d] Distance in base pairs between the midpoints of the focused gene and *Cyp6g1*.

[e] Only the coding region of the first exon of *jeb* was sequenced.

component in *piRNA*-mediated TE silencing (Brennecke *et al.* 2007; Gunawardane *et al.* 2007) and its PAZ domain, known to mediate the binding of single-strand *RNAs* associated with Argonaute/Piwi proteins (Lingel *et al.* 2003; Song *et al.* 2003; Yan *et al.* 2003), also showed an excess of amino acid fixations. Both RNA helicases surveyed, *armi* and *spn-E*, showed evidence of adaptive evolution, although the detailed mechanism of their involvement in *piRNA* biogenesis and TE suppression is still not clear. Most interestingly, all three nuage component genes (*mael*, *krimp*, and *vas*) showed strong evidence of positive selection. The nuage, where many proteins encoded by *piRNA* genes localize (reviewed in Klattenhoff and Theurkauf 2008), is considered the major battleground for the host–TE arms race (Blumenstiel 2011). In addition, *vas* is involved in the formation of pole plasm (the future germline) and is essential for the proper localization and translational control of maternally deposited *mRNAs* at the rear of developing embryos (Lasko and Ashburner 1990; Styhler *et al.* 1998; Johnstone and Lasko 2004). The RNA genome of TEs that can be preferentially incorporated into the pole plasm and have its RNA tertiary structure be detangled and translated will have large fitness benefits from increased transmission. The DEAD box of VAS, the key domain mediating the unwinding of RNAs with self-annealed structure (reviewed in Linder 2006; Arkov and Ramos 2010), had a strong enrichment of amino acid fixations under the McDonald–Kreitman framework.

Our analytical model found that host alleles suppressing *P-element* transposition enjoy strong selective benefits only during a limited time frame due to the fast spread of *P elements* through strict vertical parent–offspring inheritance. This led to negligible host allele frequency differences between pre– and post–*P-element* populations with all biologically reasonable parameters tested. This theoretical prediction may be extended to other known cases of horizontally transferred TEs in *Drosophila*, such as *I-element*, *hobo*, and *mariner* (reviewed in Silva *et al.* 2004). For DNA-based TE families that also transpose with the cut-and-paste mechanism (*hobo* and *mariner*), our model should be readily applicable. RNA-based TE families, such as *I-element*, transpose through a "copy-and-paste mechanism" (replicating through insertions of reversed transcribed *cDNA* into another position in the genome). The hybrid dysgenic syndrome observed for *I-element* is caused by the catastrophic meiosis of eggs, which includes failure to produce functional female pronucleus and developmental abnormalities after fertilization (Orsi *et al.* 2010). If such a meiotic defect is initiated with a higher than tolerable threshold of *I-elements* transposition activity and is also responsible for the hybrid dysgenic syndromes of other unobserved horizontal transfer of RNA-based TEs, our model could be a reasonable generalization for all TE families.

Consistent with our quantitative analysis, we found no strong evidence supporting recent selection imposed by *P elements* on host candidate genes. The significant temporal differentiation and haplotypic structures of *Hen1*, our only strong candidate, are likely the result of the strong genetic hitchhiking effects from the nearby insecticide-resistant gene, *Cyp6g1*. Certainly it is possible that other types of alleles or loci may have responded to the *P-element* invasion. We did not find a strong reduction in polymorphism of the post–*P-element* population 10 kb upstream and downstream of each candidate gene, providing no support for the alternative that selection has acted on polymorphisms of local *cis*-acting elements (data not shown). Yet, the possibility of strong selection on *cis*-regulatory elements outside the surveyed region or of *trans*-acting regulatory variation cannot be ruled out. The possibility that there was instead strong selection for a *P-element* variant that can reduce the deleterious impact on its host is not plausible because there

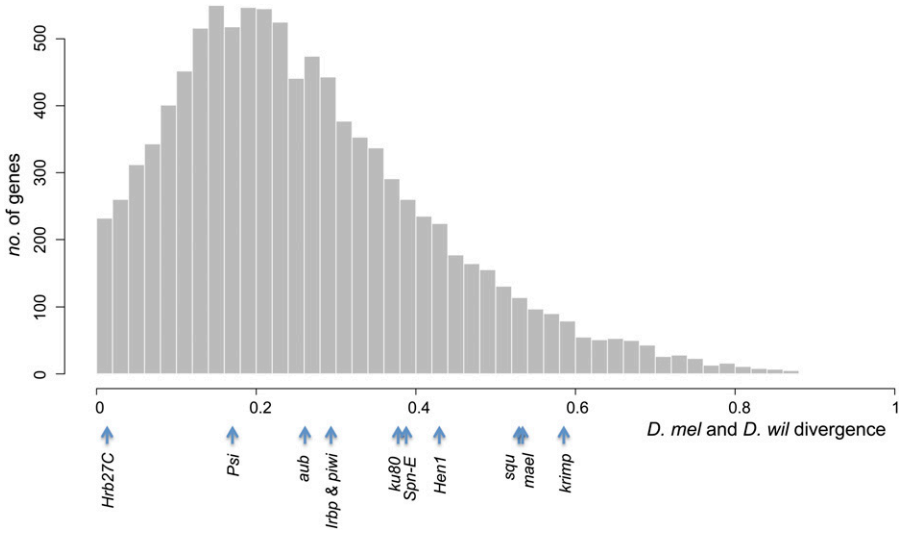

Figure 5 Divergence between *D. melanogaster* (*D. mel*) and *D. willistoni* (*D. wil*) of candidate genes among other genes. Amino acid sequence divergences were estimated for genes with an annotated *D. willistoni* ortholog (11 candidate genes and 10,029 other genes) and the genome-wide distribution of the divergence is shown. The divergences of the 11 candidate genes on the *x*-axis are shown at the bottom.

is only a 1-base difference observed between *D. willistoni* and *D. melanogaster* canonical *P elements* (Daniels *et al.* 1990). Still another alternative is that prevalent *D. melanogaster* variants of host genes segregating in the pre–*P-element* population were already similar to those of *D. willistoni*, which coevolved with *P elements* and can thus effectively reduce *P-element* transposition and minimize their deleterious effects. We found that the amino acid sequence divergence between *D. melanogaster* and *D. willistoni* orthologs (Clark *et al.* 2007) for most candidate genes is greater than the genome-wide median (Figure 5). The only two exceptions [*Hrb27C* (1.7%) and *Psi* (16.4%)] both have other essential host functions (Hammond *et al.* 1997; Labourier *et al.* 2002; Goodrich *et al.* 2004; Huynh *et al.* 2004; Yano *et al.* 2004; Blanchette *et al.* 2005) and their evolution is expected to be strongly constrained. These observations therefore do not support the scenario of *D. melanogaster* preadaptation to *P elements*.

The selection coefficient for a host allele that can reduce TE transposition rate in an outbred population with limited linkage disequilibrium is

$$s \approx \delta u\left(\frac{\bar{n}u}{2\tilde{H}}\right)$$

(Charlesworth and Langley 1986), which depends on the change in transposition rate δu and the expected number of new TE copies (average copy number $\bar{n}$ times transposition rate u). $\tilde{H}$ is the approximated harmonic mean of recombination frequency between pairs of TE insertions, which approaches $\frac{1}{2}$ in species with free recombination. This theoretical prediction suggests that the selective benefit of a host allele increases with the number of TE copies whose transposition it can suppress. Given the TE transposition rate [$10^{-5} \sim 10^{-4}$ (Nuzhdin and Mackay 1995; Nuzhdin *et al.* 1997; Maside *et al.* 2000, 2001; reviewed in Charlesworth and Langley 1989; Le Rouzic and Deceliere 2005)] and the number of active copies of individual TE families [generally <100 (Kaminker *et al.* 2002; Quesneville *et al.* 2005; Bergman *et al.* 2006)] in *Drosophila*, the selective benefits for a host allele targeting a specific TE family are small. However, *Drosophila* genomes are occupied by >100 TE families (Kaminker *et al.* 2002; Quesneville *et al.* 2005; Bergman *et al.* 2006; Clark *et al.* 2007), and TE families can be further classified into clades or superfamilies according to encoded protein products and similarities in sequences [such as Ty1-copia-like or Ty3-gypsy like (reviewed in Wicker *et al.* 2007)]. In this case, a host variant targeting attributes shared between multiple TE families can enjoy larger selective benefits than a variant targeting a single TE family and is likely to spread in the host population. For example, a host allele that can reduce half of the transposition rate of a subset of TE families with a total of 1000 copies can have a selection coefficient as large as 10^{-5}, which would be strong enough to overcome the effect of genetic drift in *D. melanogaster* [$N_e \sim 10^6$ (Langley *et al.* 1982; Kreitman 1983)]. Accordingly, unlike the arms race between host and horizontally transferred pathogens, where strong selective benefit comes from the precise targeting of the host allele to a specific pathogen variant, the antagonistic interaction between TEs and host variants that can target the aggregated influence of multiple vertically inherited TE families could be the main force driving the fast evolution of host TE-interacting genes

Acknowledgments

We thank the Bloomington *Drosophila* Stock Center and M. Itoh for providing M strains. We appreciate D. J. Begun and M. T. Levine for extensive discussion of the project and D. J. Begun, M. T. Levine, P. L. Ralph, M. Turelli, James A. Chandler, and E. I. Dietrich for critical reading of the manuscript. We also thank D. J. Obbard for helpful discussion and kindly sharing unpublished results regarding MK tests for *piRNA* genes. We thank C. M. Cardeno, P. Saelao, and T. Trejo-Cantwell for experimental assistance. Finally, we acknowledge the thoughtful, constructive, and patient criticism of two anonymous reviewers. This work was supported by National Science Foundation Doctoral Dissertation

Improvement Grant, Daphne and Ted Pengelley Award in Evolutionary Biology, and Center for Population Biology of University of California, Davis, Research Award to Y.C.G.L.

Literature Cited

Adams, M. D., R. S. Tarng, and D. C. Rio, 1997 The alternative splicing factor PSI regulates P-element third intron splicing in vivo. Genes Dev. 11: 129–138.

Aminetzach, Y. T., J. M. Macpherson, and D. A. Petrov, 2005 Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. Science 309: 764–767.

Andolfatto, P., 2001 Contrasting patterns of X–linked and autosomal nucleotide variation in Drosophila melanogaster and Drosophila simulans. Mol. Biol. Evol. 18: 279–290.

Andolfatto, P., K. M. Wong, and D. Bachtrog, 2011 Effective population size and the efficacy of selection on the X chromosomes of two closely related Drosophila species. Genome Biol. Evol. 3: 114–128.

Anxolabéhère, D., M. G. Kidwell, and G. Periquet, 1988 Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of Drosophila melanogaster by mobile P elements. Mol. Biol. Evol. 5: 252–269.

Aquadro, C. F., S. F. Desse, M. M. Bland, C. H. Langley, and C. C. Laurie-Ahlberg, 1986 Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. Genetics 114: 1165–1190.

Aquadro, C. F., K. M. Lado, and W. A. Noon, 1988 The rosy region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. Genetics 119: 875–888.

Aravin, A. A., M. S. Klenov, V. V. Vagin, F. Bantignies, G. Cavalli *et al.*, 2004 Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line. Mol. Cell. Biol. 24: 6742–6750.

Aravin, A. A., G. J. Hannon, and J. Brennecke, 2007 The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318: 761–764.

Arkov, A. L., and A. Ramos, 2010 Building RNA-protein granules: insight from the germline. Trends Cell Biol. 20: 482–490.

Beall, E. L., A. Admon, and D. C. Rio, 1994 A Drosophila protein homologous to the human p70 Ku autoimmune antigen interacts with the P transposable element inverted repeats. Proc. Natl. Acad. Sci. USA 91: 12681–12685.

Begun, D. J., A. K. Holloway, K. Stevens, L. W. Hillier, Y.-P. Poh *et al.*, 2007 Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. PLoS Biol. 5: e310.

Bergman, C. M., H. Quesneville, D. Anxolabéhère, and M. Ashburner, 2006 Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. Genome Biol. 7: R112.

Bingham, P. M., M. G. Kidwell, and G. M. Rubin, 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.

Blanchette, M., R. E. Green, S. E. Brenner, and D. C. Rio, 2005 Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. Genes Dev. 19: 1306–1314.

Blumenstiel, J. P., 2011 Evolutionary dynamics of transposable elements in a small RNA world. Trends Genet. 27: 23–31.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis *et al.*, 2007 Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128: 1089–1103.

Brennecke, J., C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark *et al.*, 2008 An epigenetic role for maternally inherited piRNAs in transposon silencing. Science 322: 1387–1392.

Brookfield, J. F., E. Montgomery, and C. H. Langley, 1984 Apparent absence of transposable elements related to the P *elements* of D. melanogaster in other species of Drosophila. Nature 310: 330–332.

Catania, F., M. O. Kauer, P. J. Daborn, J. L. Yen, R. H. Ffrench-Constant *et al.*, 2004 World-wide survey of an Accord insertion and its association with DDT resistance in Drosophila melanogaster. Mol. Ecol. 13: 2491–2504.

Charlesworth, B., and D. Charlesworth, 1983 The population dynamics of transposable elements. Genet. Res. 42: 1–27.

Charlesworth, B., and C. H. Langley, 1986 The evolution of self-regulated transposition of transposable elements. Genetics 112: 359–383.

Charlesworth, B., and C. H. Langley, 1989 The population genetics of Drosophila transposable elements. Annu. Rev. Genet. 23: 251–287.

Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson *et al.*, 2003 Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31: 3497–3500.

Chung, H., M. R. Bogwitz, C. McCart, A. Andrianopoulos, R. H. Ffrench-Constant *et al.*, 2007 *cis*-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene Cyp6g1. Genetics 175: 1071–1077.

Clark, A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver *et al.*, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature 450: 203–218.

Clark, J. B., W. P. Maddison, and M. G. Kidwell, 1994 Phylogenetic analysis supports horizontal transfer of P transposable elements. Mol. Biol. Evol. 11: 40–50.

Clark, J. B., P. C. Kim, and M. G. Kidwell, 1998 Molecular evolution of P transposable elements in the genus Drosophila. III. The melanogaster species group. Mol. Biol. Evol. 15: 746–755.

Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil *et al.*, 2002 A single P450 allele associated with insecticide resistance in Drosophila. Science 297: 2253–2256.

Daniels, S. B., K. R. Peterson, L. D. Strausbaugh, M. G. Kidwell, and A. Chovnick, 1990 Evidence for horizontal transmission of the P transposable element between Drosophila species. Genetics 124: 339–355.

Dawkins, R., and J. R. Krebs, 1979 Arms races between and within species. Proc. R. Soc. Lond., B, Biol. Sci. 205: 489–511.

Depaulis, F., and M. Veuille, 1998 Neutrality tests based on the distribution of haplotypes under an infinite-site model. Mol. Biol. Evol. 15: 1788–1790.

Dolgin, E. S., and B. Charlesworth, 2006 The fate of transposable elements in asexual populations. Genetics 174: 817–827.

Dolgin, E. S., and B. Charlesworth, 2008 The effects of recombination rate on the distribution and abundance of transposable elements. Genetics 178: 2169–2177.

Eggleston, W. B., D. M. Johnson-Schlitz, and W. R. Engels, 1988 P-M hybrid dysgenesis does not mobilize other transposable element families in D. melanogaster. Nature 331: 368–370.

Engels, W. R., 1979 Germ line aberrations associated with a case of hybrid dysgenesis in Drosophila melanogaster males. Genet. Res. 33: 137–146.

Engels, W. R., 1997 Invasions of P elements. Genetics 145: 11–15.

Engels, W. R., D. M. Johnson-Schlitz, W. B. Eggleston, and J. Sved, 1990 High-frequency P element loss in Drosophila is homolog dependent. Cell 62: 515–525.

Finn, R. D., J. Mistry, J. Tate, P. Coggill, A. Heger *et al.*, 2009 The Pfam protein families database. Nucleic Acids Res. 38: D211–D222.

Finnegan, D. J., 1992 Transposable elements. Curr. Opin. Genet. Dev. 2: 861–867.

González, J., K. Lenkov, M. Lipatov, J. M. Macpherson, and D. A. Petrov, 2008 High rate of recent transposable element-induced adaptation in Drosophila melanogaster. PLoS Biol. 6: e251.

Good, A. G., G. A. Meister, H. W. Brock, T. A. Grigliatti, and D. A. Hickey, 1989 Rapid spread of transposable P elements in experimental populations of *Drosophila melanogaster*. Genetics 122: 387–396.

Goodrich, J. S., K. N. Clouse, and T. Schüpbach, 2004 Hrb27C, Sqd and Otu cooperatively regulate gurken RNA localization and mediate nurse cell chromosome dispersion in Drosophila oogenesis. Development 131: 1949–1958.

Gossmann, T. I., B.-H. Song, A. J. Windsor, T. Mitchell-Olds, C. J. Dixon *et al.*, 2010 Genome wide analyses reveal little evidence for adaptive evolution in many plant species. Mol. Biol. Evol. 27: 1822–1832.

Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura *et al.*, 2007 A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science 315: 1587–1590.

Haldane, J. B., 1949 Disease and evolution. Ric. Sci. Suppl. A 19: 68–76.

Hammond, L. E., D. Z. Rudner, R. Kanaar, and D. C. Rio, 1997 Mutations in the hrp48 gene, which encodes a Drosophila heterogeneous nuclear ribonucleoprotein particle protein, cause lethality and developmental defects and affect P-element third-intron splicing in vivo. Mol. Cell. Biol. 17: 7260–7267.

Heger, A., and C. P. Ponting, 2007 Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. Genome Res. 17: 1837–1849.

Hiraizumi, Y., 1971 Spontaneous recombination in Drosophila melanogaster males. Proc. Natl. Acad. Sci. USA 68: 268–270.

Hoffmann, A. A., and A. R. Weeks, 2007 Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in Drosophila melanogaster from eastern Australia. Genetica 129: 133–147.

Hudson, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18: 337–338.

Hudson, R. R., D. D. Boos, and N. L. Kaplan, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. 9: 138–151.

Hudson, R. R., K. Bailey, D. Skarecky, J. Kwiatowski, and F. J. Ayala, 1994 Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. Genetics 136: 1329–1340.

Hughes, A. L., and M. Yeager, 1998 Natural selection at major histocompatibility complex loci of vertebrates. Annu. Rev. Genet. 32: 415–435.

Hughes, A. L., T. Ota, and M. Nei, 1990 Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. 7: 515–524.

Huynh, J.-R., T. P. Munro, K. Smith-Litière, J.-A. Lepesant, and D. St Johnston, 2004 The Drosophila hnRNPA/B homolog, Hrp48, is specifically required for a distinct step in osk mRNA localization. Dev. Cell 6: 625–635.

Jiggins, F. M., and K. W. Kim, 2007 A screen for immunity genes evolving under positive selection in Drosophila. J. Evol. Biol. 20: 965–970.

Johnstone, O., and P. Lasko, 2004 Interaction with eIF5B is essential for Vasa function during development. Development 131: 4167–4178.

Kaminker J. S., C. M. Bergman, B. Kronmiller, J. Carlson, R. Svirskas *et al.*, 2002 The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol. 3: RESEARCH0084.

Kidwell, M. G., and D. R. Lisch, 2001 Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55: 1–24.

Kidwell, M. G., J. F. Kidwell, and M. Nei, 1973 A case of high rate of spontaneous mutation affecting viability in Drosophila melanogaster. Genetics 75: 133–153.

Kidwell, M. G., J. F. Kidwell, and J. A. Sved, 1977 Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86: 813–833.

Kidwell, M. G., T. Frydryk, and J. Novy, 1983 The hybrid dysgenesis potential of Drosophila melanogaster strains of diverse temporal and geographical natural origins. Drosoph. Inf. Serv. 59: 63–69.

Kidwell, M. G., K. Kimura, and D. M. Black, 1988 Evolution of hybrid dysgenesis potential following P element contamination in Drosophila melanogaster. Genetics 119: 815–828.

Kim, Y., and R. Nielsen, 2004 Linkage disequilibrium as a signature of selective sweeps. Genetics 167: 1513–1524.

Klattenhoff, C., and W. Theurkauf, 2008 Biogenesis and germline functions of piRNAs. Development 135: 3–9.

Klattenhoff, C., H. Xi, C. Li, S. Lee, J. Xu *et al.*, 2009 The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell 138: 1137–1149.

Kolaczkowski, B., D. N. Hupalo, and A. D. Kern, 2011 Recurrent adaptation in RNA interference genes across the Drosophila phylogeny. Mol. Biol. Evol. 28: 1033–1042.

Kreitman, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of Drosophila melanogaster. Nature 304: 412–417.

Labourier, E., M. Blanchette, J. W. Feiger, M. D. Adams, and D. C. Rio, 2002 The KH-type RNA-binding protein PSI is required for Drosophila viability, male fertility, and cellular mRNA processing. Genes Dev. 16: 72–84.

Langley, C. H., E. Montgomery, and W. F. Quattlebaum, 1982 Restriction map variation in the Adh region of Drosophila. Proc. Natl. Acad. Sci. USA 79: 5631–5635.

Langley, C. H., J. F. Brookfield, and N. Kaplan, 1983 Transposable elements in Mendelian populations. I. A theory. Genetics 104: 457–471.

Langley, C. H., E. Montgomery, R. Hudson, N. Kaplan, and B. Charlesworth, 1988 On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. 52: 223–235.

Langley, C. H., B. P. Lazzaro, W. Phillips, E. Heikkinen, and J. M. Braverman, 2000 Linkage disequilibria and the site frequency spectra in the su(s) and su(w(a)) regions of the Drosophila melanogaster X chromosome. Genetics 156: 1837–1852.

Langley, C. H., K. Stevens, C. Cardeno, Y. C. G. Lee, D. R. Schrider *et al.*, 2012 Genomic variation in natural populations of *Drosophila melanogaster*. Genetics 192: 533–598.

Lasko, P. F., and M. Ashburner, 1990 Posterior localization of vasa protein correlates with, but is not sufficient for, pole cell development. Genes Dev. 4: 905–921.

Lazzaro, B. P., 2008 Natural selection on the Drosophila antimicrobial immune system. Curr. Opin. Microbiol. 11: 284–289.

Lee, Y. C. G., and C. H. Langley, 2010 Transposable elements in natural populations of Drosophila melanogaster. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 1219–1228.

Le Rouzic, A., and G. Deceliere, 2005 Models of the population genetics of transposable elements. Genet. Res. 85: 171–181.

Li, C., V. V. Vagin, S. Lee, J. Xu, S. Ma *et al.*, 2009 Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell 137: 509–521.

Lim, A. K., and T. Kai, 2007 Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 104: 6714–6719.

Linder, P., 2006 Dead-box proteins: a family affair—active and passive players in RNP-remodeling. Nucleic Acids Res. 34: 4168–4180.

Lingel, A., B. Simon, E. Izaurralde, and M. Sattler, 2003 Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature 426: 465–469.

Long, A. D., R. F. Lyman, C. H. Langley, and T. F. Mackay, 1998 Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. Genetics 149: 999–1017.

Loreto, E. L. S., C. M. A. Carareto, and P. Capy, 2008 Revisiting horizontal transfer of transposable elements in Drosophila. Heredity 100: 545–554.

Lu, J., and A. G. Clark, 2010 Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. Genome Res. 20: 212–227.

Mackay, T. F., 1989 Transposable elements and fitness in Drosophila melanogaster. Genome 31: 284–295.

Maside, X., S. Assimacopoulos, and B. Charlesworth, 2000 Rates of movement of transposable elements on the second chromosome of Drosophila melanogaster. Genet. Res. 75: 275–284.

Maside, X., C. Bartolomé, S. Assimacopoulos, and B. Charlesworth, 2001 Rates of movement and distribution of transposable elements in Drosophila melanogaster: in situ hybridization *vs.* Southern blotting data. Genet. Res. 78: 121–136.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23–35.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature 351: 652–654.

Miyashita, N., and C. H. Langley, 1988 Molecular and phenotypic variation of the white locus region in Drosophila melanogaster. Genetics 120: 199–212.

Montgomery, E., B. Charlesworth, and C. H. Langley, 1987 A test for the role of natural selection in the stabilization of transposable element copy number in a population of Drosophila melanogaster. Genet. Res. 49: 31–41.

Montgomery, E. A., S. M. Huang, C. H. Langley, and B. H. Judd, 1991 Chromosome rearrangement by ectopic recombination in Drosophila melanogaster: genome structure and evolution. Genetics 129: 1085–1098.

Nei M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nuzhdin, S. V., and T. F. Mackay, 1995 The genomic rate of transposable element movement in Drosophila melanogaster. Mol. Biol. Evol. 12: 180–181.

Nuzhdin, S. V., E. G. Pasyukova, and T. F. Mackay, 1997 Accumulation of transposable elements in laboratory lines of Drosophila melanogaster. Genetica 100: 167–175.

Obbard, D. J., F. M. Jiggins, D. L. Halligan, and T. J. Little, 2006 Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr. Biol. 16: 580–585.

Obbard, D. J., K. H. J. Gordon, A. H. Buck, and F. M. Jiggins, 2009a The evolution of RNAi as a defence against viruses and transposable elements. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364: 99–115.

Obbard, D. J., J. J. Welch, K.-W. Kim, and F. M. Jiggins, 2009b Quantifying adaptive evolution in the Drosophila immune system. PLoS Genet. 5: e1000698.

Obbard, D. J., F. M. Jiggins, N. J. Bradshaw, and T. J. Little, 2011 Recent and recurrent selective sweeps of the antiviral RNAi gene Argonaute-2 in three species of Drosophila. Mol. Biol. Evol. 28: 1043–1056.

O'Hare, K., and G. M. Rubin, 1983 Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. Cell 34: 25–35.

Orsi, G. A., E. F. Joyce, P. Couble, K. S. McKim, and B. Loppin, 2010 Drosophila I-R hybrid dysgenesis is associated with catastrophic meiosis and abnormal zygote formation. J. Cell Sci. 123: 3515–3524.

Pane, A., K. Wehr, and T. Schüpbach, 2007 zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev. Cell 12: 851–862.

Pasyukova, E. G., S. V. Nuzhdin, T. V. Morozova, and T. F. C. Mackay, 2004 Accumulation of transposable elements in the genome of Drosophila melanogaster is associated with a decrease in fitness. J. Hered. 95: 284–290.

Quesneville, H., C. M. Bergman, O. Andrieu, D. Autard, D. Nouaud *et al.*, 2005 Combined evidence annotation of transposable elements in genome sequences. PLOS Comput. Biol. 1: e22.

Rio, D. C., 2002 P transposable elements in Drosophila melanogaster, pp. 484–518 in *Mobile DNA II*, edited by N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz.

Rio, D. C., and G. M. Rubin, 1988 Identification and purification of a Drosophila protein that binds to the terminal 31-base-pair inverted repeats of the P transposable element. Proc. Natl. Acad. Sci. USA 85: 8929–8933.

Rose, L. E., P. D. Bittner-Eddy, C. H. Langley, E. B. Holub, R. W. Michelmore *et al.*, 2004 The maintenance of extreme amino acid diversity at the disease resistance gene, RPP13, in Arabidopsis thaliana. Genetics 166: 1517–1527.

Rozen, S., and H. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132: 365–386.

Rubin, G. M., M. G. Kidwell, and P. M. Bingham, 1982 The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell 29: 987–994.

Sackton, T. B., B. P. Lazzaro, T. A. Schlenke, J. D. Evans, D. Hultmark *et al.*, 2007 Dynamic evolution of the innate immune system in Drosophila. Nat. Genet. 39: 1461–1468.

Schaack, S., C. Gilbert, and C. Feschotte, 2010 Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol. Evol. 25: 537–546.

Schlenke, T. A., and D. J. Begun, 2003 Natural selection drives Drosophila immune system evolution. Genetics 164: 1471–1480.

Schmidt, J. M., R. T. Good, B. Appleton, J. Sherrard, G. C. Raymant *et al.*, 2010 Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. PLoS Genet. 6: e1000998.

Schmidt, P. S., L. Matzkin, M. Ippolito, and W. F. Eanes, 2005 Geographic variation in diapause incidence, life-history traits, and climatic adaptation in Drosophila melanogaster. Evolution 59: 1721–1732.

Siebel, C. W., and D. C. Rio, 1990 Regulated splicing of the Drosophila P transposable element third intron in vitro: somatic repression. Science 248: 1200–1208.

Siebel, C. W., L. D. Fresco, and D. C. Rio, 1992 The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5′ splice site control U1 snRNP binding. Genes Dev. 6: 1386–1401.

Siebel, C. W., R. Kanaar, and D. C. Rio, 1994 Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. Genes Dev. 8: 1713–1725.

Siebel, C. W., A. Admon, and D. C. Rio, 1995 Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. Genes Dev. 9: 269–283.

Silva, J. C., E. L. Loreto, and J. B. Clark, 2004 Factors that affect the horizontal transfer of transposable elements. Curr. Issues Mol. Biol. 6: 57–71.

Simmons, G. M., 1992 Horizontal transfer of hobo transposable elements within the Drosophila melanogaster species complex: evidence from DNA sequencing. Mol. Biol. Evol. 9: 1050–1060.

Siomi, M. C., K. Saito, and H. Siomi, 2008 How selfish retrotransposons are silenced in Drosophila germline and somatic cells. FEBS Lett. 582: 2473–2478.

Song, J.-J., J. Liu, N. H. Tolia, J. Schneiderman, S. K. Smith *et al.*, 2003 The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat. Struct. Biol. 10: 1026–1032.

Stephan, W., Y. S. Song, and C. H. Langley, 2006 The hitchhiking effect on linkage disequilibrium between linked neutral loci. Genetics 172: 2647–2663.

Styhler, S., A. Nakamura, A. Swan, B. Suter, and P. Lasko, 1998 vasa is required for GURKEN accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. Development 125: 1569–1578.

Takahata, N., Y. Satta, and J. Klein, 1992 Polymorphism and balancing selection at major histocompatibility complex loci. Genetics 130: 925–938.

Vagin, V. V., M. S. Klenov, A. I. Kalmykova, A. D. Stolyarenko, R. N. Kotelnikov *et al.*, 2004 The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster. RNA Biol. 1: 54–58.

Van Valen, L. M., 1973 A new evolutionary law. Evolutionary Theory 1: 1–30.

Vermaak, D., S. Henikoff, and H. S. Malik, 2005 Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. PLoS Genet. 1: 96–108.

Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370.

Welch, J. J., 2006 Estimating the genomewide rate of adaptive protein evolution in Drosophila. Genetics 173: 821–837.

Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy *et al.*, 2007 A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8: 973–982.

Yan, K. S., S. Yan, A. Farooq, A. Han, L. Zeng *et al.*, 2003 Structure and conserved RNA binding of the PAZ domain. Nature 426: 468–474.

Yang, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586–1591.

Yano, T., S. López de Quinto, Y. Matsui, A. Shevchenko, A. Shevchenko *et al.*, 2004 Hrp48, a Drosophila hnRNPA/B homolog, binds and regulates translation of oskar mRNA. Dev. Cell 6: 637–648.

Communicating editor: R. Nielsen

Appendix

In this section, we describe the derivations of r, l, and μ. Parameters and variables used in the model are listed in Table A1.

The expected fitness of offspring of a pair of parents is

$$\bar{w}(m,u)=\sum_{n=0}^{\infty}\frac{e^{-m}m^n}{n!}\sum_{i=0}^{n_{\mathrm{HD}}}\frac{e^{-nu}(nu)^i}{i!}e^{-a(n+i)-(b/2)(n+i)^2}. \tag{A1}$$

m is the mean copy number of *P elements* of the parents. u, the transposition rate of the *P* element in the offspring, depends on the type of cross [$u=u_0$ in M(female) $\times$ P(male) dysgenic cross and $u=u_1$ in other crosses]. u also depends on the host allele passed to the offspring. With the assumption that the host locus is in complete linkage equilibrium with *P*-element insertions, the expected mean fitness of offspring for a specific type of cytotype cross (ccross) when considering the effect of the host beneficial allele is

$$\begin{aligned}\bar{w}_{\mathrm{ccross}}(m,u,l)=&\,l^2\bar{w}(m,u(1-d))+2l(1-l)\bar{w}(m,u(1-hd))\\&+(1-l)^2\bar{w}(m,u).\end{aligned} \tag{A2}$$

The expected mean fitness of offspring having a specific type of genotype (geno) when considering the effect of *P-element* cytotype is

Table A1 Parameters and varables used in the model

	Parameters
a, b	Selection coefficient against *P elements* through synergistic epistasis
d	Proportional reduction of *P*-element transposition rate in homozygotes of host beneficial allele
h	Dominance coefficient for host beneficial allele
u_0	*P*-element transposition rate per copy per generation in dysgenic cross
u_1	*P*-element transposition rate per copy per generation in nondysgenic cross
n_{HD}	Maximum number of new *P*-element insertions an individual can tolerate before becoming completely sterile
	Variables
n	*P*-element copy number in a given individual
m	Mean copy number of *P element* of parents
u	*P-element* transposition rate per copy per generation
δ	Proportional reduction of *P-element* transposition rate due to the host genotype
r_t	Proportion of individuals with *P cytotype* in the population at generation t
l_t	Allele frequency of the host beneficial allele in the population at generation t
μ_t	Average *P*-element copy number among individuals with *P cytotype* at generation t
$\bar{W}_t$	Mean fitness of the population at generation t
s	Selection coefficient against host nonbeneficial allele

$$\bar{w}_{\text{geno}}(m,\delta,r) = r^2\bar{w}(m,(1-\delta)u_1) + r(1-r)\bar{w}(m/2,(1-\delta)u_1) + r(1-r)\bar{w}(m/2,(1-\delta)u_0) + (1-r)^2, \tag{A3}$$

where δ is the proportional reduction in *P-element* transposition rate due to the genotype of the host locus, which equals d and hd in the homozygote and the heterozygote of the host beneficial allele.

r of the next generation, the relative mean fitness of offspring with *P cytotype* to that of the mean population fitness, can be expressed as

$$r_{t+1} = \frac{1}{\bar{W}_{t+1}}\left\{r_t^2\bar{w}_{\text{ccross}}(\mu_t,u_1,l_t) + r_t(1-r_t)\bar{w}_{\text{ccross}}\left(\frac{\mu_t}{2},u_1,l_t\right) + r_t(1-r_t)\bar{w}_{\text{ccross}}\left(\frac{\mu_t}{2},u_0,l_t\right) - \left[r_t^2e^{-\mu_t} + 2r_t(1-r_t)e^{-\mu_t/2}\right]\right\}, \tag{A4}$$

where the mean fitness of the population $\bar{W}$ is

$$\bar{W}_{t+1} = r_t^2\bar{w}_{\text{ccross}}(\mu_t,u_1,l_t) + r_t(1-r_t)\bar{w}_{\text{ccross}}\left(\frac{\mu_t}{2},u_1,l_t\right) + r_t(1-r_t)\bar{w}_{\text{ccross}}\left(\frac{\mu_t}{2},u_0,l_t\right) + (1-r_t)^2. \tag{A5}$$

The first three terms on the right side of (A4) are for $P \times P$, $P \times M$, and $M \times P$ (dysgenic) crosses. The average parental *P*-element copy number of $P \times M$ and $M \times P$ crosses is half of the average of that of all *P-cytotype* parents ($\mu/2$). The last substraction of (A4) is cases when the offspring inherited zero *P elements*, which will have *M* cytotype under the assumption that cytocype is determined by the presence/absence of *P elements*.

The frequency of the host beneficial allele of the next generation can be expressed as

$$l_{t+1} = \frac{1}{\bar{W}_{t+1}}\left[l_t^2\bar{w}_{\text{geno}}(\mu_t,(1-d),r_t) + l_t(1-l_t)\bar{w}_{\text{geno}}(\mu_t,(1-hd),r_t)\right]. \tag{A6}$$

The expected number of newly transposed *P elements* in offspring of a pair of parents is

$$\overline{\Delta n}(m,u) = \sum_{n=1}^{\infty}\frac{e^{-m}m^n}{n!}\sum_{i=0}^{n_{\text{HD}}} i\frac{e^{-nu}(nu)^i}{i!}e^{-a(n+i)-(b/2)(n+i)^2}. \tag{A7}$$

The expected number of newly transposed *P elements* of a specific cross when considering the effect of the host beneficial allele is then

$$\overline{\Delta n}_{\text{ccross}}(m,u,l) = l^2\overline{\Delta n}(m,u(1-d)) + 2l(1-l)\overline{\Delta n}(m,u(1-hd)) + (1-l)^2\overline{\Delta n}(m,u). \tag{A8}$$

The average *P-element* copy number among *P-cytotype* individuals in the next generation is thus

$$\mu_{t+1} = \mu_t + \frac{r_t^2\overline{\Delta n}_{\text{ccross}}(\mu_t,u_1,l_t) + r_t(1-r_t)\overline{\Delta n}_{\text{ccross}}(\mu_t/2,u_1,l_t)}{r_{t+1}} + \frac{r_t(1-r_t)\overline{\Delta n}_{\text{ccross}}(\mu_t/2,u_0,l_t)}{r_{t+1}}. \tag{A9}$$

Table A2 The effects of change in μ_0 and u_1 on $r_{0.99}$, l_{1000}, μ_{1000}, and max(s)

	μ_0				u_1		
μ_0 or u_1	5	10	20	30	10^{-3}	10^{-4}	10^{-5}
$r_{0.99}$ (gen)	87	87	88	88	78	87	88
l_{1000} (10^{-3})	1.312	1.331	1.331	1.313	1.189	1.331	1.349
μ_{1000}	6.650	6.451	6.451	6.650	9.698	6.451	6.297
max(s) (10^{-3})[a]	5.776	5.788	5.777	5.776	5.804	5.788	5.786

Other not varied parameters are set as $a = 10^{-5}$, $b = 10^{-6}$, $d = 0.5$, $h = 0.5$, $u_0 = 1$, $u_1 = 10^{-4}$, $\mu_0 = 10$, and $n_{\text{HD}} = 5$.

[a] The largest s within 1000 generations.

Table A3 The effects of change in d on $r_{0.99}$, l_{1000}, μ_{1000}, and max(s)

	h = 0.5					hd = 0.5			
d	0.2	0.3	0.5	0.7	0.9	0.3	0.5	0.7	0.9
$r_{0.99}$ (gen)	87	87	87	87	87	87	87	87	87
$l_{1000}(10^{-3})$	1.066	1.201	1.331	1.448	1.543	1.331	1.331	1.331	1.331
μ_{1000}	6.454	6.451	6.451	6.451	6.451	6.451	6.451	6.451	6.451
max(s)$(10^{-3})^a$	1.850	4.476	5.788	6.189	6.229	4.475	5.788	6.190	6.232

Other not varied parameters are set as $a = 10^{-5}$, $b = 10^{-6}$, $u_0 = 1$, $u_1 = 10^{-4}$, $\mu_0 = 10$, and $n_{HD} = 5$.
[a] The largest s within 1000 generations.

The selection coefficient against the host nonbeneficial allele is

$$s_{t+1} = 1 - \frac{\bar{W}_{aa}}{\bar{W}_{AA}} = \frac{\bar{w}_{\text{geno}}(\mu, 0, r_t)}{\bar{w}_{\text{geno}}(\mu, 1-d, r_t)}. \tag{A10}$$

We iterated over (A4), (A6), and (A9) over 1000–10,000 generations for each set of parameters chosen. With the assumption that *D. melanogaster* can have ~10 generations per year, there are ~1000 generations between when *D. melanogaster* was first collected as M strains and now. We thus mainly reported r, l, and μ at generation 1000 for most cases. One aspect of the model we are interested in is the time until *P-element* fixed in the population. As our model is deterministic, we set when $r \geq 0.99$ as the approximate time of fixation of the *P cytotype* in the population ($r_{0.99}$). In a finite population, genetic drift can expediate the fixation when r is nearly fixed in the population.

μ_0 (the average *P-element* copy number among individuals with *P cytotype* at generation zero) was tested for several values (5, 10, 20, and 30; Table A2), which did not influence the population dynamics or l_{1000} significantly. This is because only individuals with low copy number will have low probability of hybrid dysgenesis during the initial spread of the *P elements*. Transposition rates estimated for other TE families range from 10^{-5} to 10^{-4} (Nuzhdin and Mackay 1995; Nuzhdin *et al.* 1997; Maside *et al.* 2000) while old estimates for *P-element* transposition rate in a dysgenic cross are ~10^{-3} (Eggleston *et al.* 1988). Similarly, for μ_1 tested (10^{-5}, 10^{-4}, and 10^{-3}), no dramatic differences in population dyanmics were observed (Table A2). We tested the effect of changing d, the proportional reduction of *P-element* transposition rate in the homozygotes of host beneficial allele, with fixed h and fixed hd (Table A3). With fixed h, l increases as d increases (Table A3 and Figure A1), yet the change is slight. a and b were set as 10^{-5} and 10^{-6}, respectively, according to previous studies (Dolgin and Charlesworth 2006, 2008). Increasing these two parameters, and thus increasing the deleterious effects of *P-element* insertions, does not have significant effects on l_{1000} either (Table A4).

To sum up, for all a, b, d, h, u_1, and μ_0 tested, there were no significant changes in the dynamics of r, l, μ, and s over generations and the relationship between r and s. In neither case tested is there temporal allele frequency change of the host beneficial allele (l_0 and l_{1000}) that will be detectable with the usual size of samples. We thus fixed a, b, d, h, u_1, and μ_0 to 10^{-5}, 10^{-6}, 0.5, 0.5, 10^{-4}, and 10.

On the other hand, the combined effects of u_0 and n_{HD} did show strong impacts on the dynamics of the host beneficial allele as discussed in the main text and shown in Table A5. When either u_0 or n_{HD} and other parameters are changed at the

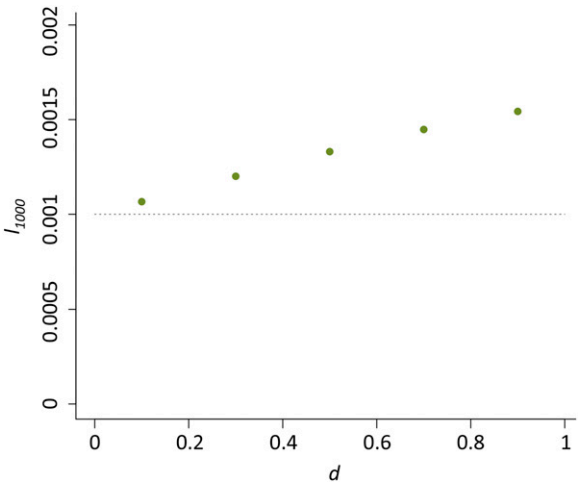

Figure A1 The impact of d on the allele frequency of the host beneficial allele at generation 1000 (l_{1000}). The dashed line is the beneficial allele frequency of host beneficial allele at generation zero (l_0).

Table A4 The effects of change in *a* and *b* on $r_{0.99}$, l_{1000}, μ_{1000}, and max(*s*)

	a			*b*		
a or *b*	10^{-5}	5×10^{-5}	10^{-4}	10^{-6}	5×10^{-6}	10^{-5}
$r_{0.99}$ (gen)	87	92	100	87	93	102
$l_{1000}(10^{-3})$	1.331	1.413	1.521	1.331	1.450	1.564
μ_{1000}	6.451	5.867	5.423	6.451	5.677	5.2829
max(*s*)$(10^{-3})^a$	5.788	5.784	5.779	5.788	5.786	5.783

Other not varied parameters are set as $u_0 = 1$, $u_1 = 10^{-4}$, $\mu_0 = 10$, and $n_{HD} = 5$.
[a] The largest s within 1000 generations.

same time, u_0 or n_{HD} is the determining factor of the population dynamics observed. For example, when the same values of a and b are tested with $u_0 = 0.1$, the observed population dynamics follow Case 3 below instead of Case 1 (see below and Tables A4 and A6). We discussed the three most characteristic cases in terms of the dynamics of r, l, μ, and s over generations and the relationship between r and s in the following.

Case 1: $u_0 = 1$, $n_{HD} \geq 3$ (Figure A2, case of $n_{HD} = 5$ is shown)

The dynamics of r, l, s, μ, and $\bar{W}$ (population mean fitness) over generations and the relationship between r and s are shown in Figure A2. The spread of the *P element* is fast (Figure A2A), even though it did lower the population mean fitness during its spread (Figure A2E). $\bar{W}$ started to bounce back when the proportion of *P cytotype* is intermediate due to the much reduced probability of hybrid dysgenesis. The change of s increased dramatically during the initial spread of *P elements* and then dropped as *P cytotype* is in high frequency in the population (Figure A2, C and F). This is also reflected in the dynamics of l, which has two phases of increase, with the former having a faster rate (Figure A2B).

Case 2: $u_0 = 1$, $n_{HD} = 2$ (Figure A3)

This is the exception of Case 1. The spread of *P elements* in the population is slowed due to the high probability of hybrid dysgenesis even when the majority of the population has *P cytotype* (Figure A3A). Because of this delay in *P-elements* spread, there is an extended period when the host repressing allele has large s (Figure A3C). However, several aspects of the population dynamics under this parameters are not consistent with known *P-element* biology (see main text). In a finite population, genetic drift may shorten the time of fixation for *P cytotype*, resulting in a shorter period for the repressing allele to have large selective benefits from lingering incidences of hybrid dysgenesis.

Case 3: $u_0 = 0.1$ and $n_{HD} \geq 2$ (Figure A4, case of $n_{HD} = 5$ is shown)

Because of the lower u_0, the spread of the *P element* is slower than the case when $u_0 = 1$ (Figure A4A) and the duration for the host beneficial allele to have large selective benefit is therefore longer (Figure A4C). However, the absolute magnitude of the host allele selection benefit is much lower due to the much reduced probability of hybrid dysgenesis. The increase of μ is also delayed until the proportion of *P cytotype* is intermediate in the population (Figure A4D), which may explain why the

Table A5 The effects of change in u_0 and n_{HD} on $r_{0.99}$, l_{1000}, μ_{1000}, and max(*s*)

$u_0 = 1$					
n_{HD}	2	3	5	7	9
$r_{0.99}$ (gen)	>10,000	3,787	87	64	57
$l_{1000}(10^{-3})$	592.2	4.701	1.331	1.093	1.016
μ_{1000}	3.441	4.637	6.451	6.994	7.261
max(*s*)[a]	2.590×10^{-2}	1.688×10^{-2}	5.788×10^{-3}	1.503×10^{-3}	1.380×10^{-4}
$u_0 = 0.1$					
n_{HD}	2	3	5	7	9
$r_{0.99}$ (gen)	687	628	621	621	621
$l_{1000}(10^{-3})$	1.032	1.003	1.000	1.000	1.000
μ_{1000}	5.138	5.240	5.255	5.255	5.255
max(*s*)[a]	1.366×10^{-4}	1.093×10^{-5}	7.066×10^{-8}	1.092×10^{-9}	1.337×10^{-12}

Other not varied parameters are set as $a = 10^{-5}$, $b = 10^{-6}$, $d = 0.5$, $h = 0.5$, $u_1 = 10^{-4}$, and $\mu_0 = 10$.
[a] The largest s within 1000 generations.

Table A6 The effects of change in *a* and *b* on $r_{0.99}$, l_{1000}, μ_{1000}, and max(*s*)

a			
a or *b*	10^{-5}	5×10^{-5}	10^{-4}
$R_{0.99}$	621	>1000	>1000
$l_{1000}(10^{-3})$	1.000	1.000	1.000
μ_{1000}	5.255	4.259	3.139
max(*s*)	7.066×10^{-8}	7.060×10^{-8}	7.018×10^{-8}
b			
a or *b*	10^{-6}	5×10^{-6}	10^{-5}
$r_{0.99}$	621	741	>1000
$l_{1000}(10^{-3})$	1.000	1.000	1.000
μ_{1000}	5.255	4.973	4.237
max(*s*)[a]	7.066×10^{-8}	7.063×10^{-8}	7.053×10^{-8}

Other not varied parameters are set as $u_0 = 0.1$, $u_1 = 10^{-4}$, $\mu_0 = 10$, and $n_{HD} = 5$.
[a] The largest s within 1000 generations.

largest *s* happened when *r* is relatively large (Figure A4E). The population mean fitness does not bounce back, suggesting that the increase in *P-element* copy number after fixation of *P cytotype* led to stronger fitness reduction than hybrid dysgenesis during the *P-element* spread.

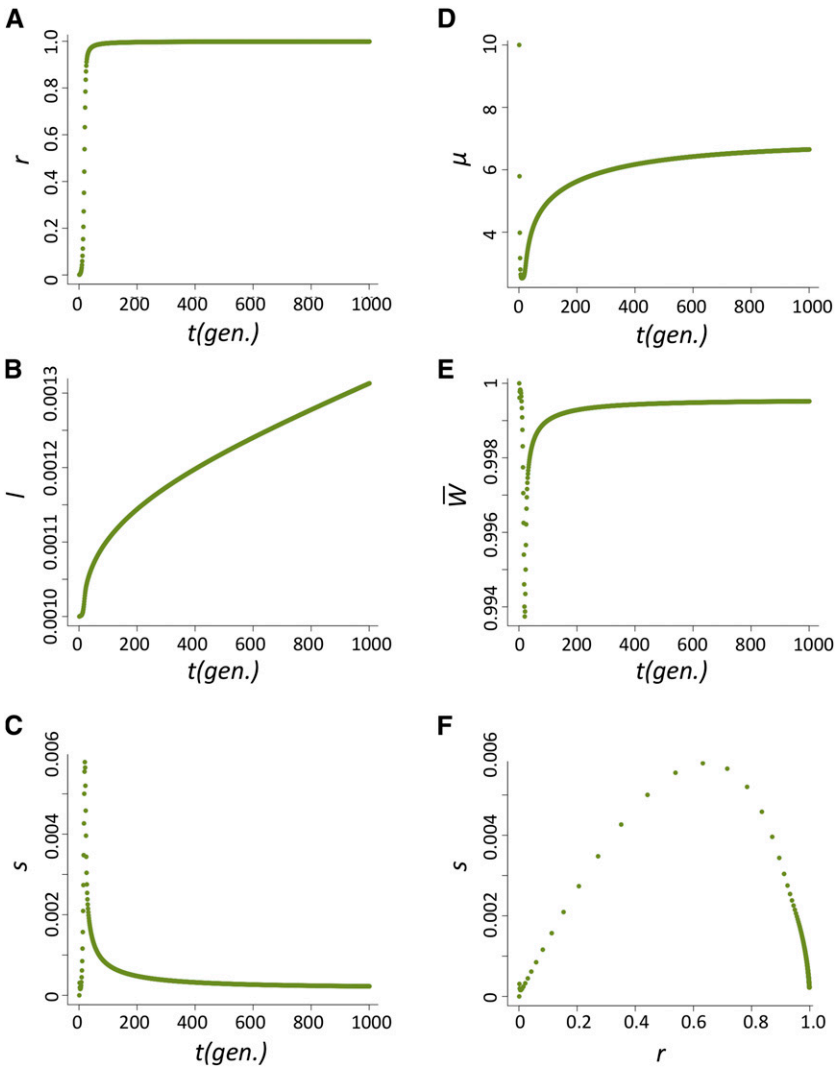

Figure A2 The dynamics of *r* (A), *l* (B), *s* (C), μ (D), and $\bar{W}$ (E) over generations when $u_0 = 1$ and $n_{HD} = 5$. (F) describes how *s* changes as *r* varies.

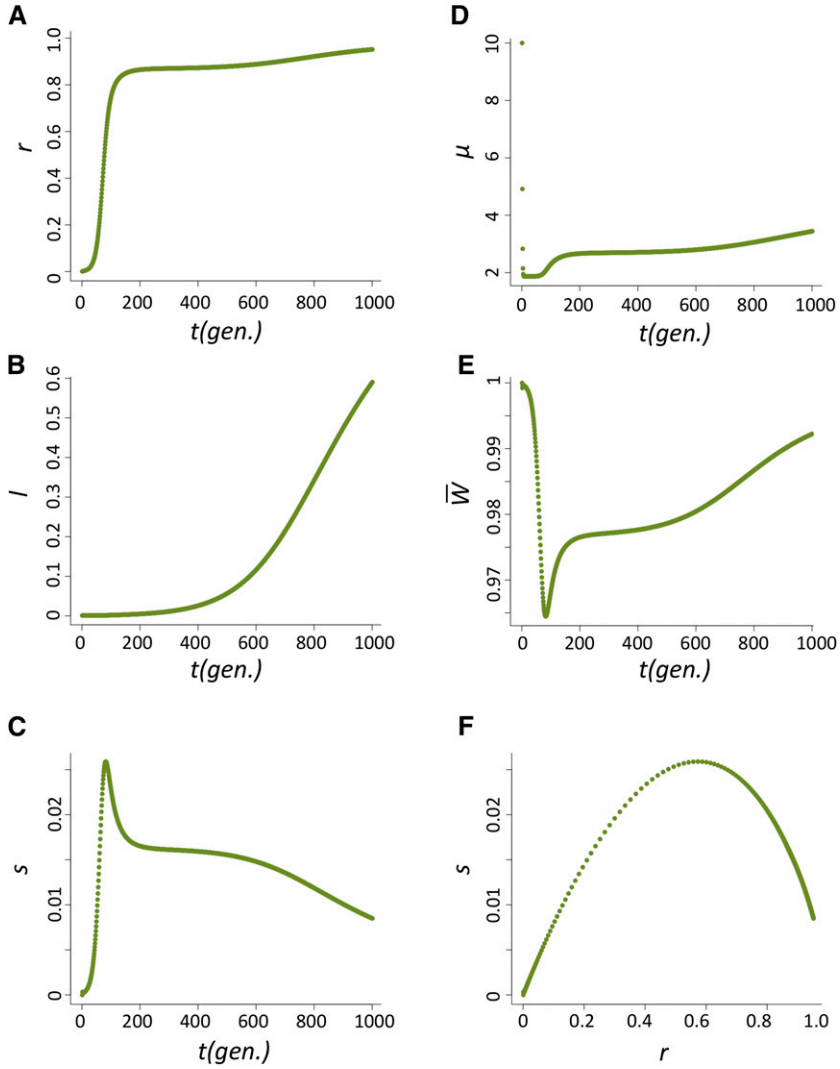

Figure A3 The dynamics of r (A), l (B), s (C), μ (D), and $\bar{W}$ (E) over generations when $u_0 = 1$ and $n_{HD} = 2$. (F) describes how s changes as r varies.

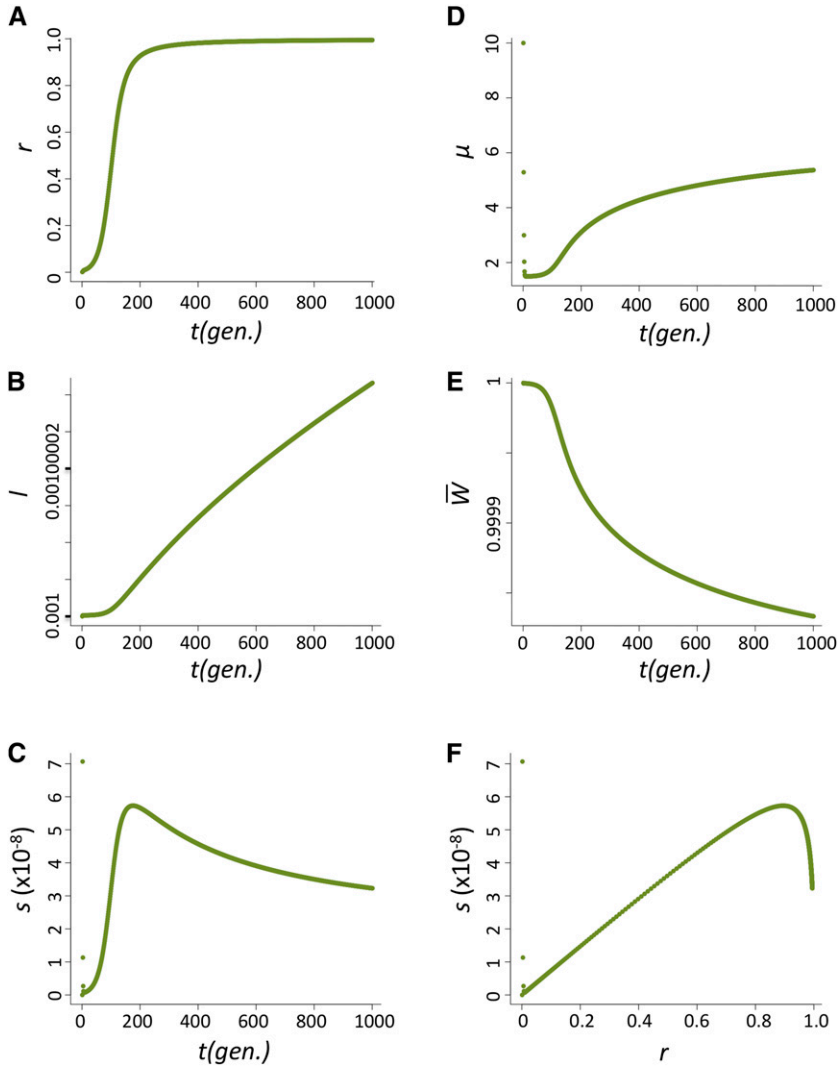

Figure A4 The dynamics of r (A), l (B), s (C), μ (D), and $\bar{W}$ (E) over generations when $u_0 = 0.1$ and $n_{HD} = 5$. (F) describes how s changes as r varies.

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.145714/-/DC1

Long-Term and Short-Term Evolutionary Impacts of Transposable Elements on *Drosophila*

Yuh Chwen G. Lee and Charles H. Langley

DOI: 10.1534/genetics.112.145714

File S1

Supporting Text: Evaluation of the *Poisson* assumption of the analytical model

We modeled the distribution of TE copy number as *Poisson,* which has been well established in the literature. However, this approximation holds true mainly when the TE population is at near equilibrium and the TE copy number is large. The key part of our model is the spread of a newly invaded TE family, during which the TE population is not at equilibrium and the copy number may be low. To investigate how the deviation from *Poisson* approximation may influence the predictions of our analytical models, we performed full Monte Carlo simulations to evaluate the potential impacts of this assumption.

Monte Carlo Simulations

We used the following Monte Carlo simulation to address this issue. The host population size is 100,000. Each host individual genome is comprised of two parental complements of three chromosomes, each of which has 1,000 potential TE insertion sites and a host locus. Crossover is modeled as *Poisson* process and the crossover rate is set as 0.001 between two potential TE insertion sites, making it averagely one crossover per chromosome per generation. At generation zero, the 0.1% of the population contain on average μ_0 copies of the TE (distribution is *Poisson*). Independently chosen 0.1% of the population have the beneficial allele at the host locus.

A new member of the next generation is simulated through the following steps. Two parents are first chosen and each parent contributes a haploid genome to the offspring (assuming independent assortment and crossing over as described above). Neither TE insertions nor the host locus influence the transmission. Each TE insertion of the offspring then independently undergoes a single replicative transposition with probability equal to the transposition rate *u*, which changes according to the cytotype of the parent (u_0 in hybrid dysgenic cross and u_1 in the other crosses) and the host locus genotype of the offspring (*u(1-d)* in homozygotes of beneficial allele, *u(1-hd)* in heterozygotes of beneficial allele and *u* for the other genotype). If the total number of transposition events in an offspring is above the hybrid dysgenic threshold (*HD*), the offspring's fitness is set to zero and the offspring is not passed to the next generation. If the total number of transposition events in an offspring is below the threshold, its fitness is calculated according to the following equation ($w(n) = e^{-an-bn^2/2}$, where *n* is the total TE copy number and *a* and *b* are 10^{-5} and 10^{-6} respectively). The offspring is transmitted to the next generation with probability equal to it fitness. This process is repeated until 100,000 offspring are generated.

According to the analyses of the analytical modeling, following parameters did not have significant impacts on the dynamics of *I* and were chosen as follows for the simulation: $d = 0.5$, $h = 0.5$, $u_1 = 10^{-4}$ and $\mu_0 = 10$. As discussed in the main text, the spread of newly invaded TE family has almost no impacts on the host gene for cases where u_0 equals 0.1, which is of course less interesting case for our analysis. We thus chose $u_0 = 1$ for our simulation. We did pilot simulations with n_{HD} equals 3, 5, 7, and 10 and found no apparent differences (data not shown) and thus only the case with the greater numbers of simulations, $n_{HD} = 5$, are presented below.

Results of Simulations

Following figures showed the averaged result of 1,000 Monte Carlo simulations for proportion of *P cytotype* (Figure S1), TE copy number (Figure S2) and the frequency of host beneficial allele (Figure S3 and Figure S4), comparing with the prediction of analytical model. The most critical part of our analytical model is from the invasion of the newly invaded TE family to its reaching equilibrium in the population, which takes approximately 100 generations after its first invasion.

Simulations showed the spread and the increase in copy number of the newly invaded TE family is slower than the analytical prediction (Figure S1 and S2). The allele frequency predicted by the analytical model based on the assumed *Poisson* distribution of copy number tends to initially exceed then fall below the simulated host allele frequency (Figure S3). However, the error between analytical approximation and the simulation is always within 2% (Figure S4). Thus, our overall conclusion that the spread of a newly invaded TE family is unlikely to drive the fast evolution of interacting host genes is not sensitive to the naïve assumptions of the analytic model.

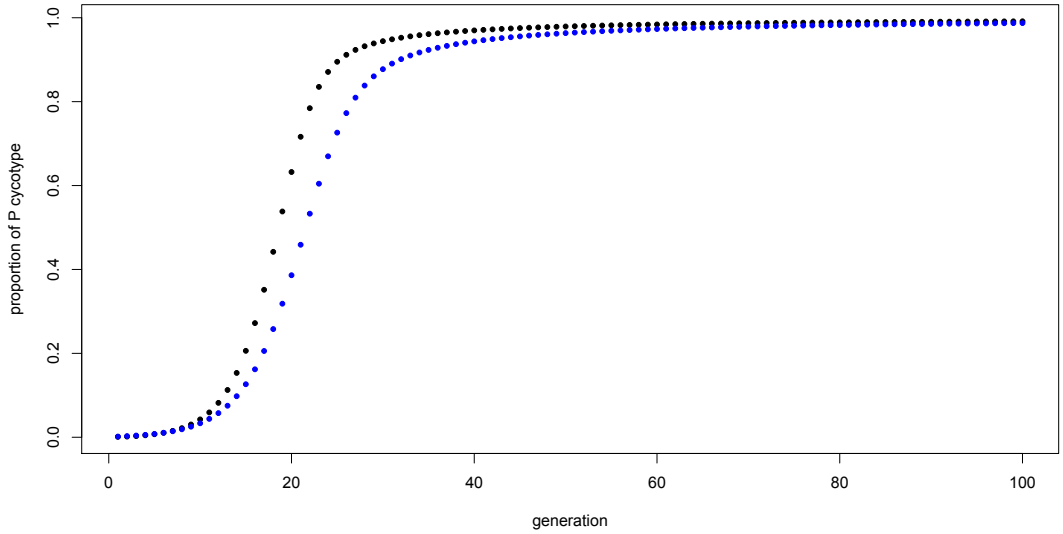

Figure S1 The proportion of *P cytotype* individuals over time. Black and blue dots are the analytical prediction and simulation results respectively.

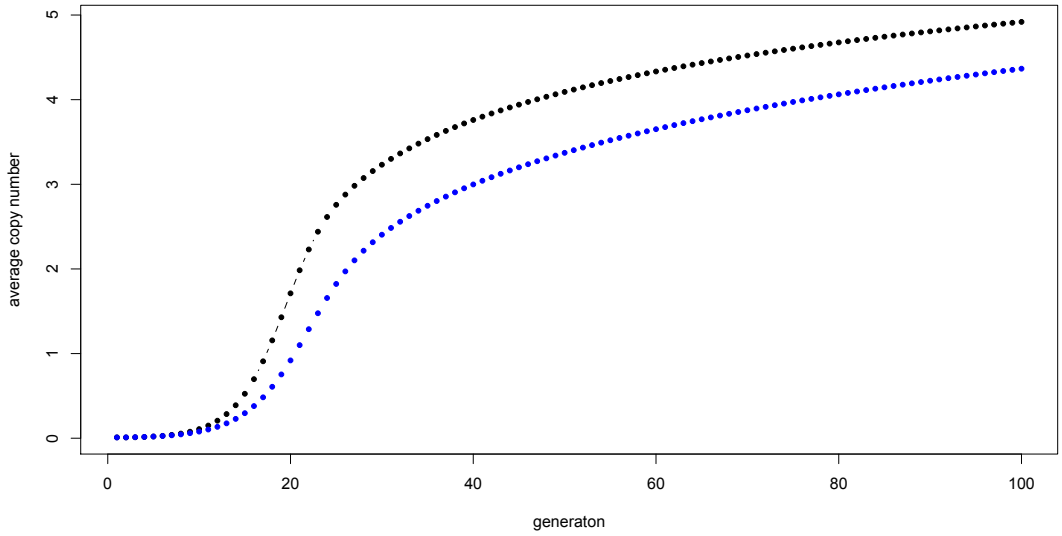

Figure S2 The averaged TE copy number over time. Black and blue dots are the analytical prediction and simulation results respectively.

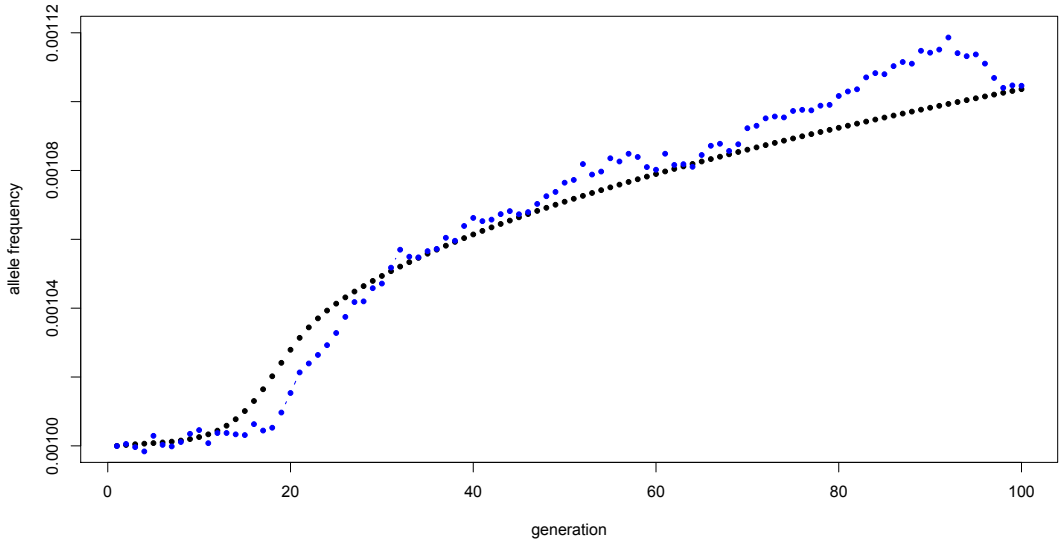

Figure S3 The frequency of host beneficial allele over time. Black and blue dots are the analytical prediction and simulation results respectively.

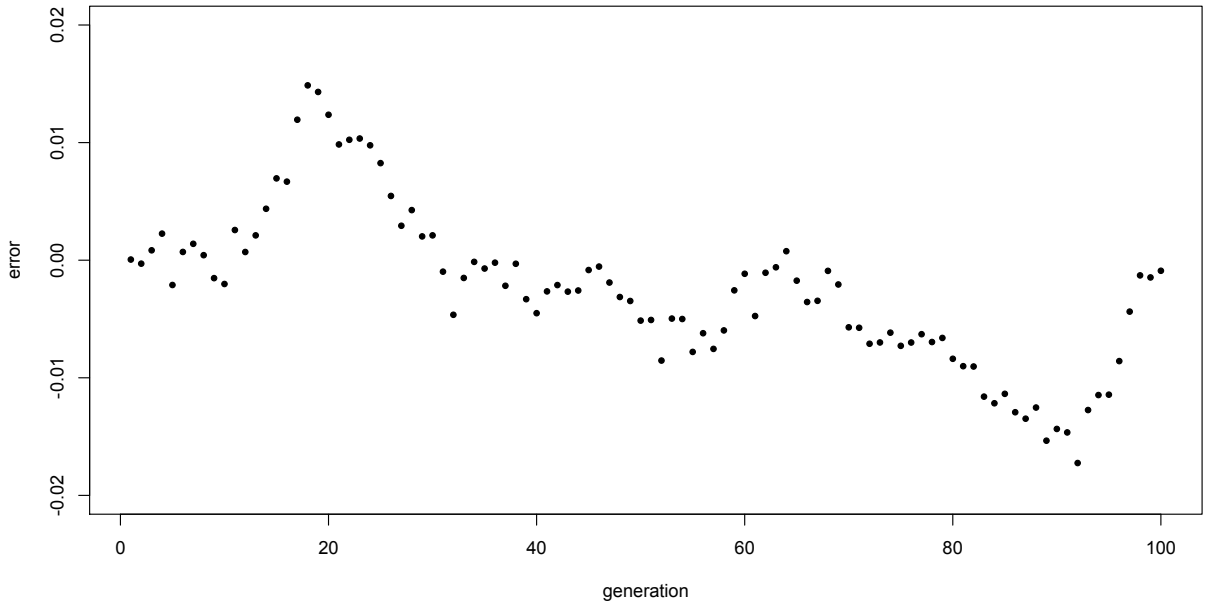

Figure S4 The errors of analytical prediction of the host beneficial allele with respect to simulations.

During the initial invasion phase, the *Poisson* distribution predicts a distribution that has larger mode than the actual simulation (Figure S5A, B). Soon after *the P cytotype* individuals in the population become common (≈ generation 35), the TE copy number distribution is nearly *Poisson* (Figure S5C, D). In addition to the fact that the *Poisson* distribution is a good approximation to the Binomial sampling when the TE copy number is large, the linkage among TE insertions also contributes to the differences between the predictions of analytical model and simulations. In simulations where there is free recombination among TE insertions, the distribution of TE copy number quickly reaches *Poisson* within 15 generations, when the *P cytotype* in the population is still rare (results not shown).

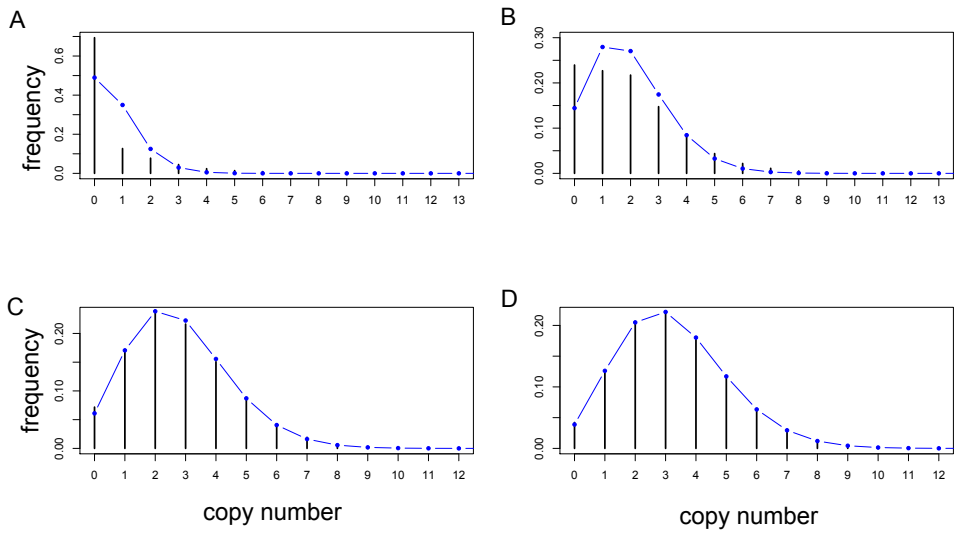

Figure S5 Distribution of TE copy number among host individuals at generation 15 (A), 25 (B), 35 (C) and 45 (D). Black bars are the simulated values while the blue dots are the *Poisson* expectation. The distribution of TE copy number reaches nearly *Poisson* around generation 35, when the *P cycotype* start being common in the population.

Our analytical model initially overestimates the host allele frequency. This is potentially caused by that fact that the *Poisson* approximation has a larger mode than the real copy number distribution. Because the copy number of an individual is generally small and the probability of hybrid dysgenic crosses happening is low, this did not lead to sever deviation between analytical predictions and the actual simulations.

After the *P element* in the population becomes common (around generation 35), our analytical model starts to predict lower host allele frequencies. This could be attributable to the transient linkage disequilibrium among TE insertions in the simulations. In this case, the simulation has a heavier right tail than the expectation from the *Poisson* approximation. The following tables (Table S5, 6, 7) show the proportion of simulations that have more individuals with a particular copy number than the predictions of the *Poisson* approximation. This proportion is universally greater than 50% for individuals with larger copy number, whose offspring are likely to have too many TE transpositions in a single generation. This can lead to stronger selection and a slightly higher host allele frequency change than the analytical model.

Table S5 The proportion of simulations that have more individuals with a particular copy number than the *Poisson* predictions at generation 40. This is the generation when the analytical model starts to underpredict the host allele frequency. The proportion of simulations that have more individuals with a particular copy number than the *Poisson* predictions are shown in the "Proportion" row, with proportion greater than 50% highlighted in blue. Individuals with large *P element* copy number may not present in all simulations and thus the "No Simulations" may not always be 1,000.

Copy Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
No Simulations	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	998	825	361	89	25	3	1
Proportion	1	0.017	0.06	0	0	0.034	0.968	1	1	1	1	0.992	0.92	0.976	1	1	1	1	1

Table S6 The proportion of simulations that have more individuals with a particular copy number than the *Poisson* predictions at generation 50.

Copy Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
No Simulations	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	979	600	215	40	7	3	2
Proportion	1	0.011	0.701	0.493	0.134	0.072	0.271	0.833	0.985	0.979	0.906	0.811	0.692	0.618	1	1	1	1	1	1

Table S7 The proportion of simulations that have more individuals with a particular copy number than the *Poisson* predictions at generation 60.

Copy Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	22
No Simulations	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	999	839	354	100	32	6	1
Proportion	0.969	0.001	0.661	0.854	0.646	0.3	0.233	0.362	0.605	0.652	0.618	0.537	0.517	0.48	0.613	1	1	1	1	1

Table S1 Sample locations of M strains

Continent	Country	specific position	name	stock number	Source
Africa	South Africa	Capetown	CA1	3846	BDSC[a]
	Zimbabwe	Kariba Dam	KSA2	3852	BDSC
		Kariba Dam	KSA3	3853	BDSC
		Kariba Dam	KSA4	3854	BDSC
Northern America	USA	South Carolina	Wild 10E	3892	BDSC
		North Carolina	Wild 11A	3893	BDSC
		North Carolina	Wild 11C	3894	BDSC
		North Carolina	Wild 11D	3895	BDSC
		New York	EV	3851	BDSC
		New York	MO1	3857	BDSC
		New York	Wild 1B	3880	BDSC
		Wisconsin	MWA1	3859	BDSC
		Wisconsin	Lausanne	4268	BDSC
		Ohio	Canton-S	1	BDSC
		Massachusetts	Amherst	4265	BDSC
		Illinoi	Urbana-S	4272	BDSC
		Oregon	Oregon-R	5	BDSC
		Riverside	RC1	3865	BDSC
		Riverside	RVC2	3869	BDSC
South America	Columbia	Bogata	BOG3	3843	BDSC
Asia	Japan	Iriomote island	IR98-01		M Itoh
		Iriomote island	IR98-06		M Itoh
		Hikone	HikoneR		M Itoh
	Russia	Uzbek Republic	Samarkand	4270	BDSC
Europe	Spain	Pyrenees	PYR3	3863	BDSC
	Portugal	Madeira	Reids1	3866	BDSC
	Greece	Athens	VAG2	3876	BDSC
		Athens	VAG3	3877	BDSC
	Sweden	Stockholm	Swedish-C	4271	BDSC

[a] Bloomington Drosophila Stock Center

Table S2 Linage-specific divergence on the *D. melanogaster* and *D. simulans* branches of TE-interacting and immunity genes

gene name	FBgn	*mel dN/dS*	*mel dN*	*mel dS*	*sim dN/dS*	*sim dN*	*sim dS*	functional class
Ago3	FBgn0250816	0.2614	0.0147	0.0563	0.2648	0.0208	0.0785	piRNA gene
armi	FBgn0041164	0.2973	0.0154	0.0518	0.3905	0.0222	0.0569	piRNA gene
aub	FBgn0000146	0.3824	0.0263	0.0687	0.5301	0.0338	0.0637	piRNA gene
krimp	FBgn0034098	0.4705	0.0342	0.0728	0.9641	0.0602	0.0625	piRNA gene
mael	FBgn0016034	0.9024	0.0324	0.0359	0.4907	0.0267	0.0545	piRNA gene
Hen1	FBgn0033686	0.2464	0.0159	0.0644	0.2658	0.0192	0.0721	piRNA gene
piwi	FBgn0004872	0.0807	0.0067	0.0836	0.1935	0.0093	0.0483	piRNA gene
rhi	FBgn0004400	1.4145	0.0764	0.054	0.5077	0.0556	0.1096	piRNA gene
Spn-E	FBgn0003483	0.2186	0.014	0.0642	0.2812	0.0166	0.0591	piRNA gene
squ	FBgn0002652	0.2399	0.0228	0.095	0.3376	0.0496	0.1469	piRNA gene
vas	FBgn0262526	0.2425	0.0453	0.1867	0.306	0.0437	0.143	piRNA gene
zuc	FBgn0261266	0.3585	0.0353	0.0985	0.2068	0.0214	0.1033	piRNA gene
Hrb27C	FBgn0004838	0.176	0.0021	0.0117	0.019	0.001	0.0538	P element gene
Irbp	FBgn0011774	0.1009	0.0088	0.0875	0.1635	0.0154	0.0944	P element gene
Ku80	FBgn0041627	0.1805	0.0125	0.069	0.245	0.0157	0.0642	P element gene
Psi	FBgn0014870	0.0908	0.0052	0.0575	0.0267	0.0011	0.0401	P element gene
AttA	FBgn0012042	0.0851	0.0151	0.178	0.1005	0.013	0.1294	effector
AttB	FBgn0041581	0.0351	0.0024	0.0677	0.0675	0.0062	0.0915	effector
AttC	FBgn0000276	0.0544	0.0085	0.1555	0.1263	0.0062	0.0489	effector
AttD	FBgn0038530	0.0948	0.0081	0.0851	0.3633	0.0158	0.0435	effector
Catsup	FBgn0002022	0.1012	0.009	0.0886	0.2507	0.0166	0.0662	effector
CecA1	FBgn0000276	0.0001	0	0.0648	NA	NA	NA	effector
CecA2	FBgn0000277	NA	NA	NA	NA	NA	NA	effector
CecB	FBgn0000278	0.0001	0	0.1231	0.3626	0.0069	0.0191	effector
CecC	FBgn0000279	0.0001	0	0.1431	NA	NA	NA	effector
CG11159	FBgn0034539	0.0001	0	0.0766	NA	NA	NA	effector
CG14823	FBgn0035734	0.1411	0.0102	0.072	0.188	0.0033	0.0176	effector
CG15293	FBgn0028526	0.4002	0.0353	0.0882	0.5249	0.0262	0.0499	effector
CG15825	FBgn0032773	0.1421	0.0152	0.1068	0.228	0.0072	0.0315	effector
CG16756	FBgn0029765	0.0854	0.0084	0.0985	0.1899	0.0173	0.0911	effector
CG16799	FBgn0034538	0.0678	0.0074	0.1092	0.1242	0.0047	0.0379	effector
CG18107	FBgn0034330	0.5364	0.0443	0.0825	0.0001	0	0.0543	effector
CG33470	FBgn0053470	NA	NA	NA	NA	NA	NA	effector
CG6421	FBgn0025827	0.1022	0.0101	0.0985	0.1508	0.0079	0.0523	effector
CG6426	FBgn0034162	0.062	0.0037	0.06	0.377	0.1255	0.3328	effector

CG6429	FBgn0046999	0.1226	0.0122	0.0999	0.0681	0.0099	0.1458	effector
CG6435	FBgn0034165	0.0001	0	0.0785	0.0581	0.0027	0.0469	effector
CG7798	FBgn0034092	0.0189	0.0029	0.1545	0.0001	0	0.0682	effector
CG8193	FBgn0033367	0.0235	0.0024	0.1001	0.0734	0.0089	0.1215	effector
CG8492	FBgn0035813	0.127	0.007	0.0552	0.1144	0.0069	0.0601	effector
Ddc	FBgn0000422	0.0184	0.0017	0.0904	0.0001	0	0.0413	effector
Def	FBgn0010385	0.1745	0.0151	0.0865	0.0001	0	0.0265	effector
Dpt	FBgn0004240	0.1002	0.011	0.1096	NA	NA	NA	effector
DptB	FBgn0034407	0.0811	0.0079	0.097	0.2091	0.0078	0.0373	effector
Dro	FBgn0010388	0.2487	0.0149	0.0601	0.3377	0.0144	0.0426	effector
Dro-2	FBgn0052279	0.0001	0	0.0665	0.1467	0.0124	0.0846	effector
Dro-3	FBgn0052283	NA	NA	NA	NA	NA	NA	effector
Dro-4	FBgn0052282	NA	NA	NA	0.0001	0	0.0186	effector
Dro-5	FBgn0035434	NA	NA	NA	0.0001	0	0.0263	effector
Dro6	FBgn0052268	0.2098	0.0123	0.0585	0.0676	0.006	0.0892	effector
Drs	FBgn0010381	NA	NA	NA	NA	NA	NA	effector
Drs-l	FBgn0052274	1.0897	0.0367	0.0337	0.0001	0	0.0615	effector
Duox	FBgn0031464	0.0028	0.0003	0.0986	0.0147	0.0011	0.0749	effector
Hml	FBgn0029167	0.0503	0.0046	0.0906	0.1561	0.0118	0.0757	effector
IM1	FBgn0034329	0.08	0.0108	0.1349	0.0001	0	0.041	effector
IM10	FBgn0033835	0.1533	0.0105	0.0686	0.2858	0.0222	0.0776	effector
IM2	FBgn0025583	0.0001	0	0.0917	0.2206	0.0109	0.0495	effector
IM23	FBgn0034328	NA	NA	NA	NA	NA	NA	effector
IM3	FBgn0040736	0.1634	0.0117	0.0717	0.0001	0	0.0339	effector
IM4	FBgn0040653	0.0001	0	0.0729	NA	NA	NA	effector
Irc	FBgn0038465	0.0771	0.0061	0.0796	0.0988	0.0069	0.0696	effector
Jafrac1	FBgn0040309	0.0001	0	0.1406	0.0001	0	0.0583	effector
Jafrac2	FBgn0040308	0.0001	0	0.0951	0.0001	0	0.0606	effector
LysB	FBgn0004425	NA	NA	NA	NA	NA	NA	effector
LysC	FBgn0004426	NA	NA	NA	NA	NA	NA	effector
LysD	FBgn0004427	NA	NA	NA	NA	NA	NA	effector
LysE	FBgn0004428	NA	NA	NA	NA	NA	NA	effector
LysP	FBgn0004429	NA	NA	NA	NA	NA	NA	effector
LysS	FBgn0004430	NA	NA	NA	NA	NA	NA	effector
LysX	FBgn0004431	0.273	0.0168	0.0614	0.1854	0.0122	0.066	effector
Mtk	FBgn0014865	0.0001	0	0.0246	0.178	0.0089	0.05	effector
ple	FBgn0005626	0.0357	0.0021	0.0598	0.0225	0.0007	0.0312	effector
Pu	FBgn0003162	0.097	0.004	0.0414	0.0001	0	0.0376	effector
Tig	FBgn0011722	0.0331	0.0043	0.1301	0.0573	0.0047	0.0822	effector

TotA	FBgn0028396	0.7779	0.0348	0.0447	0.6271	0.0222	0.0354	effector
TotB	FBgn0038838	0.435	0.0225	0.0517	0.5846	0.0278	0.0475	effector
TotC	FBgn0044812	NA	NA	NA	NA	NA	NA	effector
TotE	FBgn0053117	0.2167	0.0246	0.1137	0.1744	0.0087	0.0502	effector
TotF	FBgn0044811	NA	NA	NA	NA	NA	NA	effector
TotM	FBgn0031701	0.3909	0.0449	0.1148	0.2715	0.0193	0.0713	effector
TotX	FBgn0044810	0.3709	0.0247	0.0667	0.8708	0.0245	0.0282	effector
TotZ	FBgn0044809	0.0932	0.0033	0.0356	0.0001	0	0.0377	effector
Tsf1	FBgn0022355	0.0331	0.003	0.0915	0.0936	0.0113	0.1207	effector
Tsf2	FBgn0036299	0.0267	0.0021	0.0791	0.0272	0.002	0.0749	effector
Tsf3	FBgn0034094	0.0313	0.0029	0.0938	0.1219	0.007	0.0571	effector
yellow-f	FBgn0041710	0.1742	0.0137	0.0786	0.12	0.0141	0.1174	effector
yellow-f2	FBgn0038105	0.0778	0.0071	0.0907	0.1282	0.0078	0.061	effector
CG12780	FBgn0033301	0.1617	0.0159	0.0983	0.5244	0.0169	0.0323	recognition
CG13079	FBgn0032808	0.7221	0.0516	0.0715	0.4602	0.0294	0.0639	recognition
CG13422	FBgn0034511	0.1101	0.0176	0.1599	0.1177	0.0054	0.0456	recognition
CG30148	FBgn0050148	0.1991	0.0225	0.1131	0.463	0.0297	0.0642	recognition
CG31217	FBgn0051217	0.8493	0.0087	0.0103	0.1307	0.0068	0.0523	recognition
CG3212	FBgn0031547	0.459	0.0292	0.0636	0.463	0.0212	0.0458	recognition
CG6124	FBgn0243514	0.6973	0.048	0.0688	1.0046	0.0277	0.0276	recognition
Corin	FBgn0033192	0.0462	0.0035	0.0755	0.0799	0.0061	0.076	recognition
crq	FBgn0015924	0.1571	0.0088	0.0563	0.0875	0.0071	0.0812	recognition
emp	FBgn0010435	0.0001	0	0.0659	0.076	0.0032	0.0416	recognition
GNBP1	FBgn0040323	0.1069	0.0076	0.0706	0.0308	0.0018	0.0598	recognition
GNBP2	FBgn0040322	0.0698	0.0076	0.1086	0.1458	0.0057	0.039	recognition
GNBP3	FBgn0040321	0.0933	0.0064	0.0683	0.1986	0.0071	0.0356	recognition
He	FBgn0028430	1.0548	0.0873	0.0827	0.4288	0.0561	0.1307	recognition
Mcr	FBgn0020240	0.0186	0.0015	0.0826	0.0325	0.0015	0.0476	recognition
NimA	FBgn0261514	0.0459	0.0053	0.1156	0.0891	0.0064	0.072	recognition
NimB1	FBgn0027929	0.1212	0.0147	0.1209	0.2102	0.0112	0.0533	recognition
NimB2	FBgn0028543	0.0589	0.0051	0.0868	0.0232	0.002	0.084	recognition
NimB3	FBgn0054003	0.0799	0.0053	0.0663	0.3395	0.0163	0.048	recognition
NimB4	FBgn0028542	0.1125	0.0123	0.1092	0.1379	0.0168	0.1217	recognition
NimB5	FBgn0028936	0.1293	0.0068	0.0528	0.1955	0.0104	0.053	recognition
NimC1	FBgn0259896	0.5789	0.0335	0.0579	0.6331	0.0346	0.0547	recognition
NimC2	FBgn0028939	0.0964	0.0072	0.0746	0.1102	0.0052	0.0469	recognition
NimC3	FBgn0001967	0.1099	0.0098	0.0888	0.0001	0	0.0836	recognition
NimC4	FBgn0260011	0.0666	0.0081	0.1224	0.0627	0.0046	0.0732	recognition
pes	FBgn0031969	0.16	0.0085	0.0531	0.1808	0.0352	0.1947	recognition

PGRP-LA	FBgn0035975	0.1364	0.0113	0.0825	0.1091	0.0085	0.0778	recognition
PGRP-LB	FBgn0037906	0.1298	0.0073	0.056	0.051	0.0037	0.0731	recognition
PGRP-LC	FBgn0035976	0.1748	0.0122	0.07	0.1985	0.0119	0.0598	recognition
PGRP-LD	FBgn0260458	0.2447	0.01	0.0408	0.2587	0.0204	0.0787	recognition
PGRP-LE	FBgn0030695	0.0769	0.0052	0.0676	0.029	0.0025	0.0858	recognition
PGRP-LF	FBgn0035977	0.2456	0.0237	0.0966	0.2295	0.0157	0.0686	recognition
PGRP-SA	FBgn0030310	0.113	0.0106	0.0939	0.1491	0.0051	0.034	recognition
PGRP-SB1	FBgn0043578	0.03	0.0037	0.1248	0.274	0.0109	0.0397	recognition
PGRP-SB2	FBgn0043577	0.1484	0.0156	0.1051	0.0573	0.0022	0.0381	recognition
PGRP-SC1a	FBgn0043576	NA	NA	NA	NA	NA	NA	recognition
PGRP-SC1b	FBgn0033327	0.0001	0	0.1259	0.0314	0.002	0.0639	recognition
PGRP-SC2	FBgn0043575	0.0114	0.0022	0.1892	0.1564	0.0043	0.0277	recognition
PGRP-SD	FBgn0035806	0.1054	0.0076	0.072	0.3752	0.0073	0.0193	recognition
Sr-CI	FBgn0014033	0.3234	0.0428	0.1325	0.3326	0.0333	0.1003	recognition
Sr-CII	FBgn0020377	0.2926	0.0165	0.0562	0.1487	0.0087	0.0587	recognition
Sr-CIII	FBgn0020376	0.9783	0.0355	0.0363	0.3033	0.0136	0.0448	recognition
TepI	FBgn0041183	0.6271	0.0459	0.0732	0.7154	0.0324	0.0453	recognition
TepII	FBgn0041182	0.2017	0.0182	0.0902	0.2245	0.0169	0.0752	recognition
TepIII	FBgn0041181	0.0966	0.0053	0.0547	0.1046	0.0053	0.0507	recognition
TepIV	FBgn0041180	0.1581	0.0109	0.0692	0.2125	0.0081	0.0381	recognition
18w	FBgn0004364	0.0054	0.0004	0.0685	0.0241	0.0013	0.0521	signaling
Alk	FBgn0040505	0.0155	0.0013	0.0819	0.0148	0.0009	0.0599	signaling
aop	FBgn0000097	0.0332	0.0013	0.0383	0.1779	0.0045	0.0254	signaling
Atf-2	FBgn0050420	0.1091	0.0077	0.0703	0.1023	0.006	0.0588	signaling
ben	FBgn0000173	0.0001	0	0.0651	0.0001	0	0.0595	signaling
BG4	FBgn0038928	0.4744	0.0418	0.0881	0.7441	0.0358	0.0481	signaling
brm	FBgn0000212	0.007	0.0003	0.045	0.0204	0.0009	0.046	signaling
bsk	FBgn0000229	0.0001	0	0.033	0.0001	0	0.0187	signaling
cact	FBgn0000250	0.0418	0.0009	0.021	0.1454	0.0073	0.0502	signaling
caspar	FBgn0034068	0.0731	0.0057	0.0776	0.093	0.005	0.0539	signaling
CG11023	FBgn0031208	NA	NA	NA	NA	NA	NA	signaling
CG11501	FBgn0039666	0.2003	0.0253	0.1262	0.218	0.0431	0.1978	signaling
CG14225	FBgn0031055	NA	NA	NA	NA	NA	NA	signaling
CG16705	FBgn0039102	0.073	0.0087	0.1193	0.0924	0.0076	0.0826	signaling
CG2056	FBgn0030051	0.2411	0.0356	0.1477	NA	NA	NA	signaling
CG32382	FBgn0052382	0.2103	0.0224	0.1065	0.2277	0.0184	0.0808	signaling
CG32383	FBgn0052383	NA	NA	NA	NA	NA	NA	signaling
CG5896	FBgn0039494	0.0084	0.0014	0.1686	0.023	0.0014	0.0616	signaling
CG6361	FBgn0030925	0.2733	0.0246	0.0898	0.1544	0.0168	0.109	signaling

CG9675	FBgn0030774	0.2781	0.0205	0.0736	0.0827	0.0089	0.1077	signaling
cher	FBgn0014141	0.0029	0.0002	0.0636	0.0076	0.0004	0.0479	signaling
Dif	FBgn0011274	0.0742	0.0038	0.0517	0.3065	0.0195	0.0636	signaling
dl	FBgn0260632	0.0937	0.0055	0.0587	0.1534	0.0085	0.0553	signaling
Dnr1	FBgn0260866	0.0519	0.0051	0.0984	0.2411	0.0211	0.0877	signaling
dom	FBgn0020306	0.1153	0.0066	0.057	0.0918	0.0042	0.0458	signaling
dome	FBgn0043903	NA	NA	NA	NA	NA	NA	signaling
dpp	FBgn0000490	0.0244	0.0015	0.0628	0.1406	0.0041	0.0292	signaling
Dredd	FBgn0020381	0.2487	0.0217	0.0874	0.3217	0.018	0.0558	signaling
Dsor1	FBgn0010269	NA	NA	NA	NA	NA	NA	signaling
ea	FBgn0000533	0.0164	0.0011	0.065	0.0417	0.0032	0.0778	signaling
ECSIT	FBgn0028436	0.0694	0.0076	0.1099	0.0223	0.0021	0.0939	signaling
edl	FBgn0023214	0.1611	0.0158	0.0981	0.2673	0.013	0.0487	signaling
Egfr	FBgn0003731	0.0085	0.0009	0.1042	0.0723	0.005	0.0688	signaling
emb	FBgn0020497	0.0245	0.0008	0.0345	0.0001	0	0.0379	signaling
gcm	FBgn0014179	0.0567	0.0073	0.1279	0.1785	0.0072	0.0402	signaling
gcm2	FBgn0019809	0.0499	0.0035	0.0693	0.1858	0.019	0.1023	signaling
Hel89B	FBgn0022787	0.0745	0.0056	0.0748	0.0701	0.0032	0.0462	signaling
hep	FBgn0010303	0.1243	0.0069	0.0554	0.1565	0.0037	0.0234	signaling
hop	FBgn0004864	0.0244	0.0032	0.1333	0.1104	0.0043	0.0392	signaling
Iap2	FBgn0015247	0.0293	0.0013	0.0432	0.3387	0.0069	0.0205	signaling
imd	FBgn0013983	0.1094	0.0062	0.0566	0.0926	0.0031	0.0333	signaling
ird5	FBgn0024222	0.6424	0.0414	0.0644	0.5904	0.0264	0.0447	signaling
Jra	FBgn0001291	0.283	0.0128	0.0451	0.0218	0.0015	0.0712	signaling
kay	FBgn0001297	0.1724	0.0186	0.108	0.4523	0.0238	0.0527	signaling
key	FBgn0041205	0.6915	0.0358	0.0518	0.2065	0.0168	0.0815	signaling
kn	FBgn0001319	0.0001	0	0.0566	0.2697	0.0041	0.0153	signaling
lwr	FBgn0010602	0.0001	0	0.1845	0.0001	0	0.043	signaling
lz	FBgn0002576	0.0993	0.0054	0.0541	0.4277	0.0183	0.0427	signaling
mask	FBgn0043884	0.1113	0.0061	0.0544	0.1502	0.0064	0.0424	signaling
mbo	FBgn0026207	0.1619	0.0106	0.0656	0.3973	0.0218	0.055	signaling
Mekk1	FBgn0024329	0.1261	0.0068	0.0542	0.0374	0.0023	0.0618	signaling
Mkk4	FBgn0024326	0.0279	0.0021	0.0743	0.0001	0	0.0337	signaling
MP1	FBgn0027930	0.2143	0.0126	0.0589	0.3672	0.0184	0.0502	signaling
MP2	FBgn0037515	0.1428	0.0096	0.0671	0.2001	0.0105	0.0523	signaling
Mpk2	FBgn0015765	0.0508	0.0037	0.0722	0.0249	0.0012	0.0474	signaling
msn	FBgn0010909	0.02	0.0011	0.055	0.2116	0.0081	0.0381	signaling
MstProx	FBgn0015770	0.0001	0	0.0903	0.2794	0.0267	0.0954	signaling
mxc	FBgn0261524	0.0001	0	0.089	0.0001	0	0.0403	signaling

Myd88	FBgn0033402	0.0001	0	0.0579	0.0077	0.0008	0.1001	signaling
N	FBgn0004647	0.0348	0.0068	0.1949	0.0171	0.0016	0.0937	signaling
nec	FBgn0002930	0.1617	0.0161	0.0998	0.2151	0.0195	0.0907	signaling
Nos	FBgn0011676	0.0623	0.0042	0.0671	0.2299	0.018	0.0784	signaling
Ntf-2	FBgn0031145	0.0001	0	0.0304	0.0216	0.0033	0.1547	signaling
Ntf-2r	FBgn0032680	NA	NA	NA	NA	NA	NA	signaling
Nup214	FBgn0010660	0.2926	0.0159	0.0544	0.4797	0.0251	0.0522	signaling
os	FBgn0004956	0.0067	0.0007	0.1049	0.1633	0.0123	0.0756	signaling
p38b	FBgn0024846	0.0122	0.0012	0.0961	0.0001	0	0.0367	signaling
phl	FBgn0003079	0.042	0.0024	0.0567	0.0001	0	0.0313	signaling
pll	FBgn0010441	0.0083	0.001	0.1269	0.0693	0.0068	0.0977	signaling
pnt	FBgn0003118	0.0849	0.0071	0.0837	0.1334	0.0078	0.0583	signaling
POSH	FBgn0040294	0.0917	0.0039	0.0424	0.24	0.0081	0.0338	signaling
psh	FBgn0030926	0.0382	0.0058	0.1513	0.5135	0.0378	0.0736	signaling
puc	FBgn0243512	0.1645	0.0081	0.0494	0.2738	0.006	0.0218	signaling
Pvf1	FBgn0030964	0.017	0.0018	0.1039	0.0001	0	0.1067	signaling
Pvf2	FBgn0031888	0.1487	0.0141	0.0948	0.2669	0.0075	0.0281	signaling
Pvf3	FBgn0085407	0.0373	0.0056	0.15	0.0821	0.0022	0.0269	signaling
Pvr	FBgn0032006	0.0556	0.0051	0.0918	0.1103	0.0056	0.0505	signaling
Rac1	FBgn0010333	0.0001	0	0.054	0.0001	0	0.0509	signaling
Rac2	FBgn0014011	0.0001	0	0.0571	0.0001	0	0.0675	signaling
Ras85D	FBgn0003205	0.0001	0	0.0539	0.0001	0	0.0167	signaling
ref(2)P	FBgn0003231	0.5284	0.0209	0.0395	0.2144	0.0102	0.0475	signaling
Rel	FBgn0014018	0.4843	0.0287	0.0592	0.4329	0.0254	0.0586	signaling
RpS6	FBgn0261592	0.0232	0.0017	0.0729	0.0001	0	0.0418	signaling
SAE1	FBgn0029512	0.053	0.0054	0.102	0.0727	0.0052	0.0715	signaling
SAE2	FBgn0029113	0.1187	0.0053	0.0445	0.0989	0.0061	0.0618	signaling
Ser	FBgn0004197	0.1331	0.0046	0.0347	0.0692	0.0027	0.0386	signaling
slbo	FBgn0005638	0.0187	0.0038	0.2049	0.118	0.0115	0.097	signaling
slpr	FBgn0030018	0.0626	0.0055	0.0883	0.4856	0.0138	0.0284	signaling
smt3	FBgn0026170	0.0001	0	0.0561	0.0001	0	0.1209	signaling
Socs36E	FBgn0041184	0.1194	0.0084	0.0707	0.1435	0.0058	0.0405	signaling
Spn27A	FBgn0028990	0.0001	0	0.1094	0.1119	0.0056	0.0501	signaling
spz	FBgn0003495	0.199	0.0165	0.0827	0.2719	0.016	0.0589	signaling
srp	FBgn0003507	0.4695	0.0293	0.0624	0.3705	0.0194	0.0522	signaling
Stam	FBgn0027363	0.0571	0.0052	0.0905	0.2015	0.0122	0.0603	signaling
Stat92E	FBgn0016917	0.0263	0.0011	0.0433	0.1109	0.0057	0.0518	signaling
Su(H)	FBgn0004837	0.0001	0	0.0549	0.1309	0.0083	0.0634	signaling
Su(var)2-10	FBgn0003612	0.0804	0.0036	0.0449	0.0591	0.0008	0.0137	signaling

Tab2	FBgn0086358	0.1113	0.0076	0.0683	0.1541	0.005	0.0328	signaling
Tak1	FBgn0026323	0.1479	0.0082	0.0551	0.203	0.0066	0.0323	signaling
tamo	FBgn0041582	0.0761	0.0099	0.13	0.3117	0.0143	0.0457	signaling
Tehao	FBgn0026760	0.0413	0.0055	0.1321	0.1396	0.0081	0.0582	signaling
Thor	FBgn0261560	0.0437	0.0037	0.084	0.0001	0	0.0519	signaling
Tl	FBgn0262473	0.096	0.009	0.0935	0.0452	0.0037	0.0808	signaling
Toll-4	FBgn0032095	0.395	0.0219	0.0556	0.526	0.0236	0.0448	signaling
Toll-6	FBgn0036494	0.0191	0.0015	0.0773	0.0531	0.0021	0.0387	signaling
Toll-7	FBgn0034476	0.0118	0.0006	0.0538	0.0325	0.0013	0.0389	signaling
Toll-9	FBgn0036978	0.1592	0.0114	0.0717	0.1282	0.0083	0.0644	signaling
Tollo	FBgn0029114	0.0299	0.0017	0.0576	0.0468	0.0023	0.0501	signaling
Traf1	FBgn0026319	0.0001	0	0.1	0.0201	0.0009	0.0439	signaling
Traf2	FBgn0026318	0.0001	0	0.0548	0.0854	0.0068	0.079	signaling
Traf3	FBgn0030748	0.0523	0.0056	0.1075	0.0109	0.0009	0.0793	signaling
tub	FBgn0003882	0.2028	0.0084	0.0413	0.1457	0.0066	0.0451	signaling
Uev1A	FBgn0035601	0.0001	0	0.0239	NA	NA	NA	signaling
Ulp1	FBgn0027603	0.7093	0.0586	0.0827	0.9442	0.0671	0.0711	signaling
upd2	FBgn0030904	0.0333	0.0055	0.1655	0.2773	0.0032	0.0114	signaling
upd3	FBgn0053542	NA	NA	NA	NA	NA	NA	signaling
ush	FBgn0003963	0.1476	0.007	0.0473	0.5598	0.027	0.0482	signaling
WntD	FBgn0038134	0.0429	0.0029	0.0679	0.1115	0.0044	0.0392	signaling
ytr	FBgn0021895	0.2964	0.0024	0.0082	0.0001	0	0.0196	signaling

Table S3 Temporal and geographic differentiation of all candidate genes

	no. M strain	M strain π		*temporal differentiation*		*geographic differentiation*[b]	
		nonsyn	syn	*Fst*	*p-value*	*Fst*	*p-value*
AGO3	8	0.0000	0.0010	-0.067	0.824	0.125	0.121
armi	8	0.0015	0.0094	-0.045	0.897	0.237	**< 0.001**
aub	8	0.0009	0.0065	-0.045	0.883	0.484	**< 0.001**
Hen1	30	0.0013	0.0097	**0.303**	**0.001**	0.632	**< 0.001**
Hrb27C	8	0.0008	0.0025	0.048	0.153	0.116	0.083
Irbp	30	0.0020	0.0058	**0.170**	**0.038**	0.144	0.111
krimp	30	0.0076	0.0304	0.011	0.262	0.143	**0.004**
Ku80	8	0.0018	0.0134	-0.035	0.891	0.256	**0.001**
mael	8	0.0015	0.0000	-0.061	0.98	0.344	**< 0.001**
piwi	8	0.0011	0.0121	0.051	0.092	0.172	**0.006**
Psi	8	0.0007	0.0123	0.034	0.204	0.338	**< 0.001**
rhi	8	0.0017	0.0056	-0.013	0.469	0.236	**0.013**
Spn-E	30	0.0010	0.0033	0.053	0.08	0.459	**< 0.001**
squ	30	0.0016	0.0080	**0.202**	**0.001**	0.028	0.25
vas	8	0.0005	0.0045	0.025	0.223	0.213	**0.003**
zuc	8	0.0000	0.0054	-0.052	0.797	0.284	**0.005**

[a]Genetic differentiation between current (*post-P element*) African and North American populations
[b]Genetic differentiation between *pre-P* and *post-P* element invasion populations
All the significant results are in bold-type

Table S4 Information of Control Genes

symbol	FBgn	gene functions
jeb	FBgn0086677	visceral mesoderm development
CG8378	FBgn0027495	negative regulation of transcription
CG13178	FBgn0033685	cilium assembly
CG8878	FBgn0027504	protein serine/threonine kinase; protein phosphorylation
CG8407	FBgn0033687	microtubule-based movement
Oda	FBgn0014184	cell differentiation; embryonic development
wash	FBgn0033692	GTPase binding; signal transductions; actin filament and microtubule bundles assembly
CG33964	FBgn0053964	unknown
Cyp6t3	FBgn0033697	oxidation-reduction process
RpS11	FBgn0033699	structural constituent of ribosome; translation; mitotic spindle organization
Sr-CII	FBgn0020377	scavenger receptor activity; defense response