Horizontal Transposon Transfer in Eukarya: Detection, Bias, and Perspectives

Gabriel Luz Wallau[1,*,†], Mauro Freitas Ortiz[2,†], and Elgion Lucio Silva Loreto[3]

[1]Programa de Pós-Graduação em Biodiversidade Animal, Universidade Federal de Santa Maria, Brazil

[2]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Brazil

[3]Departamento de Biologia, Universidade Federal de Santa Maria, Brazil

*Corresponding author: E-mail: gabriel.wallau@gmail.com.

[†]These authors contributed equally to this work.

Accepted: July 5, 2012

Abstract

The genetic similarity observed among species is normally attributed to the existence of a common ancestor. However, a growing body of evidence suggests that the exchange of genetic material is not limited to the transfer from parent to offspring but can also occur through horizontal transfer (HT). Transposable elements (TEs) are DNA fragments with an innate propensity for HT; they are mobile and possess parasitic characteristics that allow them to exist and proliferate within host genomes. However, horizontal transposon transfer (HTT) is not easily detected, primarily because the complex TE life cycle can generate phylogenetic patterns similar to those expected for HTT events. The increasingly large number of new genome projects, in all branches of life, has provided an unprecedented opportunity to evaluate the TE content and HTT events in these species, although a standardized method of HTT detection is required before trends in the HTT rates can be evaluated in a wide range of eukaryotic taxa and predictions about these events can be made. Thus, we propose a straightforward hypothesis test that can be used by TE specialists and nonspecialists alike to discriminate between HTT events and natural TE life cycle patterns. We also discuss several plausible explanations and predictions for the distribution and frequency of HTT and for the inherent biases of HTT detection. Finally, we discuss some of the methodological concerns for HTT detection that may result in the underestimation and overestimation of HTT rates during eukaryotic genome evolution.

Key words: horizontal transfer, horizontal transmission, transposable elements, genome, eukaryote evolution.

Introduction

Since the discovery of DNA as the molecule that stores genetic information and governs trait inheritance from parents to their offspring, no biologist doubts that the vertical transfer of genetic material between ancestral and extant species has occurred. However, there is now growing evidence suggesting that another process also promotes the sharing of genetic material among species: horizontal transfer (HT) (Keeling and Palmer 2008).

HT events are characterized by the exchange of genetic material between species by methods other than ancestral to descendant inheritance (Schaack et al. 2010). These events are quite common among bacterial species (Gogarten and Townsend 2005), and as a result, sets of bacterial species are now being called genetic exchange communities (Skippington and Ragan 2011). In multicellular eukaryotes, HT is thought to be a rare event (Kidwell 1993; Anderson 2005). However, a growing body of evidence suggests that a particular type of HT, horizontal transposon transfer (HTT), could be a widespread process during eukaryote evolution (Schaack et al. 2010).

Transposable elements (TEs) are prone to HT compared with other coding and noncoding DNA sequences because of their parasitic characteristics and their intrinsic capacity to mobilize and reintegrate into chromosomes (Schaack et al. 2010). HT is a key step in the TE life cycle, allowing these parasites to immigrate to and colonize new genomes and escape loss by genetic drift (Le Rouzic and Capy 2006; Venner et al. 2009; Hua-Van et al. 2011). The arrival of a new TE in a host genome can have detrimental consequences

because TE mobility may induce mutation. Moreover, transposition activity increases the TE copy number and generates chromosomal rearrangement hotspots (Cáceres et al. 2001; McVean 2010). However, HTT can also introduce new genetic material into a genome and promote the shuffling of genes and TE domains among hosts, which can be co-opted by the host genome to perform new functions (Pace et al. 2008; Thomas et al. 2010).

HTT is difficult to detect because it is necessary to consider all the intrinsic features of the TE life cycle, such as sequence degeneration, stochastic loss, and any different evolutionary rates (Cummings 1994; Capy et al. 1998). In addition, the same patterns found in HTT can be observed at various stages during the natural TE life cycle, or they can be generated by the hybridization of closely related species. Since HTT was first described, many authors have suggested different approaches to obtain evidence of these events (Loreto et al. 2008). These methodologies involve looking for phylogenetic incongruence (PI) between the host and TE phylogenies, patchy TE distributions (PD), or a high similarity (HS) between TEs from different species (Silva et al. 2004).

In the last decade, new methodological approaches based on comparisons between host genes (HGs) and TEs were developed, allowing a broader evaluation of HTT events (Silva and Kidwell 2000; Lerat et al. 2000). Nevertheless, the identification of HTT events can still be difficult, even when combinations of several methodologies are used, because these methods can both overestimate and underestimate the occurrence of HTT events depending on when and in which species the HTT occurred. The astonishing number of new genome projects, in all branches of life, presents an unprecedented challenge to the field of comparative genomics. The amazing quantity of genomic data that is now available for many taxa urgently calls for the development and application of standardized methodologies that will produce widely comparable results. To date, there is no gold-standard approach to clearly discern between alternative explanations and HTT events.

The main purpose of this article is to propose a standard hypothesis test for the evaluation of HTT events. We discuss the biological bias found in the distribution of the HTT events described in the literature and caution against methodological biases in regards to inferring the number of HTT events.

HTT Detection

Currently, the most robust approach for evaluating potential HTT events is a combination of evidence supported by statistical tests (Loreto et al. 2008). However, in some cases, only one type of evidence, such as HS, PD, or PI, is necessary to support HTT. For example, the classical and unequivocal uptake of *P*-elements by *Drosophila melanogaster* from *Drosophila willistoni* is supported by the PD of this element in the *melanogaster* species group, where it is present despite being absent from the genomes of related species (Daniels et al. 1990).

One of the most promising methodologies for the detection of HTT is based on a between-species comparison of the neutral rate of evolution (assessed by synonymous substitution divergence) for both the TEs and the HGs. This approach assumes that, if TEs have been vertically transmitted and maintained by neutral evolutionary processes in the genomes of two different species since their last common ancestor, the number of synonymous substitutions per synonymous site (dS) of the TEs should be equal to or greater than that of the vertically transmitted HGs. However, if the dS obtained for the TE is significantly lower than the dS for the vertically transferred HG, the most probable explanation is that these elements were exchanged by HT between the species after their reproductive isolation. This pattern can be observed because a horizontally transferred TE has spent less time in the new host genome than the original HGs. These HGs have been in the genome since the last common ancestor of the species involved in the HT. Therefore, these TEs have had less time to accumulate synonymous substitutions than the HGs. It is noteworthy to state that even if a TE shows a dS value equal to or greater than the HG dS, it does not necessarily imply vertical transmission (VT). This pattern can also be generated by an HTT event occurring just after the split of the involved species. For these comparisons, it is necessary to choose HGs with similar codon usages to those of the TEs (Silva and Kidwell 2000; Ludwig et al. 2008). If an HG with a higher codon usage bias is chosen, it can present low dS values and results in the underestimation of the number of HTT events (Silva and Kidwell 2000; Vidal et al. 2009).

Another interesting method for evaluating HTT involves the use of the unique codon usage bias of each genome (Lerat et al. 2002; Jia and Xue 2009; Plotkin and Kudla 2011). Differences in the codon usage bias are expected to be higher among genomes from different species than among the genes within the genome of the same species. According to this premise, it should be possible to detect the recent invasion of a genome by TEs from the patterns of codon usage bias because the TE's codon usage should be more homologous to that of the donor species than that of the receptor species. Recently, Rodelsperger and Sommer (2011) showed the utility of this methodology for detecting HTT events between a beetle species and its associated nematode. It is noteworthy that the species–specific codon usage bias becomes less evident when more closely related species are considered because of their phylogenetic similarity (Sharp et al. 1995). Thus, although this methodology can be very useful in detecting HTT between distantly related species, there are limitations to its application in related species.

Multiple hypothesis testing using several methodologies could be an efficient approach for discriminating between HTT and alternative hypotheses. On the basis of recent reports

on TE characteristics and HTT events, we propose a straightforward hypothesis test to evaluate potential HTT events (fig. 1).

Hypotheses Test

Normally, the first sign of evidence to suggest HTT comes from PI, a PD, or a HS between the TEs from distantly related host species. PI is inferred if a phylogeny of TE does not match the host phylogeny (fig. 1). PD is detected when a specific TE shows a random distribution, characterized by the presence TE in one or a few species from a phylogenetic branch that otherwise lacks the TE. However, although these patterns can be generated by HTT events, they can also be the result of the natural degeneration of TEs inside the host genomes, when combined with ancestral polymorphism and stochastic loss.

The first step of the hypothesis test is the implementation of two different tests (T1 in fig. 1): 1) a comparison of the dS between the TEs and HGs and 2) a comparison of the codon usage bias between the TEs and the host genome. These two tests can be complementary in HTT detection. The codon usage bias comparison can be used to evaluate HTT in distantly related species; however, the difference in the codon usage bias among closely related species is normally low, which does not allow the donor (and TE) and the recipient species codon usages to be distinguished. TE and HG dS comparisons can be used to evaluate HTT in closely related species and in distantly related species alike. If we find that the codon usage bias is similar between the TEs and the host genome and that the TE dS values are equal to or greater compared with the HGs, then it is likely that the TEs are being inherited by vertical transfer (fig. 1*B*). Otherwise, if the TE's codon usage bias is different from that of the host genome or if the dS is significantly lower for the TEs than for the nuclear HGs, then the TEs were most likely exchanged among the species by HT (fig. 1*A*). It is necessary to perform dS and codon usage bias comparisons even if PI or PD were not detected because the absence of these evidences does not guarantee that an HTT event has not occurred.

Nevertheless, alternative hypotheses attempting to explain the observed differences in the dS values between TEs and HGs have also been suggested. For example, selective constraint can act at the RNA/DNA level as a pressure established on the mRNA structural stability or on splicing sites or if a TE is integral in the siRNA regulatory machinery (Rubinstein et al. 2011; Plotkin and Kudla 2011). However, as these constraints are expected to act on specific sites and not on the sequence as a whole, the magnitude of these constraints should be small. Therefore, these factors cannot explain the dS differences observed between HGs and TEs when the TE is conserved across the entire sequence. In fact, sometimes the dS values between TEs from different species are very low, the magnitude of which could not be easily explained by the previously described constraints. Therefore, a very low dS measurement is better explained by the occurrence of an HTT.

When HTT events among distantly related species are considered, only the T1 stage of the hypothesis testing is necessary for validation. However, HTT events can occur among

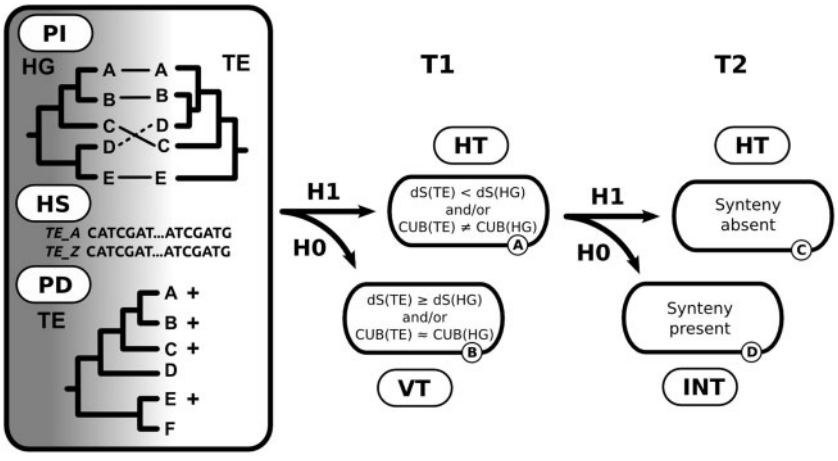

Fig. 1.—A schematic representation of a hypothesis test for discerning between HT and the natural stages of the TE life cycle. BOX: The first line of evidence for HTT: Phylogenetic incongruence (PI) between the host and TE phylogenies. Patchy distribution (PD) of a given TE within a group of species and high similarity (HS) between the TEs from different species. T1—The first test to distinguish between HTT and vertical transmission (VT)—comparing the dS between the TE and host genes (HGs) and species-specific codon usage bias (CUB). H0—vertical transfer is more probable if the dS values for the TEs are greater than or equal to the dS values of the vertically transmitted host genes and if the TE codon usage bias is similar to the codon usage bias in the host species. H1—HT will be selected if the TE's dS value is significantly lower than the dS values of the vertically transmitted host genes or if the TE codon usage bias is different from the host species codon usage bias. T2—A second step can be used to evaluate HTT between closely related species. H0—If there is synteny beyond the border of the TE copies, it is more probable that these copies were shared by hybridization among the host species (an introgression [INT] occurred). H1—If there is no synteny, it is more probable that these copies were shared by an HTT event between host species.

individuals encompassing any taxonomic level, from different phyla to closely related species (Bartolomé et al. 2009). It is very difficult to prove HTT among closely related species, and in this case, the sharing of TEs between species can be the result of the occasional cross-fertilization between species. Introgression events between closely related species can generate significantly lower dS values for the TEs compared with the nonintrogressed HGs. The analysis of synteny beyond the border of the TE copies, that is, analysis all the TE copies present in one species, and an evaluation of whether they are found at the same locus in another species is one method that has been suggested to discern between introgression and HTT (Fortune et al. 2008) (T2 in fig. 1). Introgression events normally maintain synteny among the species involved in the hybridization; in other words, homology and high identity are encountered not only in the TE sequences but also in the neighboring DNA regions (fig. 1*D*). However, when HTT events occur, only variability with the absence of synteny is typically encountered in the TE-neighboring regions (Fortune et al. 2008) (fig. 1*C*). Nevertheless, despite the fact that this methodology is consistent and straightforward, it has yet to be tested, and it could proved to be particularly difficult to evaluate the synteny of TEs because of their inherent mobility. It is likely that this methodology will be restricted to the analysis of nonautonomous TEs, but even nonautonomous elements can be mobilized by other TEs in trans, a factor that would complicate the analysis. Regardless, if synteny is found, it is taken as evidence that hybridization occurred; therefore, in the absence of synteny, the probability that the sharing of TEs between species as the result of hybridization decreases, whereas the probability of an HTT event increases.

High similarities between the TE sequences in different species can also be the result of TE domestication, where a TE region is co-opted to perform a new, useful function in the genome of the host (Gould and Vrba 1982; Huda et al. 2010). Domestication can be detected using features such as copy number, orthologous position, and evaluating the selective constraint (dN:dS ratio) acting between the TEs that are incongruent with the host species' phylogeny and comparing this constraint with the selective constraint on the HGs. Thereby, we can discern whether the HS found between the TEs from different host species is due to domestication events, different evolutionary rates, or ancestral polymorphism. Other analyses can also reveal clues as to whether a TE is domesticated, such as the presence of only one TE copy in the genome or the observation that the TEs occur at orthologous positions in different species (Sinzelle et al. 2009). Another approach that can be used to gather clues about TE domestication is the analysis of full-length TE copies (including inverted terminal repeats, long terminal repeats, and coding and noncoding regions). If there are high similarities along all the TE sequences, the best explanation for the sequence conservation is the occurrence of an HTT event. This is because TE domestication only imposes strong selective constraints on one region of a TE and not in the full-length copies (Feschotte 2008; Sinzelle et al. 2009). Even if a domestication occurred, a dS TE smaller than dS HG is unlikely to be observed, because negative selection acts only in nonsynonymous substitutions and not over neutral synonymous substitutions. Therefore, we also can evaluate if occurred HT events before the domestication event using the dS analysis (T1 in fig. 1).

Another analysis that can be useful for understanding HTT is the dating of these events along the molecular clock. One way to perform this analysis is by evaluating the molecular evolution rate of the nuclear genes with a codon usage bias similar to the TE to estimate the time of divergence between horizontally transferred TEs copies (Ludwig et al. 2008). A second type of analysis can be performed when the entire host genome is available. In this case, an ancestral sequence can be inferred when evaluating many copies of one horizontally transferred element. This analysis is based on the premise that these elements have been evolving neutrally since the HT event; therefore, we can estimate the time of the first insertion event and the subsequent amplification inside of the host genome using a neutral substitution rate (Mouse Genome Sequencing Consortium 2002; Yang et al. 2004; Khan et al. 2006; Pace and Feschotte 2007). This neutral substitution rate can be estimated from an ancestral TE present in an orthologous position (inherited vertically) in genomes where we have an estimate of the host species' divergence time (Pace et al. 2008). Therefore, with these type of data, we can evaluate whether a TE is more recent than expected for vertical transfer, and by comparing this activity estimative among different species, we also can reveal relationships between the donor and the receptor species.

HTT Distribution and Frequency

HTT Rates

Here, we analyze the HTT events previously collected from the literature by Schaack et al. (2010) along with new events to compile all the HTT events described to date (supplementary table 1, Supplementary Material online). HTT events have already been detected in three eukaryote kingdoms: Animalia, Fungi, and Plantae (fig. 2). The majority (94.37%) of the HTT events were detected in Animalia, followed by Plantae (4.30%) and Fungi (1.32%). The differences in the HTT frequencies among kingdoms may be explained by differential susceptibilities of taxas to experiencing HTT. However, these differences could also be due to a historical bias for the use of animal model organisms in TE research or the differential abilities of the studied TEs to undergo HT (Pritham 2009; Schaack et al. 2010). To date, 178 of the 330 HTT cases described in the literature were detected among *Drosophila* species (54%). This disproportionate number of HTTs in *Drosophila* could be biased because some of the pioneering studies in TEs,

including the first well-documented case of HTT (Daniels et al. 1984), were performed in these model organisms. Thus, these studies opened the door for TE research using the *Drosophila* genus. Several recent publications have shown evidence of HTT events in other Animalia taxa, such as crustaceans and mammals (Casse et al. 2006; Gilbert et al. 2010; Novick et al. 2010), further suggesting that the elevated number of HTT events described in *Drosophila* may show a historical bias.

Genome Projects, TE, and HTT Bias

Although exponentially growing, global species biodiversity is still poorly represented in current genome projects. In Eukarya, only the Animalia (270 projects), Fungi (234 projects), and Plantae (101 projects) kingdoms have a large number of genome projects (http://www.ncbi.nlm.nih.gov [cited 2011 October 12]). Many of these genomes are still undergoing sequencing or are in other steps of analysis; thus, we have differing knowledge about the TE content in these genomes (fig. 2). Moreover, many studies remove these elements to facilitate genome assembly or analysis (Bergman and Quesneville 2007; Treangen and Salzberg 2011). The lack of knowledge about the TE content in some taxa could strongly bias the descriptions of HTT distribution and frequency.

To evaluate how genomic analysis can influence the TE and HTT descriptions, we collected, for each of the aforementioned kingdoms, the number of genome projects in NCBI and the TE entries from the Repbase site (http://www.girinst.org [cited 2011 October 12]; Kohany et al. 2006) (fig. 2*A*). For this evaluation, two points should be noted: 1) the genome projects are in different stages and many have not yet analyzed the TE content and 2) the entries in Repbase are not limited to the TEs from genome projects.

Most of the HTT events described in the literature were from Animalia (fig. 2*C* and *D*). This finding likely reflects the larger number of genome projects for animals. Moreover, on the basis of TE entries available in Repbase for different taxa, we noted that animal species have been analyzed more deeply in regards to their TE content compared with the other phyla (12,565 TE entries) (fig. 2*A*).

The Plantae kingdom is an intriguing case; some species have high TE content (more than 60% in maize; Biémont and Vieira 2006), and a large number of elements have been characterized (4,638 TEs entries Repbase); however, only 13 HTT events have been detected in this kingdom (fig. 2*C*). This discrepancy could be explained by the following: 1) the smaller number of genome projects in Plantae compared with the Animalia and Fungi kingdoms; 2) some unknown, specific features of these organisms; or 3) historical bias in the HTT analysis, despite TE characterization.

In fungi, there is no apparent bias due to the number of genomes available as there are a similar number of projects when compared with Animalia (fig. 2*A*); however, to date, only four HTT events have been described for fungi (fig. 2*C*). One possible explanation for this fact could be related to the N_e (effective population size) of these organisms because they have among the largest eukaryotic N_e (Lynch and Conery 2003). It has been shown that there is a negative correlation between the N_e and TE maintenance in host genomes (Lynch and Conery 2003). Moreover, fungi present a low, and most likely poorly studied, TE content (1,603 TEs in Repbase) compared with animals (12,565) or plants (4,638) (fig. 2*A*). It is important to note that the existence of only a few described HTT events in fungi does not mean that HTT does not occur; it more likely indicates that HTT occurs but cannot be detected due to the high turnover of TEs in species with large N_e values and small genomes. However, this is not always the case. *D. melanogaster*, for example, has a small N_e compared with most fungi species but has a high turnover for retrotransposons and a high rate of HTT (Lerat et al. 2003).

Excavates, Chromalveolates, and Rhizaria are the least represented of the kingdoms in the NCBI genome projects database, and they also have fewer entries in the Repbase repository (fig. 2*A* and *C*). The lack of knowledge about the TE content in these groups, along with the high turnover of TEs in taxa with large N_e values, may explain why there have been no HTT cases reported for these groups thus far.

TE Features Influencing HTT Frequency

Despite historical bias in the evaluation of HTT among taxa, we can observe patterns in HTT distribution and frequency that are associated with different TE features. Silva et al. (2004) suggested that an effective HTT event may be related to the presence of a stable intermediate during the transposition process. Moreover, TE self-regulatory mechanisms can also influence the success of certain HTT events. HTT events appear to be more frequent for LTR retrotransposons and DNA transposons when compared with non-LTR retrotransposons (Silva et al. 2004; Loreto et al. 2008; Schaack et al. 2010).

The evolutionary relationship between LTR retrotransposons and retroviruses is well established (Xiong and Eickbush 1988, 1990; Poch et al. 1989). This evolutionary link suggests that some LTR retrotransposons can undergo HTT by themselves if they are capable of producing viral capsids and envelopes (*env* gene), hence promoting a viral-like infection and thereby eliminating the requirement for a vector. It has been shown that *gypsy* elements are capable of producing viral capsids and infecting *gypsy*-free *D. melanogaster* strains (Kim et al. 1994; Song et al. 1994). Even LTR retrotransposons that lack the *env* gene and the gene responsible for producing viral capsids can use the viral capsids from other LTR retrotransposons in trans, allowing a "helped" infection (Coffin et al. 1997). Recently, Routh et al. (2012) showed that at least 5.3% of the RNAs packaged inside of viral-like particles contain sequences derived from TEs, including DNA transposons, LTR and non-LTR retrotransposons. However, the capacity of *gypsy* viral capsids to infect other *Drosophila* species still

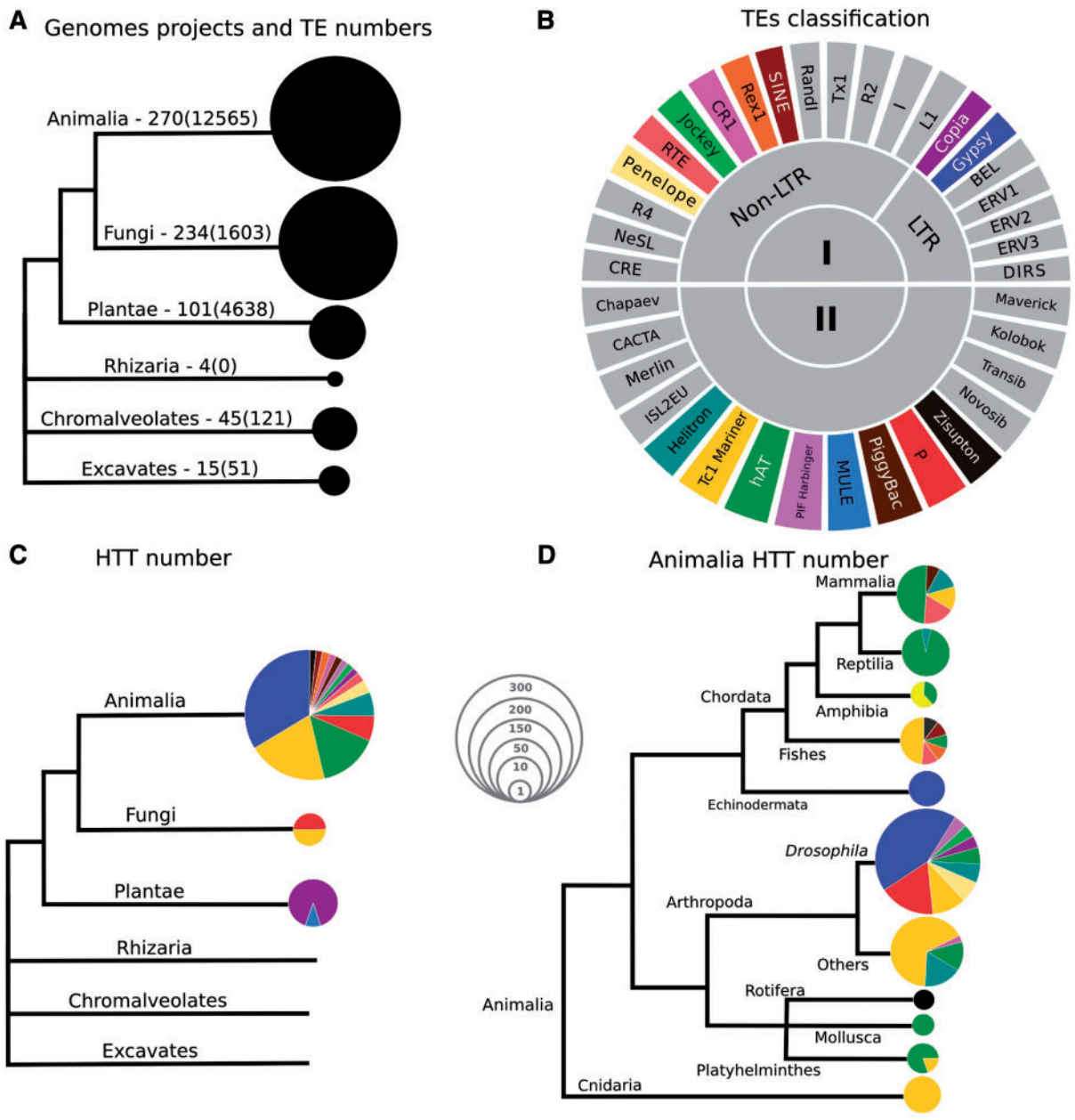

Fig. 2.—A representation of the genome projects, TEs, and number of HTT events in each major eukaryotic taxon. (*A*) The number of genome projects from the NCBI database (corresponding to cycle size and the number after the branch name, respectively) and TE RepBase entries (indicated by the number within the parentheses) in each major branch of the tree of life. (*B*) TE superfamily classifications based on RepBase. (*C*) The distribution of HTT events in each major eukaryotic taxon. (*D*) Distribution of HTT events within Animalia. The colors represent the TE superfamilies described in (*B*), and the cycle size represents the number of HTT events for each host taxon.

remains unclear and requires further elucidation. The same holds true for the in trans infection hypothesis.

If we suppose that the infective ability of LTR retrotransposons plays a significant role in promoting HTT events among species, we should expect that LTR retrotransposons would be preferentially transferred among species with cell structures that are similarly recognized by the LTR retrotransposon's capsids. This assumption is based on the premise that a retrotransposon's recognition machinery is analogous to that of a virus, which recognizes a restricted set of cellular receptors from a particular group of species. Furthermore, this analogy allows the extrapolation that HTT events should occur in waves, similar to those in a viral infection. When we look at all the previously described examples of HTT involving LTR retrotransposons, we note that 88.88% (104 of 117) of the events are among species from the same genus, 3.41% (4 of 117)

occur among species from different genera, 6.83% (8 of 117) occur among species from different orders, and only one event was observed among species from different phyla. However, the tendency for a higher frequency of LTR retrotransposon HTT events among closely related species could represent a strong taxonomic bias because 74 of the 86 described HTT events involving retrotransposons were described in the *Drosophila*. Regarding the HTT waves, one study reported that retrotransposon HTT waves occurred among *Drosophila* species (de Setta et al. 2009). Because of this, future studies are required to evaluate whether the LTR retrotransposon HTT events also occur more frequently among closely related species in other taxa.

The most widely distributed DNA transposon elements, from the *Tc1-mariner* superfamily, are simple in structure (presenting one or only a few ORFs and a primary structure rarely longer than 4 kb) (Wicker et al. 2007) and possess self-regulatory mechanisms. This structural simplicity can increase the likelihood of stable vector transportation during an HT event and is thought to represent an adaptation for HTTs (Schaack et al. 2010). O'Brochta et al. (2009) observed that *hobo/Hermes hAT* elements commonly produce stable and recombinogenic episomes; the circular extrachromosomal DNA of these transposons is a stable excision product that can reintegrate at a new site. Therefore, it is possible that these episomes could maintain TE recombinogenic properties following the transport by a vector into another species.

In addition to being carried by vectors, we cannot rule out the possibility that some DNA transposons may be self-transmissible. It is important to note that some complex DNA transposons may have originated from virophages (Fischer and Suttle 2011) and single-stranded DNA viruses (Liu et al. 2011).

As mentioned previously, some LTR retrotransposons are able to produce virus-like particles, and this has been suggested as a mechanism for HTT. Based on the common features shared between some TEs and viruses, one would assume that if the infective capacity is an important step in the HTT of LTR retrotransposons, then jumps by viral species between host species are expected to be common. In fact, there are many works reporting the jumping of viral species between host species. These events are commonly called viral-host switches (Gibbs and Weiller 1999; Nemirov et al. 2002; Vijaykrishna et al. 2007; Kang et al. 2010; Liu et al. 2010, 2011; Longdon et al. 2011) or species jumps. Viral-host switches have been primarily described in vertebrate species, the majority of which are related to human infectious diseases such as HIV, SARS, and H5N1 (Woolhouse et al. 2005; Parrish et al. 2008). Thus far, little importance has been given to these events in other taxa; however, some examples of viral species jumps have recently been described in *Drosophila* species (Liu et al. 2010). Altogether, these viral and TE data suggest that the infective capacity of some TEs is likely a key step that allows their horizontal transmission across species.

Once a host-switch event occurs, some virus integrates into the germ cells of the host genome as a provirus by a process known as endogenizaton [retroviruses—see Patel et al. (2011) and Feschotte and Gilbert (2012)]. These proviruses can be maintained from parent to offspring by VT. Even viruses that do not have a natural or obligate integration step into the host genome can also be endogenized using the LTR retrotransposon and endogenous retroviruses machinery (Holmes 2011; Patel et al. 2011). Currently, the studies have shown that the majority of eukaryotic viruses can be integrated in the host chromosomes via different pathways. Therefore, once a virus is endogenized, the host switch can be detected by analyzing the same evidence used to detect HTT, such as our hypothesis test.

The discussion of self-regulatory mechanisms yields support for the high efficiency of *mariner* elements to perform an efficient invasion strategy in a new genome (Lohe et al. 1995). Under excessive transposase production, *mariner* transposases aggregate together, causing a decrease in the transposition rate (Hartl et al. 1997; Lohe et al. 1997). When an organism acquires a new active transposon by HTT, a burst of transposition events typically follows, until all copies are mutationally inactivated or regulated by the inner host regulation mechanisms (Boer et al. 2007). High TE activity may cause detrimental changes in the host genome with TE insertions in coding or regulatory gene sequences. Self-regulatory mechanisms can be advantageous to TEs because these mechanisms can decrease the probability that a detrimental mutation will be introduced into the host genome, thereby increasing the TE's odds for inheritance by the host's descendants. Thus, TEs with self-regulatory mechanisms appear to have evolved a more effective strategy for an efficient invasion and for being maintained in host descendants reducing their harmful effects for the new host's genome.

Several characteristics of DNA transposons, such as autonomous transposition capacity (independent of the host's proteins), short length, and the presence of self-regulatory mechanisms, can enhance the probability that these elements will undergo HT. To date, only one event has been reported to have occurred among domains; 25.49% (39 of 153) of the HTT events occurred among phyla, 3.26% (5 of 153) occurred among classes, 17.54% (27 of 153) occurred among orders, 11.76% (18 of 153) occurred among families, 3.26% (5 of 153) occurred among genera, and 37.90% (58 of 153) occurred among species of the same genus. These data suggest that HTT involving DNA transposons can occur in all taxonomic levels, including the more distantly related levels.

Host Features Influencing HTT Frequency

Intrinsic host features can also influence HTT rates. The frequency of HTT events can be influenced by factors such as the

natural history or life cycle of the host species. For example, if two species have a close ecological relationship, such as a predator–prey relationship, symbiotic contact, the sharing of parasites, or even the use of the same natural resources, the chances that an HTT event will occur between these species increases (Houck et al. 1991; Yoshiyama et al. 2001; Loreto et al. 2008; Gilbert et al. 2010; Schaack et al. 2010). This scenario has been used to explain cases of HTT among sympatric crustaceans (Casse et al. 2006) and in *Drosophila* species (Mota et al. 2010; Carareto et al. 2011).

In the majority of multicellular eukaryotes, the reproductive and somatic cells are differentiated. Therefore, the TEs must be transmitted to the reproductive cells to be inherited by the descendants of a new host, that is, to gain entry into a new host genome via HTT. Thus, we might expect that HTTs should be more prevalent in unicellular eukaryotes and multicellular eukaryotes with undifferentiated reproductive and somatic cells because any cell in the body that has acquired a new TE can transmit it to future generations. Along these lines, Pritham (2009) suggested that unicellular eukaryotes should be particularly susceptible to HT due to the lack of a protected germline. Supporting these ideas, Robertson (1997) identified seven putative HTT events between insects and *Hydra* and one HTT case between the planarian *Dugesia tigrina* and the ant *Crematogaster cerasi*. In line with these findings, Chapman et al. (2010) more recently identified at least 90 potential HTT events in the *Hydra magnipapillata* genome. Hydras and planarians are animals without germ and somatic cell differentiation. Nevertheless, despite the difficulty imposed by cellular segregation, almost all HTT cases have been described in multicellular eukaryotes that have reproductive and somatic cell differentiation.

The mutation rate ($N_e\mu$) and the N_e of a receptor–host species can also influence the probability of a successful TE invasion by HT. For example, successful TE invasions by HT are less likely in species with a higher $N_e\mu$ and shorter generation times because there is an increased probability of the TE being inactivated. The influence of the N_e is a result of the balance between natural selection and genetic drift (Lynch and Conery 2003). The genomes of host species with large population sizes, as many unicellular organisms possess, are also subject to a strong purifying selection (Lynch 2007). Thus, if an HTT event occurs in these species, the probability that the TEs will be quickly eliminated by natural selection is high. On the other hand, in species with small population sizes, such as many tetrapods, genetic drift increases the probability that a new TE will be maintained in the host genome following an HTT event. Lynch and Coney (2003) reported that host species should have an N_e less than approximately equal to 7×10^7 to allow retrotransposon proliferation and an N_e less than approximately equal to 2×10^7 to allow the proliferation of DNA transposons.

As expected, each species set has unique ecological interactions (among species and among their parasites) leading to differential probabilities of HTT. However, it seems likely that there are some patterns that will be useful for predicting HTT in a broad range of species due to host reproduction and population features.

HTT Underestimation and Overestimation

Even when all the available approaches for detecting HTT are used, it is likely that many events will remain undetected. The inability to detect HTT results from the high turnover of TEs in host genomes (Lerat et al. 2003). When a TE arrives in a new genome, it usually occurs through a transposition burst that can be detrimental to the new host. The individuals bearing these detrimental changes can then be eliminated by natural selection, hence abolishing the signal of the primary invasion. When TEs successfully invade and are maintained in a new genome, the TE copies will evolve under neutral or weak natural selection (Silva and Kidwell 2000). Both low-dS measures obtained from the TEs compared with the HGs and species-specific codon usage biases from donor species tend to degenerate over the course of time. Thus, the more ancient an HTT event, the more difficult it will be to detect (fig. 3*A*). This promotes a weak signal of HTT events, leading to underestimation (fig. 3*B*).

Alternatively, HTT can also be overestimated. The overestimation of HTT events is directly related to the number of species in both the donor and the receptor clades. For example, if HTT occurs in the ancestor of two clades (fig. 3*A*–*C*), comparisons of the dS TE/dS HG and the codon usage bias could be significant for all pairwise species comparisons, suggesting many HTT events, when in reality only one has occurred. The maximum number of overestimated HTT events will be the number of analyzed species derived from the donor clade since the last common ancestor, multiplied by the number of analyzed species derived from the receptor clade since the HTT event.

A more complex scenario can also be considered. For example, members of a family of related TEs could have undergone HTT at different evolutionary times (fig. 3*D*). In this situation, overestimation will occur if we count each pairwise comparison resulting in a dS TE < dS HG as one event, as mentioned earlier. However, in the scenario depicted in figure 3*D*, for example, if we consider the observed cases as a unique, ancient HTT events, we will obtain an underestimate because three independent HTT events have occurred. In some specific cases, however, dS values can be used to date these HTT events (fig. 3*E*). For example, HTT events may be dated when the time since the occurrence of the HTT is long enough to result in differentiated in dS values, when the studied species have a well-resolved phylogeny and when a calibrated molecular clock is available. The number of dates obtained in these analyses may then be used to parsimoniously estimate the number of HTT cases.

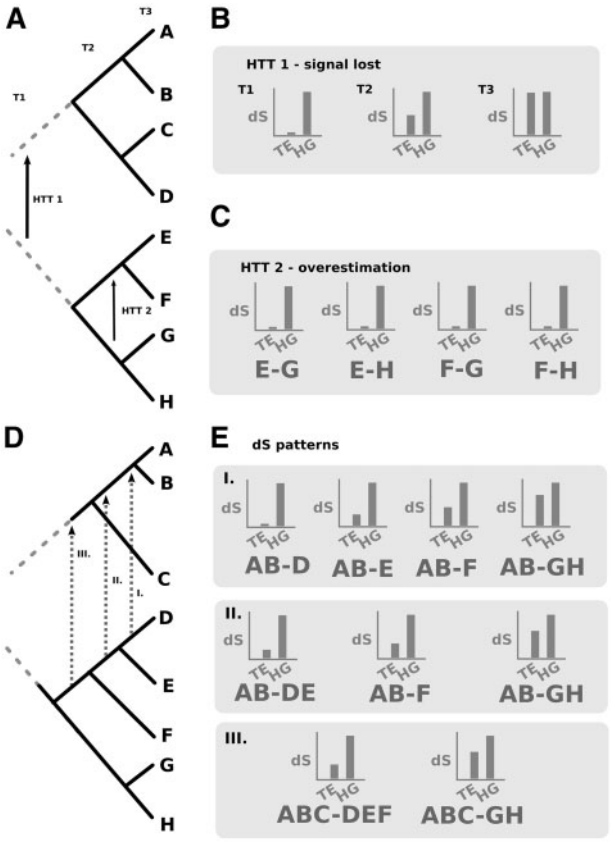

Fig. 3.—The underestimation and overestimation of HTT events. (*A*) The host species' phylogenies that represent HTTs at different evolutionary times (HTT1 and HTT2). (*B*) In HTT1, as the TEs evolve under neutral- or weak-natural selection, the dS value will increase over time (T1–T2–T3), and the species-specific codon usage bias from the donor species will be lost (resulting in the underestimation of old events). (*C*) In HTT2, all the TE-dS comparisons among species E, F, G, and H will be significantly lower than the HG's dS due to the maintenance of only one HT signal. (*D*) The host species' phylogenies that represent a more complex scenario with three HTT events (I, II, and III). (*E*) The TE dS patterns resulting from the HTT events in (*D*).

These theoretical models are simplistic compared with the complex evolution of TEs, where HTT is common. However, the use of these models can allow us to describe the degree of overestimation for a given situation. Moreover, because we observe a number of significantly lower dS values in the potential HTT events among current species, we may propose, by parsimony, the probable donor and receptor species by identifying the lower dS value.

Conclusions

Currently, there are no doubts as to the impact of TEs on eukaryotic genome evolution. There is a growing amount of data showing that HTT is a common and widespread phenomenon in eukaryote evolution. In light of the currently astonishing number of new eukaryotic genomes, it has become necessary to use a standardized methodology for the detection of HTT if these analyses are to be comparable across a wide range of eukaryotic taxa. Currently, different software is available to perform the analyses proposed in the hypothesis test (fig. 1), although one major challenge is to automate the data mining in the genomes to perform the analyses and organize the programs in a pipeline. This process can then facilitate and increase the discovery of HTT cases.

A strong HTT bias can be observed among eukaryotic taxa, primarily resulting from a historical bias for TE research in the *Drosophila* genus. However, even with this bias, we can observe trends that might be explained by the biological features of TEs and their hosts. HTT detection is a difficult task because of the high turnover of TEs inside host genomes and the number of species analyzed. These issues can lead to the underestimation or overestimation of HTT events between ancestral and current eukaryotic species; therefore, careful evaluation is warranted.

Supplementary Material

Supplementary table 1 is available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We are grateful to Dr Lizandra J. Robe, Dr Daniel Ângelo Sganzerla Graichen, and the two anonymous reviewers for valuable suggestions. This work was supported by Conselho Nacional de Desenvolvimento Científica e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Literature Cited

Anderson JO. 2005. Lateral gene transfer in eukaryotes. Cell Mol Life Sci. 62:1182–1197.

Bartolomé C, Bello X, Maside X. 2009. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. Genome Biol. 10:R22.

Bergman CM, Quesneville H. 2007. Discovering and detecting transposable elements in genome sequences. Brief Bioinform. 8:382–392.

Biémont C, Vieira C. 2006. Junk DNA as an evolutionary force. Nature 443: 521–524.

Boer J de, Yazawa R, Davidson W, Koop B. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics. 8:422.

Cáceres M, Puig M, Ruiz A. 2001. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. Genome Res. 11:1353–1364.

Capy P, Bazin C, Higuet D, Langin T. 1998. Dynamics and evolution of transposable elements. Austin (TX): Landes Bioscience.

Carareto C. 2011. Tropical Africa as a cradle for horizontal transfers of transposable elements between species of the genera *Drosophila* and *Zaprionus*. Mob Genet Elem. 1:179–186.

Casse N, et al. 2006. Species sympatry and horizontal transfers of *Mariner* transposons in marine crustacean genomes. Mol Phylogenet Evol. 40: 609–619.

Chapman JA, et al. 2010. The dynamic genome of *Hydra*. Nature 464: 592–596.
Coffin JM, Hughes SH, Varmus HE. 1997. Retroviruses. New York: Cold Spring Harbor Laboratory Press. p. 843.
Cummings MP. 1994. Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. Trends Ecol Evol. 9:141–145.
Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. 1990. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics 124:339–355.
Daniels SB, Strausbaugh LD, Ehrman L, Armstrong R. 1984. Sequences homologous to *P* elements occur in *Drosophila paulistorum*. Proc Natl Acad Sci U S A. 81:6794–6797.
de Setta N, Van Sluys MA, Capy P, Carareto CMA. 2009. Multiple invasions of *Gypsy* and *Micropia* retroelements in genus *Zaprionus* and *melanogaster* subgroup of the genus *Drosophila*. BMC Evol Biol. 9: 279.
Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. Nat Rev Genet. 9:397–405.
Feschotte C, Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 13: 283–296.
Fischer MG, Suttle CA. 2011. A virophage at the origin of large DNA transposons. Science 332:231–234.
Fortune PM, Roulin A, Panaud O. 2008. Horizontal transfer of transposable elements in plants. Commun Integr Biol. 1:74–77.
Gibbs MJ, Weiller GF. 1999. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. Proc Natl Acad Sci U S A. 96:8022–8027.
Gilbert C, Schaack S, Pace JK II, Brindley PJ, Feschotte C. 2010. A role for host–parasite interactions in the horizontal transfer of transposons across phyla. Nature 464:1347–1350.
Gogarten JP, Townsend JP. 2005. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol. 3:679–687.
Gould SJ, Vrba ES. 1982. Exaptation—a missing term in the science of form. Paleobiology 8:4–15.
Hartl D, Lohe A, Lozovskaya ER. 1997. Regulation of the transposable element mariner. Genetica 100:177–184.
Holmes EC. 2011. The evolution of endogenous viral elements. Cell Host Microbe. 10:368–377.
Houck MA, Clark JB, Peterson KR, Kidwell MG. 1991. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. Science 253:1125–1129.
Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. 2011. The struggle for life of the genome's selfish architects. Biol Direct. 6:19.
Huda A, Mariño-Ramírez L, Jordan IK. 2010. Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. Mobile DNA. 1:2.
Jia J, Xue Q. 2009. Codon usage biases of transposable elements and host nuclear genes in *Arabidopsis thaliana* and *Oryza sativa*. Genomics Proetomics Bioinforma. 7:175–184.
Kang HJ, et al. 2010. Host switch during evolution of a genetically distinct hantavirus in the American shrew mole (*Neurotrichus gibbsii*). Virology 388:8–14.
Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 9:605–618.
Khan H, Smit A, Boissinot S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 16:78–87.
Kidwell MG. 1993. Lateral transfer in natural populations of eukaryotes. Annu Rev Genet. 27:235–256.
Kim A, et al. 1994. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc Natl Acad Sci U S A. 91:1285–1289.
Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. 7:474.
Le Rouzic A, Capy P. 2006. Population genetics models of competition between transposable element subfamilies. Genetics 174:785–793.
Lerat E, Biemont C, Capy P. 2000. Codon usage and the origin of *P* elements. Mol Biol Evol. 17:467–468.
Lerat E, Capy P, Biémont C. 2002. Codon usage by transposable elements and their host genes among five species. J Mol Evol. 54:625–637.
Lerat E, Rizzon C, Biémont C. 2003. Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. Genome Res. 13:1889–1896.
Liu H, et al. 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. J Virol. 84:11876–1187.
Liu H, et al. 2011. Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. BMC Evol Biol. 11:276.
Lohe A, Hartl DL, Lozovskaya ER. 1997. Modern thoughts on an ancient marinere: function, evolution, regulation. Annu Rev Genet. 31: 337–358.
Lohe A, Moriyama E, Lidholm D, Hartl D. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. Mol Biol Evol. 12:62–72.
Longdon B, et al. 2011. Host-switching by a vertically transmitted rhabdovirus in *Drosophila*. Biol Lett. 7:747–750.
Loreto ELS, Carareto C, Capy P. 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity 100:545–554.
Ludwig A, Valente VLS, Loreto ELS. 2008. Multiple invasions of *Errantivirus* in the genus *Drosophila*. Insect Mol Biol. 17:113–124.
Lynch M. 2007. The origin of genome architecture. Sunderland (MA): Sinauer Associates, p. 494.
Lynch M, Conery JS. 2003. The origins of genome complexity. Science 302:1401–1404.
McVean G. 2010. What drives recombination hotspots to repeat DNA in humans? Philos Trans R Soc Lond B Biol Sci. 365:1213–1218.
Mota NR, Ludwig A, Silva Valente VL, Loreto ELS. 2010. Harrow: new *Drosophila hAT* transposons involved in horizontal transfer. Insect Mol Biol. 19:217–228.
Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520–562.
Nemirov K, Henttonen H, Vaheri A, Plyusnin A. 2002. Phylogenetic evidence for host switching in the evolution of hantaviruses carried by Apodemus mice. Virus Res. 90:207–215.
Novick P, Smith J, Ray D, Boissinot S. 2010. Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. Gene 449:85–94.
O'Brochta D, et al. 2009. Transpositionally active episomal *hAT* elements. BMC Mol Biol. 10:108.
Pace JK, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res. 17:422–432.
Pace JK, Gilbert C, Clark MS, Feschotte C. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci U S A. 105:17023–17028.
Parrish CR, et al. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol R. 72:457–470.
Patel MR, Emerman M, Malik HS. 2011. Paleovirology—ghosts and gifts of viruses past. Curr Opin Virol. 1:304–309.
Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 12:32–42.
Poch O, Sauvaget I, Delarue M, Tordo N. 1989. Identification of four conserved motifs among the RNA dependent polymerase encoding elements. EMBO J. 8:3867–3874.

Pritham EJ. 2009. Transposable elements and factors influencing their success in eukaryotes. J Hered. 100:648–655.
Robertson HM. 1997. Multiple mariner transposons in flatworms and hydras are related to those of insects. J Hered. 88:195–201.
Rodelsperger C, Sommer RJ. 2011. Computational archaeology of the *Pristionchus pacificus* genome reveals evidence of horizontal gene transfers from insects. BMC Evol Biol. 11:239.
Routh A, Domitrovic T, Johnson JE. 2012. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proc Natl Acad Sci U S A. 109:1907–1912.
Rubinstein ND, Doron-Faigenboim A, Mayrose I, Pupko T. 2011. Evolutionary models accounting for layers of selection in protein-coding genes and their impact on the inference of positive selection. Mol Biol Evol. 28:3297–3308.
Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol. 25:537–546.
Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF. 1995. DNA sequence evolution: the sounds of silence. Philos Trans R Soc Lond B. 349: 241–247.
Skippington E, Ragan MA. 2011. Lateral genetic transfer and the construction of genetic exchange communities. FEMS Microbiol Rev. 35: 707–735.
Silva JC, Kidwell MG. 2000. Horizontal transfer and selection in the evolution of P elements. Mol Biol Evol. 17:1542–1557.
Silva JC, Loreto ELS, Clark JB. 2004. Factors that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol. 6:57–71.
Sinzelle L, Izsvák Z, Ivics Z. 2009. Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci. 66:1073–1093.
Song SU, Gerasimova T, Kurkulos M, Boeke JD, Corces VG. 1994. An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. Gene Dev. 8:2046–2057.
Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circle transposons among animals. Genome Biol Evol. 2: 656–664.
Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 13: 36–46.
Venner S, Feschotte C, Biémont C. 2009. Dynamics of transposable elements: towards a community ecology of the genome. Trends Genet. 25:317–323.
Vidal NMM, Ludwig A, Loreto ELS. 2009. Evolution of Tom, 297, 17.6 and rover retrotransposons in Drosophilidae species. Mol Genet Genomics. 282:351–362.
Vijaykrishna D, et al. 2007. Evolutionary insights into the ecology of coronaviruses. J Virol. 81:4012–4020.
Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8:973–982.
Woolhouse MEJ, Haydon DT, Antia R. 2005. Emerging pathogens: the epidemiology and evolution of species jumps. Trends Ecol Evol. 20: 238–244.
Xiong Y, Eickbush TH. 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable element, and mitochondrial introns. Mol Biol Evol. 5:675–690.
Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9: 3353–3362.
Yang S, et al. 2004. Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes. Genome Res. 14: 517–527.
Yoshiyama M, et al. 2001. Possible horizontal transfer of a transposable element from host to parasitoid. Mol Biol Evol. 18: 1952–1958.

Associate editor: Ken Wolfe