Distinguishing ecological from evolutionary approaches to transposable elements

Stefan Linquist[1,*], Brent Saylor[2], Karl Cottenie[2], Tyler A. Elliott[2], Stefan C. Kremer[3] and T. Ryan Gregory[2]

[1] *Department of Philosophy, College of Arts, University of Guelph, Guelph, N1G 2W1, Canada*
[2] *Department of Integrative Biology, College of Biological Science, University of Guelph, Guelph, N1G 2W1, Canada*
[3] *School of Computer Science, College of Physical and Engineering Science, University of Guelph, Guelph, N1G 2W1, Canada*

ABSTRACT

Considerable variation exists not only in the kinds of transposable elements (TEs) occurring within the genomes of different species, but also in their abundance and distribution. Noting a similarity to the assortment of organisms among ecosystems, some researchers have called for an ecological approach to the study of transposon dynamics. However, there are several ways to adopt such an approach, and it is sometimes unclear what an ecological perspective will add to the existing co-evolutionary framework for explaining transposon-host interactions. This review aims to clarify the conceptual foundations of transposon ecology in order to evaluate its explanatory prospects. We begin by identifying three unanswered questions regarding the abundance and distribution of TEs that potentially call for an ecological explanation. We then offer an operational distinction between evolutionary and ecological approaches to these questions. By determining the amount of variance in transposon abundance and distribution that is explained by ecological and evolutionary factors, respectively, it is possible empirically to assess the prospects for each of these explanatory frameworks. To illustrate how this methodology applies to a concrete example, we analyzed whole-genome data for one set of distantly related mammals and another more closely related group of arthropods. Our expectation was that ecological factors are most informative for explaining differences among individual TE lineages, rather than TE families, and for explaining their distribution among closely related as opposed to distantly related host genomes. We found that, in these data sets, ecological factors do in fact explain most of the variation in TE abundance and distribution among TE lineages across less distantly related host organisms. Evolutionary factors were not significant at these levels. However, the explanatory roles of evolution and ecology become inverted at the level of TE families or among more distantly related genomes. Not only does this example demonstrate the utility of our distinction between ecological and evolutionary perspectives, it further suggests an appropriate explanatory domain for the burgeoning discipline of transposon ecology. The fact that ecological processes appear to be impacting TE lineages over relatively short time scales further raises the possibility that transposons might serve as useful model systems for testing more general hypotheses in ecology.

Key words: transposable elements, ecology, evolution, multivariate analysis, philosophy of science.

CONTENTS

  * Address for correspondence (Tel: 519-824-4120 ext 56672; E-mail: linquist@uoguelph.ca).

I. INTRODUCTION

The human genome, with its more than 3 billion nucleotides, contains a mere 20000 or so protein-coding genes – only slightly more than the number in a fly or a worm. Even more surprising is the fact that, in stark contrast to the number of genes, the human genome contains more than 1 million copies of a particular genetic element called *Alu*, and more than 500000 copies of another known as LINE-1. Indeed, whereas protein-coding genes comprise less than 2% of the DNA in the human genome, *Alu*, LINE-1, and similar sequences account for somewhere between 45 and 65% of the DNA in human cells (de Koning *et al.*, 2011). In some species, such as maize *Zea mays*, the total is upwards of 85% of the genome (Schnable *et al.*, 2009). *Alu*, LINE-1, and related sequences are known as transposable elements (TEs) because they are capable of moving within the genome and of making copies of themselves in the process.

Based on their capacity for self-replication and their frequently deleterious effects on organism fitness, TEs are often characterized as 'selfish DNA' or 'genomic parasites' that accumulate in huge quantities by evading deletion by the 'host' genome (Doolittle & Sapienza, 1980; Orgel & Crick, 1980). In keeping with this approach, patterns in TE abundance and distribution have been studied from a perspective of host-parasite co-evolution. This application of evolutionary models and concepts to TE biology is not controversial. Indeed, the accumulation or deletion of TEs and their impacts on gene regulation and function is recognized as a major factor in large-scale genome evolution (Hua-Van *et al.*, 2011). However, it is unclear whether this perspective can adequately explain all of the salient patterns in TE abundance and distribution.

More recently, the co-evolutionary framework has been expanded to include dynamics other than simple host-parasite interactions. In an important review, Kidwell & Lisch (2001) pointed out that host-parasite interactions are only one of several possible relationships that TEs could have with their host genomes. As they wrote, 'Rather than labelling TE-host associations as either selfish or parasitic, we prefer the idea of a continuum, ranging from aggressive parasitism at one extreme, through a neutral middle ground, to mutualism at the other extreme' (Kidwell & Lisch, 2001, p. 7). Moreover, the nature of interactions between a particular TE and its host genome could change and move along this continuum over time as the two co-evolve. As part of this discussion, Kidwell & Lisch (2001) introduced the concept of 'the ecology of the genome' to reflect the complex interactions that could occur between TEs and host genomes.

Although most of Kidwell & Lisch's (2001) discussion emphasized co-evolutionary processes, their suggestion that a 'genome ecology' exists in addition to genome evolution has captured the attention of several authors (Brookfield, 2005; Le Rouzic, Dupas & Capy, 2007; Venner, Feschotte & Biémont, 2009). The question, of course, is whether ecological approaches can be brought to bear on questions about TE abundance and diversity, and whether this could provide insights that are not already available through evolutionary analyses. In other words, if there is such a thing as genome ecology, it must be distinguished from the existing theories and models encompassed by genome evolution. The primary goal of this review is to explore the prospects for an ecological, in addition to an evolutionary, approach to TE dynamics.

To address this issue, we first identify three well-known, but currently unexplained patterns in TE abundance and distribution. These are good candidates for an ecological explanation because they resemble the patterns of species abundance and distribution that are routinely explained using a more traditional ecological perspective. The next step in our argument distinguishes two different kinds of ecological approach to TEs. One approach looks at ecological factors external to the host organism and considers their effects on TEs within the host genome. This approach, which we call genome ecology, must not be confused with the perspective that views TEs as analogous to species and the genome as akin to an intracellular ecosystem. The latter approach, which we call transposon ecology, is the central focus of this review. In particular, we are interested in whether this latter approach adds anything new to the already received view that TEs co-evolve with their hosts. Hence, in what follows, we offer an operational distinction between ecological and evolutionary explanations. At the heart of this distinction is the recognition that these disciplines make different kinds of idealizing assumptions about their subject matter. Ecology, in its pure form, aims to investigate the relationships between focal entities (e.g. species, populations, or TE lineages) and their environments. As a simplifying assumption, ecologists often treat these entities as fixed types, ignoring ways that interactions with the environment potentially modify those entities over time. By contrast, evolutionary biology aims to understand how entities change over successive generations.

As a simplification, evolutionary biologists often set aside questions about the kinds of (ecological) interactions that potentially drive those changes. This distinction between ecological and evolutionary perspectives is considered in more detail below. To further illustrate how it applies within the domain of transposon biology, we compare whole-genome data at two scales of host relatedness: among several species of *Drosophila*, and among several species of mammal. The main point of this example is to demonstrate that our conceptual framework allows one to empirically distinguish ecological from evolutionary influences on transposon abundance and distribution. The results of this analysis, and the prospects for transposon ecology, are described and discussed in Sections V and VI.

II. WHAT IS KNOWN AND WHAT IS NOT KNOWN ABOUT TES

TEs are major parts – if not the dominant component – of most eukaryote genomes, meaning that understanding their biology is a fundamental issue in genetics. Since their discovery by McClintock (1946 1947), much has been learned about the molecular properties of TEs. At the broadest level, TEs are classified by their modes of transposition. Class I elements, also called retrotransposons, replicate *via* a 'copy and paste' mechanism involving the production of an mRNA intermediate which is processed, reverse-transcribed into DNA, and inserted back into the genome (Eickbush, 2002; Sandmeyer, Aye & Menees, 2002). The retrotransposons include elements known as short interspersed nuclear elements (SINEs) such as *Alu*, long interspersed elements (LINEs) like LINE-1, and others known as long terminal repeat retrotransposons (LTR-retrotransposons) and endogenous retroviruses (ERVs). Class II elements, or DNA transposons, undergo a 'cut and paste' system of replication in which the elements are physically excised from the genome and reinserted elsewhere. In this case, an increase in copy number occurs during the repair of DNA transposon excision sites by the host during DNA synthesis, or by the element inserting into a site in the genome that has yet to be replicated (Engels *et al.*, 1990; Chen, Greenblatt & Dellaporta, 1992). Elements in both classes are further classified into orders, superfamilies, families, and sub-families based on their relatedness as determined by shared structures and sequence similarity. Aside from these molecular details, much remains to be understood about the basic biology of TEs. In this review, we identify three major outstanding questions in this regard.

(1) Why is there such an enormous difference in TE content among eukaryotes?

Eukaryote genome sizes are estimated to vary more than 200000-fold, with this entire range reported within single-celled protists. Even within animals and plants, genome size estimates range more than 7000-fold and 2000 fold,

respectively (Gregory, 2005; Fig. 1). By way of example, the human genome is 10× larger than that of a pufferfish and 10× smaller than those of many salamanders. This massive variability in total genomic DNA content has remained a major puzzle in genetics for more than 60 years and continues to represent an active topic of research. With the rise of entire-genome sequencing projects, it is becoming increasingly clear that differences in transposable element content contribute substantially to diversity in genome size (Gregory, 2005). This is one question that could potentially be informed by the development of transposon ecology.

(2) Why are there differences in the types of TEs that are most common in the genomes of different organisms?

The various categories of TEs differ not only in their replication and transposition mechanisms and basic structure, but also in their relative abundances in different types of organisms (Fig. 2). For example, the *Alu* element that is hyper-abundant in the human genome is found only in primates and LINE-1 is restricted to vertebrates. In humans, ERVs and DNA transposons are far less abundant than SINEs and LINEs. However, in the frog *Xenopus laevis*, DNA transposons outnumber retrotransposons by almost 2:1 (Hellsten *et al.*, 2010). In many plants, including grasses and massively genomed lilies, LTR-retrotransposons appear to dominate (Feschotte, Zhang & Wessler, 2002; Ambrožová *et al.*, 2011). At present, data are too limited to allow the identification of clear patterns among taxa, but it is evident that substantial variability exists in the composition of different genomes and this will require an explanation. Similarly, these differences can manifest at different scales. At the coarse-grained level described above, the salient differences are among TE classes (retrotransposons *versus* DNA transposons). However, one could equally investigate differences at a finer grain. For example, why is there an association between particular TE superfamilies, families, or individual elements and certain genomes? One possibility is that these associations are due to historically contingent events, such as the chance horizontal transmission of a TE into a genome. However, it is also plausible that structural features of the host genome have an effect on the sorts of TEs that will flourish. An ecological approach to transposon dynamics potentially could shed light on this question.

(3) What accounts for differences in the distribution of particular TEs within a given genome?

Eukaryotic genomes are physically subdivided into individual linear chromosomes which vary greatly in number among species, even among some close relatives. The chromosomes within a species' genome also often differ significantly from one another in size, gene density, DNA compaction level, and other features including which TEs they harbour and how abundant those TEs are. A simple example is provided by the distribution of *Alu* elements across the 23 human chromosomes, which is clearly not uniform
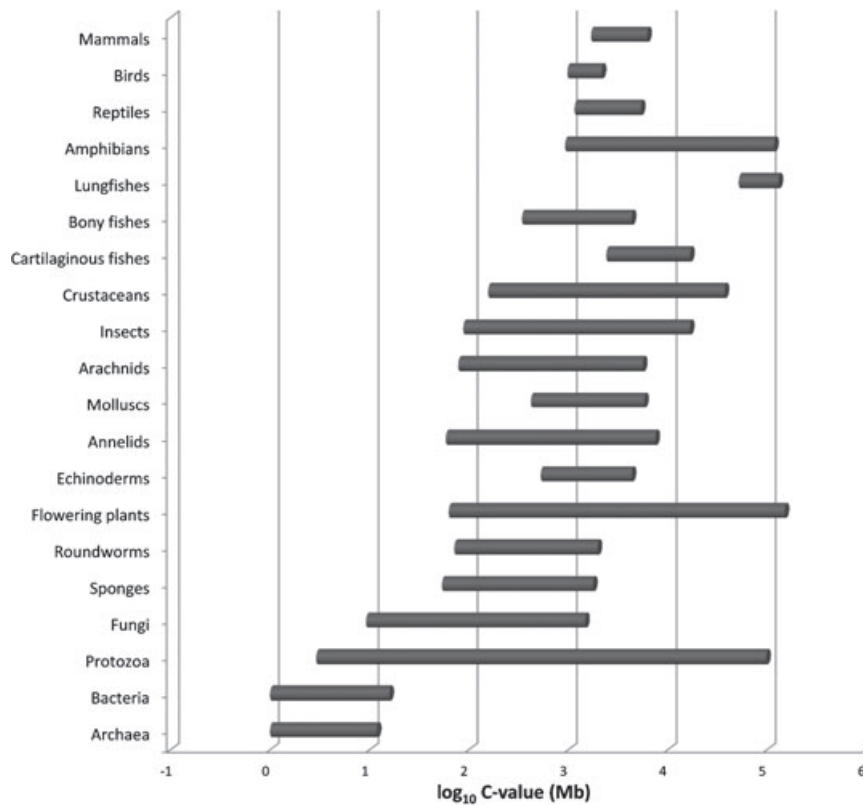
Fig. 1. Ranges in haploid nuclear DNA content ('C-value') for various groups of organisms. There is no relationship between genome size and morphological complexity or number of genes in eukaryotes. Data from the Animal Genome Size Database (http://www.genomesize.com/) and the Plant DNA C-values Database (http://data.kew.org/cvalues/).

(Bolzer *et al.*, 2005; Fig. 3). Some chromosomes exhibit a high concentration of *Alu* elements whereas others are rather desolate. Even within individual chromosomes there can be substantial variability in *Alu* abundance. Patterns at both inter- and intra-chromosomal levels should be identified and accounted for, and it is possible that the explanation lies in ecological processes unfolding within the host genome.

III. CAN AN ECOLOGICAL APPROACH SHED LIGHT ON THE OUTSTANDING QUESTIONS?

As other researchers have noticed, there are important similarities between these questions about TEs within genomes and those studied by ecologists focused on organisms within ecosystems (e.g. Kidwell & Lisch, 2001; Brookfield, 2005; Le Rouzic *et al.*, 2007; Venner *et al.*, 2009). Like transposon families in the genome, species also distribute non-randomly and in varying abundances within ecosystems. Particular species are also more or less successful at colonizing new ecosystems. The nature of these key questions therefore lends some credence to the notion that transposon ecology could be a useful approach in understanding patterns of TE abundance and distribution.

Several recent authors have begun to explore TE biology within the context of 'genome ecology', although as will be

seen, some important conceptual clarifications must still be made. In one notable example, Brookfield (2005) outlined five questions that he argued could be answered using an ecological approach. These questions can be summarized as follows: (*i*) are the pressures exerted on TEs in and/or by the host genome density dependent and in equilibrium? (*ii*) What factors determine the total proportion of a genome consisting of TEs? (*iii*) What are the effects of mutations in TEs on TE activity? (*iv*) Will mutations in host genes that reduce transposition rates spread through host populations? (*v*) To what extent is the evolution of TE lineages tied to the evolution of specific host lineages.

Although these questions contain elements of ecological concepts, they also address evolutionary questions. For example, questions 3, 4, and 5 are primarily about the molecular biology, population genetics, and co-evolution of TEs and their host genomes. Some authors clearly place these evolutionary processes in the same category with ecological processes. However, this framework makes it difficult to assess whether transposon ecology is indeed a novel approach within transposon biology, and if so, how exactly it differs from more traditional modes of investigation.

In another important contribution, Venner *et al.* (2009) sought to develop further the notion of genome ecology by focusing specifically on community ecology and questions relating to niche partitioning due to competition. They also
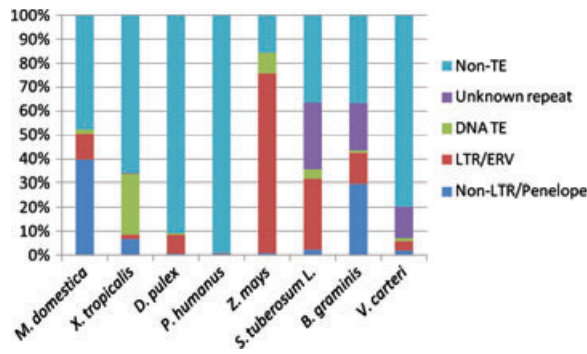
Fig. 2. Percentage of the genome composed of transposable elements (TEs) in various eukaryotic species. *Monodelphis domestica*, opossum (Gentles *et al.*, 2007); *Xenopus tropicalis*, western clawed frog (Hellsten *et al.*, 2010); *Daphnia pulex*, water flea (Colbourne *et al.*, 2011); *Pediculus humanus*, human body louse (Kirkness *et al.*, 2010); *Zea mays*, corn (Schnable *et al.*, 2009); *Solanum tuberosum*, potato (The Potato Genome Sequencing Consortium, 2011); *Blumeria graminis*, powdery mildew fungus (Spanu *et al.*, 2010); *Volvox carteri*, multicellular green alga (Prochnik *et al.*, 2010). LTR, long terminal repeat retrotransposons; ERV, endogenous retrovirus. *Penelope* elements are retrotransposons which are sister taxa to non-LTRs.

took the key step of identifying the analogous components of ecosystems and genomes, for example, linking TEs with species, transposition rates with birth rates, and so on (see Table 1 in Venner *et al.*, 2009). However, in this case there appears to have been some mixing of ecological processes at two distinct levels. For example, Venner *et al.* (2009) note that bats inhabit relatively large home ranges compared to other mammals by virtue of their unique ability to fly. The mobility of these flying mammals, they suggest, could selectively favour active transposons by providing increased opportunities for horizontal transmission among populations of bats. In this example, the relevant ecological factors are the home ranges of bats and their encounter rates with other populations. These are ecological factors external to the host organism that are conceptually distinct from ecological processes possibly occurring within the host genome. It is

important to distinguish ecological processes at these two levels. To avoid confusion, we reserve the term 'genome ecology' for the study of organism-level ecological factors that impact features of the genome, including transposons. 'Transposon ecology' shall refer to ecological processes occurring within the host genome, involving interactions among TEs and the molecular environment. The focus of this review is on the latter approach.

A third example, which addresses question three (Section II.3), is provided by Abrusán & Krambeck (2006). These authors used a modified Lotka-Volterra model to test the hypothesis that RNA interference defences of the host organism drive the diversity and abundance of TEs. This model compares genomic defence systems to predators, while TEs are viewed as a type of prey. Clearly, this research falls within the purview of transposon ecology by focusing on ecological questions within the genome. However, it also addresses the evolutionary question of whether this interaction selects for new TE lineages and increased TE diversity. Thus, even the explicit use of a well-known ecological model involved a hybrid of ecological and evolutionary approaches.

To be sure, explanations for many phenomena of interest at both organismal and genomic levels will require both ecological and evolutionary components. However, it remains a useful and important exercise to distinguish what those components are, how they contribute individually to an explanatory framework, and how their roles can be quantified in separation and in combination. For this reason, it is necessary to articulate clearly the ways in which the concepts of ecology and evolution are used operationally when studying TEs within genomes as well as organisms within ecosystems.

IV. ECOLOGY *VERSUS* EVOLUTION: OPERATIONAL DEFINITIONS

The very prospect of a genome ecology raises questions about what distinguishes an ecological approach in general. Herein we adopt a specific definition of 'ecology' because we are

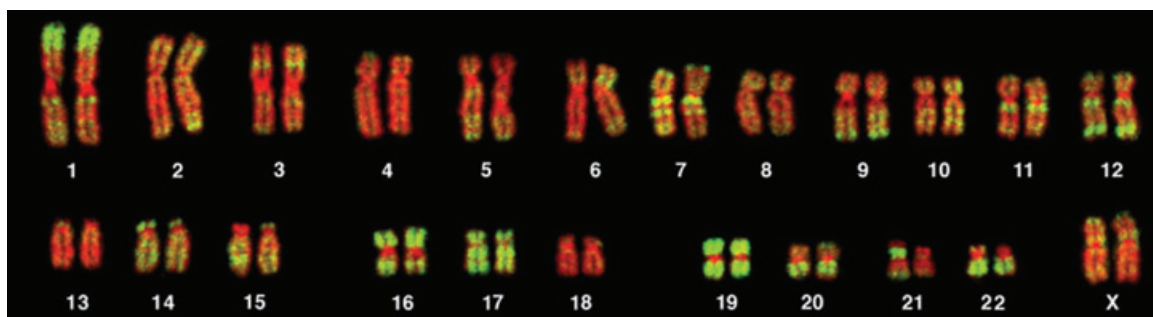**Fig. 3.** Karyotype from a human female, showing the different locations and abundances of the transposable element *Alu* on each chromosome as revealed by fluorescent *in situ* hybridization (FISH). *Alu*, a short interspersed nuclear element (SINE), is present in more than 1 million copies in the human genome. Image reproduced from Bolzer *et al.* (2005), *PLoS Biology* **3**(5): e157. DOI: 10.1371/journal.pbio.0030157.

interested in whether an ecological approach to transposons adds anything new to the existing co-evolutionary framework for explaining their dynamics. This question requires an operational distinction between ecology and evolution. In most biological systems, ecological and evolutionary processes interact. In this respect, these two kinds of process are analogous to genetic and environmental influences on trait development. It is possible to distinguish genetic from environmental influences on development, but doing so requires the right sort of data (Sober, 2000). Specifically, one requires a data set consisting of a population of individuals who vary in both genes and environments. It is then possible to tease apart the contributions of these two factors by partitioning the variance. Our approach to the question of whether ecological processes are occurring within the genome involves a similar technique. Looking at a population of genomes that vary in transposon abundance and distribution, we ask how much of this variation can be explained by ecological and evolutionary factors, respectively. Hence, we adopt the following definitions of 'evolution' and 'ecology'.

(1) A strictly evolutionary approach investigates change (or the lack thereof) in some focal entity over successive generations. The focal entities can range from genes to traits or from populations to higher taxonomic units.

(2) A strictly ecological approach assumes no change in the focal entities themselves, but focuses instead on the relationships between these entities and their environment. Here we use 'environment' in a broad sense potentially to include any of the factors with which an entity interacts.

In identifying these as distinct approaches, our point is simply that it is possible to focus on just one type of influence at a time. By analogy, one might focus specifically on whether the frequency of an allele changes over time and at what rate without asking about the ecological factors driving its change. This would be a strictly evolutionary approach to that gene. Likewise, one might investigate how organisms of a given species interact with their environment without asking about longer-term impacts on the genetic make-up of the population. This would be a strictly ecological approach to that species.

Before applying this distinction at the transposon level, the difference between ecological and evolutionary approaches can be illustrated with an example at the species level. The introduction of the North American beaver (*Castor canadensis*) in the southern tip of South America provides an illuminating case study in which ecological and evolutionary processes have been studied independently. In 1946, 50 beavers were released in Tierra del Fuego, Argentina, and consequently spread throughout the region. Some researchers have focused exclusively on the relations between this population and its environment. For example, the rate of the beavers' dispersal has been tied to the absence of predators in this habitat (Skewes *et al.*, 2006). Similarly, ecologists have investigated the influence of these beaver colonies on the local diversity of macroinvertebrates (Anderson & Rosemond, 2007) and plant communities (Anderson *et al.*, 2006). These investigations are purely ecological, in our sense, because

they explain patterns in beaver abundance and distribution exclusively in terms of relations to the local environment, while ignoring (for the sake of simplicity) recent evolutionary changes. Other investigators have focused exclusively on the genetic changes that this population has undergone over the 40 years since its introduction. For example, Lizarralde *et al.* (2008) identified 10 genetic lineages different from the original source population. In this case, the researchers investigated the evolution of this population while setting aside questions about the influence of particular ecological factors.

Although they often interact in nature, ecological and evolutionary processes can be, and often are, studied independently by different researchers adopting different approaches. This division of labour offers practical benefits over (or in addition to) a more complicated hybrid approach. For instance, in some cases ecological factors explain a negligible amount of the variance among a group of systems. This can be illustrated by imagining a slight modification to the Argentinian beaver example. Suppose that the founding population happened to contain a high frequency of a genetic variant, uncommon in the source population, that is especially industrious compared to conspecifics. If its frequency in North American populations was greater, this industrious variety would have the same impact that is currently being witnessed in Argentina. In this case, the explanation for the difference in abundance and distribution between these two regions would be purely evolutionary, in our sense, because it appeals to a change in the genetic frequencies among two populations. By contrast, natural experiments like the one unfolding in Argentina suggest that at least some of the variation in beaver abundance and distribution is due, in fact, to ecological differences.

When it comes to explaining the variation in abundance and distribution of transposons among genomes, it is an open question how much of this will involve ecological as opposed to evolutionary factors. It is also important to consider whether the distribution of explanatory effort depends on the level of grain, that is, on whether the relevant differences are among particular TEs, transposon families, or perhaps some higher classification. To demonstrate how one might address these questions, we conducted an analysis of whole-genome data for two groups of organisms: one very distantly related group of mammals and a less distantly related group of arthropods. Our analysis partitioned the amount of variation in TE abundance and distribution that can be explained by ecological and evolutionary factors, respectively, both at the level of individual TEs and at the level of transposon families. The remainder of this review describes this analysis and discusses its implications.

V. PROOF OF PRINCIPLE: DISTINGUISHING ECOLOGICAL AND EVOLUTIONARY FACTORS IN AN EXAMPLE ANALYSIS

From the perspective of a transposon, it seems plausible that the genomes of closely related hosts present similar

environments. Likewise, distantly related hosts probably present dissimilar environments. So, a bacterial genome and a mammalian genome might, to a transposon, be as ecologically dissimilar as a temperate pond and a tropical rainforest. Insofar as the host genome is the ecosystem for a TE, it is thus possible to distinguish different genomic environments in terms of host relatedness. Perhaps the most effective way to identify whether ecological factors influence TE abundance and distribution would be to compare a given lineage in maximally dissimilar (unrelated) environments. However, a potentially confounding factor is that TE lineages also diverge along with host genomes. Over large time scales it is therefore difficult to uniquely partition ecological and evolutionary influences. Hence, if one aims to identify strictly ecological (as opposed to evolutionary) processes, it is necessary to compare closely related TEs in closely related host genomes. So, we expect that the probability of finding evidence for the importance of ecological processes independently from evolutionary processes is highest when comparing closely related rather than distantly related host genomes, and when comparing individual TE lineages instead of TE families.

(1) Proxies for ecological factors

To test this prediction (and the prospects for transposon ecology generally), we used proxies for evolutionary and ecological processes that could be applied to an analysis of transposon data. In organism-level ecology, other species constitute the biotic environment and physical features (e.g. rainfall, topography, size of the habitat) make up the abiotic environment. At the TE level, the size of the host genome can be conceptually related to the size of the ecosystem, which can affect the carrying capacities of transposon lineages or families and cause competition for space. Second, as with ecosystems, the genomes of eukaryotic species are neither identical nor uniform internally. A simple parameter that illustrates this is given by variation in the quantity and distribution of different nucleotides – that is, GC or AT content. Importantly, some TEs are found preferentially in either GC-rich or GC-poor regions of the genome, which creates an analogy with ecological niches. Moreover, many protein-coding genes are found in areas of high GC content, whereas the rest of the genome tends to have a lower GC content (Galtier *et al.*, 2001). This has the potential to create areas of the genome in which insertion by a TE is relatively 'safe' and others where a deleterious effect on a gene and an associated selective pressure against insertions is more likely. Both proxies chosen for our analysis, genome size and GC content, thus represent the environment of the TEs.

(2) Proxies for evolutionary factors

An obvious parameter of interest from an evolutionary perspective is relatedness. With the exception of horizontal transfer, TEs are directly propagated by/in their host genomes (Schaack, Gilbert & Feschotte, 2010). This makes host relatedness an acceptable proxy for TE relatedness. Data on host relatedness are available from any reliable phylogeny with a distance measurement. The distances between host genomes measured in these phylogenies are representative of the relatedness of the entirety of the genomes in which they reside.

(3) Analytical approach

Our expectation is that distinctively ecological processes will be most salient over relatively short times scales. To test this, we analyzed two sets of host genomes with differing degrees of phylogenetic relatedness. We compared a dataset of genomes from 10 species in the genus *Drosophila* [maximum divergence time approximately 50 million years ago (Mya); *Drosophila* 12 Genomes Consortium, 2007] and a dataset consisting of 13 distantly related mammals including placentals, marsupials, and monotremes (maximum divergence time approximately 165 Mya; Warren *et al.*, 2008). These two sets of host genomes were used because they had published full genome sequences as well as phylogenies with available evolutionary distance measurements (Appendix 1) (Bininda-Emonds *et al.*, 2007; *Drosophila* 12 Genomes Consortium, 2007). Another expectation is that, if there are purely ecological patterns distinct from evolutionary patterns, they will be most salient at the level of the individual TE lineage compared with the TE family. Hence, after identifying all of the TEs in each genome using RepeatMasker (Smit, Hubley & Green, 2004), the amount of each genome covered by each TE family and each TE lineage was calculated. A redundancy analysis (RDA) was then used to calculate the amount of variation in the proportion of a genome made up of TEs which was explained by the ecological and evolutionary proxies. This resulted in an adjusted r^2 for both the evolutionary and ecological proxies. This process was carried out at both the TE family and the individual TE levels on both sets of genomes. For each of these host genome sets, we obtained TE abundances at two levels of resolution: the abundance of the individual TE lineages within each host genome (as determined by a high degree of sequence similarity), and the abundances of the TE families within each host genome.

We then computed the amount of variation within either the individual TE abundances or TE family abundances that is associated with either the ecological or environmental proxies (see Appendix 1 for details). If our prediction is true, we expect to find the largest explained variation for the analysis with individual TE abundances for the *Drosophila* host genomes.

(4) Evidence of ecological and evolutionary influences on TEs

The results of this study are summarized in Fig. 4. In our analysis of the more distantly related group of genomes (mammals), evolutionary proxies explained 30% of the observed variation among the TE families ($P = 0.025$), and 14% of the variation among the individual TEs ($P = 0.02$). In other words, differences in overall numbers of different TE families are largely explained by the divergence times in these mammals. That is, the more distantly related species exhibit larger differences in the relative amounts
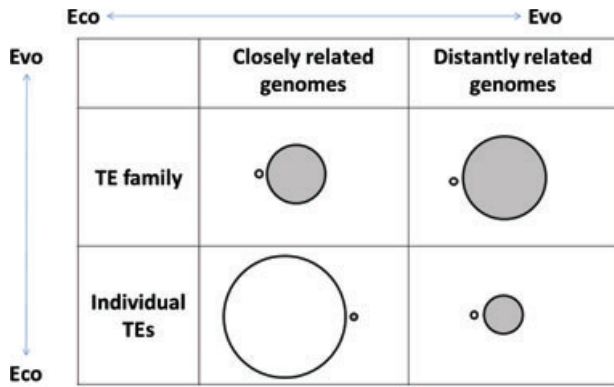
Fig. 4. Distinguishing between evolutionary and ecological processes in shaping transposable element (TE) distribution and abundance in animal genomes. Columns represent groups of host genomes [*Drosophila* species (closely related) *versus* mammals (distantly related)]; rows represent the taxonomic level at which TEs were assessed (individually or by family). Circle size represents the degree to which the evolutionary (grey) and ecological (white) proxies explain the observed variation. The lack of overlap between these circles indicates that these factors do not interact. Arrows outside the box indicate the relative contribution of ecological *versus* evolutionary factors based on the proxies chosen. See Appendix 1 for details.

of TE families and less distantly related species show fewer differences. Among these distantly related genomes, ecological processes did not explain a significant amount of variation at either the family or individual TE levels (Fig. 4).

For the more closely related group of genomes (10 *Drosophila* species), at the TE family level, evolutionary proxies were found to be a significant factor explaining 21% of the variation ($P = 0.01$) and the ecological factors were not found to be significant. However, at the individual TE level, evolutionary processes were not found to explain a significant amount of the variation among genomes, while ecological processes explained 44% of the variation among this group of genomes ($P = 0.005$). In other words, in *Drosophila*, differences in the overall numbers of different TE lineages is largely explained by the ecological factors of genome size and GC content, and not by divergence time. These results completely correspond with our prediction of finding evidence for ecological processes independent from evolutionary processes in closely related host genomes for individual TE abundances.

These findings suggest that the appropriate explanatory domain for transposon ecology is variation at the level of individual TE lineages across relatively closely related taxa. However, given that there is still substantial evolutionary divergence among *Drosophila* lineages (*Drosophila* 12 Genomes Consortium, 2007; Stark *et al.*, 2007), further analyses like this are required to establish this point more conclusively. Ideally, such studies will investigate a wide range of different taxa as data on their complete genomes become available. Another limitation of the current study is that it does not take into account the potential effects of horizontal transmission. In this analysis, we used host relatedness as a proxy for TE relatedness. However, TEs are known occasionally to jump

across taxa. Such events potentially exaggerate the influence of ecological factors, while downplaying the importance of evolutionary influences on TE abundance and distribution. Hence, future applications of this methodology might establish TE relatedness independently of host relatedness, or, attempt to remove horizontally transferred TEs from the analysis.

VI. IMPLICATIONS FOR THE OUTSTANDING QUESTIONS IN TE BIOLOGY

As other researchers have noted, interactions between TEs and host genomes are similar in many respects to the interactions between species and their environments. Like organisms, TEs enjoy a degree of mobility. TE lineages and families sort themselves in the genome in ways that resemble the distribution of particular species and genera within an ecosystem. Likewise, genomes offer resources or 'niches' that TEs might inhabit, or for which they perhaps even compete. Such similarities suggest that ecological concepts and models could illuminate our understanding of transposon dynamics. Herein, we have distinguished two distinct ways that one might adopt an ecological perspective towards TEs. Genome ecology (as defined here) considers the effect on TEs of ecological processes encountered by the host organism. Transposon ecology, the approach that we have focused on here, regards the host genome as a mini-ecosystem in which ecological processes unfold at the molecular level.

To identify the prospects of transposon ecology, we drew an operational distinction between ecology and evolution. If ecological processes are occurring within the genome, one would expect the effect to be most noticeable over relatively short time periods and at a relatively fine level of grain. Hence, comparing closely related genomes, one expects to find that a significant amount of the variation in TE abundance and distribution is explained by ecological factors. Had no such co-variation been identified, this would have indicated either that our proxies for ecological factors were unreliable or that the prospects for transposon ecology are quite grim. To the contrary, we found that a large portion of the variance in abundance and distribution among closely related TEs is explained by the ecological proxies that were selected. Let us now consider the implications of this finding for the three questions outlined at the beginning of this review.

The first question concerned differences in TE abundance across eukaryote genomes. To what extent can these differences be explained from a purely ecological perspective? Our analysis discovered that ecological processes are most discernable among closely related host organisms. Hence, differences in TE abundance across distantly related eukaryotes are unlikely to be explained at the level of transposon ecology. Such large-scale patterns are more likely explained from a purely evolutionary perspective. However, any variation in TE abundance among closely related species, such as *Alu* in primates, is a good candidate for a purely ecological explanation.

The second question concerned differences in the types of TEs that are most abundant in various host genomes. To what extent are these differences explained at the level of transposon ecology? Again, our analysis suggests that ecological concepts and models will have the greatest explanatory significance at the level of the differences in individual lineages. Differences in transposon lineages might also call for some degree of ecological explanation. However, larger differences such as the prevalence of Class I *versus* Class II transposons are unlikely to be explained in terms of genome ecology.

The third question was whether ecological processes explain the spatial distribution of TEs within a particular host genome. Our current analysis suggests only that such processes are most likely to be detected by comparing closely related hosts. One way to detect ecological processes at this level would involve comparing the spatial distribution of particular TE lineages or families among closely related genomes. Co-variation between ecological proxies and spatial location would then be indicative of ecological processes. At an even more fine-grained level, one could search for the influence of ecological factors by comparing chromosomes within a particular species genome. In a future paper we will report the results of such an investigation (B. Saylor, T.A. Elliott, S. Linquist, S.C. Kremer, T.R. Gregory & K. Cottenie, in preparation), in which ecological factors were again identified as playing a significant role in determining TE spatial distribution.

Our analysis points towards a viable future for transposon ecology. A logical next step will identify the particular kinds of ecological processes occurring at this level. For example, it would be interesting to investigate the extent to which TE abundance and distribution is governed by competition for high-quality genomic 'niches'. Chromosome number and the proportion of the genome made up of DNA that is compact (heterochromatin) or more loosely aggregated (euchromatin) could also be envisioned as providing patchiness to the environment of the genome (Kidwell & Lisch, 1997). Similarly, Abrusán & Krambeck (2006) proposed the idea that the mechanisms used by the genome to suppress TE expression could have an effect on the abundance of TEs, although obtaining reliable data on the relative importance of these different mechanisms within any given genome would be difficult. Another avenue for future research could investigate whether there are particular ecological strategies on which certain TE lineages specialize, while other TEs adopt a more generalist strategy. A further question concerns the relative influence of 'biotic' *versus* 'abiotic' factors at this level. From the perspective of transposon ecology, it makes sense to consider active TEs as akin to biotic factors, while non-mobile parts of the genome are akin to abiotic factors. These designations are complicated, however, by the fact that TEs can switch from active to dormant and (occasionally) back again. Addressing these issues will require further theoretical as well as empirical investigation.

Another interesting implication of this line of research concerns its significance for traditional questions in ecology.

A familiar challenge for ecologists operating at the whole-organism level is that ecosystems are not clearly bounded in space. This poses problems for testing ecological hypotheses and models, because it is often difficult to determine whether two systems are truly independent. Another challenge for whole-organism ecologists is that their study organisms are often difficult to track in space and time. Both of these problems are avoided by the transposon ecologist. Individual genomes constitute well-bounded and fairly independent ecosystems in which ecological investigations can be replicated. Similarly, the discrete nature of nucleotides and their location along a single, linear dimension makes it possible to track ecological changes at an extremely fine level of grain. For these reasons, the promising findings reported herein will be of interest not only to researchers in transposon biology, but to ecologists interested in testing particular hypotheses and models.

VII. CONCLUSIONS

(1) The proposal that an 'ecological approach to the genome' might shed light on some of the outstanding questions surrounding transposon abundance and distribution enjoys strong *prima facie* support. However, previous attempts do not always distinguish ecological processes occurring at different levels. Nor do these attempts explain how an ecological approach differs from the (already received) co-evolutionary approach to transposon dynamics.

(2) In an effort to avoid the first sort of confusion it is important to distinguish genome ecology from transposon ecology. Genome ecology is the study of how ecological processes that are external to the organism impact its genome. Transposon ecology, by contrast, is the study of how ecological processes unfolding within the genome of a host organism impact the abundance and distribution of TEs.

(3) In an effort to determine the explanatory prospects for transposon ecology, it is helpful to distinguish evolutionary from ecology perspectives more generally. A strictly evolutionary perspective investigates change (or the lack thereof) in some focal entity over successive generations, and tends to set aside questions about the environmental factors driving that change. A strictly ecological approach assumes no change in the focal entities themselves, but focuses instead on the relationships between the entities and their environment.

(4) Given this distinction, it is possible to assess the respective contributions of evolutionary and ecological factors by partitioning the variance among a set of entities with known relatedness. Conducting this kind of analysis requires that one is able to determine how much of the variance among those entities is explained by their relatedness (evolution), and also how much of the variance is explained by environmental variables (ecology). If there is a significant amount of variance explained by ecological and not evolutionary factors, then an ecological approach is warranted.

(5) Applying this analysis to an example of two taxonomic groups helps to demonstrate the utility of this approach. We analyzed how much of the variance in TE abundance and distribution is explained by evolutionary and ecological factors, respectively, both for distantly related (mammals) and more closely related (*Drosophila*) taxa. Our preliminary finding is that ecological factors explain variation at the level of individual TE lineages (and not at the level of families) among the closely related taxa only. Evolutionary factors were not explanatory at this level; however, they do explain significant amounts of variation at the family level and among more distantly related taxa.

(6) A logical next step for the field of transposon ecology is to identify the particular kinds of ecological processes impacting TE abundance and distribution.

(7) Transposons are potentially useful model systems for addressing more general theoretical issues in ecology. Our example analysis suggests that ecological interactions will be most salient at the level of individual TE lineages and over relatively small phylogenetic distances.

VIII. ACKNOWLEDGMENTS

IX. REFERENCES

ABRUSÁN, G. & KRAMBECK, H.-J. (2006). Competition may determine the diversity of transposable elements. *Theoretical Population Biology* **70**, 364–375.

AMBROŽOVÁ, K., MANDÁKOVÁ, T., BUREŠ, P., NEUMANN, P., LEITCH, I. J., KOBLÍŽKOVÁ, A., MACAS, J. & LYSAK, M. A. (2011). Diverse retrotransposon families and AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. *Annals of Botany* **107**, 255–268.

ANDERSON, C. B., GRIFFITH, C. R., ROSEMOND, A. D., ROZZI, R. & DOLLENZ, O. (2006). The effects of invasive North American beavers on riparian plant communities in Cape Horn, Chile. Do exotic beavers engineer differently in subantarctic ecosystems? *Biological Conservation* **128**, 467–474.

ANDERSON, C. B. & ROSEMOND, A. D. (2007). Ecosystem engineering by invasive exotic beavers reduces in-stream diversity and enhances ecosystem function in Cape Horn Chile. *Oecologia* **154**, 141–153.

BININDA-EMONDS, O. R. P., CARDILLO, M., JONES, K. E., MACPHEE, R. D. E., BECK, R. M. D., GRENYER, R., PRICE, S. A., VOS, R. A., GITTLEMAN, J. L. & PURVIS, A. (2007). The delayed rise of present-day mammals. *Nature* **446**, 507–512.

BOLZER, A., KRETH, G., SOLOVEI, I., KOEHLER, D., SARACOGLU, K., FAUTH, C., MÜLLER, S., EILS, R., CREMER, C., SPEICHER, M. R. & CREMER, T. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biology* **3**, e157.

BROOKFIELD, J. F. Y. (2005). The ecology of the genome — mobile DNA elements and their hosts. *Nature Reviews Genetics* **6**, 128–136.

CHEN, Y., GREENBLATT, I. M. & DELLAPORTA, S. L. (1992). Molecular analysis of *Ac* transposition and DNA replication. *Genetics* **130**, 665–676.

COLBOURNE, J. K., PFRENDER, M. E., GILBERT, D., THOMAS, W. K., TUCKER, A., OAKLEY, T. H., TOKISHITA, S., AERTS, A., ARNOLD, G. J., BASU, M. K., BAUER, D. J., CACERES, C. E., CARMEL, L., CASOLA, C., CHOI, J.-H., DETTER, J. C., DONG, Q., DUSHEYKO, S., EADS, B. D., FROHLICH, T., *et al*. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555–561.

DE KONING, A. P. J., GU, W., CASTOE, T. A., BATZER, M. A. & POLLOCK, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics* **7**, e1002384.

DOOLITTLE, W. F. & SAPIENZA, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603.

Drosophila 12 Genomes Consortium (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218.

EICKBUSH, T. H. (2002). R2 and related site-specific non-long terminal repeat retrotransposons. In *Mobile DNA II* (eds N. L. CRAIG, R. CRAIGIE, M. GELLERT and A. M. LAMBOWITZ), pp. 813–835. American Society of Microbiology Press, Washington.

ENGELS, W. R., JOHNSON-SCHLITZ, D. M., EGGLESTON, W. S. & SVEDT, J. (1990). High-frequency P element loss in *Drosophila* is homolog dependent. *Cell* **62**, 515–525.

FESCHOTTE, C., ZHANG, X. & WESSLER, S. R. (2002). Miniature inverted-repeat transposable elements and their relationship to established DNA transposons. In *Mobile DNA II* (eds N. L. CRAIG, R. CRAIGIE, M. GELLERT and A. M. LAMBOWITZ), pp. 1147–1158. American Society for Microbiology Press, Washington.

GALTIER, N., PIGANEAU, G., MOUCHIROUD, D. & DURET, L. (2001). GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* **159**, 907–911.

GENTLES, A. J., WAKEFIELD, M. J., KOHANY, O., GU, W., BATZER, M. A., POLLOCK, D. D. & JERKA, J. (2007). Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Research* **17**, 992–1004.

GREGORY, T. R. (2005). Synergy between sequence and size in large-scale genomics. *Nature Reviews Genetics* **6**, 699–708.

HELLSTEN, U., HARLAND, R. M., GILCHRIST, M. J., HENDRIX, D., JURKA, J., KAPITONOV, V., OVCHARENKO, I., PUTNAM, N. H., SHU, S., TAHER, L., BLITZ, I. L., BLUMBERG, B., DICHMANN, D. S., DUBCHAK, I., AMAYA, E., DETTER, J. C., FLETCHER, R., GERHARD, D. S., GOODSTEIN, D., GRAVES, T., *et al*. (2010). The genome of the western clawed frog *Xenopus tropicalis*. *Science* **328**, 633–636.

HUA-VAN, A., LE ROUZIC, A., BOUTIN, T. S., FILÉE, J. & CAPY, P. (2011). The struggle for life of the genome's selfish architects. *Biology Direct* **6**, 19.

KIDWELL, M. G. & LISCH, D. R. (1997). Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 7704–7711.

KIDWELL, M. G. & LISCH, D. R. (2001). Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**, 1–24.

KIRKNESS, E. F., HAAS, B. J., SUN, W., BRAIG, H. R., PEROTTI, M. A., CLARK, J. M., LEE, S. H., ROBERTSON, H. M., KENNEDY, R. C., ELHAIK, E., GERLACH, D., KRIVENTSEVA, E. V., ELSIK, C. G., GRAUR, D., HILL, C. A., VEENSTRA, J. A., WALENZ, B., MANUEL, J., TUBÍO, C., RIBEIRO, J. M. C., *et al*. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 12168–12173.

LEGENDRE, P. & LEGENDRE, L. (1998). *Numerical Ecology*. Second Edition. Elsevier Science, Amsterdam.

LE ROUZIC, A., DUPAS, S. & CAPY, P. (2007). Genome ecosystem and transposable elements species. *Gene* **390**, 214–220.

LIZARRALDE, M. S., BAILLIET, G., POLJAK, S., FASANELLA, M. & GIULIVI, C. (2008). Assessing genetic variation and population structure of invasive North American beaver (*Castor canadensis* Kuhl, 1820) in Tierra Del Fuego (Argentina). *Biological Invasions* **10**, 673–683.

MCCLINTOCK, B. (1946). Maize genetics. *Carnegie Institute of Washington Yearbook* **45**, 176–186.

MCCLINTOCK, B. (1947). Cytogenetic studies of maize and *Neurospora*. *Carnegie Institute of Washington Yearbook* **46**, 146–152.

ORGEL, L. E. & CRICK, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.

PERES-NETO, P. R., LEGENDRE, P., DRAY, S. & BORCARD, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87**, 2614–2625.

PROCHNIK, S. E., UMEN, J., NEDELCU, A. M., HALLMANN, A., MILLER, S. M., NISHII, I., FERRIS, P., KUO, A., MITROS, T., FRITZ-LAYLIN, L. K., HELLSTEN, U., CHAPMAN, J., SIMAKOV, O., RENSING, S. A., TERRY, A., PANGILINAN, J., KAPITONOV, V., JURKA, J., SALAMOV, A., SHAPIRO, H., *et al*. (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* **329**, 223–226.

SANDMEYER, S. B., AYE, M. & MENEES, T. (2002). Ty3, a position-specific *gypsy*-like element in *Saccharomyces cerevisiae*. In *Mobile DNA II* (eds N. L. CRAIG, R. CRAIGIE, M. GELLERT and A. M. LAMBOWITZ), pp. 663–683. American Society of Microbiology Press, Washington.

SCHAACK, S., GILBERT, C. & FESCHOTTE, C. (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in Ecology & Evolution* **25**, 537–546.

SCHNABLE, P. S., WARE, D., FULTON, R. S., STEIN, J. C., WEI, F., PASTERNAK, S., LIANG, C., ZHANG, J., FULTON, L., GRAVES, T. A., MINX, P., REILY, A. D., COURTNEY, L., KRUCHOWSKI, S. S., TOMLINSON, C., STRONG, C., DELEHAUNTY, K., FRONICK, C., COURTNEY, B., ROCK, S. M., *et al*. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115.

SKEWES, O., GONZALES, F., OLAVE, R., ÁVILA, A., VARGAS, V., PAULSEN, P. & KÖNIG, H. E. (2006). Abundance and distribution of American beaver, *Castor*

canadensis (Kuhl 1820), in Tierra del Fuego and Navarino Islands, Chile. *European Journal of Wildlife Research* **52**, 292–296.

SMIT, A., HUBLEY, R. & GREEN, P. (2004). *RepeatMasker*. Institute for Systems Biology, Seattle.

SOBER, E. (2000). Appendix one: the meaning of genetic causation. In *From Chance to Choice – Genetics and Justice* (eds A. BUCHANAN, D. BROCK, N. DANIELS and D. WIKLER), pp. 349–373. Cambridge University Press, New York.

SPANU, P. D., ABBOTT, J. C., AMSELEM, J., BURGIS, T. A., SOANES, D. M., STUBER, K., LOREN VAN THEMAAT, E. V., BROWN, J. K. M., BUTCHER, S. A., GURR, S. J., LEBRUN, M.-H., RIDOUT, C. J., SCHULZE-LEFERT, P., TALBOT, N. J., AHMADINEJAD, N., AMETZ, C., BARTON, G. R., BENJDIA, M., BIDZINSKI, P., BINDSCHEDLER, L. V., *et al.* (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**, 1543–1546.

STARK, A., LIN, M. F., KHERADPOUR, P., PEDERSEN, J. S., PARTS, L., CARLSON, J. W., CROSBY, M. A., RASMUSSEN, M. D., ROY, S., DEORAS, A. N., RUBY, J. G., BRENNECKE, J., Harvard Flybase Curators, Berkeley Drosophila Genome Project, HODGES, E., HINRICHS, A. S., CASPI, A., PATEN, B., PARK, S. W., HAN, M. V., *et al.* (2007). Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**, 219–232.

The Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195.

VENNER, S., FESCHOTTE, C. & BIÉMONT, C. (2009). Dynamics of transposable elements: towards a community ecology of the genome. *Trends in Genetics* **25**, 317–323.

WARREN, W. C., HILLIER, L. W., MARSHALL GRAVES, J. A., BIRNEY, E., PONTING, C. P., GRÜTZNER, F., BELOV, K., MILLER, W., CLARKE, L., CHINWALLA, A. T., YANG, S.-P., HEGER, A., LOCKE, D. P., MIETHKE, P., WATERS, P. D., VEYRUNES, F., FULTON, L., FULTON, B., GRAVES, T., WALLIS, J., *et al.* (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183.

X. APPENDIX 1: METHODS USED IN WHOLE-GENOME ANALYSIS

The names, abundances, and classification information of the TEs in each genome were obtained using RepeatMasker
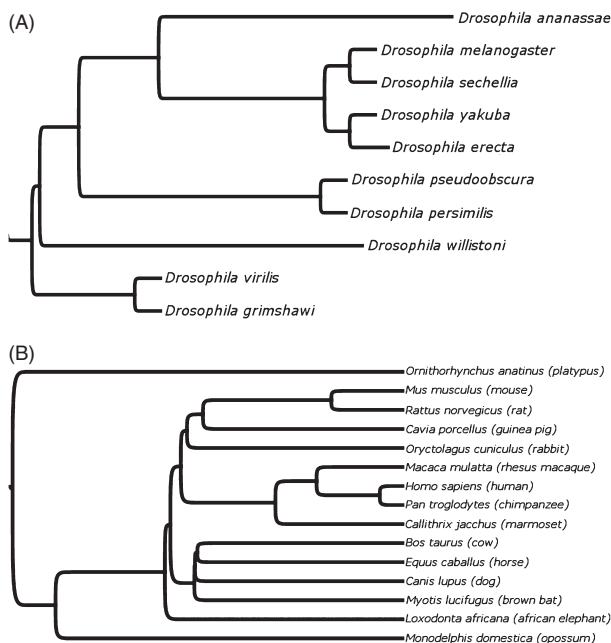

Fig. A1 . Phylogenies of the two groups of genomes used in this analysis. (A) The more closely related *Drosophila* genomes; (B) the more distantly related mammal genomes. The branch lengths in these phylogenies represent evolutionary distance, and were obtained from *Drosophila* 12 Genomes Consortium (2007) and Warren *et al.* (2008).

Table A1. *Drosophila* genomes used in the whole-genome analysis

Species	Accession/location
Drosophila ananassae	AAPP01000001-AAPP01020550
Drosophila erecta	AAPQ01000001-AAPQ01007621
Drosophila grimshawi	AAPT01000001-AAPT01024168
Drosophila melanogaster	AE013599, AE014134, AE014135, AE014296:AE014298, FA000001
Drosophila persimilis	AAIZ01000001-AAIZ01026813
Drosophila pseudoobscura	AADE01000001-AADE01012826
Drosophila sechellia	AAKO01000001-AAKO01021425
Drosophila virilis	CH940647-CH954176
Drosophila willistoni	AAQB01000001-AAQB01020812
Drosophila yakuba	CM000157-CM000162

Table A2. Mammal genomes used in the whole-genome analysis

Species	Accession/location
Bos taurus	CM000177-CM000206
Callithrix jacchus	CM000856-CM000879
Canis lupus familiaris	CM000001-CM000039
Cavia porcellus	AAKN02000001-AAKN02061603
Equus caballus	CM000377-CM000408
Loxodonta africana	AAGU03000001-AAGU03095866
Macaca mulatta	AANU01000001-AANU01301039
Monodelphis domestica	http://www.broadinstitute.org/ftp/pub/ assemblies/mammals/monodelphis/ monDom5/
Mus musculus	CAAA01000001-CAAA01224713
Myotis lucifugus	AAPE02000001-AAPE02072785
Ornithorhynchus anatinus	CM000409-CM000427
Oryctolagus cuniculus	CM000790-CM000811
Rattus norvegicus	AABR05000001-AABR05187024

software with default settings (Smit *et al.*, 2004). These TEs were parsed into one dataset for the coarse and one for the fine-grained entity level. The dataset for the coarse entity level contains the proportions of each TE family in each genome, and the matrix for the fine-grained entity level contains the proportions of each individual TE in each genome. The evolutionary distances between the different host genomes from each phylogeny were used to create a distance matrix for each of the two host genome phylogenies (Fig. A1). The distances in these matrices served as the evolutionary proxy. These distance matrices were reduced into two continuous variables using the isoMSD function in R (http://www.r-project.org/). These variables were used as a proxy for the evolutionary relationship between the host species in our analysis. The ecological proxies in the analysis are also properties of the host genomes, but relate more to the immediate 'environment' of the TEs. These ecological variables are genome size and GC content.

The amounts of variation caused by ecological and evolutionary variables were computed using redundancy analysis

(RDA). This is a multivariate extension of multiple regression, with more than one dependent variable (proportion of host genome made up of TEs) and several independent variables (Legendre & Legendre, 1998) either ecological or evolutionary proxies. Analogous to a multiple regression, one can compute the amount of variation explained by the different groups of explanatory variables (adjusted r^2) (Peres-Neto *et al.*, 2006), ecological or evolutionary proxies, and the unique variation associated with each group of explanatory variables (Legendre & Legendre, 1998). For example, the amount of variation explained by the ecological proxies after removing the phylogenetic signal and *vice versa*.

This approximates the amount of variation associated with evolutionary and ecological processes independently (Fig. 4). To test predictions of the relative importance of ecological *versus* evolutionary processes, a permutation procedure (Legendre & Legendre, 1998) was used to test the significance of each of the proxies. This procedure was carried out at both the coarse and fine-grained entity level for each group of genomes. That is, we considered both the individual TE level and the higher taxonomic level of TE families, in addition to conducting the analysis among closely related host taxa (10 *Drosophila* species; Table A1) and more distantly related host taxa (a selection of mammal genomes; Table A2).