Neutral Theory, Transposable Elements, and Eukaryotic Genome Evolution

Irina R. Arkhipova*,[1]

[1]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA

*Corresponding author: E-mail: iarkhipova@mbl.edu.

Associate editor: Sudhir Kumar

Abstract

Among the multitude of papers published yearly in scientific journals, precious few publications may be worth looking back in half a century to appreciate the significance of the discoveries that would later become common knowledge and get a chance to shape a field or several adjacent fields. Here, Kimura's fundamental concept of neutral mutation-random drift, which was published 50 years ago, is re-examined in light of its pervasive influence on comparative genomics and, more specifically, on the contribution of transposable elements to eukaryotic genome evolution.

Key words: **mobile genetic elements, selfish DNA, molecular parasites.**

Perspective

Scientific papers were much more sparse at the time of publication of the initial Kimura's description of the neutral theory (Kimura 1968), which aimed to explain the seemingly excessive number of mutations in protein-coding sequences that surprised him. In late 1960s, little was known about the organization of DNA within genomes or the different types of mutation. By 1981, inspired by the "selfish DNA" concept, he was already investigating distribution of "selfish" repeated DNA in genomes, deeming it likely to be selectively neutral (Ohta and Kimura 1981). Today, when whole-genome sequencing (WGS) became routine and can be performed not only by genome centers but also by individual investigators, it is particularly fitting to re-evaluate the connections between the neutral theory and transposable elements (TEs), which in their "junk DNA" status can be expected to serve as a quintessential example of neutrality, as their presence is not seemingly associated with the operation of any basic host functions—except of course for the cases when it is.

Neutrality is always the null hypothesis in any evolutionary investigations involving TEs, whether we are considering evolution of TE families within genomes, intragenomic distribution of TE insertions, or the total numbers of TEs occupying a fraction of a given genome. While the original Kimura theory considers molecular evolution at the population level, with regard to TEs neutrality can be manifested at several levels of organization: subgenomic, whereby TEs evolve as distinct molecular species populating the genome; genomic, in which TEs can be viewed as bona fide, fully integrated components of the genome; and supragenomic, whereby TE activity may influence the evolutionary trajectories in populations and species as a whole. Below, each of these levels is considered in view of the neutral theory and its sister, the near neutral theory (Ohta 1992), and departures from neutrality are discussed with the emphasis on the most recent and most significant cases.

TEs as Insertional Mutagens

As causative agents of insertional mutations, TEs in eukaryotic genomes typically operate in full agreement with the neutral paradigm. That is, the majority of TE insertions is selectively neutral or slightly deleterious; some insertions impair a host gene function or induce deleterious chromosomal rearrangements and would be subjected to negative selection; and only a minor fraction, deemed negligible in most theoretical calculations, would be of adaptive significance and subjected to positive selection. From the earliest days of TE discovery, their role as insertional mutagens was at the forefront, as evident from the title of the contemporary book "Eukaryotic transposable elements as mutagenic agents" (Lambert et al. 1988). Indeed, TEs were recognized as causal agents of most spontaneous mutations in Drosophila laboratory strains (Finnegan 1992). Even the first mutation ever described, the wrinkled phenotype of Mendel's peas, was caused by insertion of a Ds-like DNA TE into the *r* locus, inactivating the starch-branching enzyme (Bhattacharyya et al. 1990).

Whether causing gene disruption, activation, or neither, TEs undoubtedly represent a major source of genetic variation in natural populations, some of which may turn out to be of adaptive value in changing environmental conditions. A textbook case of industrial melanism in peppered moths, first recorded in Manchester in 1848 and cited by Kimura as a prime example of adaptive mutation (Kimura 1983), was recently found to represent a 22-kb DNA TE insertion in the first intron of the *cortex* gene, associated with an increase in its transcript abundance (van't Hof et al. 2016). This insertion, responsible for melanization in the *carbonaria* morph, was dated back to the beginning of the 19th century and was likely segregating in the population at low frequency before it swept to near-fixation during the industrial boom, followed by a rapid postindustrialization demise. This scenario is fully

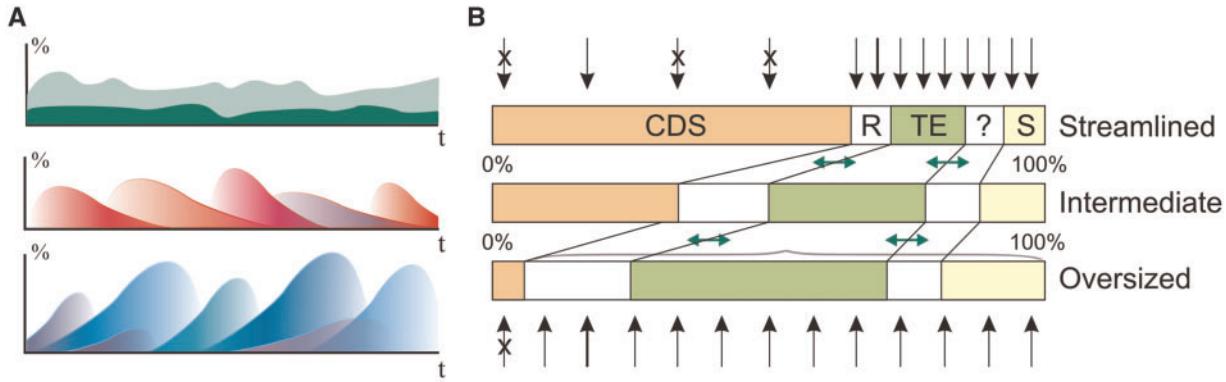

FIG. 1. Transposable element dynamics and insertion patterns in eukaryotic genomes. (*A*) Examples of differing modes of intragenomic TE proliferation and maintenance over time (*t*), influenced by the strength of host response. Green, "benign" TEs adapted to intragenomic "safe havens" with copy numbers at equilibrium. Red, "aggressive" TEs which periodically invade, amplify, get suppressed, and undergo slow decay (e.g. by point mutation). Blue, "dormant" TEs subject to waves of amplification, suppression, and faster decay (e.g. by hypermutation or deletion). If amplification and decay rates are equal, there is no skew. TE content is typically measured in percentage of the genome (*Y* axis). (*B*) Major types of eukaryotic genome organization on a 100% scale, regardless of actual size. CDS, % genome covered by coding sequences (including introns); R, regulatory regions; TE, transposable element sequences; S, high-copy repeats and satellites; ?, sequences of unknown origin. New TE insertions are shown by arrows; crossed arrows, subject to negative selection; thicker arrows, insertions with adaptive potential. Double-headed arrows denote the possibility of TE conversion into regulatory regions or decay beyond recognition; curly bracket, the possibility of noncoding DNA removal from oversized germline genomes (as in ciliates). Streamlined genomes (as in yeast) have few TEs, which are mostly confined to preferred targets. In oversized genomes, TEs occupy most of the genome, and their turnover rate may be either low (as in mammals) or high (as in most plants). All generalizations are for illustrative purposes only; individual genomes and TE types may combine different features.

consistent with Kimura's view of Darwinian selection acting on the vast numbers of pre-existing neutral or nearly neutral variants. In general, a TE insertion into a genic region is highly likely to disrupt it, and genic insertions are mostly confined to introns and UTRs. Insertions into regulatory regions, however, may have a strong adaptive potential and often reshape gene regulatory networks (see below).

Population Genomics of TEs

Studies of TE population dynamics received a huge boost with the advent of WGS analysis methods. Genome-wide TE insertion rates were recently measured at $\sim 2 \times 10^{-9}$ per-site per-generation in eight *Drosophila melanogaster* mutation–accumulation lines, in the absence of natural selection, and were confirmed to exceed deletion rates by 1–2 orders of magnitude (Adrion et al. 2017). Insertion rate constitutes an important parameter in the classical model of TE population dynamics, which assumes that TE frequencies in the population are at equilibrium, that is, the rate of transposition is counterbalanced by the rate of excision and by selection against deleterious insertions or chromosomal rearrangements (Charlesworth et al. 1994). However, the equilibrium state is rarely achieved: a typical TE life cycle involves initial amplification, which, after reaching a peak, subsides when the host defense systems come into effect, and eventually undergoes mutational decay via base substitution and/or deletion (fig. 1*A*). Host control mechanisms may be family-specific or may act on repeats in general, through DNA modification and RNA-mediated silencing. Interestingly, components of the RNA-mediated silencing pathways often display signatures of adaptive evolution, indicating their involvement in the molecular arms race with invading TEs and viruses (Palmer et al. 2018).

In Drosophila populations, 50–80% of analyzed TE insertions segregate at low frequencies and are slightly deleterious or neutral, whereas ~0.1–3% insertions segregating at high frequencies in derived populations were designated as putatively adaptive, with several cases corroborated by fitness and mechanistic analyses (Barron et al. 2014). A study of pooled TE insertions in *D. melanogaster* and *Drosophila simulans* populations, limited to insertions in orthologous euchromatic sites, placed the emphasis on the demography, with habitat expansion triggering TE invasion and rapid evolution (Kofler et al. 2015). In natural accessions of the Mediterranean grass *Brachypodium distachyon*, recent activity in expanding populations and purifying selection against deleterious insertions have shaped the TE landscape, and the absence of massive invasions and bursts may indicate efficient host control (Stritt et al. 2018).

To discern between selection and drift under nonequilibrium conditions, a method taking into account the age of insertions, which relies on the number of terminal branch substitutions accumulated in retroelement sequences since insertion, has been proposed (Blumenstiel et al. 2014). This method may overestimate the age of TEs, since the host-based molecular clock may be augmented by reverse transcriptase-generated errors during replication; nevertheless, it also detects negative selection for most insertions and a few putatively adaptive insertions. In mammals, most LINE and SINE insertions are fixed, 5′-truncated, and selectively neutral; weak deleterious effects were proposed to result mostly from ectopic recombination between longer copies (Song and Boissinot 2007). However, in other vertebrates, such as the *Anolis* lizard, insertions mostly occur as singletons and rarely reach fixation; considering their large population sizes,

negative selection most likely limits their expansion (Ruggiero et al. 2017).

TEs and Eukaryotic Genome Organization

In eukaryotes, genomic TE content may vary wildly, from only a few per cent to over 80%, with several orders-of-magnitude variation observed at all levels of taxonomic hierarchy from protists to plants to animals (the *C*-value paradox, Thomas 1971). Whereas some of the genome size differences may be attributed to ploidy changes, the most drastic changes can be explained by differential accumulation of TE families (Rodriguez and Arkhipova 2018; Wendel et al. 2018). Except for the relatively few eukaryotes with streamlined genomes, protein-coding genes typically occupy a relatively small fraction of the genome, and the rest of noncoding DNA should be largely indifferent to additional insertions (fig. 1*B*). In oversized genomes, the majority of TE insertions would not damage any genes, and the contribution of ectopic recombination to the genetic load would be expected to dominate, unless it is suppressed. In more compact genomes, TE compartmentalization is likely to develop through a combination of TE insertion specificity and selection purging TEs from euchromatic regions. It should be emphasized that, for the most part, large repetitive genomic regions are not included in next-generation WGS assemblies, and it is only from third-generation long-read assemblies that we will eventually obtain a comprehensive picture of TE organization not restricted to euchromatin. So far, such analysis has been performed only in maize, revealing that most TE copies are intact and showing marked expansion of distinct families (Jiao et al. 2017). Interestingly, the percentage of maize genome assembly occupied by TEs was reduced to 64% in comparison with the previous estimate of 75% (Baucom et al. 2009).

Since multiple factors can contribute to TE distribution, nearly any observed pattern could fit one of the widely accepted theoretical assumptions, while not necessarily agreeing with the others. Let us compare TE distribution in two best-studied model species, *D. melanogaster* and *Caenorhabditis elegans*, and limit consideration to natural populations, rather than laboratory strains not subject to selective pressures encountered in nature. In agreement with previous theoretical and experimental observations, TEs in reference panels derived from nearly 200 natural populations of *D. melanogaster* were concentrated in the regions of low recombination (Cridland et al. 2013). This pattern agrees with the studies of the Y chromosomes in humans and other mammals showing massive TE accumulation in nonrecombining regions (Skaletsky et al. 2003). Surprisingly, in the next-best-studied model organism, *C. elegans*, the pattern of TE distribution in 208 wild strains was close to opposite: TEs tend to be excluded from the core genomic regions with low recombination rates, and to concentrate on the genome periphery, in gene-poor regions highly prone to recombination, apparently placing more weight on the insertion-mutagenic properties of TEs (Laricchia et al. 2017). In plants with large genomes, such as maize and conifers, the rates of recombination-mediated TE removal are much lower than in plants with smaller genomes, but gene conversion rates are increased, possibly as a result of heterochromatin-mediated bias in resolution of recombination (Cossu et al. 2017).

Whereas the patterns of TE evolution within genomes are largely dictated by the host, the dual nature of TEs should be always kept in mind. On the one hand, they represent bona fide genomic components and complement other intrinsic mutagenic forces, for example, polymerase errors or the efficiency of DNA repair. On the other hand, they may form only transient associations with the hosts, due to their nature as molecular invaders, and can speed up their evolution during replication cycles. As the participants in the host–parasite evolutionary arms races, TEs are subject to suppression of their activity by the host, can evolve to counteract and escape this pressure, and eventually face the newly evolved host defense systems. A common view is that TEs often evolve strategies for minimizing damage to the host, so that they could survive within that host and still spread on a limited basis. Ways to reduce damage to the host may include developing insertional specificity (Sultana et al. 2017) or self-limitation mechanisms (Charlesworth and Langley 1986; Tucker et al. 2015). However, these strategies may backfire on TEs: a target may disappear from the genome, or a self-limiting TE may have a higher chance of losing out to its competitors, especially in asexuals (Arkhipova and Meselson 2005). In "oversized" genomes (fig. 1*B*), "benign" TEs are less likely to evolve than in "streamlined" genomes, such as the 400-My old yeasts, which apparently co-existed with Ty elements inserting near multicopy gene promoters or telomeric heterochromatin during their co-evolution.

Molecular Parasites, Commensals, and Symbionts

This is the wide spectrum that emerges when we accept dual roles of TEs as genome ingredients subject to the rules imposed by the host biology, and at the same time as independently proliferating units which can come and go, be countered by the host defenses, and reshape the host genomes in the process (Kidwell and Lisch 2001). The analogy between TEs as members of the genome "ecosystem" has periodically been invoked, with application of the principles of community ecology to the different components inhabiting eukaryotic genomes (Brookfield 2005; Venner et al. 2009). More recently, an attempt was made to assess the fit of the unified neutral theory of biodiversity (Hubbell 2001), which was in turn originally inspired by Kimura's neutralist vision of population genetics, to the distribution of "molecular species" abundance across eukaryotic chromosomes, whereby "molecular species" are represented by diverse types of TEs, satellite repeats, multicopy RNAs, etc. (Serra et al. 2013). While the distribution of very few molecular species along the chromosomes agreed with the random expectation, the overall molecular species abundance and diversity were surprisingly similar when various molecular species were allowed to compensate for each other by shifts in the ranking order of abundances. The neutral model was sufficient to explain the overall abundance and diversity of genetic elements in each

chromosome of the 31 eukaryotic genomes analyzed, from protists to humans. While these findings cannot be regarded as evidence in favor of a neutral process behind the observed distribution patterns, the contribution of neutral drift cannot be underestimated either. Across large evolutionary distances, the long-term intragenomic patterns of TE distribution are likely to be guided by genetic drift, as was demonstrated in a recent comparison of 42 sequenced genomes in the phylum Nematoda spanning 500 My of evolution (Szitenberg et al. 2016).

Regulatory Novelties

The overwhelming dominance of neutral DNA notwithstanding, we are always going to be fascinated by the small proportion of TE-mediated adaptive changes that may serve as a basis for Darwinian selection. Eukaryotic evolution has been progressing for over a billion years in the long term, and while each group followed its own path, some features may have been selected repeatedly in a convergent fashion. While numerous examples of TE-mediated novelties have been described in recent years (Chuong et al. 2017), it is worth reiterating that TEs represent ready-to-use building blocks which can be co-opted (exapted, domesticated) by the host, either in their protein-coding capacity (entire ORFs, ORF assemblies, or separate functional domains) or as noncoding regulatory elements (enhancers, regulatory RNAs, epigenetic modification carriers, etc.). Their functional significance may not be instantly confirmed by experimental studies, as in the well-known example of ultra-conserved elements in vertebrate genomes, many of which have originated from TEs such as SINE or MER, and display tissue-specific enhancer activity (Bejerano et al. 2004, 2006; Nishihara et al. 2006; Pennacchio et al. 2006; Notwell et al. 2015). Initially, targeted deletion of four ultra-conserved enhancers failed to yield detrimental effects, leading the authors to conclude that they play no functional role (Ahituv et al. 2007). Ten years later, a more thorough inspection revealed that such removal does cause profound developmental defects, which may not be critical in the laboratory environment, but would be essential for normal development and survival in natural habitats (Dickel et al. 2018). Thus, it is reassuring to know that the most widely used approach to define functionality, that is, targeted disruption, can eventually validate findings made by the comparative approach based on evolutionary conservation.

Recruitment of TE Components

As previously argued, most of the complex traits can be explained by increased efficiency of genetic drift in times of population bottlenecks (Lynch and Conery 2003). Major novelties may be easily overlooked by selection in the first place, but several well-known innovations, including but not limited to telomerase protecting chromosome ends in eukaryotes, RAG subunits of V(D)J recombinase responsible for adaptive immunity in vertebrates, or syncytins repeatedly captured from retroviral envelope genes for placentation in mammals, owe their origin to TEs (Nakamura and Cech 1998; Kapitonov and Jurka 2005; Lavialle et al. 2013). More recent examples include the pan-eukaryotic gamete fusion protein HAP2 taking its origin from viral membrane fusion proteins (Fédry et al. 2017), neuronal gene Arc derived from retrotransposon Gag protein to form capsid-like structures for trafficking RNA across synapses (Ashley et al. 2018; Pastuzyn et al. 2018), or generation of L1 retrotransposon-induced somatic mosaicism in the mouse brain in response to experience (Bedrosian et al. 2018). Recruitment of transposases by different ciliates to eliminate most of the DNA from the germline genome to give rise to the expressed somatic macronuclei is complemented by the genome defense system employing small RNAs to distinguish germline from soma (Baudry et al. 2009; Nowacki et al. 2009). Finally, evolution of the core spliceosome component Prp8 from a catalytically disabled reverse transcriptase (Galej et al. 2013) may be the ultimate example of "constructive neutral evolution" (Stoltzfus 1999), whereby an incredibly complex biological machine with excess capacity has evolved and needs to be maintained for precise removal of noncoding intronic DNA, sometimes represented by as few as three introns per genome (Morrison et al. 2007).

How Much of the Genome Is Important?

To understand what proportion of the genome is vulnerable to mutations, it is important to have an estimate of the fraction of the genome that is involved in its functionality. In other words, how much of a given genome is functional and how much is junk—the question we all remember from the ENCODE project estimating "functional" human DNA at 80%, which stimulated lively debates five years ago. On the basis of mutational load consideration, an upper limit of 25% on "functional" DNA for the human genome has been proposed (Graur 2017). Such numbers of course depend on the genome size and complexity. In a relatively simple genome of *Saccharomyces cerevisiae*, synthetic biology teams intend to contribute designer chromosomes to a fully synthetic Sc2.0 genome presumably free of all junk, after all "unnecessary" sequences are removed by design, shrinking the genome by 8% (Richardson et al. 2017). By analogy to ordered gene deletion libraries, ordered intergenic deletion libraries may be entertained in the future to interrogate segments of DNA between each pair of genes.

Fortunately, less expensive mutagenesis approaches are also being developed, aiming to define the importance of each genomic region (or lack thereof) on a genome-wide scale under different conditions (e.g. by varying temperature, concentration of added compounds, or other stresses). In this case, TEs themselves serve as the most appropriate tools. A saturated mutagenesis approach, initially developed in bacteria (Tn-seq; van Opijnen et al. 2009) and more recently applied to yeast (Guo et al. 2013; Michel et al. 2017), involves generating insertions at high density and sequencing the flanking regions *en masse*. Heterologous TEs (insect mariner in *Escherichia coli*, insect Hermes in *Schizosaccharomyces pombe*, and plant Ac/Ds in *S. cerevisiae*) are used to achieve high-density uniform distribution of inserts and to avoid

pre-existing targeting effects or host-specific suppression. If a locus is important for growth (and growth conditions may vary), the density of insertions falls dramatically, revealing gaps in insertion coverage that may also help to dissect the critical functional domains; nonessential genes may display reduced coverage, and analysis of deletion mutants would reveal interactions between loci. In higher organisms, such methods might be adapted to reveal haplo-insufficient or dominant-negative loci. Whether by synthetic or analytic means, we will eventually learn the adaptive value of most intergenic regions.

Conclusion

TEs are virtually ubiquitous in eukaryotes, having apparently been lost only from the greatly reduced genomes of apicomplexan parasites (DeBarry and Kissinger 2011). Their full impact, however, is still greatly underestimated, with studies of repetitive regions likely to be propelled by future technology developments. Without TEs, eukaryotic genomes might look more orderly, but evolution would be much less eventful if it were limited to traditionally considered changes such as those resulting from errors in the basic mechanisms of DNA replication or repair, or duplication and diversification of existing genes. TEs more than any other factors appear suited for bringing about unexpected shake-ups of eukaryotic genomes. Such evolutionary perturbations are always a thrill to disentangle, even though not every species can preserve enough molecular evidence to serve as proof. Nevertheless, in search for departures from the ordinary, the presumption of neutrality will remain the default starting point: everything is neutral until proven otherwise. Thus, the neutral theory will always continue to bring the necessary sense of balance into our investigations of multiple forces shaping eukaryotic genomes, and for that we will always be thankful to Motoo Kimura.

Acknowledgment

Work in the laboratory is supported by R01GM111917 (US National Institutes of Health).

References

Adrion JR, Song MJ, Schrider DR, Hahn MW, Schaack S. 2017. Genome-wide estimates of transposable element insertion and deletion rates in *Drosophila melanogaster*. *Genome Biol Evol.* 9(5):1329–1340.

Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* 5(9):e234.

Arkhipova I, Meselson M. 2005. Deleterious transposable elements and the extinction of asexuals. *Bioessays* 27(1):76–85.

Ashley J, Cordy B, Lucia D, Fradkin LG, Budnik V, Thomson T. 2018. Retrovirus-like Gag protein Arc1 binds RNA and traffics across synaptic boutons. *Cell* 172(1–2):262–274.

Barron MG, Fiston-Lavier AS, Petrov DA, Gonzalez J. 2014. Population genomics of transposable elements in Drosophila. *Annu Rev Genet.* 48:561–581.

Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon J-M, Westerman RP, SanMiguel PJ, Bennetzen JL. 2009. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLOS Genet.* 5(11):e1000732.

Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Betermier M. 2009. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.* 23(21):2478–2483.

Bedrosian TA, Quayle C, Novaresi N, Gage FH. 2018. Early life experience drives structural variation of neural genomes in mice. *Science* 359(6382):1395–1399.

Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441(7089):87–90.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. 2004. Ultraconserved elements in the human genome. *Science* 304(5675):1321–1325.

Bhattacharyya MK, Smith AM, Ellis THN, Hedley C, Martin C. 1990. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* 60(1):115–122.

Blumenstiel JP, Chen X, He M, Bergman CM. 2014. An age-of-allele test of neutrality for transposable element insertions. *Genetics* 196(2):523–538.

Brookfield JF. 2005. The ecology of the genome – mobile DNA elements and their hosts. *Nat Rev Genet.* 6(2):128–136.

Charlesworth B, Langley CH. 1986. The evolution of self-regulated transposition of transposable elements. *Genetics* 112(2):359–383.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371(6494):215–220.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet.* 18(2):71–86.

Cossu RM, Casola C, Giacomello S, Vidalis A, Scofield DG, Zuccolo A. 2017. LTR retrotransposons show low levels of unequal recombination and high rates of intraelement gene conversion in large plant genomes. *Genome Biol Evol.* 9(12):3449–3462.

Cridland JM, Macdonald SJ, Long AD, Thornton KR. 2013. Abundance and distribution of transposable elements in two Drosophila QTL mapping resources. *Mol Biol Evol.* 30(10):2311–2327.

DeBarry JD, Kissinger JC. 2011. Jumbled genomes: missing Apicomplexan synteny. *Mol Biol Evol.* 28(10):2855–2871.

Dickel DE, Ypsilanti AR, Pla R, Zhu Y, Barozzi I, Mannion BJ, Khin YS, Fukuda-Yuzawa Y, Plajzer-Frick I, Pickle CS, et al. 2018. Ultraconserved enhancers are required for normal development. *Cell* 172(3):491–499.

Fédry J, Liu Y, Péhau-Arnaudet G, Pei J, Li W, Tortorici MA, Traincard F, Meola A, Bricogne G, Grishin NV, et al. 2017. The ancient gamete fusogen HAP2 is a eukaryotic class II fusion protein. *Cell* 168(5):904–915.

Finnegan DJ. 1992. Transposable elements. In: Lindsley DL, Zimm G, editors. The genome of Drosophila melanogaster. New York: Academic Press. p. 1096–1107.

Galej WP, Oubridge C, Newman AJ, Nagai K. 2013. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* 493(7434):638–643.

Graur D. 2017. An upper limit on the functional fraction of the human genome. *Genome Biol Evol.* 9(7):1880–1885.

Guo Y, Park JM, Cui B, Humes E, Gangadharan S, Hung S, FitzGerald PC, Hoe KL, Grewal SI, Craig NL. 2013. Integration profiling of gene function with dense maps of transposon integration. *Genetics* 195(2):599–609.

Hubbell SP. 2001. The unified neutral theory of biodiversity and biogeography. Princeton (NJ): Princeton University Press.

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C-S, et al. 2017. Improved maize reference genome with single-molecule technologies. *Nature* 546(7659):524.

Kapitonov VV, Jurka J. 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* 3(6):e181.

Kidwell MG, Lisch DR. 2001. Transposable elements, parasitic DNA, and genome evolution. *Evolution* 55(1):1–24.

Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* 217(5129):624–626.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge (United Kingdom): Cambridge University Press.
Kofler R, Nolte V, Schlötterer C. 2015. Tempo and mode of transposable element activity in Drosophila. *PLOS Genet.* 11(7):e1005406.
Lambert ME, McDonald JF, Weinstein IB. 1988. Eukaryotic transposable elements as mutagenic agents. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
Laricchia KM, Zdraljevic S, Cook DE, Andersen EC. 2017. Natural variation in the distribution and abundance of transposable elements across the *Caenorhabditis elegans* species. *Mol Biol Evol.* 34(9):2187–2202.
Lavialle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, Heidmann T. 2013. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond B Biol Sci.* 368(1626):20120507.
Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302(5649):1401–1404.
Michel AH, Hatakeyama R, Kimmig P, Arter M, Peter M, Matos J, De Virgilio C, Kornmann B. 2017. Functional mapping of yeast genomes by saturated transposition. *Elife* 6:e23570.
Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, et al. 2007. Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. *Science* 317(5846):1921–1926.
Nakamura TM, Cech TR. 1998. Reversing time: origin of telomerases. *Cell* 92(5):587–590.
Nishihara H, Smit AF, Okada N. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* 16(7):864–874.
Notwell JH, Chung T, Heavner W, Bejerano G. 2015. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat Commun.* 6:6644.
Nowacki M, Higgins BP, Maquilan GM, Swart EC, Doak TG, Landweber LF. 2009. A functional role for transposases in a large eukaryotic genome. *Science* 324(5929):935–938.
Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst.* 23(1):263–286.
Ohta T, Kimura M. 1981. Some calculations on the amount of selfish DNA. *Proc Natl Acad Sci U S A.* 78(2):1129–1132.
Palmer WH, Hadfield JD, Obbard DJ. 2018. RNA-interference pathways display high rates of adaptive protein evolution in multiple invertebrates. *Genetics* 208(4):1585–1599.
Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, Yoder N, Belnap DM, Erlendsson S, Morado DR, et al. 2018. The neuronal gene Arc encodes a repurposed retrotransposon Gag protein that mediates intercellular RNA transfer. *Cell* 172(1–2):275–288.
Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444(7118):499–502.
Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CL, Chandrasegaran S, Cai Y, et al. 2017. Design of a synthetic yeast genome. *Science* 355(6329):1040–1044.
Rodriguez F, Arkhipova IR. 2018. Transposable elements and polyploid evolution in animals. *Curr Opin Genet Dev.* 49C:115–123.
Ruggiero RP, Bourgeois Y, Boissinot S. 2017. LINE insertion polymorphisms are abundant but at low frequencies across populations of Anolis carolinensis. *Front Genet.* 8:44.
Serra F, Becher V, Dopazo H. 2013. Neutral theory predicts the relative abundance and diversity of genetic elements in a broad array of eukaryotic genomes. *PLoS ONE.* 8(6):e63915.
Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423(6942):825–837.
Song M, Boissinot S. 2007. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene* 390(1–2):206–213.
Stoltzfus A. 1999. On the possibility of constructive neutral evolution. *J Mol Evol.* 49(2):169–181.
Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. 2018. Recent activity in expanding populations and purifying selection have shaped transposable element landscapes across natural accessions of the Mediterranean grass *Brachypodium distachyon*. *Genome Biol Evol.* 10(1):304–318.
Sultana T, Zamborlini A, Cristofari G, Lesage P. 2017. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat Rev Genet.* 18(5):292–308.
Szitenberg A, Cha S, Opperman CH, Bird DM, Blaxter ML, Lunt DH. 2016. Genetic drift, not life history or RNAi, determine long-term evolution of transposable elements. *Genome Biol Evol.* 8(9): 2964–2978.
Thomas CA Jr. 1971. The genetic organization of chromosomes. *Annu Rev Genet.* 5:237–256.
Tucker JM, Larango ME, Wachsmuth LP, Kannan N, Garfinkel DJ. 2015. The Ty1 retrotransposon restriction factor p22 targets Gag. *PLoS Genet.* 11(10):e1005571.
van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534(7605):102–105.
van Opijnen T, Bodi KL, Camilli A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods.* 6(10):767.
Venner S, Feschotte C, Biemont C. 2009. Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet.* 25(7):317–323.
Wendel JF, Lisch D, Hu G, Mason AS. 2018. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr Opin Genet Dev.* 49:1–7.