Multilevel Selection Theory and the Evolutionary Functions of Transposable Elements

Tyler D.P. Brunet and W. Ford Doolittle*

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

*Corresponding author: E-mail: ford@dal.ca.

Abstract

One of several issues at play in the renewed debate over "junk DNA" is the organizational level at which genomic features might be seen as selected, and thus to exhibit function, as etiologically defined. The intuition frequently expressed by molecular geneticists that junk DNA is functional because it serves to "speed evolution" or as an "evolutionary repository" could be recast as a claim about selection between species (or clades) rather than within them, but this is not often done. Here, we review general arguments for the importance of selection at levels above that of organisms in evolution, and develop them further for a common genomic feature: the carriage of transposable elements (TEs). In many species, not least our own, TEs comprise a large fraction of all nuclear DNA, and whether they individually or collectively contribute to fitness—or are instead junk— is a subject of ongoing contestation. Even if TEs generally owe their origin to selfish selection at the lowest level (that of genomes), their prevalence in extant organisms and the prevalence of extant organisms bearing them must also respond to selection within species (on organismal fitness) and between species (on rates of speciation and extinction). At an even higher level, the persistence of clades may be affected (positively or negatively) by TE carriage. If indeed TEs speed evolution, it is at these higher levels of selection that such a function might best be attributed to them as a class.

Key words: evolution, multi-level selection, transposable elements, genomes, ENCODE.

Introduction

A central task of evolutionary biology is to explain why organisms exhibit the traits they do, especially complex ones. In general, neo-Darwinists couch such explanation in terms of selection on genes or individuals (organisms) within populations or species (microevolution). A complex trait is taken to be the product of successive fixations of individually advantageous incremental changes occurring sequentially within species in a lineage, in aid of some organism-level function. But the successive fixations that led to a trait might just as plausibly have been macroevolutionary in their selectively significant consequences; conferring fitness in reproductive competition at the level of species, not that of individual organisms or their genes. Here, speciation and extinction assume the parts played by individual reproduction and death within species. This alternative higher-level context for selective explanations allows us to make some sense of otherwise seemingly teleological claims about transposable elements (TEs) as "repositories" for or "facilitators" of future evolution, or even as "nature's tools for genetic engineering" (Kleckner 1981). Multilevel selection (MLS) theory allows us to recast teleology as etiology

Function and Selection at Multiple Levels

The ENCODE debate (Doolittle 2013; Eddy 2013; Graur et al. 2013) makes it clear that biologists differ among themselves in their understanding of the meaning of "function," including as it applies to the activities of the myriad mostly moribund TEs making up the majority of our own genomes (de Koning et al. 2011). Genomicists often seem to entertain an operational ("causal role") view, in which the functions of an element or process are what its current effects in the organism are, what it in fact does. Many evolutionary biologists, however, would restrict function to those effects that have been favored by natural selection—why it is there—an "etiological" or "selected effect" definition. We also endorse this conception, but hold that selected effects can arise at any level at which selection is deemed to occur (Doolittle et al. 2014).

Any replicating entities that are born and die and show heritable variation in fitness will experience natural selection over time, fitter types tending to increase in frequency. MLS theory attempts to identify levels of the biological hierarchy (genes, genomes, organisms, groups/populations, species, and possibly higher taxa) at which such entities exist (Okasha 2006; Doolittle 2014). At supra-organismal levels, species are especially good candidate entities, exhibiting heritable variation in fitness affecting their own forms of birth and death—events of speciation and extinction. The differences between rates of speciation and extinction—"diversification" rates—will inevitably vary between species within a clade, depending on both environmental and organismal factors. These latter are thus fitness determinants at the species level. Fitter species will, all else being equal, give rise to more descendant species, and organismal (or genomic) features that contribute thus serve "species functions."

As Borrello (2005) recounts the history of MLS theory, pre-New Synthesis Darwinians lacked a well-developed hierarchical view and allowed a broad reading of selection and function, encompassing individuals and species indiscriminately. The narrowing focus of the New Synthesis on individuals—and beginning in the mid-1960s on genes—encouraged a default view that selection between populations or species ("species-selection") is at best a weak force. It was in particularly bad repute because of the exaggerated group-selectionist claims of Wynne-Edwards (1986), according to which individuals sacrifice their own reproductive interests in order to control population growth for "the good of the species."

Although the persuasive rhetorical force of opponents of this view (Maynard Smith 1964; Trivers 1971) targeted altruistic "behaviors" within populations, other species-level phenomena came under suspicion. Explanatory reductionism was thus always to be preferred: when selection pressures at higher and lower levels were aligned (either both positive or both negative), processes at the lower were to be given causal priority. Only in very rare cases in which species-level selection appears to (improbably) triumph over oppositely directed organismal selection was it to be invoked. Moreover, some argued that only emergent or population/species-level traits—features that are not ascribable to individual organisms—could be used in models of higher-level selection, and that such traits are few (Vrba 1984; Vrba and Gould 1986).

There are now, though, increasingly many acknowledged cases of species-level traits that cannot but affect species diversification rates, and species selection is no longer anathema in organismal evolutionary biology. Jablonski (2008) lists a dozen broad categories of such traits, geographic range and patchiness, population size, structure, and genetic variability among them. He calls the evolutionary process that these participate in "strict-sense species selection." And he notes an increasing willingness to accept a looser ("broad-sense") model, in which some "aggregate" traits (nonemergent characters fixed within populations because they are advantageous to individuals) also serve as determinants of species selection—body size, ecological specialization, and reproductive mode being examples. Of possible roles for aggregate traits fixed at a still lower level, including selfish genomic elements, there has been little discussion (Doolittle 1987).

"Upward causation," in which an organism-level aggregate trait has species- or higher-level effects, Jablonski (2008) also calls "effect macroevolution." Problematic claims about a greater evolutionary potential of eukaryotes vis-à-vis prokaryotes being attributable to differences in cell structural complexity invoke positive effect macroevolution at the highest possible (domain) level (Booth and Doolittle 2015). There will also be "downward causation," where species-level dynamics influence the frequency of organismal traits either negatively or positively. We do not expect to find fixed within many extant species traits that, though individually beneficial, drastically reduce rates of speciation or increase rates of extinction. Traits like this might indeed persist and display increases in frequency in the short term, but without subsequent speciation and adaptive radiation the only place to go is down into extinction. The relative infrequency of asexuality is often explained in this way, in spite of the immediate 2-fold advantage it offers to individuals (West et al. 1999).

It is important to bear in mind that selection processes at different levels, even when similarly directed, are causally distinct (fig. 1). Though levels may by hard to define and delimit, it is not the case that individual-level selection grades seamlessly into species-level selection. There is a disconnect. As Rabosky and McCune (2010) put it...

> Selection at the individual level contributes to trait variation between species by transforming intraspecific variation into species differences that might result in species selection. However, the mechanism by which a trait becomes fixed within a species, whether through selection or drift, need not be the same as the mechanism by which the trait influences diversification (p.70).

A good example, according to these authors, is floral symmetry. This is selected for both "within" species, because it reduces waste by pollinators, and "between" species, because it increases the specificity of those pollinators. We also note, further to Rabosky and McCune (2010), that the intensities of selection at different levels are incommensurable—have different units—so it is profoundly problematic to say which is stronger. The in-principle answerable question may be, for any genome, what fraction of its nucleotides owe their presence and nature (A, T, C, or G) to selection or drift at one level versus another. We submit that this remains an open—and too infrequently asked—question.

The effects of drift will also surely be seen at every level of the biological hierarchy, and will have interlevel consequences (Lynch 2007). Organismal or genomic features fixed by drift in small populations will potentially impact speciation and
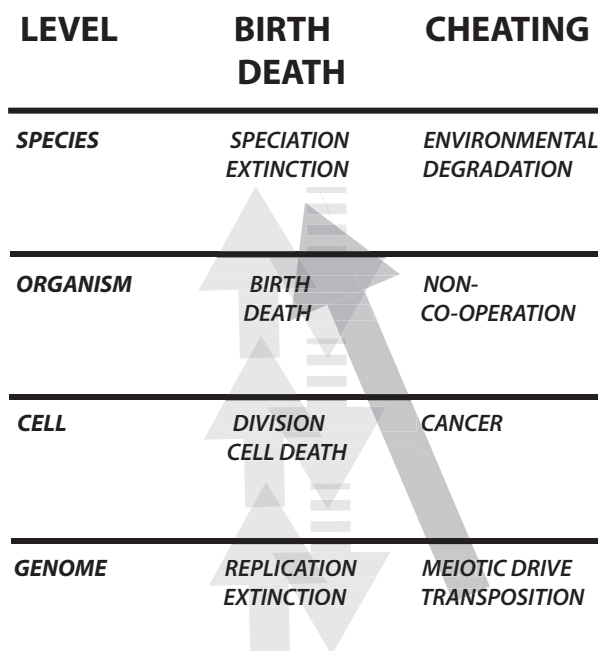
LEVEL	BIRTH DEATH	CHEATING
SPECIES	*SPECIATION EXTINCTION*	*ENVIRONMENTAL DEGRADATION*
ORGANISM	*BIRTH DEATH*	*NON-CO-OPERATION*
CELL	*DIVISION CELL DEATH*	*CANCER*
GENOME	*REPLICATION EXTINCTION*	*MEIOTIC DRIVE TRANSPOSITION*

Fig. 1.— Levels of selection. At each level there are birth/death processes affecting entities at that level, and "fitness" determinants differentially affecting these processes. Paler solid and dashed arrows indicate upward or downward causation, by which differential replication of entities at one level affects differential replication of entities at one or more other levels. Although levels and directions of causation are in principle distinct they will often be hard to distinguish. At the three lower levels, "cheating" comprises all activities that compromise fitness of elements at a higher level or levels. Indeed, that cheaters—which we here define as "elements whose competition at one level jeopardizes survival at the next higher level"—exist is proof of the semi-independence of levels. In this manuscript, we have not emphasized the real and relevant distinction between cellular and organismal levels, but cancer provides a particularly good example of upward and downward causation. Our claim is that TEs, which are cheaters that compromise fitness at cellular and organismal levels, nevertheless might promote speciation or forestall species extinction (darker arrow). Other upward and downward influences crossing several levels are not precluded. The reality of levels above species is contentious (Doolittle 2014).

extinction, and species selection itself will often be difficult to distinguish from what might be called species drift. Though the number of species in a clade will surely be even fewer than the number of individuals in an effectively small population, it remains the case that traits accelerating speciation or delaying extinction will inevitably be more strongly represented in the world of species at large. (Similarly, although mildly favorable organismal traits may by chance not be fixed in individual small populations, such traits will be differentially represented in a sufficiently large assemblage of such small populations.)

The history of the thinking of molecular biologists (and genomicists who are their intellectual heirs) about the function of TEs should have paralleled that in whole organism biology, one would think. That is, an initial enthusiasm to invoke a

vaguely delimited "good of the species" (historical stage 1), chastened by the debates over selfish genes and selfish DNA (stage 2), should by now have given rise to a more formally articulated MLS theory (stage 3). But evolutionary thinking at the molecular level generally trails behind that at organismal levels, perhaps because few theorists and philosophers focus down to it. Some molecular biologists and genomicists are still stuck in stage 1, some are at stage 2, and only a few have entered stage 3. Thus TE functionality is often discussed teleologically, as if TEs' functions were anticipatory in nature. Passages quoted later in this essay will illustrate this.

Some Ancient History

From their initial discovery as "interspersed or middle repetitive DNAs" in mammalian genomes and "controlling elements" in maize, eukaryotic TEs have been assigned both regulatory and evolutionary functions, not just as individuals but as a class. In 1969, Britten and Davidson (1969) proposed that the vast collection of interspersed repetitive DNAs just then being uncovered by mammalian DNA reassociation studies (and now known, in our own species, to be mostly TEs of the *Alu* and Line1 families) were regulatory in nature.

Specifically, they were to comprise redundant *batteries* of *regulator, integrator*, and *receptor* genes whose combinatorial interactions served to integrate the transcription of unique-sequence *producer* (structural) genes in the complex and flexible modular ways needed during the development of such sophisticated organisms. Moreover, there was an important evolutionary spinoff. Because their repetitiveness facilitated rearrangements and quick changes in copy number, it was thought that interspersed repetitive DNAs could facilitate evolutionary innovation.

> It is known that new repeated sequence families have originated periodically in the course of evolution. The new families of repeated sequences might well be utilized to form integrator and receptor gene sets specifying novel batteries of producer genes. Thus saltatory replications can be considered the source of new regulatory DNA. (Britten and Davidson 1969, p. 356)

The mobility of controlling elements in maize was also seen as vital to their function in development, and parallels to bacterial insertion sequences (IS) (both comprising interspersed repeats in their respective genomes) were noted early, by McClintock herself (McClintock 1961) and by bacterial geneticists such as Nevers and Saedler (1977). The latter wrote that

> The prevalence of IS sequences and controlling elements in organisms as diverse as *E. coli*, *Zea mays* and *Drosophila* suggest that they may be of general biological significance . . . Whether they exert control functions at these positions or are simply kept in reserve as

prefabricated units for the evolution of new control circuits remains unclear. (p. 114)

This second, then and still quite widely held, notion—that complex genomic features would be taken on or "kept in reserve" by organisms simply because they might prove useful in the future—struck many who had read Dawkin's 1976 classic, *The Selfish Gene*, as fanciful and wrongheaded. Benefit to the future could only be for the good of the species, and there is nothing in it for individuals or their genes. Moreover, the insistence that TEs must at least be "good for something" betrayed the panadaptationism then and still typical in molecular genetics.

Indeed, Brenner (1998) linked these common but misguided intellectual predispositions in expressing his scorn for both . . .

There is a strong and widely held belief that all organisms are perfect and that everything within them is there for a function. Believers ascribe to the Darwinian natural selection process a fastidious prescience that it cannot possibly have and some go so far as to think that patently useless features of existing organisms are there as an investment for the future. (p. R669)

The "Selfish DNA" Era

The 1980 selfish DNA papers (Doolittle and Sapienza 1980; Orgel and Crick 1980) argued that such future-directed evolution was in any case logically unnecessary to explain the existence of TEs in either prokaryotic or eukaryotic genomes. The propensity of TEs to increase their copy numbers makes them replicating entities with heritable variation in fitness—and thus subject to natural selection at their own intragenomic level, below the levels of cells and organisms in the MLS hierarchy (fig. 1). TEs will inevitably create an upward pressure on genome size, though this might be opposed by selection at the organismal level (for metabolic and or regulatory efficiency).

Selfish DNAs differ from "selfish genes" in that the latter promote their own replication only through expression in organismal phenotype: selfish gene G (or selfish allele G_a) has a selective advantage only in so far as it promotes—through its expression in phenotype—the differential reproductive success of organisms that bear it. All genes under positive or stabilizing selection are selfish in this sense. Selfish DNAs on the other hand are successful because their inherent properties allow for differential replication and spread within lineages or populations of genomes, independently of any effect—positive or negative—on organismal or species phenotype. Segregation distorters, alleles that bias chromosome segregation in their own favor, are one sort, though often with collateral individual or species-level effects. TEs, which enhance their own chances of spreading within a population by increasing the number of chromosomal sites they occupy, are

another. Replication is essential to TE selfishness and before it was understood that elements that move by "cut and paste" (as opposed to "copy and paste") mechanisms can by various tricks increase in numbers (e.g., Spradling et al. 2011), they would not have been considered selfish.

Since its inception, debate over the possibility of DNA selfishness has been polarized. As summarized by Kleckner in 1981 (Kleckner 1981).

Two types of evolutionary explanations for the existence of [prokaryotic TEs] are being debated. Transposons may have evolved as nature's tools for genetic engineering: their ability to rearrange other DNA sequences would thus be a directly selectable phenotype which could lead to the increased survival of individual organisms, individual replicons or populations of organisms or replicons harboring such elements. The discovery of transposition as a fundamentally replicative process has also led to the radically different suggestion that the existence of transposons is attributable solely to their ability to overreplicate the host; their ability to replicate and move would permit them to escape the normal mechanisms that would eliminate DNA sequences for which no direct phenotypic selection exists. (p. 343)

Early and continuing arguments against the claim for TE "selfishness" are of three sorts. First, many investigators (perhaps stuck in our historical stage 1) simply continue to insist that TEs are kept in reserve to "facilitate or speed evolutionary innovation," without addressing how or at what level this could be a selected effect—how evolution as normally understood might possibly be able to look ahead into the future. If anything, the last decade has witnessed resurgence in this loosely teleological way of thinking, although its problematic nature is well recognized by more sophisticated students of evolvability (Kirschner 2013).

A second objection, to the anthropomorphism of the term "selfish," is often coupled to the default panadaptationism derided by Brenner (1998). Natural selection, God-like in its oversight, should prevent any such genomic subversion. Elements of such a viewpoint are also still very much alive, as the debate over ENCODE (and the ready acceptance of ENCODE's claims by intelligent design creationists) reveals (Tomkins 2012; Doolittle 2013; Eddy 2013).

A third and more empirically grounded objection was championed by Cavalier-Smith (1978), who had for some time held that an organism's *C* value (the size in picograms or base pairs of its haploid genome) is itself under selection, determining various features of cell biology (Cavalier-Smith 1978, for a more modern and thorough analysis, see Gregory 2001). From an MLS perspective, such a structural role is not incompatible with selfishness: if selection requires DNA in excess of that needed to encode and regulate gene products, that excess could be made up of replicating TEs competing for space. Such DNA might be considered

"clean fill," or in the recent neologism of Graur et al. (2015) "indifferent DNA"—or the very much earlier one of Zuckerkandl (1986), "polite DNA." Genomic elements may have functions at several levels, functions that can be opposed or reinforcing. As we argue later, the adoption of MLS thinking will facilitate such a more nuanced appreciation, retrospectively illuminating such early understandings as Nevers and Saedler's (1977).

Evolutionary Roles of TE Carriage

ENCODE investigators and many who promoted ENCODE's results claimed that because 80.4% of the human genome is "functional," the notion of junk DNA is overthrown. For instance Kolata (2012), wrote in the *New York Times* . . .

> . . . [T]he human genome is packed with at least four million gene switches that reside in bits of DNA that were once dismissed as 'junk' but that turn out to play critical roles in controlling how cells, organs and other tissues behave . . . At least 80% of this DNA is active and needed. (p. A1)

Very many if not most of these "gene switches" must reside in TEs, since these, in various stages of decay, make up as much as two-thirds of our genome (de Koning et al. 2011). ENCODE investigators seem not to differentiate between switches underwriting the fitness of the TE's host (selected at the organismal level) and switches that reside in TEs and serve their selfish evolutionary interests (selected at the genomic level). Or perhaps they assume without proof that most of the latter have been co-opted into organismal roles: in any case they do not even implicitly engage with MLS theory. Though admitting, in a partial rebuttal to heavy criticism, that small population sizes permit . . .

> . . . proliferation of transposable elements and other neutrally evolving DNA. If repetitive DNA elements could be equated with non-functional DNA, then one would surmise that the human genome contains vast non-functional regions because nearly 50% of nucleotides in the human genome are readily recognizable as repeat elements, often of high degeneracy. (p. 6134)

Kellis et al. (2014) go on to claim that . . .

> Genome-wide biochemical studies, including recent reports from ENCODE, have revealed pervasive activity over an unexpectedly large fraction of the genome, including noncoding and nonconserved regions and repeat elements [including TEs]. Such results greatly increase upper bound estimates of candidate functional sequences. (p. 6134)

In thus extending the upper bounds of an undifferentiated "functionality" into territory occupied by TEs without acknowledging that there are other (genome level) selective

processes at play to explain the presence of so much DNA, ENCODE investigators dismiss much previous theory in genome evolution (Doolittle 2013; Palazzo and Gregory 2014; Elliott et al. 2014). Some ENCODE supporters even claim to have at last exposed ignorant and possibly willful bias on the part of evolutionary theorists who have argued for junk and selfish DNA. Francis Collins for instance ventured in a recent public lecture that . . .

> I would say, in terms of junk DNA, we don't use that term any more 'cause I think it was pretty much a case of hubris to imagine that we could dispense with any part of the genome as if we knew enough to say it wasn't functional . . . most of the genome that we used to think was there for spacer turns out to be doing stuff and most of that stuff is about regulation and that's where the epigenome gets involved, and is teaching us a lot (Collins 2015)

Others, the admonishments of Brenner notwithstanding, continue to imagine that DNA not currently functional may prove so in future, and that this potential use really is a sort of function—an explanation of the DNA's presence. Barroso (2012), in an editorial accompanying several primary ENCODE papers, writes

> . . . [T]here is a good reason to keep this DNA. Results from the ENCODE project show that most of these stretches of DNA harbor regions that bind proteins and RNA molecules, bringing these into positions from which they cooperate with each other to regulate the function and level of expression of protein-coding genes. In addition, it seems that widespread transcription from non-coding DNA potentially acts as a reservoir for the creation of new functional molecules, such as regulatory RNAs. (p. 54)

In very much the same vein, Taft et al. (2007) earlier attributed both current and future function to the RNA transcribed from TEs, which make so large a fraction of the DNA of large genomed organisms. They wrote that . . .

> . . . there may be a vast hidden layer of RNA regulatory information in complex organisms and that increasing amounts of genetic information in these organisms is expressed as and transacted by RNA. This suggestion is supported by the finding that many genetic phenomena in the higher organisms, such as imprinting, co-suppression, RNA interference and chromatin modification, involve RNA signaling . . . (p. 297)

They do not claim that "all transcribed sequences are necessarily functional," however, and venture that "Indeed, there may be a reservoir of such transcripts that are themselves simply *raw material for evolution*" (emphasis added).

We see these passages as unconscious conflation of levels of selection and of functions engendered and maintained by

selection. That TEs are transcribed can be understood either as noise or as a consequence of earlier or ongoing selection at the genomic level. While transcription of individual TEs may have been co-opted as "regulatory information," a role as "raw material" or as "nature's tools for genetic engineering" is better cast as a function for species or clades. Species will (or will not) speciate more often or go extinct less often because of TE carriage, which is selected for (or against) at that level, just as geographic range or population size might be (Jablonski 2008). It is surely not the case that TEs as a class (or most TEs as individuals) are selected within or are maintained within species because they contribute to the fitness—the reproductive advantage—of the individual organisms that bear them. Nor indeed can the carriage of a particular TE family be seen as "hitchhiking" within populations of organisms on the beneficial effects of a mutational innovation that one of that family's members has caused—at least not in sexually reproducing species. The vast majority of the family's member will quickly become separated by recombination from any positive mutation caused by one of them, and it will only be remnants of this causative element that might be said to hitchhike.

It is of course one thing to recount past evolutionary contributions of TEs and quite another to infer that natural selection embedded TEs in genomes so that they might thus contribute. Not all authors are guilty of such teleological elision: the problem is a too narrow reading of Darwinian evolutionary processes. What MLS theory would do for all authors would be to provide a context in which the "evolutionary roles" of TEs could be posited and tested. It is a remarkable fact that transposases (in particular prokaryotic transposases) are "the most abundant genes in nature" (Aziz et al. 2010) and unquestionably the case that TEs have made many key contributions to eukaryotic evolution. The literature on TEs is rich and diverse. We can only sample this richness, picking out four general sorts of TE-related process potentially relevant to speciation or extinction: the creation of new genes, the nuancing of RNA-based regulatory networks, host-mediated effect, and chromosomal (or genomic) speciation.

New Genes

There are many documented cases in which TEs (or parts thereof) as individual elements have been co-opted or domesticated to serve host functions which might also, through effect macroevolution, affect species diversification rate. Many such cases involve exaptation of TE-encoded proteins, TE promoters and other regulatory elements (Feschotte 2008).

The protein-coding genes of long terminal repeat (LTR) retrotransposons and endogenous retroviruses (ERVs) are particularly suitable for such "neofunctionalization" (Lynch 2007), although DNA transposons and even *Alu* elements can be exapted by their hosts (Feschotte 2008; Oliver and Greene 2011; Alzohairy et al. 2013). Feschotte (2008) in

particular extols the possibility of co-optation of transposase (TPase) genes, which often encode DNA-binding domains and various chromatin modifying effects ripe for repurposing into host functions (and ripe for confusion with host functional elements by approaches like ENCODE's). A good documented evolutionary example would be the *Kat1* gene of the yeast *Kluyveromyces*, which participates in mating type switching and is derived from a transposition event at the base of this genus (Rajaei et al. 2014). In mammalian evolution, envelope genes of ERVs have been converted to syncytin genes (essential for placenta formation) multiple times (Dupressoir et al. 2012). And quite spectacularly, *Drosophila* telomeres are believed to be maintained by repurposed non-LTR retrotransposon reverse transcriptases (Belfort et al. 2011).

More global assertions about the evolutionary roles of TEs in gene creation are grounded in observations that TEs preferentially lie in transcriptionally active regions (open chromatin), do (or may) affect transcription of neighboring regions, and correlate in their family composition and dynamics with major evolutionary events, trends or divergences. Jacques et al. (2013) for instance used ENCODE DNase I hypersensitivity data to place 80% of ERVs within open chromatin, and TEs in 63% of primate-specific regions. In a recent paper suggestively titled "Widespread contribution of TEs to the innovation of gene regulatory networks," Sundaram et al. (2014) report a massive comparative survey of human and mouse transcription factor binding sites resident in TEs, concluding that . . .

> TEs have been described as parasitic or junk DNA. However there is mounting evidence for their significant evolutionary contribution to the wiring of gene regulatory networks, a theory rooted in Barbara McClintocks's discovery that TEs can control gene expression. (p. 1963)

Oliver and Greene (2011) list close to 100 individual cases "implicated in primate-specific traits" of one sort or another. Fitness consequences will of course have been in most cases hard to assess, though there can be no question that some instances will have been under positive selection in primates. More important to remember, however, is these authors' calculation that our genomes harbor at least three million TEs. Their 100 positive cases represent only about 0.0003% of that multitude. Surely even the most optimistic and up-to-date estimate of proven contributions to organismal fitness would not yet reach 1%. It seems premature to imagine that a function for TEs as a class has been found!

Quite commonly in the history of molecular biology and genomics a mysterious collection of elements (TEs, introns, long noncoding RNAs [lncRNAs]) is operationally defined as one kind of thing. Researchers then ask themselves "What is the function of this class of elements"—as if there *must* be such an underlying (essential) function and/or a common origin. They then collect evidence for the function of the

class by looking at individuals within it. In a few cases, most members of such a grouping do in fact share a property or function: tRNAs or spliceosomal introns (in so far as they all need to be spliced out) would be good examples. But in others, class membership is defined by a single property—such as transposability, having a particular size or degree of secondary structure, being transcribed but not translated (lncRNAs)—that is a very poor guarantor of homology or of the possession of any other shared properties.

One of evolutionary biology's more robust lessons is that evolution is a tinkerer, co-opting elements created by one process in the service of another (Jacob 1977). It should be no surprise to evolutionary biologists that some individual TEs (or lncRNAs or introns) have been co-opted for regulatory function and no surprise to sociologists that researchers will seek out and document such cases, claiming generality. There seems little warrant for the belief that gene creation is the "organismal" function for TEs as a class, or indeed that either the class or most of its members must have any function at all at the level of organisms. But MLS theory will allow us to expand our notions of tinkering: elements selected at one level to serve a level-specific function may be tinkered at another level, and have quite different functions there.

Nuancing Networks

In the last few years, lncRNAs have been shown to be transcribed (though often at very low levels) from the majority our own and other complex genomes. Collectively, lncRNAs are often seen as 1) regulatory, 2) generated directly (by transcription from TE promoters) or indirectly by TEs, and 3) because of their evolutionary instability and malleability, sources of evolutionary innovation [Rebollo et al. 2012; Cowley and Oakey 2013; Kapusta et al. 2013]). Kapusta and Feschotte (2014), for instance, suggest that...

> ...species with high TE content and activity, and thus more dynamic genomes, also have more complex and malleable transcriptomes, thereby increasing their capacity to evolve newly functional lncRNA molecules. It is tempting to further speculate that in these organisms with high lncRNA turnover, to which humans likely belong, variation in lncRNA content and expression could occupy a prominent position among the regulatory layers underlying trait variation. (p. 448)

Beyond this concomitance of TEs and transcribed regions, Feschotte (2008) proposes a number of other regulatory roles for TEs and TE-derived sequences. He highlights the role of post-transcriptional regulation by production of small regulatory RNA families, such as micro RNAs, and small interfering RNAs (siRNA). Some TE families (for instance MITEs) have a palindromic structure that when transcribed forms hairpins that are excellent precursors to siRNA, and since these TEs are located throughout the genome, the siRNAs

have the potential to effect the degradation of many other mRNAs containing the original TE—regulation by mRNA silencing. The potential for regulatory regions to be copied throughout the genome and link together sets of genes is clear, again as Britten and Davidson speculated (1969). Feschotte (2008) observes...

> Regardless of whether the regulatory elements arise 'de novo' by a few mutations or are pre-existing within TE sequences, the dispersal of expanding TE families throughout genomes potentially allows the same regulatory motif(s) to be recruited at many chromosomal locations, drawing multiple genes into the same regulatory network. (p. 399)

Volff (2006) in his discussion of the "raison d'être" of TEs adds "exonization" to the explanation of TE prevalence as a reservoir of variation, another nuance that depends on the complexities of eukaryotic RNA manipulations.

> TE-mediated mutations can be beneficial to the host under certain conditions. Furthermore, mobilization in *cis* or in *trans* of host gene sequences by several types of TEs suggests an involvement in exon shuffling and gene duplication.
> There is now substantial evidence from a variety of organisms that TE-derived DNA has an additional important role in evolution: *it can serve as a reservoir of sequences for genetic innovation.* (p. 913, emphasis added)

Host-Mediated Effects

A lesson taken in part from the selfish DNA era was the possibility not only of a suborganismal level adaptive explanation for TE prevalence, but of a rich host-TE antagonistic evolutionary interplay. Transcriptional suppression mechanisms like methylation of promoter regions, methylation specifically targeted to highly similar repeated sequences, and posttranscriptional interactions involving short RNA (endo-siRNAs and piRNAs) are well studied, and other specific mechanisms targeted at both LTR and non-LTR retrotransposons have been reported (Okamoto and Hirochika 2001; Levin and Moran 2011).

Of course mechanisms thought of as defense might alternatively (or also) aid TEs as beneficial agents of evolution. Huda et al. (2010) endeavored to test whether epigenetic histone modifications function as a transcriptional host defense mechanism against transposition, or whether they facilitate exaptation of TE sequences—keeping them suppressed yet present at certain exaptive genomic loci. Finding both that families of TEs are enriched with active histone modifications, and old families more so than younger, they conclude, "'With a few exceptions', most of our findings support the exaptation hypothesis" (Huda et al. 2010).

Fedoroff (2012) presents a more fully developed hypothesis of this sort, bringing the explanation of TE prevalence firmly into active modern debates about epigenetics and evolvability. Fedoroff (2012) proposes that epigenetic TE silencing mechanisms, typically thought to have evolved to combat rampant transposition, actually preexisted in prokaryote ancestors and functioned to combat deleterious lateral gene transfer—and that these same taming devices actually facilitated transposition and thus genome growth. Eukaryotes have subsequently taken advantage of their expanded genomes by selectively regulating sequences provided by TEs, as envisioned in many earlier scenarios. Taking herself to be arguing against both the "host-defense" hypothesis of host-TE interaction, and the interpretation of TEs as mere parasites, Fedoroff (2012) contends that . . .

> It is precisely the elaboration of epigenetic mechanisms from their prokaryotic origins as suppressors of genetic exchanges that underlies both the genome expansion and the proliferation of TEs characteristic of higher eukaryotes. This is the inverse of the prevailing view that epigenetic mechanisms evolved to control the disruptive potential of TEs. (p. 758)

And, she writes . . .

> It is becoming increasingly difficult to escape the conclusion that eukaryotic genome evolution is driven from within not just by the gentle breeze of the genetic mechanisms that replicate and repair DNA, but by the stronger winds (with perhaps occasional gale-force gusts) of transposon activity. The ability to evoke rapid genome restructuring is at the heart of eukaryotic evolvability—the capacity of organisms with larger and larger genomes to maintain evolutionary flexibility. (p. 766)

Formulations such as Fedoroff's explain TE prevalence without being teleological (i.e., assuming that TEs are kept on reserve for the future). Even though the advantage of a large, TE filled, genome only obtains after some period of growth and elaboration of epigenetic regulatory mechanisms, the initial growth and suppression can be explained as an exaptation of (preexistent, prokaryotic) mechanisms for suppression of genetic exchange.

It is not necessary, however, to cast this view as being in opposition to selfish DNA theory, as Fedoroff seems to do, writing that her purpose was . . .

> [T]o challenge the current, somewhat pejorative, view of TEs as genomic parasites with the mounting evidence that TEs and transposition play a profoundly generative role in genome evolution. (p. 758)

These views are not necessarily in conflict: Fedoroff's exaptive explanation for TE prevalence is still consistent with the possibility that individually selfish (transposition) events are commonly maladaptive for their hosts. In other words, genome growth can be a species-level adaptive consequence of ancestral TE suppression mechanisms, while the majority of transposition events remain mildly deleterious at the level of organisms.

Etiological views attentive to host-TE evolutionary interplay are a significant advancement over earlier explanations involving "only" TE selfishness or host functions—they move beyond initial either/or thinking. Nonetheless, the expectation that transcriptional suppression of TEs (as a class) should be explained by a single evolutionary story—and at that one which is only attentive to host and TE-level selection—is precisely the kind of pretheoretical assumption eschewed by MLS. The search for a unitary raison d'être for TEs is parochial when viewed from an MLS standpoint.

Genomic Speciation

In order for TEs to have functions at the level of species they must influence speciation (or extinction). This could be accomplished by disrupting or modifying characteristics of the genome or gross karyotypic features—a theory many authors have put forward in some form or other. As dispersed repeats, TE families facilitate duplication, inversions, translocations, and transpositions of TE-bounded genes and "exon shuffling," all with the aid of host-encoded (not element-specific) recombinational machinery. Moreover, bulk chromosomal and nuclear properties, karyotype, chromosome stability, and pairing at meiosis must all be affected by TE content and TE activity. These latter properties are likely to have species-level consequences. There is undeniable overlap between what Oliver and Greene (2011) call "passive TE thrust," what Rose and Doolittle (1983) much earlier described as "molecular biological mechanisms of speciation" and Dobzhansky's still earlier model of "chromosomal speciation" (Dobzhansky 1935). In this, chromosomal rearrangements in diverging populations precede and cause reduced hybrid fertility through mispairing at meiosis and the generation of unbalanced gametes ("underdominance"). Alternative ("genic speciation" or Dobzhansky-Muller) models see failures in more complex epistatic interactions between diverging gene products as primary (Brown and O'Neill 2010).

The effects of TEs on genomic speciation events occupy a special place in the biological hierarchy, not reducible in either quantity or type to the adaptive organism- or genome-level explanations for TE carriage. TE contribution to speciation, whether by chromosomal speciation, TE thrust, or the mechanisms described in Rose and Doolittle (1983), is casually distinct from their contribution to organismal fitness. For example, transposition into a genomic location that interferes with reproduction and induces speciation is not only casually distinct from any organism-level adaptive hypotheses for TE carriage, but in most cases is maladaptive at that level.

An appealing chromosomal speciation scenario tailored to TEs was entailed in Dover's (1982) molecular drive model. TE families exhibit concerted evolution, changing in sequence together within a population while diverging cohesively between populations—through the combined actions of element replication and loss, and/or gene conversion. Dover inferred that . . .

> . . . if in addition, future studies confirm that many non-genic families affect DNA transcription and transcript processing, or chromatin structure and chromosome behaviour, then the dual process of generating intra-population cohesion and interpopulation discontinuities would be of considerable evolutionary significance. (p. 115)

Such effects are indeed what subsequent studies (for instance those cited above on lncRNAs) might be seen to confirm. Chromosomal speciation in its simpler forms (hybrid infertility consequent to gross chromosomal rearrangements) seems now a less popular speciation model than those invoking complex epistatic and epigenetic interactions at multiple sites. Nevertheless, Brown and O'Neill note in their 2010 review on chromosomes and speciation that . . .

> A recurring theme . . . is the appearance of mobile DNA as participants in the establishment of species barriers. Whether through repeat divergence, piRNA or crasiRNA divergence, rDNA mobility, or chromosome restructuring, mobile DNAs can mediate ectopic chromosome exchanges, thus facilitating mobile DNA-mediated homologous or nonhomologous chromosome rearrangements. McClintock noted that major restructuring of the genome could occur in hybrids, mediated by the activity of mobile DNA. More recently, the activity of mobile DNA has also been implicated in the functioning of centromeres and the fragility of breakpoints in the genome, pointing to a complex role for such selfish elements in chromosome change. (p. 303)

How We Should Think About TEs' Evolutionary Functions

Our position is that the various proposed evolutionarily beneficial consequences of TE carriage (whether instantiated or not) are best looked at as species-level traits, accountable to selection at that level (fig. 1). TE carriage will be advantageous insofar as it promotes speciation or forestalls extinction, by any or all of the means we summarize above. Thus it can exert downward causation (species level to individual), and is not different in this regard from other strict-sense species-level features listed by Jablonski (2008), such as population size and geographical range. But, "within" species, TE carriage is in general not a positively selected property, we argue. Indeed TEs may usually be detrimental to the individual organisms

that carry them, and in which their initial presence is best seen as a consequence of selection at a still lower level, that of genomes. In other words, it is differential reproduction of TEs within genomes and (potentially) the differential reproduction of TE-bearing species within genera that account for TE prevalence in existing genomes and the prevalence of genomes that bear them. Differential reproduction of individuals within species—the most commonly invoked force for the attribution of a trait's function—is seldom the explanation and presumably works against TE carriage most of the time.

Selection at the species level need not always or even often be positive, of course. How frequently it is a matter for empirical investigation, as discussed below. But at least MLS theory is a formalism that brings claims that TEs serve evolutionary functions into a Darwinian framework, and is in principle testable. A case might also be made for selection at a level above species. To be sure, clades higher than species do not "cladiate" in the way that species speciate, and indeed it is only through speciation that new clades arise. Thus, Okasha (2006) does not endorse "clade selection." But clades, like the dinosaurs, do go extinct. And some clades, because of aggregate properties of the organisms or species that comprise them (like the small size of early mammals, apocryphally), are less likely to do so than others.

Evolution cannot "look ahead," and it is the implication that it can that so upset Brenner. But if we "look back," species that have spawned larger and more robust multispecies clades will (sometimes) be those that have had TEs forced upon them, as selfish or junk DNA. This may not often happen: TE carriage may be in general detrimental in species selection. Our point is that if and when it is positive for any of the reasons cited above, it is species (or clades) that benefit, and MLS theory that we must use to explain the benefit in anything resembling Darwinian language. There are several advantages to thinking this way.

TE Carriage as a Species Level Trait

A consensus among supporters of selected effect concepts of function is that relevant selection must be appropriately recent. Thus, the human appendix once had a function, but (at least in popular wisdom) does not now. Feathers may have first arisen to keep the reptilian ancestors of birds warm, but their current function is flight. Though such examples seem clear, a conundrum Kraemer (2014) calls the "no variation problem" arises for recently selected traits that are so useful that trait-determining alleles are 100% fixed within a species. If selection requires heritable variation in fitness, then when there is no variation selection stops—an ironic but seemingly unavoidable consequence, unless the "potential" of purifying selection is taken on board.

There are other ways to sidestep this inconvenience. MLS theory is one of them, of especial relevance to TE carriage. TE

family number and sequence—element mobility notwithstanding—may vary relatively little between organisms within a species, but vary importantly between species and between populations during speciation, especially when bottlenecked or under stress (Dover 1982; Lockton et al. 2008; Zeh et al. 2009; Kim et al. 2014; Chalopin et al. 2015). When TE complements differentially affect speciation and extinction so as to produce more species, they will have selected effect functions for species. Whatever we think about whether or how recently selection must have been operative within species for functional attribution at that level, this will be so.

Three further, related, arguments can be made. First, within-species trait fixation or relative uniformity coupled with between-species differences make a good recipe for species-level selection or effect macroevolution quite generally. Jablonski (2008) argues...

> The key requirements [for what Jablonski calls "broad sense species selection"] are that (a) a trait exhibits little or no variation within species relative to the variation among species... and (b) speciation and/or extinction covaries consistently across one or more clades with that trait. Mammalian body size is often viewed in this light: Species tend to exhibit modal sizes, and a cross-level discordance may exist in the evolutionary consequences of size in that short-term organismic selection might often favor larger body size... but larger bodied species or clades may be more extinction-prone over longer time scales. Of course, broad-sense species selection need not oppose selection at the organismic level, although this is analytically more tractable; selection might as readily operate in the same direction at multiple levels... such concordance assumes some intraspecific variation in a focal trait, which still could be modest relative to among-species variation... (p. 503)

Second, we need to remember that although individual TEs may be responsible for innovations, it is the trait "possession of many TEs, or of TE family X, or of many different and active TE families" to which evolvability-promoting, anticipatory, "future," "reserve," or "genetic engineering" function is generally attributed, albeit sometimes only tacitly. No one has seriously claimed that a particular individual element—which by its fortuitous insertion near a particular gene has resulted in a phenotypic innovation—was earlier selected for in its host genome so that it would have that beneficial effect in the future!

Third, Dover's "molecular drive" hypothesis anticipates and conveniently explains within-species TE family cohesion and between-species divergence, providing a mechanism for genomic speciation independent of selection on organismal phenotype (Dover 1982). Dover notes that the "concerted evolution" of multigene families, whether by unequal crossing over, gene conversion or gain and loss (as with TEs) will effect both within-population homogeneity and between-population divergence. If rates of within-population

recombination are high with respect to intragenomic transposition and loss...

> ...the fixation of variant sequences on different chromosomes by stochastic and directional mechanism would induce a concerted phenotypic change in a group of individuals without the concomitant effect of disturbing relative differences in individual fitness... (p. 115)

Should the repeat elements thus homogeneously altered be involved in coordinated regulation or chromosome mechanics, postzygotic isolation, which Dover deemed "accidental speciation," may result.

Getting Past either/or Thinking

Both public (including creationist) and scientific discourse around ENCODE has seemed fixated on the question of whether most of the DNA that makes up our own and other large genomes "is" or "is not" junk. It is as if, collectively, TEs must either be functional or junk as a class. Biémont and Viera (2006) for instance write that...

> What was once dismissed as junk DNA must now be regarded as a major player in many of the processes that shape the genome and control the activity of its genes. (p. 521)

Or, as Jacques et al. (2013) recently put it...

> Although the parasitic behavior of TEs was initially put forward as a sufficient explanation for their maintenance within genomes there is growing evidence to support the alternative view that TEs have facilitated genomic innovations and contributed critical regulatory elements to their hosts. (p. 1)

And addressing physicians, Guttmacher and Collins (2002) as editors of the New England Journal of Medicine, assert that...

> Much is known, but much remains mysterious. We know that less than 2 percent of the human genome codes for proteins, while over 50 percent represents repeat sequences of several types, whose function is less well understood. These stretches of repetitive sequences, sometimes wrongly dismissed as "junk DNA," constitute an informative historical record of evolutionary biology, provide a rich source of information for population genetics and medical genetics, and by introducing changes into coding regions, are active agents for change within the genome. (p. 1514)

Brenner (and more recently Graur et al. [2015]) have sought to separate the various sorts of DNA not under selection into "garbage" (harmful) and junk (not harmful, and potentially recruitable into function) and insist that it is only current utility that is at issue.

... function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact a handful of *Alu* elements have become functional cannot be taken as support for the hypothesis that all *Alu* elements are functional. The Aristotelian distinction between potentiality and actuality is crucial. (p. 643)

Recognition that TE carriage might most often be selfish at the genomic level, detrimental if not neutral at the organismal and—perhaps sometimes—beneficial at the species level, replaces "either/or" thinking with a more realistic pluralist ontology.

Amenability to Empirical Test

Claims of species-level or clade level function can "in principle" be tested quantitatively. Teleological imputations of future utility—as Graur et al. (2015) note—cannot. Rabosky and McCune (2010) review how phylogenetic trees supporting species selection may show heritable differences in speciation or extinction rates between lineages and or reliable associations between such diversification rates and the possession of relevant traits. Relevant traits at the population level might include geographic range and patchiness, population size, structure, and genetic variability. At the organismal level body size and floral symmetry, might drive effect macroevolution by upward causation. Our point here is to add—at the still lower, genomic, level—the relevant quantitative and qualitative characteristics of TE families and their activities. Increasingly sophisticated methods are being developed to perform tests like those described by Rabosky and McCune (2010), The studies of Brawand et al. (2014) on the African cichlid radiation or of Oliver et al. (2013) comparing angiosperms and gymnosperms are steps in this direction. Statistical methods for avoiding misattribution of species-level functions are under active development (Magnuson-Ford and Otto 2012; Ng and Smith 2014; Rabosky and Goldberg 2015).

Of particular relevance here, Kraaijeveld (2010) reported that species richness and TE content are inversely correlated, not what one would expect from imputations of TEs' roles in evolutionary innovation. It is what one might expect if population decline prior to extinction often accompanies TE accumulation (Lynch 2007; Kraaijeveld 2010). In any case our point is not that abundant TE carriage always or even most often benefits species, but that it is at the level of species that evolutionary benefit should be sought. At the individual level, the "junk" designation remains defensible.

What is needed is final transition (into historical stage 3) in the evolution of thinking and talking about the evolution of TEs. As Brenner (1998) noted, this transition will not come easily.

Even today, long after the discovery of repetitive sequences and introns, pointing out that 25% of our genome consists of millions of copies of one boring sequence, fails to move audiences. They are all convinced by the argument that if this DNA were totally useless, natural selection would already have removed it. Consequently, it must have a function that still remains to be discovered. Some think that it could even be there for evolution in the future — that is, to allow the creation of new genes. As this was done in the past, they argue, why not in the future? (p. R669)

Only if such invocations of "future function" are cast in more formal hierarchical terms do they become in principle testable and potentially consistent with Darwinian theory. The traits of an individual can be rationalized not only as to how they helped an individual's ancestors compete with conspecifics but how they helped its species compete with other species. Though levels of selection may be difficult to delimit, conceptually they are discrete. The vagueness with which many of the invocations of evolutionary roles for TEs cited above are expressed does not do justice to this.

Questions about why particular species have abundant TEs have different answers, even different kinds of answers, than do questions about why there are so many species with abundant TEs. Both are at issue in any genomic characterization. Real-life cases will be fraught with complexities and difficulties, and our point is not one of practice but of principle. It is at the level of species that popular claims about the evolutionary functions of TEs should be formulated and tested. We could then move beyond unsophisticated rhetoric of the typical form "TEs, once derided as 'selfish junk' are now known to be vital drivers of genome evolution."

Acknowledgments

Literature Cited

Alzohairy AM, Gyulai G, Jansen RK, Bahieldin A. 2013. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid 69:1–15.

Aziz RK, Breitbart M, Edwards RA. 2010. Transposases are the most abundant genes, most ubiquitous genes in nature. Nucleic Acids Res. 38:4207–4217.

Barroso I. 2012. Non-coding but functional. Nature 489:54.

Belfort M, Curcio MJ, Lue NF. 2011. Telomerase and retrotransposons: reverse transcriptase that shaped genomes. Proc Natl Acad Sci U S A. 108(51) 20304–20310.

Biémont C, Viera C. 2006. Genetics: junk DNA as an evolutionary force. Nature 443:521–524.

Booth A, Doolittle WF. 2015. Eukaryogenesis, how special really? DOI: 101073/pnas.1421376112.

Borrello ME. 2005. The rise, fall and resurrection of group selection. Endeavour 29((1):43–47.

Brawand D, et al. 2014. The genomic subtrate for adaptive radiation on African cichlid fish. Nature 513(7518); 375–381.

Brenner S. 1998. Refuge of spandrels. Curr Biol. 8(19) R669.

Britten RJ, Davidson EH. 1969. Gene regulation for higher cells: a theory. Science 165(3891):349–357.

Brown JD, O'Neill RJ. 2010. Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. Annu Rev Genomics Hum Genet. 11:291–316.

Cavalier-Smith T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the C-value paradox. J Cell Sci. 34:247–278.

Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. 2015. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol. 7:567–580.

Collins F. 2015. Speech at the JP Morgan Healthcare Conference on 2015 Jan 13; [cited 2015 Jun 3]. Available from: http://sandwalk.blogspot.ca/2015/01/francis-collins-rejects-junk-dna.html.

Cowley M, Oakey RJ. 2013. Transposable elements re-wire and fine-tune the transcriptome. PLoS Genet. 9(1):e1003234.

Dawkins R. 1976. The selfish gene. Oxford: Oxford University Press.

de Koning APJ, Gu W, Castoe TA, Batzer MA, PollockDD. 2011. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 7(12):e100238.

Dobzhansky T. 1935. The Y chromosome of Drosophila pseudoobscura, Genetics 20:366–376.

Doolittle WF. 1987. What introns have to tell us: hierarchy in genome evolution. Cold Spring Harb Symp Quant Biol. 52:907–913.

Doolittle WF. 2013. Is junk DNA bunk? A critique of ENCODE. Proc Natl Acad Sci U S A. 110(14):5294–5300.

Doolittle WF. 2014. Natural selection through survival alone, and the possibility of Gaia. Biol Philos. 29:415–423.

Doolittle WF, Brunet TD, Linquist S, Gregory TR. 2014. Distinguishing between "function" and "effect" in genome biology. Genome Biol Evol. 6(5)l 1234-1237.

Doolittle WF, Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284(5757):601–603.

Dover G. 1982. Molecular drive: a cohesive model of species evolution. Nature 299:111–117.

Dupressoir A, Lavaille C, Heidmann T. 2012. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. Placenta 33:663–671.

Eddy S. 2013. The ENCODE project: missteps oveshadowing a success. Curr Biol. 23(7):R529–R561.

Elliott TA, Linquist S, Gregory TR. 2014. Conceptual and empirical challenges of ascribing function to transposable elements. Am Nat. 184:14–24.

Fedoroff NV. 2012. Presidential address. Transposable elements, epigenetics and genome evolution. Science 338(6108):758–767.

Feschotte C. 2008. The contribution of transposable elements to the evolution of regulatory networks. Nat Rev Genet. 9(5):397–405.

Graur D, et al. 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. Genome biology and evolution, 5(3):578–590.

Graur D, Zheng Y, Azvedo RBR. 2015. An evolutionary classification of genomic function. Genome Biol Evol. 7(3):642–645.

Gregory TR. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol Rev Camb Philos Soc. 76(1):65–101.

Guttmacher AE, Collins FS. 2002. Genomic medicine—a primer. N Engl J Med. 347(19):1512–1520.

Huda A, Mariño-Ramirez L, Jordan IK. 2010. Epigenetic histone modification of human transposable elements: genome defense versus exaptation. Mobile DNA 1(1):2.

Jablonski D. 2008. Species selection: theory and data. Ann Rev Ecol Ecol Syst. 39:501–524.

Jacob F. 1977. Evolution and tinkering. Science 196(4295):1161–1166.

Jacques P-É, Jeyakani J, Bourque G. 2013. the majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet. 9(5):e1003504.

Kapusta A, et al. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. 9(4):e10034.

Kapusta A, Feschotte C. 2014. Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications, Trends Genet. 20(10):439–452.

Kellis M, et al. 2014. Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A. 111(17):6131–6138.

Kim YB, et al. 2014. Divergence of Drosophila melanogaster repeatomes in response to a sharp microclimate contrast in Evolution Canyon, Israel. Proc Natl Acad Sci U S A. 111(29):10630–10635.

Kirschner M. 2013. Beyond Darwin: evolvability and the generation of novelty BMC Biol. 11:110.

Kleckner N. 1981. Transposable elements in prokaryotes. Ann Rev Genet. 15:341–404.

Kolata G. 2012. Bits of mystery DNA, far from 'junk', play crucial role. New York Times, 2012 Sep 5.

Kraaijeveld K. 2010. Genome size and species diversification. Evol Biol. 37(4):227–233.

Kraemer DM. 2014. Revisting recent etiological theories of function. Biol Philos. 29:747–759.

Levin HL, Moran JV. 2011. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet. 12(9):615–627.

Lockton S, Ross-Ibarra J, Gaut BS. 2008. Demography and weak selection drive patterns of transposable element diversity in natural populations of Arabidopsis lyrata. Proc Natl Acad Sci U S A. 105(37):13965–13970.

Lynch M. 2007. Origins of genome architecture. New York: WH Freeman.

Magnuson-Ford K, Otto SP. 2012. Linking the investigations of character evolution and species diversification. Am Nat. 180(2):225–245.

Maynard Smith J. 1964. Group selection and kin selection. Nature 201(4924):1145–1147.

McClintock B. 1961. Some parallels between gene control systems in maize and in bacteria. Am Nat. 95(884):265–277.

Nevers PJ, Saedler H. 1977. Transposable elements as agents of gene instability and chromosomal rearrangements. Nature 268(5616):109–115.

Ng J, Smith SD. 2014. How traits shape trees: new approaches for detecting character state-dependent lineage diversification. J Evol Biol. 27:2035–2045.

Okamoto H, Hirochika H. 2001. Silencing of transposable elements in plants. Trends Plant Sci. 6(11):527–534.

Okasha S. 2006. Evolution and the levels of selection. Oxford: Oxford University Press.

Oliver KR, Greene WK. 2011. Mobile DNA and the TE-Thrust hypothesis: supporting evidence from primates. Mobile DNA 2:8.

Oliver KR, McComb JA, Greene WK. 2013. Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol Evol. 5(10):1886–1901.

Orgel LE, Crick FHC. 1980. Selfish DNA: the ultimate parasite. Nature 284(5757):604–607.

Palazzo AF, Gregory TR. 2014. The case for junk DNA. PLoS Genet. 10(5):e1004351.

Rabosky DL, Goldberg EE. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst Biol. 64(2):340–355.

Rabosky DL, McCune AR. 2010. Reinventing species selection with molecular phylogenies. Trends Ecol Evol. 25(2):68–74.

Rajaei N, Chiruvella KK, Lin F, Åström SU. 2014. Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast. Proc Natl Acad Sci U S A. 111(43):15491–15496.

Rebollo R, Romanish MT, Mager DL. 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 2012;46:21–42.

Rose MR,Doolittle WF. 1983. Molecular biological mechanisms of speciation. Science 220(4593):157–162.

Spradling AC, Bellen HJ, Hoskins RA. 2011. *Dorsophila* P elements preferentially transpose to replication origins. Proc Natl Acad Sci U S A. 108(38):15948–15953.

Sundaram V, et al. 2014. Widespread contribution of transposable elements to the innovation of regulatory networks. Genome Res. 24(12):1963–1976.

Taft RJ, Pheasant M, Mattick JS. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 29(3):288–299.

Tomkins J. 2012. Junk DNA myth continues its demise. Acts and Facts 41(11):11–13.

Trivers R. 1971. The evolution of reciprocal altruism, Quart Rev Biol. 46(1):58–65.

Volff J-N. 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28:913–22.

Vrba ES. 1984. What is species selection? Syst Zool 33(3):318–328

Vrba ES, Gould SJ. 1986. The hierarchical expansion of sorting and selection: sorting and selection cannot be equated. Paleobiol 12(2):217–228.

West SA, Lively CM, Read AF. 1999. A pluralist approach to sex and recombination. J Evol Biol. 12:1003–1012.

Wynne-Edwards VC. 1986. Evolution through group selection. Oxford: Blackwell Scientific Publications.

Zeh DW, Zeh JA, Ishida Y. 2009. Transposable elements and an epigenetic basis for punctuated equilibria. Bioessays 31:715–726.

Zuckerkandl E. 1986. Polite DNA: functional density and functional compatibility in genomes. J Mol Evol. 24(1-2):12–27.

Associate editor: Bill Martin