Current Biology
Dispatches

Genome Size Evolution: Small Transposons with Large Consequences

Alexander Suh
Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Science for Life Laboratory, Uppsala University, SE-752 36, Uppsala, Sweden
Correspondence: alexander.suh@ebc.uu.se
https://doi.org/10.1016/j.cub.2019.02.032

Transposable elements (TEs) heavily influence genome size variation between organisms. A new study on larvacean tunicates now shows that even non-autonomous TEs — small TEs that parasitize the enzymatic machinery of large, autonomous TEs — can have a large impact on genome size.

Genome sizes vary massively across the tree of life. Among animals, extreme genome sizes range from ~0.02 Gb in a nematode to ~130 Gb in a lungfish [1,2]. This >6,600-fold genome size variance in animals alone has been puzzling researchers for decades. Genome size does not simply reflect the number of genes or organismal complexity, but instead strongly correlates with the abundance of transposable elements (TEs) in animals and other eukaryotes [3]. How much of genome size variation between organisms results from adaptive versus non-adaptive processes has been debated elsewhere [4,5], but in the light of population genetics, at least a large aspect of genome size variation can be attributed to non-adaptive processes [6,7].

Just like their organismal hosts, TEs come in many shapes and flavors [8]. TEs are selfish genetic elements that propagate via a cut-and-paste (DNA transposons) or copy-and-paste mechanism (retrotransposons). Transposition involves one or several proteins encoded by the TEs themselves (Figure 1) — at least a transposase protein in DNA transposons and a reverse transcriptase protein in retrotransposons. However, some TEs lack their own coding capacity — they are 'non-autonomous' and instead rely on proteins encoded by other ('autonomous') TEs. Considering that TEs propagate as parasites of their hosts, non-autonomous TEs in principle are 'parasites of parasites'. They exist in all major groups of TEs (Figure 1), can be of diminutive size (<100 bp) [9], and hijack the enzymes of their autonomous TE counterparts through specific sequences needed for (retro) transposition [8]. In non-autonomous non-LTR retrotransposons (e.g., SINEs), these are 3' sequence tails with strong sequence similarity to their autonomous counterparts (LINEs) [10]. In LTR retrotransposons and DNA transposons, the *trans*-mobilization of non-autonomous TEs occurs through the high sequence similarity of the terminal repeats to their respective autonomous versions [8]. A new non-autonomous TE can thus emerge from an autonomous TE simply through the loss of its own coding capacity [8]. With regards to their consequences for genome size evolution, one may expect intuitively that autonomous TEs affect genome size expansion more strongly than smaller non-autonomous ones, simply due to their larger size and their enzymatic autonomy (Figure 2A). It is therefore not surprising that, for example, the gigantic >32-Gb genomes of salamanders are heavily enriched in autonomous LTR retrotransposons which are each 6–16 kb in size [11].

In a new study in this issue of *Current Biology*, Naville *et al.* [12] report an alternative route to genome size expansion — one through small, non-autonomous TEs (Figure 2B,C). The authors studied larvacean tunicates, a group of planktonic, tadpole-resembling invertebrates that is very closely related to vertebrates [13]. These tunicates have genome sizes of 72–874 Mb that may appear rather small compared to many vertebrates, but this notably corresponds to 12-fold variation in genome size across relatively short evolutionary timescales.

Naville *et al.* [12] analyzed in detail the genomes of seven larvacean tunicates for repetitive elements using various *de novo* prediction and homology-based methods. This approach should identify both known as well as novel families of TEs in each sampled genome. The authors further estimated the relative age of TE accumulation for each genome to decide whether the genome size differences are due to TE expansion in larger genomes or instead a lack of TE accumulation in smaller genomes. The data suggest that larger species have larger genomes, which in turn have larger densities of TEs resulting from recent bursts of (retro)transposition.

Interestingly, while Naville *et al.* [12] detected low numbers of autonomous TEs in all genomes, the authors found that a diversity of non-autonomous TE families makes up large genome proportions. SINEs alone account for 83% of the genome size variation. All of the identified non-autonomous TE families appear to be species-specific, i.e., they are each present in only one of the sampled genomes. This further indicates that the genome size differences result from differential TE accumulation on relatively recent evolutionary timescales. The authors note, however, that it is difficult to establish which autonomous TE families mobilized their non-autonomous parasites due to little shared sequence homology between them (*cf.* Figure 1).

The results of Naville *et al.* [12] are counterintuitive for two reasons. As mentioned above, non-autonomous TEs are much shorter than their autonomous counterparts and, more importantly, depend on parasitizing their enzymatic machinery. A continuous expansion of non-autonomous TEs should thus either coincide with a comparable expansion of the corresponding autonomous TEs, or happen strongly at the expense of the autonomous TEs, ultimately leading to the extinction of the latter through mutations

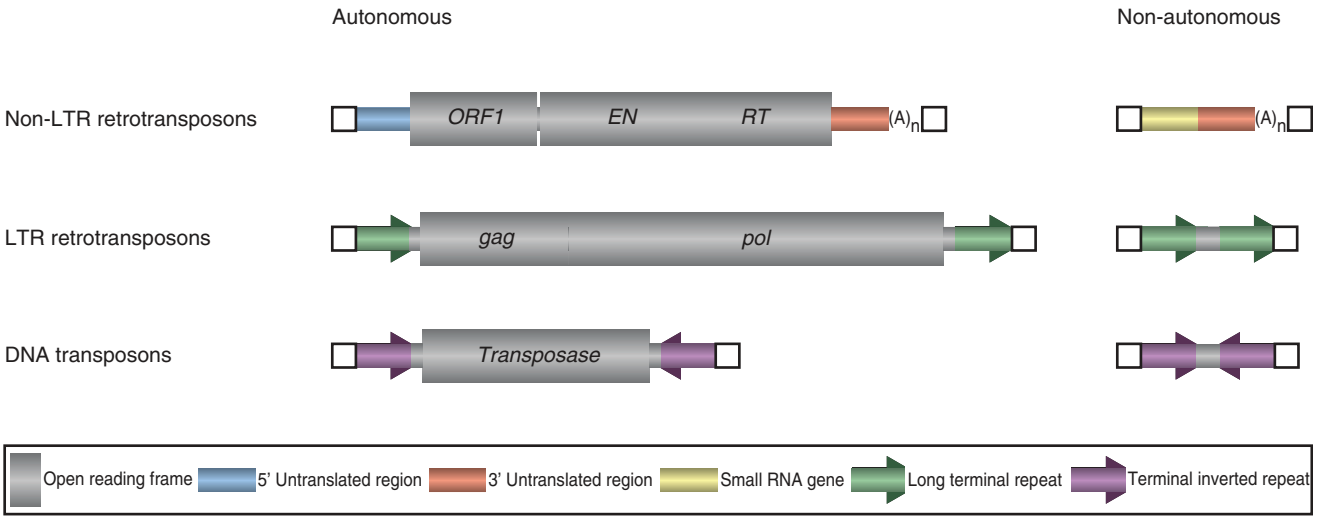

Figure 1. Non-autonomous transposable elements are parasites of autonomous transposable elements.
The three major groups of transposable elements (TEs) contain non-autonomous versions, short TEs which are *trans*-mobilized by the enzymatic machinery of their autonomous counterparts [8]. Among non-LTR retrotransposons, the most widespread non-autonomous TEs are short interspersed elements (SINEs) mobilized by long interspersed elements (LINEs). LINEs and SINEs often have a poly(A) tail or a different simple repeat at their very 3' ends. White squares indicate target site duplications.

(Figure 2C). Such an extinction would obviously stop the expansion of non-autonomous TEs and likewise lead to their extinction. Genomes thus usually have a mix of significant copy numbers of non-autonomous and autonomous TEs. Take, for example, our own human genome, arguably one of the best-studied animal genomes. SINEs are by far the most numerous group of TEs (>1.5 million copies) but make up only 13% of the genome, while less numerous autonomous TEs account for nearly 30% of the genome [14]. Why are larvacean tunicates so different in this regard?

One may speculate about these findings in the light of population genetics. Larvacean tunicates likely have very large population sizes due to their marine planktonic life style, which might allow active autonomous TEs to persist at very low frequencies in the population. These low frequencies might still be sufficient for

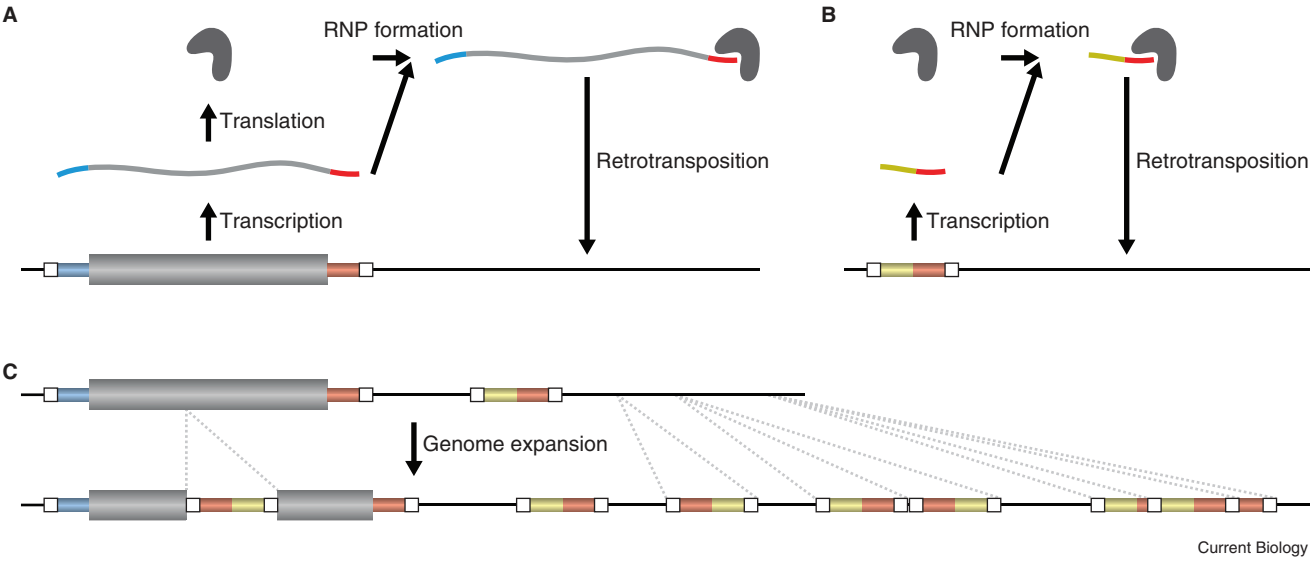

Figure 2. Transposable elements are drivers of genome size evolution.
(A) Direct genome expansion through retrotransposition of autonomous TEs. The example shows a long interspersed element (LINE). (B) Indirect genome expansion through retrotransposition of non-autonomous TEs after hijacking the enzymatic machinery of autonomous TEs. The example shows a short interspersed element (SINE). (C) Schematic illustration of genome expansion via a SINE, which might ultimately lead to the interruption of the mobilizing LINE. Dashed lines indicate SINE insertion events in the expanded genome (below) relative to the pre-expansion situation (above). RNP, ribonucleoprotein. LINE/SINE poly(A) tails are not shown for simplicity.

TE protein production and mobilization of non-autonomous TEs in significant numbers. Under this scenario, it seems counterintuitive that newly inserted small TEs would drift to fixation in such large numbers that they lead to genome expansion. However, it is plausible that new insertions of small TEs are less deleterious than those of large autonomous ones, and might thus more easily accumulate. This would be especially the case in large populations with a high efficacy of selection against deleterious mutations [15].

It is also important to keep in mind that genome size is the net result of gain and loss of DNA [16]. In organisms such as birds and mammals with relatively low genome size variation, the rate of TE accumulation covaries with the rate of deletions in an 'accordion'-like process [17]. It remains to be seen whether the extensive genome size variation in larvacean tunicates arises only from differential TE accumulation or additionally from differential covariation with DNA deletion rates. Future studies on larvacean tunicates will hopefully elucidate these open questions by sequencing additional species and looking at population-level variation of genome sizes.

In summary, it now emerges that not only large TEs can have a significant impact on genome size, but also small, non-autonomous TEs through continuous hijacking of proteins from autonomous TEs. This phenomenon further adds to the diverse ways that TEs shape the genomes of their hosts [18]. One may wonder, however, whether larvacean tunicates are an exceptional case of genome size expansion through non-autonomous TEs. Maybe more such cases will be unearthed once we dive deeper into the genomes of other understudied organisms.

REFERENCES

1. Gregory, T.R. (2017). Animal Genome Size Database. http://www.genomesize.com.
2. Doležel, J., Bartoš, J., Voglmayr, H., and Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. Cytometry A *51A*, 127–128.
3. Elliott, T.A., and Gregory, T.R. (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos. Trans. R. Soc. B *370*, 20140331.
4. Petrov, D.A. (2001). Evolution of genome size: new approaches to an old problem. Trends Genet. *17*, 23–28.
5. Cavalier-Smith, T. (2005). Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. Ann. Bot. *95*, 147–175.
6. Lynch, M., and Conery, J.S. (2003). The origins of genome complexity. Science *302*, 1401–1404.
7. Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. Proc. Natl. Acad. Sci. USA *104*, 8597–8604.
8. Feschotte, C., Jiang, N., and Wessler, S.R. (2002). Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. *3*, 329.
9. Wessler, S.R. (2006). Transposable elements and the evolution of eukaryotic genomes. Proc. Natl. Acad. Sci. USA *103*, 17600–17601.
10. Ohshima, K., Hamada, M., Terai, Y., and Okada, N. (1996). The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. Mol. Cell. Biol. *16*, 3756–3764.
11. Nowoshilow, S., Schloissnig, S., Fei, J.-F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., *et al*. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature *554*, 50.
12. Naville, M., Henriet, S., Warren, I., Sumic, S., Reeve, M., Volff, J.-N., and Chourrout, D. (2019). Massive changes of genome size driven by expansions of non-autonomous transposable elements. Curr. Biol. *29*, 1161–1168.
13. Delsuc, F., Brinkmann, H., Chourrout, D., and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature *439*, 965.
14. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., *et al*. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.
15. Wolf, J.B.W., and Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. Nat. Rev. Genet. *18*, 87–100.
16. Petrov, D.A., Sangster, T.A., Johnston, J.S., Hartl, D.L., and Shaw, K.L. (2000). Evidence for DNA loss as a determinant of genome size. Science *287*, 1060–1062.
17. Kapusta, A., Suh, A., and Feschotte, C. (2017). Dynamics of genome size evolution in birds and mammals. Proc. Natl. Acad. Sci. USA *114*, E1460–E1469.
18. Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., *et al*. (2018). Ten things you should know about transposable elements. Genome Biol. *19*, 199.

Behavior: Why Male Flies Sing Different Songs

Dana S. Galili[1] and Gregory S.X.E. Jefferis[1,2,*]
[1]Neurobiology Division, MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK
[2]Drosophila Connectomics Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
*Correspondence: jefferis@mrc-lmb.cam.ac.uk
https://doi.org/10.1016/j.cub.2019.02.054

A new study investigates the distinct male courtship songs of two related *Drosophila* species and the neurons controlling this behavior, localizing a site of evolutionary divergence to the motor system, downstream of the central brain.

How do brain circuits evolve to produce distinct patterns of behavior across species? Answering this question is likely to have a significant impact on both neuroscience and the study of evolution. A well-studied example of species-specific behavior is acoustic communication, used as a courtship signal from insects to frogs, birds and humans. Sexual selection through female mate choice has driven the evolution of a huge variety of male courtship songs across closely related species. Divergent evolution of song production (and