

Transposable element invasions

Elizabeth H.B. Hellen and John F.Y. Brookfield*

Centre for Genetics and Genomics; School of Biology; University of Nottingham; Nottingham, UK

Transposable elements have an ongoing, largely parasitic interaction with their hosts. We are interested in the timescale of this interaction. In a recent publication, we have examined the sequence divergence between class II DNA transposons from mammalian genomes. We asked whether these sequences undergo a continuing process of turnover, keeping a family as an integrated whole, as members of the family are continually created and lost. Alternatively, we envisaged that elements might have been involved in a burst of amplification, soon after they first occupied a mammalian genome, and the shared ancestry of present-day elements harks back to this initial amplification, a process that we termed a “life cycle.” We resolved between these processes by estimating the time to common ancestry predicted from the genetic diversity of sequences found in a transposon family, and also estimating, from the mammalian orders that currently possess copies of the family, the time when the family first entered the mammalian genome. These times are approximately the same, supporting the “life cycle” model. This casts light on how far we can infer genetic changes in the past through the study of DNA sequences from the present.

Transposable Elements and the Genomic Fossil Record

Transposable elements are, by the standards of living things, described using unusual terminologies. Thus, in considering these DNA sequences from our human chromosomes, one reads that these are “extinct,” following “mass extinctions”.¹ This is, at first sight, an unexpected way

to describe DNA sequences replicated in our chromosomes every time a cell divides and passed on faithfully to our offspring. But this reflects the unusual capacity that transposable element biologists have to reconstruct the past. “Extinct” elements, in this context, are not elements that have been lost from our chromosomes, they have merely become transpositionally inactive and live on in a slow process of decay. Thus, our chromosomes are said to contain “genomic fossils,” remnants from once active elements which continue to offer clues that allow us to reconstruct genomic events in the past.

The Search for Equilibria

The study of the history of transposable elements, which these genomic fossils allow, attracts the attention of evolutionary biologists. Evolutionary biologists, as a class, are particularly fascinated by the possibility of stable equilibrium states. Much of evolutionary biologists’ activities consist of studying dynamic processes of change, such as mutation, migration, selection and genetic drift. As a result of the constant operation of these processes, the genotype, the phenotype, and indeed the genetic diversity of populations, is expected to evolve toward an equilibrium state, and this expected equilibrium is taken as a prediction of the patterns that we expect to see. Indeed, levels of diversity observed today can be used in the context of models to estimate population characteristics in the past, as when levels of human molecular diversity are used to predict a human effective population size, over the last million years, of around 10,000,^{2,3} compared with the billions of individuals seen today. This search for

Keywords: mammals, transposons, class II, molecular dating, evolution

Submitted: 12/10/12

Revised: 02/04/13

Accepted: 02/07/13

Citation: Hellen EHB, Brookfield JFY. Transposable element invasions. *Mobile Genetic Elements* 2013; 3:e23920; <http://dx.doi.org/10.4161/mge.23920>

*Correspondence to: John F.Y. Brookfield;
Email: John.Brookfield@nottingham.ac.uk

Commentary to: Hellen EHB, Brookfield JFY. The diversity of class II transposable elements in mammalian genomes has arisen from ancestral phylogenetic splits during ancient waves of proliferation through the genome. *Mol Biol Evol* 2013; 30:100–8; PMID:22923465; <http://dx.doi.org/10.1093/molbev/mss206>

stable equilibria is at the very heart of evolutionary theory, in that the adaptation to the environment that is observed throughout the living world is explained through this state being a stable equilibrium state in the dynamic process that is evolution by natural selection.

It was in this tradition that one of us has sought to predict equilibrium states for families of transposable elements.^{4,5} But the expectation that systems will occupy the equilibrium states defined by their dynamic properties rests on the assumption that the parameters that determine the dynamic process change, if at all, on a timescale that is slow relative to the rate at which the dynamic process that they define moves toward its equilibrium state. For transposable elements, the dynamics of their genomic spread and copy number changes are slow, and, throughout these slow processes, the state parameters of their evolutionary processes are themselves changing due to host adaptation.⁶

No Equilibria in Mammalian Class II Elements

Previous work has indicated that there is a lack of homogeneity in the spread of transposable elements in mammalian genomes. The human genome project⁷ used genomic fossils to catalog the activities of class I and class II transposable elements at various times in the past and saw times when elements were active and other times when they were relatively quiescent. But, as always, the picture of the past that genomic fossils supply is incomplete, since comparisons can be used only for those elements that have left descendants today.

We have investigated the spread of transposon (Class II) families through mammalian genomes by estimating times to common ancestry of the elements. Following an initial study as a proof of principle using the Golem family of transposons⁸ a more extensive analysis examined the DNA sequence variation between sequence copies of 29 families of class II transposons from mammalian genomes.⁹

The idea behind our analysis is that, if we examine the collection of copies of a given transposable element family in a given genome, these elements have a Most Recent Common Ancestor (MRCA) at

some time in the past. If there is a process of turnover of the elements within a family, elements will be replicating to produce new elements, and other element copies will be lost. Over time, in such a continuous process of gain and loss, the element that is the MRCA of the family will move forward in time. This is the same process that is seen systematically at the population genetics level for alleles within a population of organisms. If we were to look at the collection of alleles present at a human autosomal locus today, then these would have a common ancestor at some time in the past— for humans the average time to common ancestry is approximately 800,000 y ago. But if we were to go back 400,000 y to the population from which we are descended and in which the ancestral alleles of the present variation existed, the time to common ancestry of all the alleles in such a population would not be 400,000 y earlier (i.e., 800,000 y before now). The time would be much more likely to be around 800,000 y earlier than our 400,000 y old sampling, since the alleles present 400,000 y ago that have had the good fortune to still have descendants today would be just a subset of the variation that existed at that time.

This simple picture of element turnover is, of course, just one end of a continuum, when applied to transposable elements. It could be that, as time proceeds, elements are still being created and lost, and the MRCA is moving forward, but not as fast as is time itself, such that the time to MRCA is systematically lengthening.

For each of the element families from the human genome, the time to the most recent common ancestor of the copies of the family was estimated, using the software BEAST,¹⁰ more typically used for population genetic diversity. Any estimate of the time of common ancestry using sequence diversity requires some estimate of the rate of evolutionary change and in our analysis this was derived by looking at the sequence divergence between human and chimpanzee orthologs (members of the same sequence family at the same chromosomal location and derived from a unique insertion event in an ancestor of the human and chimpanzee). In a further analysis, we replaced the chimpanzee sequences with orang utan orthologs. The

times to common ancestry of the elements within a family differed from element family to element family. The key question was the relationship between this predicted time to common ancestry of element copies from the human genome and the time when the element first proliferated through the genome of an ancestor. For human element families that were also found outside the primates, it was possible to estimate the time to common ancestry of elements in two other genomes, those of the cow and the dog, with the evolutionary rate being estimated by finding element orthologs in the pig and in the cat and panda, respectively. The consistent finding was that, approximately, the estimated times to common ancestry for a given family, from the three types of comparison, using Primates, Artiodactyls and Carnivores, were approximately the same and occurred before the divergence of these mammalian orders. This result is not what would be expected if the elements within a family had undergone a process of turnover that had continued after the divergence of the ancestors of the three orders studied. It appeared, on the contrary, that the time to common ancestry of elements sampled from the three descendant orders were sufficiently similar that all were reflecting a single process of genomic proliferation in a shared ancestor. But what is the relationship between the time to common ancestry of the element copies within a family and the time when that family first proliferated through its mammalian host genome? To investigate the latter, we observed the distribution of the presence and absence of each element family across mammalian species and orders and, from these data, we estimated the time when the element family first resided in an ancestral mammalian genome. In doing this, we assumed that if an element family is seen, albeit possibly being sparsely distributed, in two mammalian orders, then that family would have existed in the genome of the mammalian group that was the ancestor of those orders. Thus, we explicitly modeled a patchy distribution of a family as being the result of stochastic losses of the family in some descendant lineages of ancestral groups that possessed it, rather than the patchy distribution reflecting horizontal

transfers between groups. This choice to exclude horizontal transfer appears arbitrary, but turned out to be justified by the strong correlation that we found between the age of the most recent common ancestor of elements from a transposon family and the age of the first mammalian host inferred for the family, implying that patchy distributions are indeed the result of family losses (or incomplete genome information) and not the result of horizontal transfers. It is hard to assess whether such wholesale losses (one might call these “extirpations” to distinguish them from the “extinctions” where it is only transposition activity that is lost) are likely. An analysis, using repeat masker,¹¹ of the class II transposon family MER63B’s numbers in primate, rodent, carnivore and artiodactyls shows that this may be possible. Here, we find 2399 elements in the human genome, 2121 elements in the dog genome, 2141 elements in the cow genome, 1733 in the squirrel genome, but only 50 in the rat genome, the majority of which are partial fragments. The reduced number of elements in the rat genome may be due to the increased rate of rearrangements found in murid rodents compared to other orders of mammals¹² and, as such, may be a special case. It does, however, suggest that if an ancestral number of thousands can diminish to 50, complete extirpation will also be possible. It is likely that extirpation of already extinct elements will be by stochastic loss, possibly accompanied by very weak selection for reduction in genome size. Host resistance to transposable elements, such as that arising from the RNAi and Piwi pathways,^{6,13} counters the expression and activity of transposable elements, rather than removing their inactive DNA remains.

If the MRCA of copies of a given class II element family in the human genome existed soon after the initial occupation of the genome by this family, and similar times are estimated for dog and cow elements’ MRCAs, the implication is that, if a phylogenetic tree of elements sampled from all three orders were constructed, the root of the tree may be the MRCA for all three sets of sequences, and we attempted to see if this was generally the case. Unfortunately, however, the roots of trees constructed in this way were very poorly

resolved, with weak bootstrap support for all bifurcations in the early topology.

Inferring the Past

The conclusion from our data was that the model of constant turnover of elements, such that the element diversity might indeed represent an equilibrium state in a dynamic model, is wrong. Rather, the diversity of the elements has a historical cause and reflects a proliferation of the elements soon after they were first introduced into their host genome. We referred to this scenario as a “life cycle” model to distinguish it from a model of continuing turnover.

Our results and analysis, inferring the past of transposable elements, form an interesting contrast with the way data of genetic diversity between sequences is used in population genetics. Population geneticists use information about genetic variation today to reconstruct species population sizes in the past. Their ability to do so is based on the application of a set of population genetics tools called the coalescent. Theories of the coalescent are based on the idea that if we sample a set of sequences of shared ancestry today, these sequences are connected by a phylogeny. This phylogeny has mathematical properties that are dictated by coalescent theory—specifically, the expected times of changes in the number of ancestral lineages are determined by the effective population size at the time. Thus, the mitochondrial DNA variation in humans throws light on changes in human population numbers and population movements over the last two hundred thousand years. A most recent ancestor of the mitochondrial DNAs is defined and confusingly referred to as the mitochondrial “Eve”.¹⁴ This is confusing because, of course, this woman, possessing the common ancestral mitochondrial DNA, had a mother, and there would have been many thousands of other females in the population, each with their own, diverse, mitochondrial DNAs. The specialness of the time and identity of this particular female is defined only by stochastic events tens or hundreds of thousands of years after she lived. Since, at the time of “Eve” only one mitochondrial DNA ancestral to any modern DNAs

existed, present day mitochondrial variation is silent about any even earlier human demographic changes.

For transposable elements under the “life-cycle” model, we are faced with a different situation. In its strongest form, as we trace ancestry back to a single common ancestor, we are indeed tracing back to a historically special event. So here, our ignorance of earlier events arises not just from the diversity at the present day ceasing to reveal anything about those earlier events. Rather, those events were different in kind and represent either introduction of the family into the genome by horizontal transfer, or the evolution of the family from progenitors so diverged that their descendants can no longer be recognized.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We would like to thank the BBSRC (research grant reference: BB/H009884/1) for funding this study.

References

1. Pace JK 2nd, Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res* 2007; 17:422-32; PMID:17339369; <http://dx.doi.org/10.1101/gr.5826307>
2. Relethford JH. Genetic evidence and the modern human origins debate. *Heredity (Edinb)* 2008; 100:555-63; PMID:18322457; <http://dx.doi.org/10.1038/hdy.2008.14>
3. Eller E, Hawks J, Relethford JH. Local extinction and recolonization, species effective population size, and modern human origins. *Hum Biol* 2004; 76:689-709; PMID:15757241; <http://dx.doi.org/10.1353/hub.2005.0006>
4. Brookfield JFY. A model for DNA sequence evolution within transposable element families. *Genetics* 1986; 112:393-407; PMID:3000869
5. Brookfield JFY, Johnson LJ. The evolution of mobile DNAs: when will transposons create phylogenies that look as if there is a master gene? *Genetics* 2006; 173:1115-23; PMID:16790583; <http://dx.doi.org/10.1534/genetics.104.027219>
6. Thomson T, Lin H. The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annu Rev Cell Dev Biol* 2009; 25:355-76; PMID:19575643; <http://dx.doi.org/10.1146/annurev.cellbio.24.110707.175327>
7. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001; 409:860-921; PMID:11237011; <http://dx.doi.org/10.1038/35057062>
8. Hellen EHB, Brookfield JFY. Investigation of the origin and spread of a Mammalian transposable element based on current sequence diversity. *J Mol Evol* 2011; 73:287-96; PMID:22222953; <http://dx.doi.org/10.1007/s00239-011-9475-y>

9. Hellen EHB, Brookfield JFY. The diversity of class II transposable elements in mammalian genomes has arisen from ancestral phylogenetic splits during ancient waves of proliferation through the genome. *Mol Biol Evol* 2013; 30:100-8; PMID:22923465; <http://dx.doi.org/10.1093/molbev/mss206>
10. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007; 7:214; PMID:17996036; <http://dx.doi.org/10.1186/1471-2148-7-214>
11. Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0. 1996-2010 <http://www.repeatmasker.org>
12. Bourque G, Pevzner PA, Tesler G. Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome Res* 2004; 14:507-16; PMID:15059991; <http://dx.doi.org/10.1101/gr.1975204>
13. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 2007; 318:761-4; PMID:17975059; <http://dx.doi.org/10.1126/science.1146484>
14. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature* 1987; 325:31-6; PMID:3025745; <http://dx.doi.org/10.1038/325031a0>