Evolution of Genome Size

Stephen I Wright, *Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada*

Advanced article

Article Contents

Online posting date: 16th January 2017

The size of the genome represents one of the most strikingly variable yet poorly understood traits in eukaryotic organisms. Genomic comparisons suggest that most properties of genomes tend to increase with genome size, but the fraction of the genome that comprises transposable elements (TEs) and other repetitive elements tends to increase disproportionately. Neutral, nearly neutral and adaptive models for the evolution of genome size have been proposed, but strong evidence for the general importance of any of these models remains lacking, and improved understanding of factors driving the activity of TEs should also be considered. Fine-scale investigation of the mutational and population-genetic properties of both small and large insertions and deletions should help advance our understanding of how and why genome size evolution has occurred.

Introduction

Genomes show remarkable variation in size, often even among closely related species (Bennett and Leitch, 2011; Gregory, 2005). From the earliest studies of this variation, it was clear that, beyond the broadest differences in genome size between prokaryotes and eukaryotes, this variation does not generally correlate well with the number of genes or with organismal complexity (Thomas, 1971). Why species should vary across orders of magnitude in the amount of heritable material transmitted across generations has been among the most puzzling and elusive questions for evolutionary and molecular biologists in the last century. While considerable progress has been made in the characterisation of the extent of genome size variation, the dominant evolutionary processes driving genome size evolution remain subject to considerable debate. Large-scale genome sequencing is enabling new insights into both the proximate causes and evolutionary forces governing genome size differences.

eLS subject area: Evolution & Diversity of Life

How to cite:
Wright, Stephen I (January 2017) Evolution of Genome Size. In: eLS. John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0023983

How Much Variation Is There?

Because determining the amount of DNA (deoxyribonucleic acid) in a cell has been much more straightforward and cheaper than determining the complete sequence of genomes, characterisation of the 'C-value' (the amount of DNA in a haploid cell) began much earlier than whole genome sequencing, and in fact the measurement of DNA content per cell and the recognition of its constancy within an organism provided important evidence that DNA was the basis of inheritance (Gregory, 2005). Given this extensive history and ease of quantification, databases of genome sizes now contain information from many thousands of species (Bennett and Leitch, 2011; Gregory, 2016; Islas *et al*., 2004). These estimates reveal striking variation; flowering plants, for example range in genome size from roughly 60 MB in the carnivorous plant *Genlisea margaretae* to *Paris japonica*, the organism with the largest known genome, at roughly 150 GB (Bennett and Leitch, 2011). Studies within families and genera typically reveal less variation but still show considerable change across short evolutionary distances (Greilhuber *et al*., 2006), and even within species (Long *et al*., 2013).

What Types of DNA Drive Genome Size Variation?

Genome size differences can be determined by several major properties. The most obvious source of genome size variation is polyploidy; organisms with more than two sets of chromosomes will have higher DNA content simply because of increased numbers of copies of genes. While the effect of ploidy on genome

size differences is often corrected for by controlling for ploidy in the estimation (Soltis *et al*., 2003), there are a number of reasons why this does not fully resolve the issue. First, it has become clear that in many organisms, ancient whole genome duplication events have occurred that may no longer be obvious by chromosome counts (Soltis and Soltis, 2016). Following polyploidisation, organisms typically undergo a process of duplicate gene loss (Leitch and Bennett, 2004; De Smet *et al*., 2013; Lynch and Conery, 2000), in a process known as 'diploidisation', which may often also include chromosome rearrangement events. Because these rearrangements will make it difficult to identify polyploids based on chromosome number, many ancient whole genome duplication events can only be uncovered through large-scale genome or transcriptome sequencing. Thus, unknown ancient polyploidisation can still contribute to genome size variation, and the extent to which the time since the last whole genome duplication contributes to genome size variation is not yet resolved (Lysak *et al*., 2009). Furthermore, by increasing DNA content, polyploids are expected (and often observed) to experience similar phenotypic effects on the organism as other genome size increases, namely larger cell sizes (Tsukaya, 2013) and possible increases in development time (Gregory, 2002). This means that for some models of genome size evolution, polyploidy has similar effects as other mechanisms mediating genome size evolution. **See also**: **Polyploidy and Paralogous Chromosome Regions**; **Polyploidy**

Beyond differences in ploidy, genomic sequence comparisons have helped assess the major factors that drive genome size differences among species. On the one hand, recent genome sequence comparisons suggest that many factors tend to correlate with genome size, including copy numbers of repetitive elements, the length of introns and, contrary to previous suggestions, even the number of genes (Elliott and Gregory, 2015; Hu *et al*., 2011). However, the composition of large genomes is typically very different from that of small genomes (**Figure 1**). In particular, transposable elements (TEs) and other repetitive sequences tend to dominate sequence differences among species (Tenaillon *et al*., 2010; Chalopin *et al*., 2015), and the proportion of the genome that is protein coding is strongly negatively correlated with genome size. However, some cases of genome size increase appear to be predominantly driven by repeat classes other than TEs, such as the number of ribosomal DNA gene copies (Long *et al*., 2013) or the abundance of centromeric repeats (Slotte *et al*., 2013). **See also**: **Repetitive Elements: Bioinformatic Identification, Classification and Analysis**; **Microsatellites**

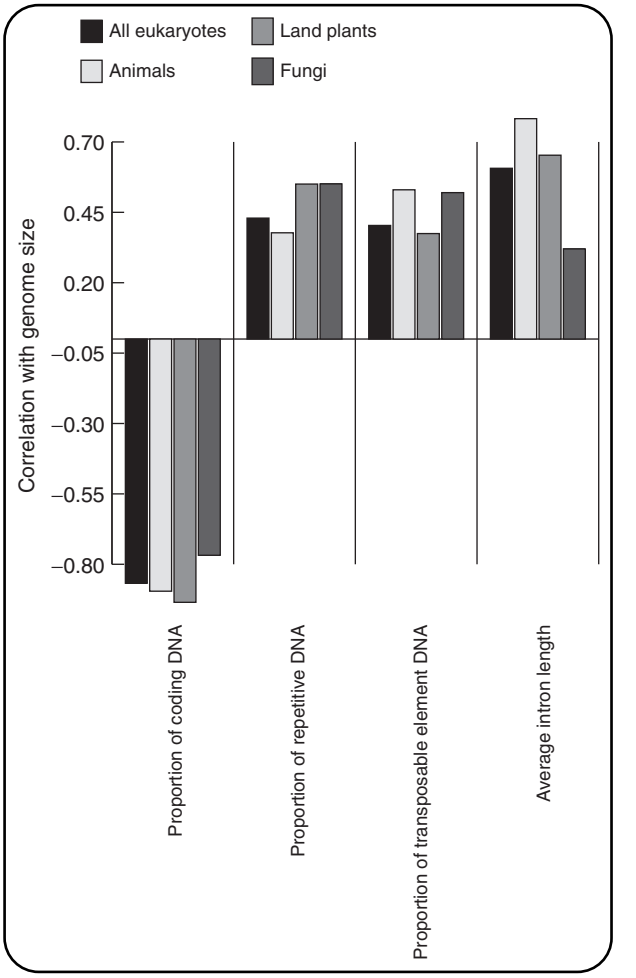

Figure 1 Phylogenetically corrected correlations (independent contrasts) between total genome size and genomic properties. From Elliott and Gregory (2015).

Neutral Model

Perhaps the simplest explanation for genome size evolution is that differences across taxa in the profile and relative rate of insertion and deletion drive species to different genome sizes. If some taxa experience much higher deletion rates, or larger average deletion sizes than others, they could converge on a smaller genome than other taxa. This hypothesis generates some clear predictions; species with larger genomes should show a distinct mutational spectrum of insertions and deletions that leads to a higher rate of DNA removal (and/or lower rate of accumulation) than species with smaller genomes.

Consistent with the hypothesis that species differ considerably in their relative rates of insertion and deletion, an experimental study examining the repair of double-strand breaks in *Arabidopsis* and tobacco highlighted that species may differ qualitatively in the length distribution of deletions, as well as whether or not double-strand breaks are accompanied by insertions (Kirik *et al*., 2000). This highlights that species differences in the rate and spectrum of insertion/deletion events may play an important role in driving the evolution of genome size. **See also**: **Mutational Biases**

Early analyses testing this possibility aimed at identifying the spectrum of small insertion and deletion events at neutral genomic regions using between-species comparisons (Petrov *et al*., 1996, 2000; Bensasson *et al*., 2001). These studies indeed suggested that the rate of deletion relative to insertion may vary across species, where species such as *Drosophila* showed

evidence for a higher rate of deletion relative to insertion in comparison with species with larger genomes such as humans. However, using between-species comparisons, it is difficult to completely rule out a role for natural selection in the fixation of these indel events (Charlesworth, 1996). With the lowering costs of genome sequencing, direct estimates of genome-wide mutation rate are becoming much more feasible, enabling potentially unbiased estimates of rates and patterns of insertion and deletion. Some of these direct estimates have confirmed results from interspecific comparisons; for example, a deletion bias in *Drosophila melanogaster* has been observed from direct observation of spontaneous short insertion–deletion events (Haag-Liautard *et al*., 2007; Keightley *et al*., 2009; Schrider *et al*., 2013). However, in other cases, very different profiles have been observed from direct estimation; *Caenorhabditis elegans*, for example, shows a bias towards insertion events from observations of spontaneous mutations, in contrast with patterns of substitution observed at pseudogenes (Denver *et al*., 2004). Furthermore, factoring in larger duplication events, the *Drosophila* genome also appears to show a net gain in DNA each generation, rather than loss (Schrider *et al*., 2013). More generally, direct estimates to date do not suggest a generally positive correlation between genome size and the net gain/loss of sequence caused by spontaneous short insertions and deletions (**Figure 2**). However, few estimates are available to date from species with very large genomes, and these estimates generally focus on small indels due to the limits of short-read sequencing technologies, so the spectrum for rarer but larger insertion/deletion events is less well characterised. Given their length, the potential for small indels to drive rapid genome size differences seems limited (Gregory, 2003).

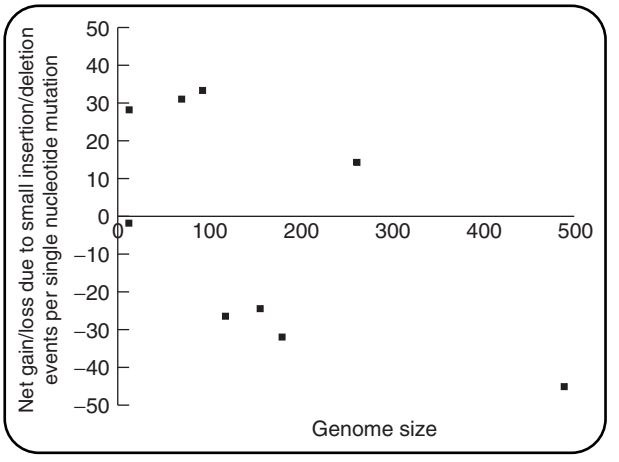

Figure 2 Relation between amount of DNA (deoxyribonucleic acid) gained/lost due to small indels (calculated as the total size of small insertions minus the total size of small deletions) per single nucleotide mutation and genome size, from direct estimation of insertion and deletion events in mutation studies. Estimates from *Caenorhabditis elegans* (Denver *et al.*, 2004), *Drosophila melanogaster* (Schrider *et al.*, 2013), *Arabidopsis thaliana* (Yang *et al.*, 2015), *Apis mellifera* (Yang *et al.*, 2015), *Oryza sativa* (Yang *et al.*, 2015), *Chlamydomonas rheinhardtii* (Ness *et al.*, 2015), *Schizosaccharomyces pombe* (Farlow *et al.*, 2015), *Saccharomyces cerevisiae* (Zhu *et al.*, 2014) and *Paramecium tetraurelia* (Sung *et al.*, 2012).

Nearly Neutral Model

Many insertion/deletion events may have slightly deleterious effects on organismal fitness on average rather than being neutral and therefore they are subject to effects of both genetic drift and natural selection. In particular, for mutations subject to weak natural selection, slightly deleterious insertion/deletion events may spread through populations due to genetic drift. Under this model, species' differences in effective population size, which controls the relative importance of drift versus selection, can have a major impact on the outcome. Small populations may be more likely to fix slightly deleterious insertions in comparison with larger populations, driving their genomes larger (Lynch and Conery, 2003). **See also**: **Molecular Evolution: Nearly Neutral Theory**

Population genetic analyses in *Drosophila* do suggest that small indels may on average be slightly deleterious (Leushkin *et al*., 2013), consistent with this nearly neutral model of genome size evolution. Furthermore, large-scale comparative studies (Lynch and Conery, 2003) have shown that effective population size correlates with genome size across a large taxonomic scale. However, studies on smaller taxonomic scales and/or that use comparative phylogenetic approaches to control for evolutionary relationships have often found weaker or no association between genome size and effective population size (Whitney *et al*., 2010; Whitney and Garland, 2010; Ai *et al*., 2012). Furthermore, recent population genomic analyses in *Drosophila* suggest that small deletions may be more deleterious than small insertions (Leushkin *et al*., 2013), making it unclear whether less efficient selection will necessarily drive genome increase. More investigation of the distribution of fitness effects of insertions and deletions of varying sizes are needed to better understand the interplay between changes in effective population size and genome size evolution.

Adaptive Hypotheses

While genomes are typically considered the 'information content' in the cell without direct phenotypic effects, there is a clear correlation between DNA content and cell size (Tsukaya, 2013), suggesting that differences in genome size may have direct phenotypic effects that could be selected on. Genome size across species correlates with numerous traits including development time and body size (Gregory, 2003). While correlative, such patterns raise the possibility of directional selection in some species on genome size. Weedy plants (Bennett *et al*., 2008) and animals with flight (Wright *et al*., 2014), for example, may experience selection for rapid cell division and/or small cell size, driving genomes to shrink over evolutionary time. Correlations between genome size and flower size in *Silene* (Meagher and Vassiliadis, 2005), and clines in genome size associated with environmental variables (Rayburn and Auger, 1990), provide further evidence that the phenotypic effects of genome size may be under selection. However, given that environmental variables can also influence the concentration of compounds that can affect genome size estimation, it is crucial to control for possible technical artefacts when assessing environmental effects (Bennett *et al*., 2008).

Alternatively, there may be indirect selection on genome size as a result of phenotypic effects of copy number of a specific genome compartment. For example, in *Arabidopsis* the number of rDNA copies varies between Northern and Southern Sweden, and this is the major predictor of genome size differences across these populations (Long *et al*., 2013). It is possible that this difference in rDNA content has been driven by environmental selection on thermal regulation, with higher rDNA copy number then incidentally driving larger genome size.

Transposable Element Evolution

While neutral and nearly neutral hypotheses typically treat all insertion–deletion events as equivalent in their evolutionary dynamics, differences across species in TE copy number are typically the most dominant determinant of genome size differences (Tenaillon *et al*., 2010; Chalopin *et al*., 2015) and so consideration of the specific factors driving TE evolution may be important for a better understanding of the evolution of genome size (**Figure 3**). **See also**: **Transposons in Eukaryotes (Part B): Genomic Consequences of Transposition**

As TEs are self-replicating genetic elements, their success and abundance can be governed by distinct processes. For example, sexual reproduction and outcrossing enhance the spread of TEs through populations, while a lack of sexual reproduction and high rates of self-fertilisation can prevent their spread (Hickey, 1982; Wright *et al*., 2008). Highly self-fertilising and asexual species, which typically have strong reductions in effective population size and are expected to accumulate slightly deleterious mutations, may thus experience a loss rather than gain of TEs, and a reduction in genome size. Furthermore, species with smaller effective population sizes may also be more likely to experience stochastic loss of active TE families, leading to higher rates of transpositional activity in populations with larger effective population sizes (Le Rouzic *et al*., 2007), again suggesting that reduced effective population sizes may not always drive genome expansion.

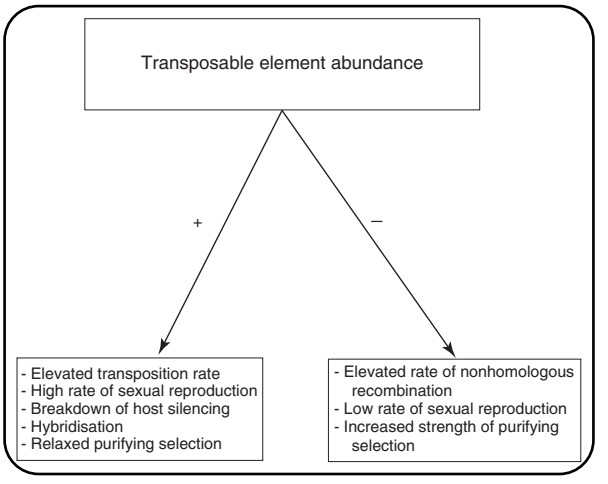

Figure 3 Factors driving changes in transposable element abundance.

Beyond the rate of sexual reproduction and effective population size, TE abundance can be determined by the balance between their increase due to transposition and their loss caused by forces acting to remove them (Charlesworth and Charlesworth, 1983). Note that TE removal can be mediated not only by excision and deletion events but also by the action of natural selection against the deleterious effects of insertions. Thus, between-species differences in the rate of transposition, deletion and/or the strength of natural selection against insertions can all contribute to changes in TE abundance and genome size.

Species' differences in the rate of transposition could be a key driver of genome size evolution; however, very few direct estimates of transposition rate exist (Nuzhdin and Mackay, 1994) and there is essentially no direct information on how they vary across species. There are several reasons for the absence of estimates of transposition rate. First, host-mediated TE silencing tends to be quite efficient, meaning that per generation transposition rates may be very low, making them difficult to observe, even though their rates may be sufficiently high to contribute to rapid genome size evolution given their size. Second, short-read genome sequencing technologies create a challenge for resolving the sequences of repetitive elements such as TEs, so large-scale identification of TE insertions from mutation accumulation experiments is difficult.

However, comparisons of the abundance of TEs and their inferred age distributions from high-quality reference genome sequences have been used to make inferences about shifts in transposition rate across species. Note, however, that such comparisons can reflect contrasts in not only the rate of transposition but also differences in the strength of natural selection against new insertions, as stronger selection will act to prevent old insertions from accumulating. Nevertheless, such studies frequently suggest strong differences between species and across element families in their history of activity (Chalopin *et al*., 2015), likely reflecting complex histories of coevolution between TEs and their hosts.

In some cases, consistent differences in TE copy number have been observed between closely related species, potentially due to differences in the efficacy of host silencing of TEs (Hu *et al*., 2011; Hollister *et al*., 2011). However, in other cases, large genome sizes tend to be associated primarily with the expansion of a small number of TE families (El Baidouri and Panaud, 2013). A growing number of cases of apparent bursts of transposition tied to genome size evolution have been observed (Piegu *et al*., 2006) suggesting that particular TE families may have evaded host silencing, driving increases in copy number and genome size. Furthermore, between-species hybridisation has been shown to be associated with increases in TE copy number in a number of taxa, consistent with the hypothesis that a history of TE-host coevolution can be broken down in hybrid crosses (Ågren and Wright, 2011), and that horizontal transfer may contribute to increases in transposition rate.

Similarly, changes in the strength of selection against TEs may be important factors governing genome evolution. Whole-genome duplication events, for example may relax selection against TE activity due to gene redundancy, and this could lead to the maintenance of large genome size in duplicated genomes even following the process of diploidisation. Evidence

for TE expansion following whole-genome duplication in maize is consistent with this hypothesis (Baucom *et al*., 2009), as is evidence for higher TE abundance in gene-rich regions in tetraploid *Capsella* (Ågren *et al*., 2016).

While changes in selection and transposition rate may be important factors governing genome size evolution, species differences in the rate of TE removal can also play an important role. For example, illegitimate recombination between the long-terminal repeats of retrotransposons can drive the loss of TE insertions (Ma *et al*., 2004), and it is possible that the rate of this removal differs across species. In this regard, it is important to note that some species with genomes that comprised large numbers of TEs do not have recently elevated rates of transposition; the massive genomes of conifers, for example, appear to have many ancient TE insertions, rather than a high rate of recent transposition (Nystedt *et al*., 2013), and therefore broad-scale differences in genome size may often reflect differences in DNA removal rather than recent increases in TE activity or a recent loss in the efficacy of natural selection.

Conclusion

Many factors can drive the evolution of genome size, and it is likely that genome expansion and contraction has multifaceted contributing factors. While important progress has been made in characterising the proximate causes of genome size evolution, the evolutionary processes that dominate remain mostly unresolved. However, new long-read sequencing technologies and data sets that utilise comparative genomics, population genomics and mutational studies should enable important advances in our ability to quantify the relative roles of mutation, selection, genetic drift and TE-host coevolution in the evolution of genome size.

Acknowledgements

I am grateful to Tyler Elliott, Rob Ness and T. Ryan Gregory for discussion and information about their published data sets. I also thank Tyler Kent, Julia Kreiner, Kim GIlbert, and Felix Beaudry for their comments on the manuscript.

References

Ågren JA and Wright SI (2011) Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? *Chromosome Research* **19** (6): 777–786.

Ågren JA, Huang HR and Wright SI (2016) Transposable element evolution in the allotetraploid *Capsella bursa-pastoris*. *American Journal of Botany* **103** (7): 1197–1202.

Ai B, Wang ZS and Ge S (2012) Genome size is not correlated with effective population size in the *Oryza* species. *Evolution* **66** (10): 3302–3310.

Baucom RS, Estill JC, Chaparro C, *et al.* (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genetics* **5** (11): e1000732.

Bennett MD, Leitch IJ and Hanson L (2008) DNA amounts in two samples of angiosperm weeds. *Annals of Botany* **82** (A): 121–134.

Bennett MD, Price HJ and Johnston JS (2008) Anthocyanin inhibits propidium iodide DNA fluorescence in *Euphorbia pulcherrima*: implications for genome size variation and flow cytometry. *Annals of Botany* **101** (6): 777–790.

Bennett MD and Leitch IJ (2011) Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. *Annals of Botany* **107** (3): 467–590.

Bensasson D, Petrov DA, Zhang DX, Hartl DL and Hewitt GM (2001) Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Molecular Biology and Evolution* **18** (2): 246–253.

Chalopin D, Naville M, Plard F, Galiana D and Volff JN (2015) Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biology and Evolution* **7** (2): 567–580.

Charlesworth B and Charlesworth D (1983) The population dynamics of transposable elements. *Genetics Research* **42**: 1–27.

Charlesworth B (1996) The changing sizes of genes. *Nature* **384** (6607): 315–316.

De Smet R, Adams KL, Vandepoele K, *et al.* (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 2898–2903.

Denver DR, Morris K, Lynch M and Thomas WK (2004) High mutation rate and Predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430** (7000): 679–682.

El Baidouri M and Panaud O (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biology and Evolution* **5** (5): 954–965.

Elliott TA and Gregory TR (2015) What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370** (1678): 20140331.

Farlow A, Long H, Arnoux S, *et al.* (2015) The Spontaneous mutation rate in the fission yeast *Schizosaccharomyces pombe*. *Genetics* **201** (2): 737–744.

Greilhuber J, Borsch T, Müller K, *et al.* (2006) Smallest angiosperm genomes found in lentibulariaceae, with chromosomes of bacterial size. *Plant Biology* **8** (6): 770–777.

Gregory TR (2002) Genome size and developmental complexity. *Genetica* **115** (1): 131–146.

Gregory TR (2003) Is small indel bias a determinant of genome size? *Trends in Genetics* **19** (9): 485–488.

Gregory TR (2005) Genome size evolution in animals. In: Gregory TR (ed) *The Evolution of the Genome*, pp. 3–87. San Diego: Elsevier.

Gregory TR (2016) *Animal Genome Size Database*. http://www.genomesize.com

Haag-Liautard C, Dorris M, Maside X, *et al.* (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* **445** (7123): 82–85.

Hickey DA (1982) Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101** (3–4): 519–531.

Hollister JD, Smith LM, Guo YL, *et al.* (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences of the United States of America* **108** (6): 2322–2327.

Hu TT, Pattyn P, Bakker EG, *et al.* (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics* **43** (5): 476–481.

Islas S, Becerra A, Luigi Luisi P and Lazcano A (2004) Comparative genomics and the gene complement of a minimal cell. *Origins of Life and Evolution of the Biosphere* **34**: 243–256.

Keightley PD, Trivedi U, Thomson M, *et al.* (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Research* **19** (7): 1195–1201.

Kirik A, Salomon S and Puchta H (2000) Species-specific double-strand break repair and genome evolution in plants. *EMBO Journal* **19** (20): 5562–5566.

Le Rouzic A, Boutin TS and Capy P (2007) Long-term evolution of transposable elements. *Proceedings of the National Academy of Sciences of the United States of America* **104** (49): 19375–19380.

Leitch IJ and Bennett MD (2004) Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* **82**: 651–663.

Leushkin EV, Bazykin GA and Kondrashov AS (2013) Strong mutational bias toward deletions in the *Drosophila melanogaster* genome is compensated by selection. *Genome Biology and Evolution* **5** (3): 514–524.

Long Q, Rabanal FA, Meng D, *et al.* (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics* **45** (8): 884–890.

Lynch M and Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290** (5494): 1151–1155.

Lynch M and Conery JS (2003) The origins of genome complexity. *Science* **302** (5649): 1401–1404.

Lysak MA, Koch MA, Beaulieu JM, Meister A and Leitch IJ (2009) The dynamic ups and downs of genome size evolution in Brassicaceae. *Molecular Biology and Evolution* **26** (1): 85–98.

Ma J, Devos KM and Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Research* **14** (5): 860–869.

Meagher TR and Vassiliadis C (2005) Phenotypic impacts of repetitive DNA in flowering plants. *New Phytologist* **168** (1): 71–80.

Ness RW, Morgan AD, Vasanthakrishnan RB, Colegrave N and Keightley PD (2015) Extensive de novo mutation rate variation between individuals and across the genome of *Chlamydomonas reinhardtii*. *Genome Research* **25** (11): 1739–1749.

Nuzhdin SV and Mackay TF (1994) Direct determination of retrotransposon Transposition rates in *Drosophila melanogaster*. *Genetics Research* **63** (2): 139–144.

Nystedt B, Street NR, Wetterbom A, *et al.* (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* **497** (7451): 579–584.

Petrov DA, Lozovskaya ER and Hartl DL (1996) High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384** (6607): 346–349.

Petrov DA, Sangster TA, Johnston JS, Hartl DL and Shaw KL (2000) Evidence for DNA loss as a determinant of genome size. *Science* **287** (5455): 1060–1062.

Piegu B, Guyot R, Picault N, *et al.* (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Research* **16** (10): 1262–1269.

Rayburn AL and Auger JA (1990) Genome size variation in *Zea mays* ssp. *mays* adapted to different altitudes. *Theoretical and Applied Genetics* **79** (4): 470–474.

Schrider DR, Houle D, Lynch M and Hahn MW (2013) Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* **194** (4): 937–954.

Slotte T, Hazzouri KM, Ågren JA, *et al.* (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genetics* **45** (7): 831–835.

Soltis DE, Soltis PS, Bennett MD and Leitch IJ (2003) Evolution of genome size in the angiosperms. *American Journal of Botany* **90** (11): 1596–1603.

Soltis PS and Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* **30**: 159–165.

Sung W, Tucker AE, Doak TG, *et al.* (2012) Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. *Proceedings of the National Academy of Sciences of the United States of America* **109** (47): 19339–19344.

Tenaillon MI, Hollister JD and Gaut BS (2010) A triptych of the evolution of plant transposable elements. *Trends in Plant Science* **15** (8): 471–478.

Thomas CA (1971) The genetic organization of chromosomes. *Annual Review of Genetics* **5**: 237–256.

Tsukaya H (2013) Does ploidy level directly control cell size? Counterevidence from Arabidopsis genetics. *PLoS One* **8** (12): e83729.

Whitney KD, Baack EJ, Hamrick JL, *et al.* (2010) A role for nonadaptive processes in plant genome size evolution? *Evolution* **64** (7): 2097–2109.

Whitney KD and Garland T Jr, (2010) Did genetic drift drive increases in genome complexity? *PLoS Genetics* **6** (8). pii: e1001080.

Wright SI, Ness RW, Foxe JP and Barrett SCH (2008) Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences* **169** (1): 105–118.

Wright NA, Gregory TR and Witt CC (2014) Metabolic 'engines' of flight drive genome size reduction in birds. *Proceedings of the Royal Society of London B: Biological Sciences* **281** (1779): 20132780.

Yang S, Wang L, Huang J, *et al.* (2015) Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* **523**: 463–467.

Zhu YO, Siegal ML, Hall DW and Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **111** (22): E2310–E2318.

Further Reading

Gregory TR (ed) (2005) *The Evolution of the Genome*. San Diego: Elsevier.

Lynch M (2007) *The Origin of Genome Architecture*. Sunderland, MA: Sinauer Associates Inc..