CellPress

Opinion

Is There an Upper Limit to Genome Size?

Oriane Hidalgo,[1,‡] Jaume Pellicer,[1,‡] Maarten Christenhusz,[2] Harald Schneider,[3,4] Andrew R. Leitch,[5] and Ilia J. Leitch[1,*]

At 50-fold the size of the human genome (3 Gb), the staggeringly huge genome of 147.3 Gb recently discovered in the fern *Tmesipteris obliqua* is comparable in size to those of the other plant and animal record-holders (i.e., *Paris japonica*, a flowering plant with a genome size of 148.8 Gb, and *Protopterus aethiopicus*, a lungfish with a genome of 130 Gb). The synthesis of available information on giant genomes suggests that the biological limit to genome size expansion in eukaryotes may have been reached. We propose several explanations for why the genomes of ferns, flowering plants, and lungfish, all of which have independently undergone dramatic increases in genome size through a variety of mechanisms, do not exceed 150 Gb.

The Extent of Genome Size Diversity Across Eukaryotes So Far

Eukaryotes exhibit an astonishing diversity of **genome sizes** (see Glossary), with data for over 15 000 species showing they vary in size by a factor of 64 000 ([1], Animal Genome Size Database, www.genomesize.com; Fungal Genome Size Database, www.zbi.ee/fungal-genomesize). The smallest genome so far reported is in the microsporidian *Encephalitozoon intestinalis*, which parasitizes a range of mammals including humans. Its genome comprises only 0.0023 Gb of DNA (= **1C-value**) and is considered to have reached the lower size limit for a fully functional eukaryotic genome [2]. At the other end of the scale, the largest genome reported using best-practice techniques is found in the flowering plant *Paris japonica* at 148.8 Gb [3] (Box 1). Given that one nucleotide is estimated to be ~0.34 nm in length, this diversity translates into only ~1.5 mm of DNA per somatic nucleus in *E. intestinalis* to ~100 m in *P. japonica*, with our own genome (1C = 3 Gb) measuring ~2 m. Such enormous variation and the lack of apparent correlation with organismal complexity have long caught the attention of biologists [4,5]. Although we now know the major contributors to genome size diversity are non-protein coding and often highly **repetitive DNA** sequences [6,7], why their amounts vary so much remains enigmatic.

Despite such diversity, species possessing enormous genomes are the exception because most eukaryotes possess small or very small genomes (Figure 1). Indeed, eukaryotes with genomes larger than 100 Gb are currently known in only 10 species (corresponding to 0.09% of species for which genome size data are available), belonging to only five eukaryotic orders: one in the ferns (Psilotales [8]), two in flowering plants (Liliales and Santalales [1]), and two in vertebrates [Lepidosireniformes (lungfish) and Urodela (salamanders); Animal Genome Size Database, www.genomesize.com)] (Table 1). Although there is increasing awareness that studying gigantic genomes is essential for providing a more complete picture of eukaryotic genome evolution [9], such species have been omitted from whole-genome sequencing projects owing to the analytical challenges that such large genomes pose. Insights into how their genomes are structured and function therefore remain limited (e.g., [10–13]).

Trends

Eukaryote genomes range in size by a factor of 64 000, but the parts comprising genes, regulatory regions, and other functional components typically account for only a small fraction of total genome size. The huge range largely arises from differences in the amount of repetitive, parasitic, and often selfishly accumulating DNA and their degraded products.

Despite the diversity, most species have small genomes, and those with giant genomes are the exception and belong to only a few phylogenetically distinct lineages.

The recent reports of giant genomes in flowering plants and ferns (the largest so far for any eukaryote) join the similarly giant genomes previously noted for lungfish and salamanders.

[1]Royal Botanic Gardens, Kew, Richmond TW9 3DS, UK
[2]Plant Gateway, 5 Talbot Street, Hertford SG13 7BX, UK
[3]Department of Life Sciences, Natural History Museum, London W7 5BD, UK
[4]Xishuangbanna Tropical Botanical Garden, Centre for Integrative Conservation, Chinese Academy of Sciences, Menglun, Mengla, Yunnan, PR China
[5]School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK
[‡]These authors contributed equally to this work.

*Correspondence:
i.leitch@kew.org (I.J. Leitch).

Box 1. What About the Giant Genomes Reported to Exist in Amoebae and Dinoflagellates – Are They Merely Technical Artifacts?

Previous C-value reports for some amoebae and dinoflagellates exceed those of *Paris japonica* and *Tmesipteris obliqua* (e.g., *Amoeba dubia*/1C = 685 Gb, *Lingulodinium polyedrum* [*Gonyaulax polyedra*]/1C = 195 Gb [47,48]). However, these estimates were not obtained using best-practice methodology (e.g., [49,50]). Potential technical issues which are likely to have compromised their accuracy are discussed below.

(i) Isolation of Nuclei. Both the report by Holm-Hansen of a giant genome in the dinoflagellate *Lingulodinium polyedrum* [48] and the measurements by Friz [47] of amoebae used whole cells and biochemical approaches that are now considered too unreliable for genome size determination [50]. Indeed, the measurements of Friz were questioned by Byers [51] whose own estimates for *Amoeba* were an order of magnitude smaller. In dinoflagellates, some very high genome size values were also based on analysing whole cells rather than isolated nuclei, and such values are highly variable (e.g., 112–268 Gb/cell in *Prorocentrum micans* [52,53]).

(ii) Selection of Calibration Standards. Most dinoflagellate measurements have used chicken red blood cells (2.2–2.9 Gb/1C) or *Arabidopsis thaliana* (0.16 Gb/1C) as calibration standards. Best-practice approaches recommend that the difference between the genome size of the target and standard should not exceed a factor of three because this can impact on the linearity of the response of the instrument [49]. While this is sometimes difficult to fulfil, the use of such small calibration standards for estimating very large genomes will undoubtedly introduce errors.

(iii) Selection of Fluorochromes. Some **fluorochromes** used in dinoflagellate studies are unreliable because they preferentially bind to AT-rich regions of the DNA (e.g., DAPI) and can artefactually increase genome size estimates by >40% [54]. In addition, the saline conditions in which dinoflagellates live can impact on the fluorescence of fluorochromes such as PicoGreen and SYTOX and hence on genome size values [52]. Further, the unusual helicoidal organization and DNA sequence composition of dinoflagellates [55] are considered likely to distort the quantitative binding of the fluorochrome to DNA – an essential requirement for robust genome size estimations.

(iv) Impact of Fixation/Drying of Cells. The use of fresh samples is essential for accurate genome size estimates because fixed or dried tissues can alter DNA staining properties and hence fluorescence intensity [49]. For example, dramatic differences in genome size estimates between live [11 Gb] and fixed [232 Gb] samples of *Prorocentrum micans* have been reported [52]. Nevertheless, dinoflagellate genome sizes are predominantly estimated using fixed material.

Overall, although it is clear that the genomes in these eukaryotic lineages are large, only by estimating their sizes using best-practice techniques will we know just how big their genomes are compared with *Paris japonica* and *Tmesipteris obliqua*.

Glossary

1C-value: the amount of DNA in the unreplicated gametic nucleus. C-values are usually reported in terms of mass in picograms (pg) or the number of base pairs in gigabasepairs (Gb); 1 pg = 0.978 Gb [45].

Genome size: the total amount of DNA in the nucleus of a cell. This can vary depending, for example, on the stage of the cell cycle and ploidy level [45].

Fluorochromes: chemicals that have the capacity to fluoresce when irradiated with light of the appropriate wavelength. The fluorochromes used to estimate genome size and ploidy by flow cytometry bind specifically and quantitatively to DNA, and include, for example, propidium iodide (PI), 4′,6-diamino-2-phenylindole (DAPI), SYTOX, and PicoGreen.

Polyploidy: the presence of more than two sets of chromosomes in the nucleus (genome); for example, tetraploid (4×) = possessing four sets of chromosomes.

Repetitive DNA: DNA sequence motifs that are repeated hundreds or thousands of times across the genome, including tandem repeats (e.g., DNA satellites) and dispersed repeats (e.g., DNA transposons and retroelements).

Recent Discovery of Genomic Gigantism Among Ferns

The most recent discovery of genomic gigantism is in the fern *Tmesipteris obliqua* (1C = 147.3 Gb [8]). This species belongs to the whisk-fern family, Psilotaceae [14], and provides further evidence that, although scarce, genomic gigantism is scattered across the tree of life. Tetraploid representatives of *Tmesipteris elongata* and *Psilotum nudum* (both $2n = 4\times = 208$ chromosomes) have 1C-values of 73.19 Gb and 71.08 Gb, respectively [15,16], suggesting that genome expansion in *T. obliqua* to gigantic proportions involved a **polyploidy** event (also termed whole-genome duplication), making it octoploid (i.e., $2n = 8\times = 416$). Ferns are reported to usually retain chromosome numbers following polyploidy [15], in contrast to large-scale chromosome restructuring that frequently reduces chromosome numbers to a diploid-like number in other polyploid lineages [17,18]. Such retention of chromosomes might explain why *T. obliqua* has far more chromosomes than the other eukaryotes with 1C > 100 Gb (Table 1). In addition, *T. obliqua* belongs to a family of ferns that already have large genomes because they possess substantially larger chromosomes than most other ferns [15], likely carrying large amounts of non-coding, repetitive DNA sequences. This suggests that its exceptional genome size has arisen from the combined effects of amplified repeats and polyploidy. Such a scenario is similar to that proposed for the octoploid *Paris japonica*, which belongs to a lineage of terrestrial geophytes that possesses some of the largest flowering plant chromosomes so far reported [19]. Indeed, all plants with genomes larger than 100 Gb are polyploid (Table 1), with the possible exception of the mistletoe *Viscum album* where ploidy level remains unknown [20]. Such observations contrast with those of giant genomes in animals

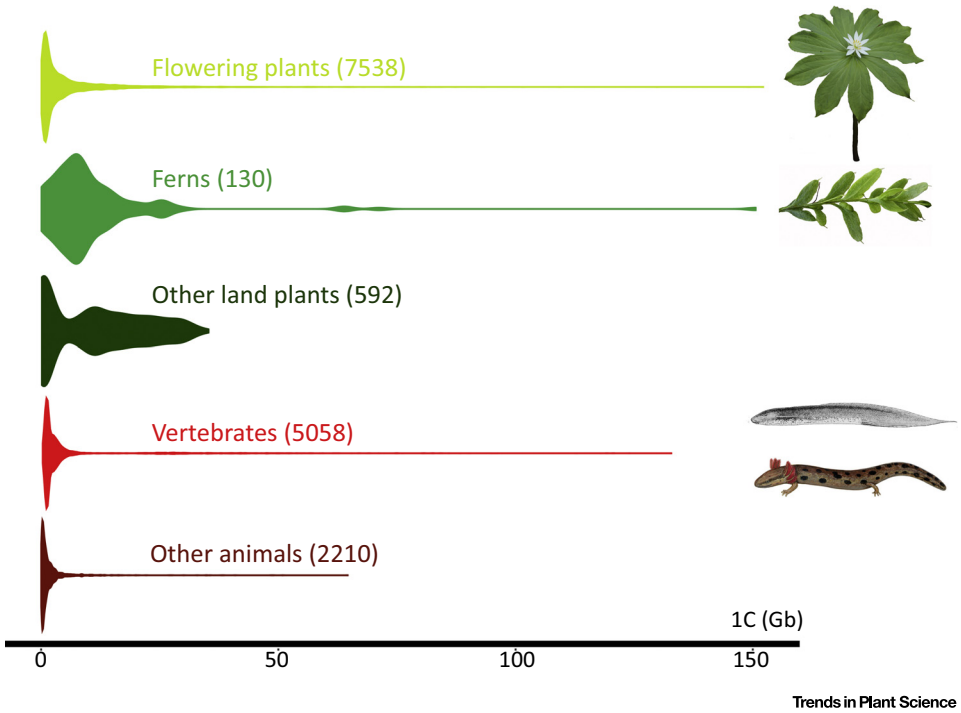

Figure 1. Violin Plots Showing the Frequency and Range of Genome Sizes in Different Eukaryote Groups, Together with Illustrations (Right) for Some of the Species with the Largest Genome Sizes. From top to bottom, *Paris japonica* (1C = 148.8 Gb), *Tmesipteris obliqua* (1C = 147.3 Gb), *Protopterus aethiopicus* (1C = 130.0 Gb), and *Necturus lewisi* (1C = 118.0 Gb). Data are taken from the Plant DNA C-values Database (http://data.kew.org/cvalues/), the Animal Genome Size Database (www.genomesize.com), and published data not yet included in these databases. Numbers in brackets following eukaryotic group names refer to the number of genome size estimates incorporated in each plot. Photographs from the top: Wikimedia commons/Maarten Christenhusz/Wikimedia commons/Joseph E. Trumpey.

where recent polyploidy is not involved. Instead, genome expansion in salamanders and lungfish has most likely been reached through the gradual accumulation of non-coding DNA sequences over a long period of evolutionary time, combined with their inactivation and decay but not elimination from the genome [12,13,21].

What Mechanisms Prevent Genomes from Uncontrolled Expansion?

Nowadays it is widely recognized that the mechanisms increasing genome size such as transposable element amplification and polyploidy (particularly in some plant lineages) are exceedingly common across many eukaryotes and may be crucial for generating evolutionary novelties [22,23]. Why then are there not more giant genomes? In most species studied to date it has been shown that processes leading to genome expansion are usually counter-balanced by recombination-based mechanisms (e.g., illegitimate and unequal homologous recombination) that result in genome downsizing [23]. The genome size of most organisms thus predominantly reflects the relative contributions of these two dynamic but opposing sets of processes. If so, the existence of giant genomes suggests that the genomic and epigenetic regulatory processes influencing genome size are operating differently, leading to the accumulation of DNA well beyond the usual limits [12,13,24]. Certainly, recent studies (e.g., genome-skimming approaches using high-throughput sequencing technologies) of giant genomes in animals and plants suggest that the composition, regulation, and evolution of their genomes may be following different trajectories compared to species possessing smaller genomes [10–13,25]. Nevertheless, whether this is due to changes at the genomic level (e.g., reduced recombination or altered epigenetic regulation [10,13,24,26]) and/or is driven

Table 1. Eukaryote Species with Genome Sizes (1C-values) Greater than 100 Gb (Arranged in Order of Decreasing Genome Size)[a,b]

Eukaryotic group Order – number of species recognised	Species	GS[c]	Method[d]	2n[e]	Ploidy[f]
Flowering plants Liliales (lilies and relatives) – 1712 spp.	*Paris japonica* (Japanese canopy plant)	148.8	FC:PI	40	8
Ferns Psilotales (whisk-ferns) – 12 spp.	*Tmesipteris obliqua* (long fork fern)	147.3	FC:PI	416	8
Vertebrates Lepidosireniformes (lungfish) – 5 spp.	*Protopterus aethiopicus* (marbled lungfish)	130.0	Fe	n.d.	n.d.
Flowering plants Liliales (lilies and relatives) – 1712 spp.	*Trillium* × *hagae* (Japanese hybrid wakerobin)	129.5	FC:PI	30	6
Vertebrates Lepidosireniformes (lungfish) – 5 spp.	*Lepidosiren paradoxa* (South American lungfish)	121.2	Fe	38	n.d.
Vertebrates Urodela (salamanders) – 655 spp.	*Necturus lewisi* (Neuse River waterdog)	118.0	Fe	38	n.d.
Vertebrates Urodela (salamanders) – 655 spp.	*Necturus punctatus* (dwarf waterdog)	116.6	Fe	38	n.d.
Flowering plants Liliales (lilies and relatives) – 1712 spp.	*Trillium rhombifolium* (Kamchatka wakerobin)	109.0	Fe	30	6
Flowering plants Liliales (lilies and relatives) – 1712 spp.	*Fritillaria elwesii* (green fritillary)	101.4	FC:PI	n.d.	n.d.
Flowering plants Santalales (mistletoes and sandalwoods) – 2373 spp.	*Viscum album* (European mistletoe)	100.6	FC:PI	20	n.d.

[a]Abbreviations: n.d., not determined or unclear.
[b]Data taken from the Plant DNA C-values Database (http://data.kew.org/cvalues/), the Animal Genome Size Database (www.genomesize.com), and Hidalgo *et al.* [8].
[c]Genome size (1C-value, Gb).
[d]Method used to estimate genome size: Fe, Feulgen microdensitometry; FC:PI, flow cytometry using the fluorochrome propidium iodide.
[e]Chromosome number.
[f]Ploidy level (x).

by a relaxation of selection pressures against giant genomes (e.g., in some geophytic [27,28], epiphytic, and parasitic plants [8]) remains unclear.

Why Might There be an Upper Limit for Genome Size?

Despite years of intense genome size prospecting that has generated records for over 15 000 animals and plants, the number of species with truly giant genomes remains negligible. It is noteworthy that those species of ferns, flowering plants, and vertebrates, each with very different life strategies, evolutionary histories, and relationships, have independently undergone such extensive genome expansions but stopped at relatively similar giant genome sizes. It is therefore tempting to speculate that ~150 Gb may be a biological upper limit for genome size – if so, why?

As our understanding of the evolution of eukaryotic genomes continues to expand, several explanations may contribute to that upper limit, either acting together or alone.

(i) The biochemical and energy costs associated with maintaining a functioning genome much over about 150 Gb are perhaps simply too great to be handled efficiently. Certainly, the elemental costs (particularly nitrogen and phosphorus) associated with copying and transcribing the DNA, and synthesizing sufficient numbers of histones to package the genome, will be substantial [29], as will be the energetic costs associated with regulating the activity of non-coding DNA sequences such as transposable elements [30].

(ii) There are also likely to be considerable energy costs associated with sustaining genome integrity in the face of ongoing DNA damage from both external and internal sources. Even in the human genome, at only 3 Gb, it is estimated that there are >10 000 endogenous DNA damage events per cell per day and that the repair of a single double-stranded break requires more than 10 000 ATPs [31]. Extrapolating to the upper end of the genome size scale, cost – both in terms of direct energy requirements (ATPs) and those associated with synthesizing sufficient amounts of the protein repair machinery – will no doubt escalate substantially, and above 150 Gb may simply be too great a cost to maintain the integrity of a viable genome.

(iii) Geometric constraints (arising from a decreasing surface area to volume ratio of the cell as genome size increases [32]) and timing constraints (arising from the longer duration of mitosis and meiosis as genome size increases [33–35]) may also play a role in setting the upper limits of genome size via their impact on key cellular processes such as those involving membrane transport and gas exchange [32], as well as their broader impact on various growth and ecological parameters [36–39].

(iv) Finally, evolutionary constraints on giant genomes may contribute to limiting genome size expansion much beyond 150 Gb. Recent studies have shown that, as genomes expand, DNA becomes increasingly partitioned into islands of gene space separated by large seas of epigenetically silenced, non-coding repetitive DNA [40]. One consequence of this arrangement is that repetitive DNAs, which can be removed by recombination-based processes in smaller genomes, become increasingly locked down into highly condensed chromatin where they can survive for millions of years, gradually mutating towards long tracks of unique/low-copy DNA sequences [10,12,41]. Thus, paradoxically, the gene space of giant genomes may be less affected by surrounding repeats than it is in species with small genomes [40]. If so, gene expression diversity upon which selection can act may simply become too limited for giant genomes to survive in the face of environmental or ecological change.

Concluding Remarks and Future Perspectives

To date, and thanks to the advent of high-throughput sequencing technologies, it is now possible to generate representative amounts of sequence data to delve into the genomic and epigenetic mechanisms responsible for the evolution of genomes of all sizes. Certainly our knowledge of the composition and epigenetic control of giant genomes has increased in recent years, and these studies have started to hint that there may well be distinctive differences in the way that giant genomes are organized, function, and evolve compared to their relatives with a smaller genome size. For example, analyses of the giant genomes in salamanders [13], lungfish [12], and *Fritillara* [10] show that they are dominated by highly heterogeneous degenerate repeats, suggesting slower rates of recombination-based elimination of repetitive DNA. In addition, relationships between genome size and cell size in distinct eukaryotic groups [42], and cell-cycle times in angiosperms [43], follow substantially different regression slopes in species with large genomes compared to those with smaller genomes. Furthermore, substitution rates have been shown to be lower in the giant genomes of salamanders compared to frogs which have smaller genomes [11]. Nevertheless, despite these tantalizing insights, there are still significant gaps in our knowledge of giant genomes (see Outstanding Questions). To tackle these, future research needs to build up a more comprehensive view of the genomic and epigenetic landscape across the diversity of genome sizes encountered in eukaryotes. This will enable us to target lineages of interest and hence identify through comparative analyses which genomic processes and mechanisms are unique to specific groups, and which are universal attributes of giant genomes.

Outstanding Questions

What is the extant diversity of genome sizes in eukaryotes? Current understanding of nuclear DNA contents across eukaryotes has revealed a staggering diversity, but data are relatively scarce or missing for many lineages.

What is the tempo, timing, and rate of genome expansion in lineages containing species with giant genomes?

Is there an ecological cost for a large genome, especially in terms of the resources required (e.g., nitrogen and phosphorus) to build them?

Why are some groups of plants and animals more prone to genome size expansion than others? Does genomic gigantism impose constraints on their ability to diversify and speciate?

Can we push the upper limits of genome size still further by artificially increasing the ploidy level of species with giant genomes?

Given the established relationship between genome size and cell size, and the proposed limit on chromosome size depending on cell size [44], how does chromosome size scale with genome size?

To what extent do population genetic processes such as genetic drift versus selection contribute to the diversity of genome sizes encountered?

How distinctive are giant genomes in terms of how they function, are regulated, and evolve compared to species with smaller genomes?

Species with giant genomes are typically rare; are they less resilient to environmental change because of their large genomes? To what extent does environmental stress such as climate change contribute to genome size diversity?

Acknowledgments

The authors thank Richard Bateman, Quentin Cronk, Steven Dodsworth, Paula Rudall, Sonja Siljak-Yakovlev, Martin Lysak, and two anonymous referees for useful discussion and/or helpful comments, and Joseph E. Trumpey for the illustration of *Necturus lewisi*. O.H. was supported by the Marie Skłodowska Curie Action Individual Fellowship program (CAPITULA, grant agreement 657918).

References

1. Garcia, S. *et al.* (2014) Recent updates and developments to plant genome size databases. *Nucleic Acids Res* 42, D1159–D1166
2. Corradi, N. *et al.* (2010) The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat. Commun.* 1, 77
3. Pellicer, J. *et al.* (2010) The largest eukaryotic genome of them all? *Bot. J. Linn. Soc.* 164, 10–15
4. Thomas, C.A. (1971) The genetic organization of chromosomes. *Annu. Rev. Genet.* 5, 237–256
5. Gregory, T.R. (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev.* 76, 65–101
6. Grover, C. and Wendel, J.F. (2010) Recent insights into mechanisms of genome size change in plants. *J. Bot.* 2010, 382732
7. Canapa, A. *et al.* (2015) Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* 147, 217–239
8. Hidalgo, O. *et al.* (2017) Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica*. *Bot. J. Linn. Soc.* 183, 509–514
9. Kelly, L.J. *et al.* (2012) Why size really matters when sequencing plant genomes. *Plant Ecol. Divers.* 5, 415–425
10. Kelly, L.J. *et al.* (2015) Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. *New Phytol.* 208, 596–607
11. Mohlhenrich, E.R. and Mueller, R.L. (2016) Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. *Evolution* 70, 2865–2878
12. Metcalfe, C.J. and Casane, D. (2013) Accommodating the load – the transposable element content of very large genomes. *Mob. Genet. Elem.* 3, e24775
13. Frahry, M.B. *et al.* (2015) Low levels of LTR retrotransposon deletion by ectopic recombination in the gigantic genomes of salamanders. *J. Mol. Evol.* 80, 120–129
14. Pryer, K.M. *et al.* (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409, 618–622
15. Clark, J. *et al.* (2016) Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytol.* 210, 1072–1082
16. Obermayer, R. *et al.* (2002) Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Ann. Bot.* 90, 209–217
17. Mandáková, T. *et al.* (2010) Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell* 22, 2277–2290
18. Dodsworth, S. *et al.* (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot. J. Linn. Soc.* 180, 1–5
19. Pellicer, J. *et al.* (2014) A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. *New Phytol.* 201, 1484–1497
20. Zonneveld, B.J.M. (2010) New record holders for maximum genome size in eudicots and monocots. *J. Bot.* 2010, 527357
21. Metcalfe, C.J. *et al.* (2012) Evolution of the Australian lungfish (*Neoceratodus forsteri*) genome: a major role for CR1 and L2 LINE elements. *Mol. Biol. Evol.* 29, 3529–3539
22. El Baidouri, M. and Panaud, O. (2013) Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biol. Evol.* 5, 954–965
23. Wendel, J.F. *et al.* (2016) Evolution of plant genome architecture. *Genome Biol.* 17, 1–14
24. Ma, L. *et al.* (2015) Angiosperms are unique among land plant lineages in the occurrence of key genes in the RNA-directed DNA methylation (RdDM) pathway. *Genome Biol. Evol.* 7, 2648–2662
25. Madison-Villar, M.J. *et al.* (2016) Small RNAs from a big genome: the piRNA pathway and transposable elements in the salamander species *Desmognathus fuscus*. *J. Mol. Evol.* 83, 126–136
26. Michael, T.P. (2014) Plant genome size variation: bloating and purging DNA. *Brief. Funct. Genomic.* 13, 308–317
27. Veselý, P. *et al.* (2013) Nutrient reserves may allow for genome size increase: evidence from comparison of geophytes and their sister non-geophytic relatives. *Ann. Bot.* 112, 1193–1200
28. Grime, J.P. and Mowforth, M.A. (1982) Variation in genome size – an ecological interpretation. *Nature* 299, 151–153
29. Hessen, D.O. *et al.* (2010) Genome streamlining and the elemental costs of growth. *Trends Ecol. Evol.* 25, 75–80
30. Bousios, A. and Gaut, B.S. (2016) Mechanistic and evolutionary questions about epigenetic conflicts between transposable elements and their plant hosts. *Curr. Opin. Plant Biol.* 30, 123–133
31. Hoeijmakers, J.H.J. (2009) DNA damage, aging, and cancer. *N. Engl. J. Med.* 361, 1475–1485
32. Šímová, I. and Herben, T. (2012) Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. *Proc. R. Soc. B Biol. Sci.* 279, 867–875
33. Bennett, M.D. (1971) The duration of meiosis. *Proc. R. Soc. Lond. Ser. B* 178, 277–299
34. Van't Hof, J. and Sparrow, A.H. (1963) A relationship between DNA content, nuclear volume, and minimum mitotic cycle time. *Proc. Natl. Acad. Sci. U. S. A.* 49, 897–902
35. Bennett, M.D. and Leitch, I.J. (2005) Genome size evolution in plants. In *The Evolution of the Genome* (Gregory, T.R., ed.), pp. 89–162, Elsevier
36. Herben, T. *et al.* (2012) Ecological effects of cell-level processes: genome size, functional traits and regional abundance of herbaceous plant species. *Ann. Bot.* 110, 1257–1267
37. Knight, C.A. *et al.* (2005) The large genome constraint hypothesis: evolution, ecology and phenotype. *Ann. Bot.* 95, 177–190
38. Knight, C.A. and Beaulieu, J.M. (2008) Genome size scaling through phenotype space. *Ann. Bot.* 101, 759–766
39. Greilhuber, J. and Leitch, I.J. (2013) Genome size and the phenotype. *In Plant genome diversity (Vol. 2): Physical Structure, Behaviour and Evolution of Plant Genomes* (Leitch, I.J. *et al.* eds), pp. 323-344, Springer-Verlag
40. Gottlieb, A. *et al.* (2013) Insular organization of gene space in grass genomes. *PLoS One* 8, e54101
41. Dodsworth, S. *et al.* (2015) Genome size diversity in angiosperms and its influence on gene space. *Curr. Opin. Genet. Dev.* 35, 73–78
42. Organ, C.L. *et al.* (2011) Paleogenomic data suggest mammal-like genome size in the ancestral amniote and derived large genome size in amphibians. *J. Evol. Biol.* 24, 372–380
43. Francis, D. *et al.* (2008) A strong nucleotypic effect of DNA C-value on the cell cycle regardless of ploidy level. *Ann. Bot.* 101, 747–757
44. Schubert, I. and Oud, J.L. (1997) There is an upper limit of chromosome size for normal development of an organism. *Cell* 88, 515–520
45. Doležel, J. *et al.* (2003) Nuclear DNA content and genome size of trout and human. *Cytometry* 51A, 127–128
46. Greilhuber, J. *et al.* (2005) The origin, evolution and proposed stabilization of the terms 'Genome size' and 'C-value' to describe nuclear DNA contents. *Ann. Bot.* 95, 255–260

47. Friz, C.T. (1968) The biochemical composition of free-living amoebae *Chaos chaos*, *Amoeba dubia* and *Amoeba proteus*. *Comp. Biochem. Physiol.* 26, 81–90
48. Holm-Hansen, O. (1969) Algae: amounts of DNA and organic carbon in single cells. *Science* 163, 87–88
49. Doležel, J. *et al.* (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* 2, 2233–2244
50. Gregory, T.R. (2005) Genome size evolution in animals. In *The Evolution of the Genome* (Gregory, T.R., ed.), pp. 3–87, Elsevier
51. Byers, T.J. (1986) Molecular biology of DNA in *Acanthamoeba*, *Amoeba*, *Entamoeba*, and *Naegleria*. *Int. Rev. Cytol.* 99, 311–341
52. Veldhuis, M.J.W. *et al.* (1997) Cellular DNA content of marine phytoplankton using two new fluorochromes: taxonomic and ecological implications. *J. Phycol.* 33, 527–541
53. LaJeunesse, T.C. *et al.* (2005) *Symbiodinium* (Pyrrhophyta) genome sizes (DNA content) are smallest among dinoflagellates. *J. Phycol.* 41, 880–886
54. Doležel, J. *et al.* (1992) Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol. Plant.* 85, 625–631
55. Shoguchi, E. *et al.* (2013) Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr. Biol.* 23, 1399–1408