Trends in Plant Science

Is there an upper limit to genome size?

--Manuscript Draft--

Manuscript Number:	
Article Type:	Opinion
Corresponding Author:	Ilia J Leitch, Ph.D. Royal Botanic Gardens Richmond, Surrey UNITED KINGDOM
First Author:	Oriane Hidalgo, Ph.D.
Order of Authors:	Oriane Hidalgo, Ph.D.
	Jaume Pellicer, Ph.D.
	Maarten Christenhusz, Ph.D.
	Harald Schneider, Ph.D.
	Andrew R Leitch, Ph.D.
	Ilia J Leitch, Ph.D.
Abstract:	At 50 times 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 with 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 size 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.
Suggested Reviewers:	Juri Macas, Ph.D. Principle Investigator, Cytogenetics Institute of Plant Molecular Biology macas@umbr.cas.cz He has expertise in the field of repetitive DNA in plants and their contribution to genome size diversity
	Rachel Lockridge Mueller, Ph.D. Colorado State University Rachel.mueller@colostate.edu She has expertise in the molecular and evolutionary processes which have led to the giant genomes in salamanders.
	Laura Kelly, Ph.D. Post-doctoral research associate, Queen Mary, University of London l.kelly@qmul.ac.uk She has expertise in the analysis of the molecular composition and evolution of large genome in plants, especially Fritillaria.
	Martin Lysak, Ph.D. Research Group Leader, CEITEC: Central Euopean Institute of Technology, Brno, Czech Republic lysak@sci.muni.cz He has expertise in genome size diversity and evolution, and the underpinning molecular processes.
Opposed Reviewers:	

OPINION

Title: Is there an upper limit to genome size?

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

Affiliations:

[1]Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK.

[2]Plant Gateway, 5 Talbot Street, Hertford, Hertfordshire, SG13 7BX, UK.

[3]Department of Life Sciences, Natural History Museum, London W7 5BD, UK.

[4]School of Life Sciences, Sun Yat-Sen University, Guangzhou 510265, P.R. 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 to:** i.leitch@kew.org (I.L.)

Keywords: C-value – genomic gigantism – polyploidy – repetitive DNA

Abstract:

At 50 times 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 with 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 size 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.

Main Text:

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 ca. 64,000-fold [1-3]. The smallest genome so far reported is in the microsporidian *Encephalitozoon intestinalis*, which parasitizes a range of mammals including humans. Its genome comprises just 0.0023 Gb of DNA (=**1C-value**) and is considered to have reached the lower size limit for a fully functional eukaryotic genome [4]. 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 [5] (see also **Box 1**). Given that one nucleotide is estimated to be ca. 0.34 nm long, this diversity translates into just ca. 1.5 mm of DNA per somatic nucleus in *E. intestinalis* to ca. 100 m in *P. japonica*, with our own genome (1C=3 Gb) measuring ca. 2 m. Such enormous variation and the lack of apparent correlation with organismal complexity has long caught the attention of biologists [6, 7]. Although we now know the major contributors to genome size diversity are non-protein coding, often highly repetitive DNA sequences [8, 9], why their amounts vary so much still remains enigmatic.

Despite such diversity, species possessing enormous genomes are the exception, as most eukaryotes possess small or very small genomes (Fig. 1). Indeed, eukaryotes with genomes larger than 100 Gb are currently known in only 10 species (corresponding to 0.09% of species that have genome size data), belonging to just five eukaryotic orders; one in the ferns (Psilotales [10]), two in flowering plants (Liliales and Santalales [1]), and two in vertebrates (Lepidosireniformes [lungfish] and Urodela [salamanders][2]) (Table 1). Although there is increasing awareness that studying gigantic genomes is essential for providing a more complete picture of eukaryotic genome evolution [11], such species have been omitted from whole-genome sequencing projects due to the analytical challenges that such large genomes pose. Insights into how their genomes are structured and function therefore remain limited [e.g. 12, 13-15].

Recent Discovery of Genomic Gigantism Among Ferns

The most recent discovery of genomic gigantism is in the fern *Tmesipteris obliqua* (1C=147.3 Gb [10]). This species belongs to the phylogenetically distinct whisk-fern family, Psilotaceae [16] and provides further evidence that although scarce, genomic gigantism is scattered across the tree of life. Tetraploid representatives of *Psilotum nudum* and *Tmesipteris elongata* (both 2n=4x=208 chromosomes) have 1C-values of 71.08 Gb and 73.19 Gb, respectively [17, 18], 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=8x=416). Ferns are reported to usually retain chromosome numbers following polyploidy [17], in contrast to large-scale chromosome restructuring that frequently reduces chromosome numbers to a diploid-like number in other polyploid lineages [19, 20]. Such a retention of chromosomes might explain why *T. obliqua* has far more chromosomes than observed in the other eukaryotes with 1C>100 Gb (Table 1). In addition, *T. obliqua* belongs to a family of ferns that already has large genomes as they possess substantially larger chromosomes than most other ferns [17], 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 [21]. 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 [22]. Such observations contrast with those of giant genomes in animals 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 [14, 15, 23].

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 [24, 25]. So why 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 [25]. The genome size of most organisms thus predominantly reflects the relative contributions of these two dynamic yet opposing sets of processes. If so, the existence of giant genomes suggests that their genomic and epigenetic regulatory processes influencing genome size are operating differently, leading to the accumulation of DNA well beyond the usual limits [14, 15, 26]. 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 with species possessing smaller genomes [12-15, 27]. Nevertheless, whether this is due to changes at the genomic level [e.g. reduced recombination or altered epigenetic regulation, 12, 15, 26, 28] and/or driven by a relaxation of selection pressures against giant genomes (e.g. in some geophytic [29, 30], epiphytic and parasitic plants[10]) 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 still 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 and stopped at relatively similar giant genome sizes. It is therefore tempting to speculate that *ca*. 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 *ca*. 150 Gb are perhaps simply too great to be handled efficiently. Certainly,

the elemental costs (particularly N and P) associated with copying and transcribing the DNA and synthesizing sufficient numbers of histones to package the genome will be substantial [31] as will be the energetic costs associated with regulating the activity of non-coding DNA sequences such as transposable elements [32].

(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 just 3 Gb, it is estimated that there are >10,000 endogenous DNA damage events per cell per day and that the repair of just a single double-stranded break requires more than 10,000 ATPs [33]. 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 [34]) and timing constraints (arising from the longer duration of mitosis and meiosis as genome size increases [35-37]) 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 [34], as well as their broader impact on various growth and ecological parameters [38-41].

(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 [42]. 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 [12, 14, 43]. Thus, paradoxically, the gene space of giant genomes may be less impacted by surrounding repeats than it is in species with small genomes [42]. 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 with their relatives that have a smaller genome size. For example, (i) analyses of the repetitive DNA content of giant genomes of salamanders [15], lungfish [14] and *Fritillara* [12] suggest that DNA loss through recombination is slower than in relatives with smaller genomes; (ii) relationships involving cell size/genome size [44], and cell cycle time/genome size [45], have been noted to follow significantly different regression slopes in species with larger versus smaller genomes, and (iii) substitution rates have been shown to be lower in the giant genomes of salamanders compared with frogs which have smaller genomes [13]. Nevertheless, despite these tantalizing insights, there are still significant gaps in our knowledge of giant genomes (e.g. see **Outstanding Questions**). In order 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.

Acknowledgements:

The authors thank, Richard Bateman, Quentin Cronk, Steven Dodsworth, Paula Rudall, and Sonja Siljak-Yakovlev for useful discussion, 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 n°657918).

References:

1. Garcia, S. *et al.* (2014) Recent updates and developments to plant genome size databases. *Nucleic Acids Res* 42, D1159-D1166

2. Gregory, T.R. (2016) Animal Genome Size Database. http://www.genomesize.com, (accessed).

3. Kullman, B. *et al.* (2005) Fungal Genome Size Database. http://www.zbi.ee/fungal-genomesize., (accessed).

4. Corradi, N. *et al.* (2010) The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nature Communications* 1, Article 77

5. Pellicer, J. *et al.* (2010) The largest eukaryotic genome of them all? *Bot J Linn Soc* 164, 10-15

6. Thomas, C.A. (1971) The genetic organization of chromosomes. *Annual Review of Genetics* **5**, 237-256.

7. Gregory, T.R. (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews* 76, 65-101

8. Grover, C. and Wendel, J.F. (2010) Recent insights into mechanisms of genome size change in plants. *Journal of Botany* 2010, Article ID 382732

9. Canapa, A. *et al.* (2015) Transposons, genome size, and evolutionary insights in animals. *Cytogenetic and Genome Research* 147, 217-239

10. Hidalgo, O. *et al.* (2017) Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica*. *Bot J Linn Soc*, (in press)

11. Kelly, L.J. *et al.* (2012) Why size really matters when sequencing plant genomes. *Plant Ecology & Diversity* 5, 415-425

12. 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 Phytologist* 208, 596-607

13. Mohlhenrich, E.R. and Mueller, R.L. (2016) Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. *Evolution* 70, 2865-2878

14. Metcalfe, C.J. and Casane, D. (2013) Accommodating the load - the transposable element content of very large genomes. *Mobile Genetic Elements* 3, e24775

15. Frahry, M.B. *et al.* (2015) Low levels of LTR retrotransposon deletion by ectopic recombination in the gigantic genomes of salamanders. *Journal of Molecular Evolution* 80, 120-129

16. 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

17. Clark, J. *et al.* (2016) Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytologist* 210, 1072-1082

18. Obermayer, R. *et al.* (2002) Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Annals of Botany* 90, 209-217

19. Mandáková, T. *et al.* (2010) Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell* 22, 2277-2290

20. Dodsworth, S. *et al.* (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot J Linn Soc* 180, 1-5

21. Pellicer, J. *et al.* (2014) A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. *New Phytologist* 201, 1484-1497

22. Zonneveld, B.J.M. (2010) New record holders for maximum genome size in eudicots and monocots. *Journal of Botany* 2010, Article ID 527357

23. Metcalfe, C.J. *et al.* (2012) Evolution of the Australian Lungfish (*Neoceratodus forsteri*) genome: A major role for CR1 and L2 LINE elements. *Molecular Biology and Evolution* 29, 3529-3539

24. 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, 954-965

25. Wendel, J.F. *et al.* (2016) Evolution of plant genome architecture. *Genome Biology* 17, 1-14

26. 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 Biology and Evolution* 7, 2648-2662

27. 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*. *Journal of Molecular Evolution* 83, 126-136

28. Michael, T.P. (2014) Plant genome size variation: bloating and purging DNA. *Briefings in Functional Genomics* 13, 308-317

29. Veselý, P. *et al.* (2013) Nutrient reserves may allow for genome size increase: evidence from comparison of geophytes and their sister non-geophytic relatives. *Annals of Botany* 112, 1193-1200

30. Grime, J.P. and Mowforth, M.A. (1982) Variation in genome size - an ecological interpretation. *Nature* 299, 151-153

31. Hessen, D.O. *et al.* (2010) Genome streamlining and the elemental costs of growth. *Trends Ecol Evol* 25, 75-80

32. Bousios, A. and Gaut, B.S. (2016) Mechanistic and evolutionary questions about epigenetic conflicts between transposable elements and their plant hosts. *Current Opinion in Plant Biology* 30, 123-133

33. Hoeijmakers, J.H.J. (2009) DNA damage, aging, and cancer. *New England Journal of Medicine* 361, 1475-1485

238 34. Šímová, I. and Herben, T. (2012) Geometrical constraints in the scaling relationships between
239 genome size, cell size and cell cycle length in herbaceous plants. *Proceedings of the Royal*
240 *Society B: Biological Sciences* 279, 867-875

241 35. Bennett, M.D. (1971) The duration of meiosis. *Proceedings of the Royal Society of London Series B*
242 178, 277-299

243 36. Van't Hof, J. and Sparrow, A.H. (1963) A relationship between DNA content, nuclear volume, and
244 minimum mitotic cycle time. *Proceedings of the National Academy of Sciences (USA)* 49, 897-902

245 37. Bennett, M.D. and Leitch, I.J. (2005) Genome size evolution in plants. In The evolution of the genome
246 (Gregory, T.R. ed), pp. 89-162, Elsevier.

247 38. Herben, T. *et al.* (2012) Ecological effects of cell-level processes: genome size, functional traits and
248 regional abundance of herbaceous plant species. *Annals of Botany* 110, 1257-1267

249 39. Knight, C.A. *et al.* (2005) The large genome constraint hypothesis: evolution, ecology and phenotype.
250 *Annals of Botany* 95, 177-190

251 40. Knight, C.A. and Beaulieu, J.M. (2008) Genome size scaling through phenotype space. *Annals of*
252 *Botany* 101, 759-766

253 41. Greilhuber, J. and Leitch, I.J. (2013) Genome size and the phenotype. In Plant genome diversity, vol
254 2, Physical structure, behaviour and evolution of plant genomes (Leitch, I.J. et al. eds), pp. 323-
255 344, Springer-Verlag.

256 42. Gottlieb, A. *et al.* (2013) Insular organization of gene space in grass genomes. *PLoS ONE* 8, e54101

257 43. Dodsworth, S. *et al.* (2015) Genome size diversity in angiosperms and its influence on gene space.
258 *Current Opinion in Genetics & Development* 35, 73-78

259 44. Organ, C.L. *et al.* (2011) Paleogenomic data suggest mammal-like genome size in the ancestral
260 amniote and derived large genome size in amphibians. *Journal of Evolutionary Biology* 24, 372-
261 380

262 45. Francis, D. *et al.* (2008) A strong nucleotypic effect of DNA C-value on the cell cycle regardless of
263 ploidy level. *Annals of Botany* 101, 747-757

264 46. Doležel, J. *et al.* (2003) Nuclear DNA content and genome size of trout and human. *Cytometry Part A*
265 51A, 127-128

266 47. Greilhuber, J. *et al.* (2005) The origin, evolution and proposed stabilization of the terms 'Genome
267 size' and 'C-value' to describe nuclear DNA contents. *Annals of Botany* 95, 255-260

268 48. Friz, C.T. (1968) The biochemical composition of free-living Amoebae *Chaos chaos, Amoeba dubia*
269 and *Amoeba proteus*. *Comparative Biochemistry and Physiology* 26, 81-90

270 49. Holm-Hansen, O. (1969) Algae: amounts of DNA and organic carbon in single cells. *Science* 163, 87-
271 88

50. Doležel, J. *et al.* (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* 2, 2233-2244

51. Gregory, T.R. (2005) Genome size evolution in animals. In The evolution of the genome (Gregory, T.R. ed), pp. 3-87, Elsevier.

52. Byers, T.J. (1986) Molecular biology of DNA in *Acanthamoeba*, *Amoeba*, *Entamoeba*, and *Naegleria*. *International Review of Cytology* 99, 311-341

53. Veldhuis, M.J.W. *et al.* (1997) Cellular DNA content of marine phytoplankton using two new fluorochromes: taxonomic and ecological implications *Journal of Phycology* 33, 527-541

54. LaJeunesse, T.C. *et al.* (2005) *Symbiodinium* (Pyrrhophyta) genome sizes (DNA content) are smallest among dinoflagellates. *Journal of Phycology* 41, 880-886

55. Doležel, J. *et al.* (1992) Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum* 85, 625-631

56. Shoguchi, E. *et al.* (2013) Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology* 23, 1399-1408

Figure 1. Violin plots showing the frequency and range of genome sizes in different eukaryote groups, together with illustrations on the 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 taken from the Plant DNA C-values database (http://data.kew.org/cvalues/), the Animal Genome Size database [2], 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.

299 **Table 1. Eukaryote species with genome sizes (1C-values) greater than 100 Gb**

300

Eukaryotic group Order - number of species recognised	Species	GS[a]	Method[b]	2n[c]	x[d]
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.

Data taken from the Plant DNA C-values database (http://data.kew.org/cvalues/), the Animal Genome Size database [2] and Hidalgo *et al.* [10]. [a]Genome size (1C-value, Gb); [b]Method used to estimate genome size: Fe = Feulgen microdensitometry, FC:PI = flow cytometry using the fluorochrome propidium iodide; [c]Chromosome number; [d]Ploidy, n.d. = Not determined or unclear.

301

302

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 number of base pairs in Gigabase pairs (Gb); 1 pg=0.978 Gb [46].

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 [47].

Fluorochromes: Chemicals that bind to DNA and have the capacity to fluoresce when irradiated with light of the appropriate wavelength. Several different fluorochromes have been used to estimate genome size, including 3, 5-diaminobenzoic acid dihdrochloride, 4', 6-diamidino-2-phenylindole (DAPI), propidium iodide (PI), Hoechst (HO-33342), SYTOX and PicoGreen.

Polyploidy: Presence of more than two sets of chromosomes in the nucleus (genome), e.g. tetraploid (4*x*) = possessing four sets of chromosomes.

Repetitive DNA: Highly repetitive DNA sequences, which include tandem repeats (e.g. DNA satellites) and dispersed repeats (e.g. transposable elements – of which the most common are DNA transposons and retroelements).

Box 1: What about the giant genomes reported to exist in amoebae and dinoflagellates - are they just 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 [48, 49]). However, these estimates were not obtained using best practice methodology [e.g. 50, 51]. Potential technical issues which may have compromised their accuracy include:

(i) **Nuclei isolation**. Friz's [48] measurements of amoebae used whole cells and biochemical approaches now considered too unreliable for genome size determination [51]. Indeed, Friz's measurements were questioned by Byers [52] 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* [53, 54]).

(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 genome size of the target and standard should not exceed 3x since this can impact on the linearity of the instrument's response [50]. While this is sometimes difficult to fulfil, the use of such small calibration standards for estimating very large genomes will undoubtable introduce errors.

(iii) **Selection of fluorochromes**. Some fluorochromes used in dinoflagellate studies are unreliable as they preferentially bind to AT-rich regions of the DNA (e.g. DAPI) and can artefactually increase genome size estimates by >40% [55]. In addition, the saline conditions in which dinoflagellates live can impact the fluorescence of fluorochromes such as PicoGreen and SYTOX and hence genome size values [53]. Further, the unusual helicoidal organization and DNA sequence composition of dinoflagellates [56], are considered likely to distort the quantitative binding of the fluorochrome to DNA - an essential requirement for robust genome size estimations.

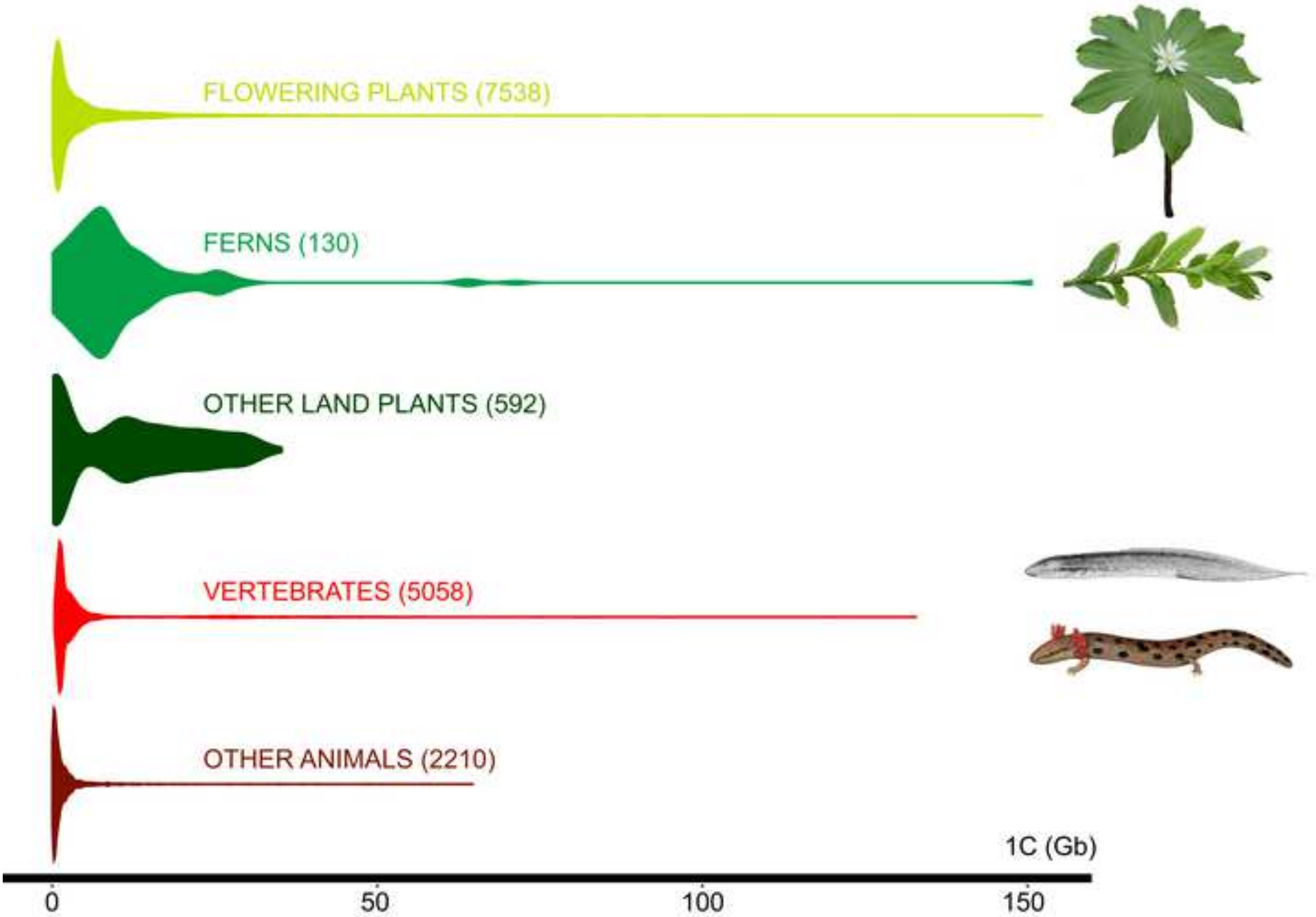
(iv) **Impact of fixation/drying of cells**: Use of fresh samples is essential for accurate genome size estimates, as fixed or dried tissues can alter DNA staining properties and hence

fluorescence intensity [50]. For example, dramatic differences in genome size estimates between live [11 Gb] and fixed [232 Gb] samples of *Prorocentrum micans* have been reported [53]. Nevertheless, dinoflagellate genome sizes are predominantly estimated using fixed material.

Overall, while it is clear that the genomes in these eukaryotic lineages are big, 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*.

Figure

Click here to download Figure TIPS OPINION-FIGURE 1.jpg



Title: Is there an upper limit to genome size?

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

Trends Box

- Eukaryote genomes range ca. 64,000-fold in size, yet the parts comprising genes, regulatory regions and other functional components typically account for just a small fraction of the 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, those with giant genomes are the exception and belong to just 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. Realizing that the largest genomes in these lineages are all similarly massive, despite coming from distinct eukaryote groups, suggests an upper limit to genome size - the theme of this paper.

Title: Is there an upper limit to genome size?

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

Outstanding Questions Box

- What is the extant diversity of genome sizes in eukaryotes? Current understanding of nuclear DNA contents across eukaryotes has revealed a staggering diversity, yet data are relatively scarce or missing for most lineages.
- Is there an ecological cost for a large genome, especially in terms of the resources required (e.g. nitrogen and phosphate) 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?
- 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 with species with smaller genomes?
- Species with giant genomes are typically rare, are they less resilient to environmental change because of their large genome? To what extent does environmental stress such as climate change contribute to genome size diversity?

Royal Botanic Gardens
Kew

Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB
020 8332 5000 | www.kew.org | info@kew.org

20th January 2017

Dear Susanne,

Submission of Opinion article entitled
"Is there an upper limit to genome size?"

Further to your kind invitation to submit an Opinion article on genome size (Presubmission #781), I have pleasure in attaching our proposed article. I hope this will be of interest, and we look forward to hearing from you when you have time.

With best wishes,

I. J. Leitch

Dr Ilia J Leitch
Senior Research Leader in Character Evolution
Department of Comparative Plant and Fungal Biology
Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB
Tel: 0208 332 5329

Kew
ROYAL BOTANIC GARDENS