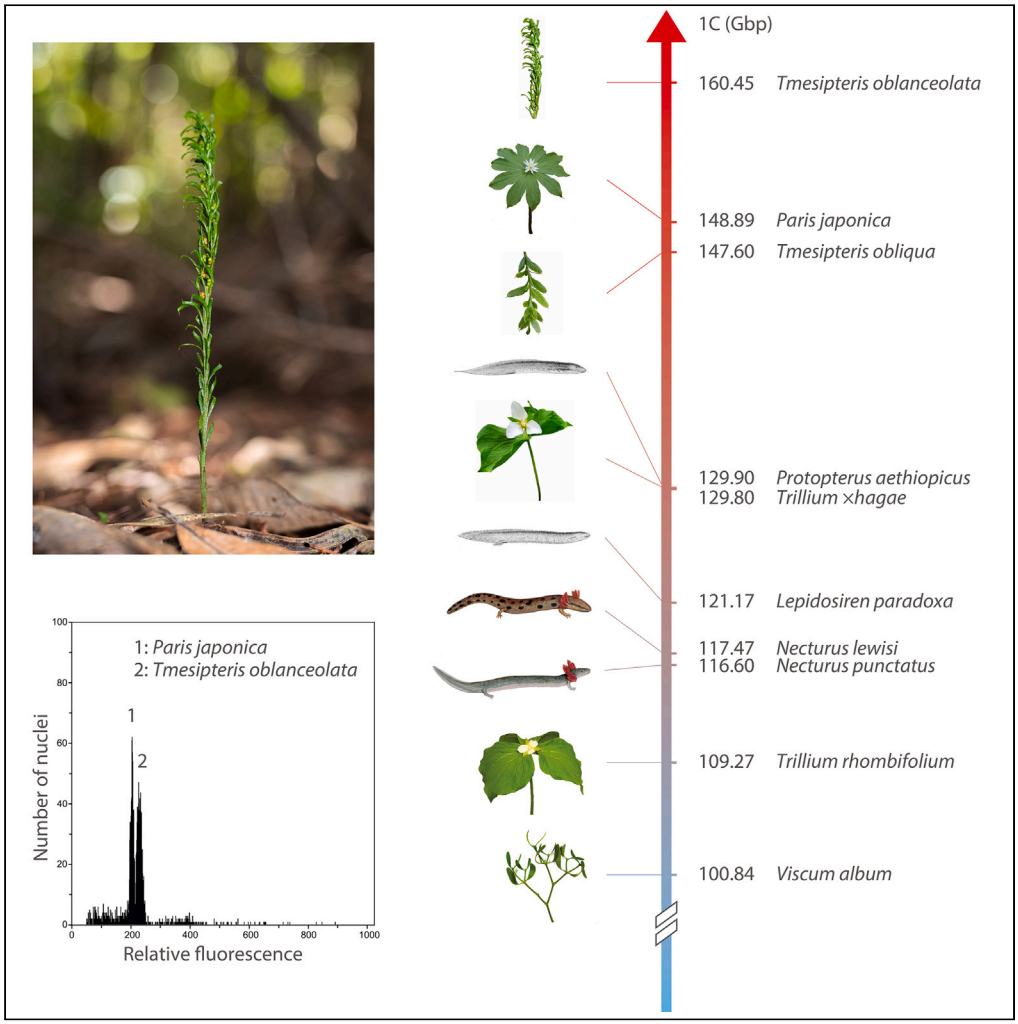
iScience

CellPress
OPEN ACCESS

Article

A 160 Gbp fork fern genome shatters size record for eukaryotes



Pol Fernández, Rémy Amice, David Bruy, ..., Lisa Pokorny, Oriane Hidalgo, Jaume Pellicer

jaume.pellicer@ibb.csic.es

Highlights

Giant genomes are restricted across the eukaryotic Tree of Life

The genome of *T. oblanceolata* is over 50 times larger than the human genome

Genome size variation among eukaryotes expands over 61,000-fold

Fernández et al., iScience ■■, 109889
■■, 2024
https://doi.org/10.1016/j.isci.2024.109889

Please cite this article in press as: Fernández et al., A 160 Gbp fork fern genome shatters size record for eukaryotes, iScience (2024), https://doi.org/10.1016/j.isci.2024.109889

iScience

CellPress
OPEN ACCESS

Article

A 160 Gbp fork fern genome shatters size record for eukaryotes

Pol Fernández,[1,2] Rémy Amice,[3] David Bruy,[4,5] Maarten J.M. Christenhusz,[6,7] Ilia J. Leitch,[6] Andrew L. Leitch,[8] Lisa Pokorny,[6,9] Oriane Hidalgo,[1,6,10] and Jaume Pellicer[1,6,10,11,*]

SUMMARY

Vascular plants are exceptional among eukaryotes due to their outstanding genome size diversity which ranges ~2,400-fold, including the largest genome so far recorded in the angiosperm *Paris japonica* (148.89 Gbp/1C). Despite available data showing that giant genomes are restricted across the Tree of Life, the biological limits to genome size expansion remain to be established. Here, we report the discovery of an even larger eukaryotic genome in *Tmesipteris oblanceolata*, a New Caledonian fork fern. At 160.45 Gbp/1C, this record-breaking genome challenges current understanding and opens new avenues to explore the evolutionary dynamics of genomic gigantism.

INTRODUCTION

The ever expanding exploration of nearly 20,000 eukaryotic genomes has revealed an astounding array of genome sizes distributed across the eukaryotic Tree of Life (Figure 1A), influencing cell sizes, life cycles, physiology and morphology, and ultimately, impacting the ecology and evolution of species.[1,2] Miniature-sized fungal genomes, including the smallest eukaryotic genome found in the microsporidian *Encephalitozoon intestinalis* (2.6 Mbp/1C,[3] i.e., 1C = nuclear DNA content in a gametic nucleus), contrast with those found in other groups where genomes have expanded up to five orders of magnitude. Yet it is only within a few animal and plant lineages that truly extreme genomic expansions beyond 100 Gbp/1C are known.[4] Among animals, obese genomes exist in just two chordate lineages—lungfishes (class Dipnoi), where they reach up to 129.90 Gbp/1C in *Protopterus aethiopicus*[5] and proteid salamanders (class Amphibia), with genomes up to 117.47 Gbp/1C in *Necturus lewisi*.[6] In contrast, several vascular plant groups have successfully expanded into the upper end of the genome size spectrum and comprise six of the top ten largest eukaryotic genomes known (Figure 1B). While most of these are angiosperms, especially within the monocot families Liliaceae and Melanthiaceae (e.g., *Paris japonica;* 148.89 Gbp/1C[7]), extreme genomes have also been reported in the parasitic eudicot *Viscum album* (Santalaceae; 100.84 Gbp/1C[8]) and within ferns in the Psilotaceae family (*Tmesipteris obliqua*, 147.60 Gbp/1C[9]). This vast scope of genome size variation, seemingly disconnected from organismal complexity and known as the "C-value paradox" or "C-value enigma",[10] has intrigued biologists for over half a century. Rapid advances in DNA sequencing are now providing compelling evidence showing that variation in DNA amount arises predominantly from differences in the frequency of polyploidy, abundance of non-coding repetitive DNA and the dynamics of the processes that amplify, erode and delete DNA.[11] Yet the question arises as to whether we have uncovered the full extent of genome size diversity.

As part of ongoing research addressed to enhance our understanding of how and why giant genomes evolve and function, and based on previous evidence of genomic gigantism in the small genus *Tmesipteris*,[9] we conducted a survey using propidium iodide flow cytometry to robustly quantify genome size variation in the genus. Here, we present the discovery of the largest eukaryotic genome so far reported.

RESULTS

The nuclear DNA content of *T. oblanceolata*

Using *Allium cepa* "Ailsa Craig" as internal standard, the species *Tmesipteris oblanceolata* subsp. *linearifolia* (Figure 2A), a rare species present in New Caledonia and some neighboring archipelagos, was found to have a genome size of 160.75 Gbp/1C (Figure 2B). An additional

[1]Institut Botànic de Barcelona (IBB), CSIC-CMCNB, Passeig del Migdia s.n, 08038 Barcelona, Spain
[2]Facultat de Farmàcia i Ciències de l'alimentació, Campus Diagonal, Universitat de Barcelona, Av. de Joan XXIII, 27-31, 08028 Barcelona, Spain
[3]Independent researcher, Nouméa, New Caledonia
[4]AMAP, IRD, Herbier de Nouvelle-Calédonie, Nouméa 98848, New Caledonia
[5]UMR AMAP, Université de Montpellier, IRD, CIRAD, CNRS, INRAE, F-34000 Montpellier, France
[6]Royal Botanic Gardens, Kew, Richmond TW9 3AE, UK
[7]Department of Environment and Agriculture, Curtin University, 6845 Perth, WA, Australia
[8]School of Biological and Behavioral Sciences, Queen Mary University of London, London E1 4NS, UK
[9]Real Jardín Botánico (RJB-CSIC), Plaza de Murillo 2, 28014 Madrid, Spain
[10]Senior authors
[11]Lead contact
*Correspondence: jaume.pellicer@ibb.csic.es
https://doi.org/10.1016/j.isci.2024.109889

Please cite this article in press as: Fernández et al., A 160 Gbp fork fern genome shatters size record for eukaryotes, iScience (2024), https://doi.org/10.1016/j.isci.2024.109889

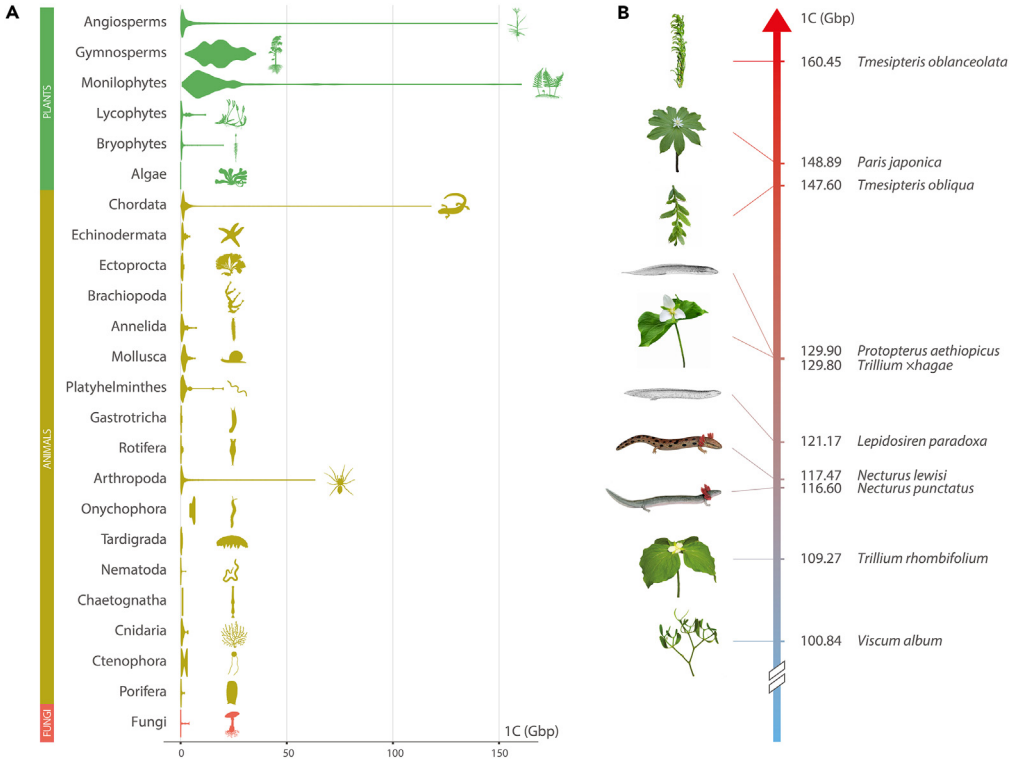


Figure 1. Genome size diversity across eukaryotes

(A) Current distribution of genome sizes across major lineages of plants, animals, and fungi.

(B) Top 10 of the largest genome size records available in eukaryotes. Image and silhouette credits are provided in supplemental information.

analysis using an alternative internal standard (*Fritillaria lusitanica*, 50.47 Gbp/1C), with a larger genome than that of *A. cepa*, was carried out with a resulting 1C-value of 160.45 ± 0.81 Gbp, that is c. 160,000 Mbp (Figure 2C, see Table S1 and STAR methods for detailed procedures).

DISCUSSION

Extending the range of genome size diversity in eukaryotes

Tmesipteris (Psilotaceae) is a relatively understudied small genus made up of ~15 species,[12] which are mainly epiphytic ferns occurring in Oceania and several Pacific Islands. Until now, the genome sizes of only two species had been reported in the genus, i.e., for the tetraploid *T. tannensis* (1C = 73.19 Gbp[13]) and the octoploid *T. obliqua* (1C = 147.29 Gbp[9]), both with giant genomes, and with the only two known ploidy levels reported for the genus (based on x = 52[14]). The size of the genome of *T. oblanceolata* subsp. *linearifolia* is larger than either of these estimates, and ~7% larger than the previous record holder for an eukaryote (i.e., the monocot angiosperm *P. japonica*; Figures 2D and 2E). In fact, the scale of magnitude of the *T. oblanceolata* genome is such that ~1,190 copies of the genome of the model plant *Arabidopsis thaliana* could be housed within it (assuming an average *A. thaliana* genome of 1C = 135 Mb[15]). Overall, this exceptional discovery extends the range of genome sizes known for eukaryotes to over 61,000-fold, and ~2,500-fold among plants. But, how likely is this genome size range to be further expanded?

At the upper end of the scale of angiosperms, *P. japonica* stands out by being the only recognized octoploid bearing large chromosomes within a group of extensively studied monocot families with large genomes in Liliales (i.e., Liliaceae and Melanthiaceae) and Asparagales (i.e., Amaryllidaceae and Asparagaceae). Therefore, the prospect of identifying larger monocot genomes within these lineages seems improbable. By contrast, the presence of high chromosome numbers is more frequent in ferns, also known as the "chromosome hoarding syndrome",[16] which is a primary consequence of the numerous rounds of whole genome multiplication (WGM). However, unlike angiosperms where polyploidy is also prevalent,[17] post-polyploid diploidization mechanisms in ferns typically involves gene silencing without significant DNA elimination, resulting in high chromosome numbers but reflecting a diploid-like gene expression.[18] *Tmesipteris oblanceolata* subsp. *linearifolia* has been reported, like *P. japonica*, to be an octoploid, but it has a much higher chromosome number ($2n = 416$ versus $2n = 40$[7,19]). Its massive genome is thus considered to have arisen through the combined effects of repetitive DNA accumulation and polyploidy, as in other species of the genus.[20] Unfortunately, our attempts to obtain cytological data from synangia were unsuccessful and did not yield analyzable meiotic chromosomes. However, previous research in the genus by Braithwaite[19,21] reported that the two subspecies known for *T. oblanceolata* were octoploid (i.e., $2n = 8x = 416$). Taking into account these records and our previous

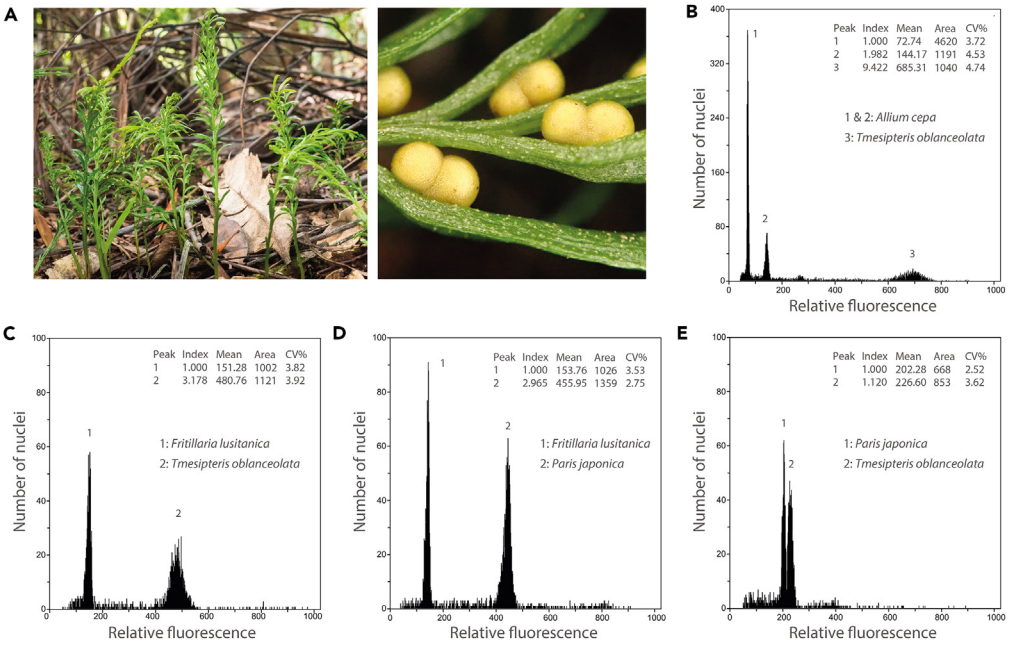


Figure 2. Largest eukaryote genome in *Tmesipteris oblanceolata*

(A) Wild population of the fork fern *Tmesipteris oblanceolata* sourced for genome size analysis (left), including details of synangia and stomata (right).

(B) Relative fluorescence histogram from the flow cytometry analysis of a combined sample comprising nuclei from the internal calibration standard *Allium cepa* (peaks 1 and 2) and *Tmesipteris oblanceolata* (peak 3). Relative fluorescence histograms from the flow cytometry analyses of *T. oblanceolata* (C) and *Paris japonica* (D) with the internal standard *Fritillaria lusitanica*. Peak 1 in both images indicates the relative fluorescence peak of the G_1 stained nuclei from the internal standard *Fritillaria lusitanica*, and peak 2 for the target sample, i.e., *P. japonica* or *T. oblanceolata*.

(E) Relative fluorescence histograms from a combined analysis of *T. oblanceolata* with the previous record-holder *P. japonica*, which confirms the larger size of the genome of *T. oblanceolata*. Image credits as follows: © Oriane Hidalgo.

genome size estimates (see aforementioned section), we believe with relatively high confidence, that the individual analyzed in this study is also octoploid.

Continuous efforts are being made to increase our understanding of the existing genome size diversity across the major fern (i.e., monilophytes) lineages, with estimates currently available for nearly 552 species (out of 10,388 sp. *sensu* PPGI,[12] i.e., ~6% of species). These data reveal that fern genome sizes range ~630-fold.[22] While increasing the representation of species with genome sizes data will be invaluable, based on available cytogenetic and karyological data the chances of discovering genomes significantly larger than that of *T. oblanceolata* in the group seem unlikely. Within Psilotales, octoploidy is the highest known ploidy level, suggesting that genomes of similar size, but not significantly larger than that of *T. oblanceolata*, may be found in other octoploid species in the order. In other ferns, genomes as large as 63–64 Gbp/1C have been reported in *Ophioglossum*, which belong to the order Ophioglossales, with extremely high chromosome numbers ($2n$ = 720–960).[23] Nevertheless, while even higher chromosome numbers exist in the genus (e.g., *O. reticulatum*, $2n$ = 1,020–1,400[24]), based on the existing genome size data it also seems unlikely that their genomes will supersede that of *T. oblanceolata*.

Among animals, the expansion of genomes to gigantic scales has so far been attributed to the accumulation of repetitive elements and retention of non-coding DNA, rather than WGM[25,26] (although polyploidy has been reported in some salamanders and lungfish with smaller genomes[27]). Without reports of associated WGM in the biggest proteid salamanders and lungfish genomes, it seems unlikely that animal genomes exceeding that of *T. oblanceolata* will be found. At the other end of the scale, the smallest eukaryotic genomes, as found in *Encephalitozoon* species, stand out by being extremely compact, with reductions at multiple levels including in the number and size of genes and minimal amounts of non-coding repetitive DNA.[3] Given these observations, the prospect of finding much smaller eukaryotic genome sizes appears also improbable.

Biodiversity genomics across the whole genome size spectrum

Recently, Armstrong et al.[28] proposed that the question "What controls plant genome size and complexity?" was among the top "One hundred important questions facing plant science." This is a critical question, not only for plants, but for eukaryotes as a whole, given that a substantial body of research has shown that there are clear biochemical, metabolic, and regulatory costs associated with increasing DNA amounts.[4] These costs, in turn, exert cascading effects impacting the biology of organisms, their ecology and their potential to persist over evolutionary timescales.[29] The results here, coupled with knowledge of genome size diversity at the lower end of the spectrum, suggest that we now know the limits, or are most likely very close to the limits of eukaryotic genome size diversity.

The unceasing production of novel genome size data, complemented with an increasingly robust understanding of the worldwide distribution of organisms and their evolutionary relationships, has opened unprecedented prospects for investigating the role of genome size in shaping and influencing the survival of extant biodiversity.[1] Likewise, the significant growth in knowledge arising from large-scale genomic approaches is opening up the horizons for exploring the vast genomic variation that underpins the variety of life on Earth. The emerging field of biodiversity genomics is helping to unravel the complex interplay between the architecture of the genome, environmental factors, and evolutionary processes from a holistic perspective.[30,31] However, most studies are still focused almost exclusively on species with small to medium size genomes.[32] Given that rapid progress in genomic technologies shows it is now possible to assemble a ~90 Gbp/1C genome to chromosome level (e.g., in the mistletoe *Viscum album*; https://www.darwintreeoflife.org/), it is clear that critical knowledge gaps can finally be filled in our understanding of eukaryotic genomes at the upper end of genome size diversity. Embracing such opportunities will shed new light on the evolutionary forces leading to the rise of giant genomes across taxonomic lineages. Our discovery of the largest genome in *T. oblanceolata* represents a unique milestone in the field, identifying this species as a critical asset for future detailed genomic analysis so we can fully understand the ecological and evolutionary consequences of genome size diversity and architecture across eukaryotes, especially in the face of ongoing biodiversity loss and climate change.

Limitation of the study

It has taken over a decade of research exploring the limits of plant genome size diversity to discover a larger genome than that of *Paris japonica*, but it cannot be completely ruled out that even larger genomes may be uncovered in the future. Nevertheless, the multiple physiological, ecological, and evolutionary costs associated with genomic expansions at such gigantic scales most likely suggest that if the upper limit has not been reached yet, that of *Tmesipteris oblanceolata* must be very close to it.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Distribution of genome sizes in eukaryotes
  - Flow cytometry and genome size data
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109889.

ACKNOWLEDGMENTS

We thank the Spanish Research Council (Ref: PID2019-108173GA-I00) funded by MCIN/AEI/10.13039/501100011033, and the Southern Province of New Caledonia for issuing collection permit (authorization No. 2879-2020/ARR/DDDT) in compliance with the requirements of use of genetic material under the Nagoya Protocol. J.P. benefited from a Ramón y Cajal grant (Ref: RYC-2017-2274) funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future". P.F. benefited from an FPU Fellowship (Ref: FPU21/03564) funded by the Spanish Ministry of Universities. We thank Laurence Hill and Miquel Veny for providing samples of *Paris japonica* and *Fritillaria lusitanica* for the flow cytometry analyses, and Joseph E. Trumpey for the illustrations of salamanders and lungfishes used.

AUTHOR CONTRIBUTIONS

J.P. and O.H. conceived and designed the study. J.P., O.H., P.F., R.A., and D.B. performed fieldwork and supplied samples for research. P.F., J.P., and O.H. analyzed data. O.H. created the figures. J.P. acquired funding. J.P., I.J.L., and A.R.L. prepared an initial version of the manuscript with input from P.F., L.P., M.J.M.C., and O.H. All authors were involved in writing and reviewing the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 12, 2024
Revised: January 31, 2024
Accepted: April 30, 2024

REFERENCES

1. Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I.J. (2018). Genome size diversity and its impact on the evolution of land plants. Genes *9*, 88. https://doi.org/10.3390/genes9020088.
2. Blommaert, J. (2020). Genome size evolution: towards new model systems for old questions. Proc. Biol. Sci. *287*, 20201441. https://doi.org/10.1098/rspb.2020.1441.
3. Mascarenhas dos Santos, A.C., Julian, A.T., Liang, P., Juárez, O., and Pombert, J.-F. (2023). Telomere-to-Telomere genome assemblies of human-infecting *Encephalitozoon* species. BMC Genom. *24*, 237. https://doi.org/10.1186/s12864-023-09331-3.
4. Hidalgo, O., Pellicer, J., Christenhusz, M., Schneider, H., Leitch, A.R., and Leitch, I.J. (2017). Is there an upper limit to genome size? Trends Plant Sci. *22*, 567–573. https://doi.org/10.1016/j.tplants.2017.04.005.
5. Pedersen, R.A. (1971). DNA content, ribosomal gene multiplicity, and cell size in fish. J. Exp. Zool. *177*, 65–78. https://doi.org/10.1002/jez.1401770108.
6. Liedtke, H.C., Gower, D.J., Wilkinson, M., and Gomez-Mestre, I. (2018). Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nat. Ecol. Evol. *2*, 1792–1799. https://doi.org/10.1038/s41559-018-0674-4.
7. Pellicer, J., Fay, M.F., and Leitch, I.J. (2010). The largest eukaryotic genome of them all? Bot. J. Linn. Soc. *164*, 10–15. https://doi.org/10.1111/j.1095-8339.2010.01072.x.
8. Zonneveld, B.J.M. (2010). New record holders for maximum genome size in eudicots and monocots. J. Bot. Le *2010*, 1–4. https://doi.org/10.1155/2010/527357.
9. Hidalgo, O., Pellicer, J., Christenhusz, M.J.M., Schneider, H., and Leitch, I.J. (2017). Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica*. Bot. J. Linn. Soc. *183*, 509–514. https://doi.org/10.1093/botlinnean/box003.
10. Elliott, T.A., and Gregory, T.R. (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20140331. https://doi.org/10.1098/rstb.2014.0331.
11. Schubert, I., and Vu, G.T.H. (2016). Genome stability and evolution: Attempting a holistic view. Trends Plant Sci. *21*, 749–757. https://doi.org/10.1016/j.tplants.2016.06.003.
12. PPGI (2016). A community-derived classification for extant lycophytes and ferns. J. Systemat. Evol. *54*, 563–603. https://doi.org/10.1111/jse.12229.
13. Clark, J., Hidalgo, O., Pellicer, J., Liu, H., Marquardt, J., Robert, Y., Christenhusz, M., Zhang, S., Gibby, M., Leitch, I.J., and Schneider, H. (2016). Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. New Phytol. *210*, 1072–1082. https://doi.org/10.1111/nph.13833.
14. Perrie, L.R., Brownsey, P.J., and Lovis, J.D. (2010). *Tmesipteris horomaka*, a new octoploid species from Banks Peninsula. N. Z. J. Bot. *48*, 15–29. https://doi.org/10.1080/00288251003640010.
15. Lian, Q., Huettel, B., Walkemeier, B., Mayjonade, B., Lopez-Roques, C., Gil, L., Roux, F., Schneeberger, K., and Mercier, R. (2024). A pan-genome of 69 *Arabidopsis thaliana* accessions reveals a conserved genome structure throughout the global species range. Nat. Genet. 1–10. https://doi.org/10.1038/s41588-024-01715-9.
16. Li, F.-W. (2023). The chromosome hoarding syndrome of (some) ferns and lycophytes. Nat. Rev. Genet. *24*, 737. https://doi.org/10.1038/s41576-023-00639-0.
17. Landis, J.B., Soltis, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C., and Soltis, P.S. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. Am. J. Bot. *105*, 348–363. https://doi.org/10.1002/ajb2.1060.
18. Li, Z., McKibben, M.T.W., Finch, G.S., Blischak, P.D., Sutherland, B.L., and Barker, M.S. (2021). Patterns and processes of diploidization in land plants. Annu. Rev. Plant Biol. *72*, 387–410. https://doi.org/10.1146/annurev-arplant-050718-100344.
19. Braithwaite, A.F. (1988). Cytological and anatomical observations on *Tmesipteris* (Tmesipteridaceae: Pteridophyta) species from New Caledonia. Fern Gaz. *13*, 199–208.
20. Fernández, P., Leitch, I.J., Leitch, A.R., Hidalgo, O., Christenhusz, M.J.M., Pokorny, L., and Pellicer, J. (2023). Giant fern genomes show complex evolution patterns: A comparative analysis in two species of *Tmesipteris* (Psilotaceae). Int. J. Mol. Sci. *24*, 2708. https://doi.org/10.3390/ijms24032708.
21. Braithwaite, A.F. (1986). *Tmesipteris* in Vanuatu (New Hebrides). Fern Gaz. *13*, 87–96.
22. Fujiwara, T., Liu, H., Meza-Torres, E.I., Morero, R.E., Vega, A.J., Liang, Z., Ebihara, A., Leitch, I.J., and Schneider, H. (2023). Evolution of genome space occupation in ferns: linking genome diversity and species richness. Ann. Bot. *131*, 59–70. https://doi.org/10.1093/aob/mcab094.
23. Obermayer, R., Leitch, I.J., Hanson, L., and Bennett, M.D. (2002). Nuclear DNA C-values in 30 species double the familial representation in Pteridophytes. Ann. Bot. *90*, 209–217. https://doi.org/10.1093/aob/mcf167.
24. Khandelwal, S. (1990). Chromosome evolution in the genus *Ophioglossum* L. Bot. J. Linn. Soc. *102*, 205–217. https://doi.org/10.1111/j.1095-8339.1990.tb01876.x.
25. Meyer, A., Schloissnig, S., Franchini, P., Du, K., Woltering, J.M., Irisarri, I., Wong, W.Y., Nowoshilow, S., Kneitz, S., Kawaguchi, A., et al. (2021). Giant lungfish genome elucidates the conquest of land by vertebrates. Nature *590*, 284–289. https://doi.org/10.1038/s41586-021-03198-8.
26. Wang, K., Wang, J., Zhu, C., Yang, L., Ren, Y., Ruan, J., Fan, G., Hu, J., Xu, W., Bi, X., et al. (2021). African lungfish genome sheds light on the vertebrate water-to-land transition. Cell *184*, 1362–1376.e18. https://doi.org/10.1016/j.cell.2021.01.047.
27. Vervoort, A. (1980). Tetraploidy in *Protopterus* (Dipnoi). Experientia *36*, 294–296.
28. Armstrong, E.M., Larson, E.R., Harper, H., Webb, C.R., Dohleman, F., Araya, Y., Meade, C., Feng, X., Mukoye, B., Levin, M.J., et al. (2023). One hundred important questions facing plant science: an international perspective. New Phytol. *238*, 470–481. https://doi.org/10.1111/nph.18771.
29. Gómez, M.S., Brown, M.J., Pironon, S., Veselý, P., Bureš, P., Elliott, T.L., Zedek, F., Pellicer, J., Forest, F., Lughadha, E.N., and Leitch, I.J. (2023). Genome size is positively correlated with extinction risk in herbaceous angiosperms. Preprint at. bioRxiv, 1–38. https://doi.org/10.1101/2023.09.10.557053.
30. Novák, P., Guignard, M.S., Neumann, P., Kelly, L.J., Mlinarec, J., Koblížková, A., Dodsworth, S., Kovařík, A., Pellicer, J., Wang, W., et al. (2020). Repeat-sequence turnover shifts fundamentally in species with large genomes. Nat. Plants *6*, 1325–1329. https://doi.org/10.1038/s41477-020-00785-x.
31. Lewin, H.A., Richards, S., Lieberman Aiden, E., Allende, M.L., Archibald, J.M., Bálint, M., Barker, K.B., Baumgartner, B., Belov, K., Bertorelle, G., et al. (2022). The Earth BioGenome Project 2020: Starting the clock. Proc. Natl. Acad. Sci. USA *119*, e2115635118. https://doi.org/10.1073/pnas.2115635118.
32. Kress, W.J., Soltis, D.E., Kersey, P.J., Wegrzyn, J.L., Leebens-Mack, J.H., Gostel, M.R., Liu, X., and Soltis, P.S. (2022). Green plant genomes: What we know in an era of rapidly expanding opportunities. Proc. Natl. Acad. Sci. USA *119*, e2115640118. https://doi.org/10.1073/pnas.2115640118.
33. Wickham, H. (2016). Elegant Graphics for Data Analysis (Springer-Verlag).
34. R Core Team (2020). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
35. Dolezel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. Nat. Protoc. *2*, 2233–2244. https://doi.org/10.1038/nprot.2007.310.
36. Ebihara, A., Ishikawa, H., Matsumoto, S., Lin, S.-J., Iwatsuki, K., Takamiya, M., Watano, Y., and Ito, M. (2005). Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the Vandenboschia radicans complex (Hymenophyllaceae) in Japan and adjacent areas. Am. J. Bot. *92*, 1535–1547. https://doi.org/10.3732/ajb.92.9.1535.
37. Pellicer, J., Powell, R., and Leitch, I.J. (2021). The application of flow cytometry for estimating genome size, ploidy level endopolyploidy, and reproductive modes in plants Molecular Plant Taxonomy. In Methods in Molecular Biology (Humana Press), pp. 325–361.

Please cite this article in press as: Fernández et al., A 160 Gbp fork fern genome shatters size record for eukaryotes, iScience (2024), https://doi.org/10.1016/j.isci.2024.109889

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Triton X-100	Fluka	CAT#93420
2-Mercaptoetanol	Sigma-Aldrich	CAS-60-24-2
Trizma base	Sigma-Aldrich	CAS-77-86-1
Propidium Iodide	Invitrogen	P3566
Polyvinylpyrrolidone	Sigma-Aldrich	CAS-9003-39-8
Sodium sulfite	Sigma-Aldrich	CAS-7757-83-7
Experimental models: Organisms/strains		
Allium cepa "Ailsa Craig"	M. D. Bennett, Royal Botanic Gardens Kew	N/A
Software and algorithms		
FlowMax v.2.11	FlowMax Operating and Analysis Software 2016© QA GmbH	http://www.quantum-analysis.com
ggplot2	Wickham[33]	https://ggplot2.tidyverse.org/
R Software v.3.6.3	R CoreTeam[34]	https://www.r-project.org/
Deposited data		
Flow cytometry output statistics	This paper/Mendeley data	Table S1 https://doi.org/10.17632/hdpgxwpmt5.1

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to the lead contact, Jaume Pellicer (jaume.pellicer@ibb.csic.es).

Materials availability

This study did not generate new physical materials.

Data and code availability

- The source of reference nuclear DNA contents used to map genome size diversity across eukaryotes have been compiled from the Plant DNA C-values Database release 7.1 (https://cvalues.science.kew.org/), the Animal Genome Size Database (www.genomesize.com/) and the Fungal Genome size Database (www.zbi.ee/fungal-genomesize/).
- Accession data based on wild samples and details are provided as follows: Individuals of *Tmesipteris oblanceolata* subsp. *linearifolia* A.F.Braitwh. were collected from a wild population nearby Rivière Bleu de Prony (OH 843, Grande Terre, South Province, New Caledonia). Further details on locality of provenance and collection details are listed on the voucher accession NOU110434, which is deposited in Herbarium NOU (http://publish.plantnet-project.org/project/nou). Fresh samples were kept in sealed plastic bags at 4°C and analyzed by flow cytometry within 7–10 days following collection.
- The article does not report any new code.
- Any additional information required to reanalyze the data reported in this article is available from the lead contact on request.

METHOD DETAILS

Distribution of genome sizes in eukaryotes

The source of reference nuclear DNA contents (i.e., 1C-values) used to plot genome size diversity across eukaryotes depicted in Figure 1 have been compiled from the Plant DNA C-values Database (https://cvalues.science.kew.org/), the Animal Genome Size Database (https://www.genomesize.com/) and the Fungal Genome size Database (http://www.zbi.ee/fungal-genomesize/). Violin plots for each lineage were constructed with ggplot2[33] in R software.[34] Image silhouettes were sourced from PhyloPic (www.phylopic.org).

Please cite this article in press as: Fernández et al., A 160 Gbp fork fern genome shatters size record for eukaryotes, iScience (2024), https://doi.org/10.1016/j.isci.2024.109889

Flow cytometry and genome size data

The nuclear DNA content of *T. oblanceolata* was determined based on the analysis of three individuals. We used flow cytometry and propidium iodide staining of the nuclei, following best practice methods as described in Doležel et al.[35] Briefly, about 1 cm^2 of fresh leaf material of the study species and the calibration standard species (see below) were chopped together in 2 mL of nuclei isolation buffer,[36] filtered through a 100 μm Celltric filter (Sysmex), and stained by adding 100 μL of 1 mg mL^{-1} propidium iodide solution. Nuclei suspensions were then incubated for c. 20 min on ice prior to analysis. The relative fluorescence of stained nuclei was acquired using a CyflowSL Sysmex flow cytometer fitted with a 100 mW green (532 nm) solid-state Cobalt Samba laser. The resulting flow histograms were analyzed using the software FloMax v.2.7. Polygon gating was applied only to exclude debris outside the G_1 fluorescence range after visual inspection of nuclei populations. A minimum of 1,000 nuclei per fluorescence peak were recorded in each analysis. Genome sizes were estimated by calculating the ratio between the mean fluorescence of the G_1 peak for the calibration standard species for which the genome size is known, and the mean fluorescence of the G_1 peak for the nuclei from the target species (i.e., *T. oblanceolata*).

A preliminary analysis was carried out using *Allium cepa* "Ailsa Craig" (1C = 17.06 Gbp[13]) as the calibration standard since this species has the largest genome of the widely used and commonly accepted calibration standards.[37] A fluorescence ratio of at least 9-fold (Figure 2B) was recovered, indicating that the genome size of *T. oblanceolata* is at least 9 times larger than that of *A. cepa*. Since it is recommended that the genome size of the internal calibration standard should be as close as possible to the species of interest to avoid any risk of technical artifacts arising from the non-linearity of the flow cytometer over such a large range, we used *Fritillaria lusitanica* Wikstr as an alternative calibration standard (Figure 2C). Its genome size was estimated to be 50.47 Gbp/1C using *A. cepa* as an internal standard and following the same procedure as outlined above. Finally, to eliminate any further potential technical issues arising from the selection of alternative calibration standards, and to confirm the larger genome size of *T. oblanceolata* compared with the previous record-holder *Paris japonica*,[5] we (i) re-analyzed a sample of *P. japonica* with the newly used calibration standard (i.e., *F. lusitanica*; Figure 2D), and (ii) analyzed a combined sample comprising *P. japonica* and *T. oblanceolata* to confirm that the mean relative peak fluorescence of nuclei from *T. oblanceolata* (peak 2) was higher than that of *P. japonica* (peak 1) (Figure 2E). Note that previous investigations have reported the existence of even larger genomes than that of *T. oblanceolata* in some protists. These older measurements are nowadays considered doubtful due to the sub-optimal techniques used to estimate nuclear DNA content, including the analysis of whole cells (not isolated nuclei) and/or use of fixed versus fresh samples, among others, as already thoroughly discussed in Hidalgo et al.[4]

QUANTIFICATION AND STATISTICAL ANALYSIS

For each run, the nuclear DNA content of *T. oblanceolata* was calculated by dividing the relative fluorescence intensity of the *T. oblanceolata* peak sample by the relative fluorescence intensity of the internal standard peak sample, and the resulting value was then multiplied by the known genome size of the internal standard used (see Table S1). The ultimate genome size of *T. oblanceolata* was obtained by calculating the mean and standard deviation from all the replicates analyzed.