

RESEARCH ARTICLE

Evidence for an ancient whole genome duplication in the cycad lineage

Danielle Roodt^{1,2}, Rolf Lohaus^{3,4,5}, Lieven Sterck^{3,4,5}, Riaan L. Swanepoel^{1,2}, Yves Van de Peer^{2,3,4,5}, Eshchar Mizrachi^{1,2*}

1 Department of Genetics, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Private bag X20, Pretoria, South Africa, **2** Centre for Bioinformatics and Computational Biology, Genomics Research Institute, University of Pretoria, Private bag X20, Pretoria, South Africa, **3** Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, **4** VIB Center for Plant Systems Biology, Ghent, Belgium, **5** Bioinformatics Institute Ghent, Ghent University, Ghent, Belgium

* eshchar.mizrachi@fabi.up.ac.za



OPEN ACCESS

Citation: Roodt D, Lohaus R, Sterck L, Swanepoel RL, Van de Peer Y, Mizrachi E (2017) Evidence for an ancient whole genome duplication in the cycad lineage. PLoS ONE 12(9): e0184454. <https://doi.org/10.1371/journal.pone.0184454>

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: March 9, 2017

Accepted: August 24, 2017

Published: September 8, 2017

Copyright: © 2017 Roodt et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw reads of both transcriptomes are available at the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>) under the submission number SUB2337915.

Funding: The authors acknowledge the UP Research Development Programme (RDP), the Genomics Research Institute (GRI), Multidisciplinary Research Partnership 'Bioinformatics: from nucleotides to networks' Project (no. 01MR0310W) of Ghent University, and funding from the European Union Seventh

Abstract

Contrary to the many whole genome duplication events recorded for angiosperms (flowering plants), whole genome duplications in gymnosperms (non-flowering seed plants) seem to be much rarer. Although ancient whole genome duplications have been reported for most gymnosperm lineages as well, some are still contested and need to be confirmed. For instance, data for ginkgo, but particularly cycads have remained inconclusive so far, likely due to the quality of the data available and flaws in the analysis. We extracted and sequenced RNA from both the cycad *Encephalartos natalensis* and *Ginkgo biloba*. This was followed by transcriptome assembly, after which these data were used to build paralog age distributions. Based on these distributions, we identified remnants of an ancient whole genome duplication in both cycads and ginkgo. The most parsimonious explanation would be that this whole genome duplication event was shared between both species and had occurred prior to their divergence, about 300 million years ago.

Introduction

Whole genome duplications (WGDs) have been prevalent during the evolutionary history of flowering plants, and have even been linked to their origin as well as their fast rise to ecological dominance [1–3]. Furthermore, although the duplication of entire genomes is mostly regarded as an evolutionary dead-end [4–7], it has been proposed that, in times of rapid environmental change, WGDs can confer an important evolutionary advantage [8–11]. This is, for instance, suggested by the fact that many angiosperm lineages show evidence for independent WGD events around the Cretaceous-Paleogene (K-Pg) extinction ~66 million years ago (Mya) [11, 12].

Contrary to the many WGD events recorded for angiosperms, the history of the non-flowering gymnosperms paints a very different picture. Although far fewer gymnosperm species exist today compared to the angiosperms, and as such many lineages containing evidence for WGD events could have been lost, polyploidy events, ancient or more recent, in these seed

Framework Programme (FP7/2007-2013) under European Research Council Advanced Grant Agreement 322739 – DOUBLEUP for support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

plants seem rare. Thus far, *Welwitschia mirabilis* is the only gymnosperm showing evidence for a relatively recent WGD event [13, 14], possibly also overlapping the K-Pg boundary. In any case, this event occurred more recently than the divergence of *Welwitschia* from its closest relative, *Gnetum* (135–110 Mya) [15–17], the genome of which shows no sign of a WGD [13]. Furthermore, *Ephedra*, the third Gnetales genus, also lacks evidence of WGD events [13], excluding very recent duplication events that resulted in the widespread polyploidy seen in extant species of this genus [18–20]. Li *et al.* [13] also provided evidence for independent ancient WGDs in the conifer lineage that may have coincided with the more ancient Permian-Triassic boundary, ~250 Mya. Similarly, as with the angiosperms, these conifer-specific WGDs might have contributed to the survival and success of the conifer lineage during periods of drastic environmental change [13]. The same study found evidence for an ancient WGD in the *Ginkgo* lineage, attributing it to the ancient WGD event proposedly shared by all seed plants [1]. Clear remnants of WGDs in cycads were not uncovered, likely due to the dearth of available public EST data [21], resulting in insufficient resolution to call an ancient WGD event in this lineage.

The cycads were widespread during the Jurassic–Cretaceous, reaching their greatest diversity ~200–65 Mya [22–24]. Today, however, only a mere 348 extant species in ten genera remain [25]. The dramatic decrease in diversity was likely due to challenges such as at least three mass extinction events, as well as the arrival of, and major competition from, the angiosperms. Although the lineage itself dates back ~270 million years, most extant cycad species originated much more recently, most likely within the past 65 million years [22, 26, 27]. Therefore, the popular referral to cycads as living fossils is not entirely accurate, as the lineage itself is ancient but most species originated relatively recently. Their continued survival is somewhat paradoxical, as they have particularly slow growth and cannot compete with the fast growing, rather short-lived angiosperms. Here, we confirm that cycads have undergone an ancient WGD and show that this event was likely shared with *Ginkgo biloba*, preceding the divergence of these lineages.

Results and discussion

We sequenced transcriptome data from two tissues (see [Materials and Methods](#)) of representatives of both *Encephalartos natalensis* (a native cycad species from the Kwazulu-Natal province of South Africa) and *Ginkgo biloba*, and assembled high quality low-redundancy transcriptomes [28]. The *E. natalensis* and *G. biloba* assemblies contained 22,204 and 23,845 transcripts with average sequence lengths of 1,097 and 1,259 bases and average GC contents of 44.23 and 42.52%, respectively. Based on age distributions of paralogs inferred from synonymous substitutions per synonymous site, or so-called K_S distributions [29], a distinct peak with a median K_S of ~0.8 was identified for *E. natalensis* (Fig 1), a clear signature of an ancient WGD event. *G. biloba* showed a similar K_S distribution, and also contained a peak at a K_S of ~0.8 (Fig 1). This distribution is consistent with data reporting the presence of a WGD event in the evolutionary history of ginkgo [13]. Since the K_S peaks for both *E. natalensis* and *G. biloba* were at similar K_S values (Fig 1), this could suggest either a similar timeframe for both WGDs or a shared WGD event in the ancestor of the two lineages.

We built a K_S age distribution of one-to-one orthologs between *E. natalensis* and *G. biloba* to investigate whether the duplication peaks identified in both indicated a shared WGD. Assuming similar substitution rates in the two lineages [30], the time of speciation seems to be slightly younger than the polyploidy events in both *G. biloba* and *E. natalensis* (Fig 1). Cycad and ginkgo are thought to have diverged from one another, and from the other gymnosperms, ~330–270 Mya [16, 17, 27, 31, 32]. Although several studies suggest that cycads diverged

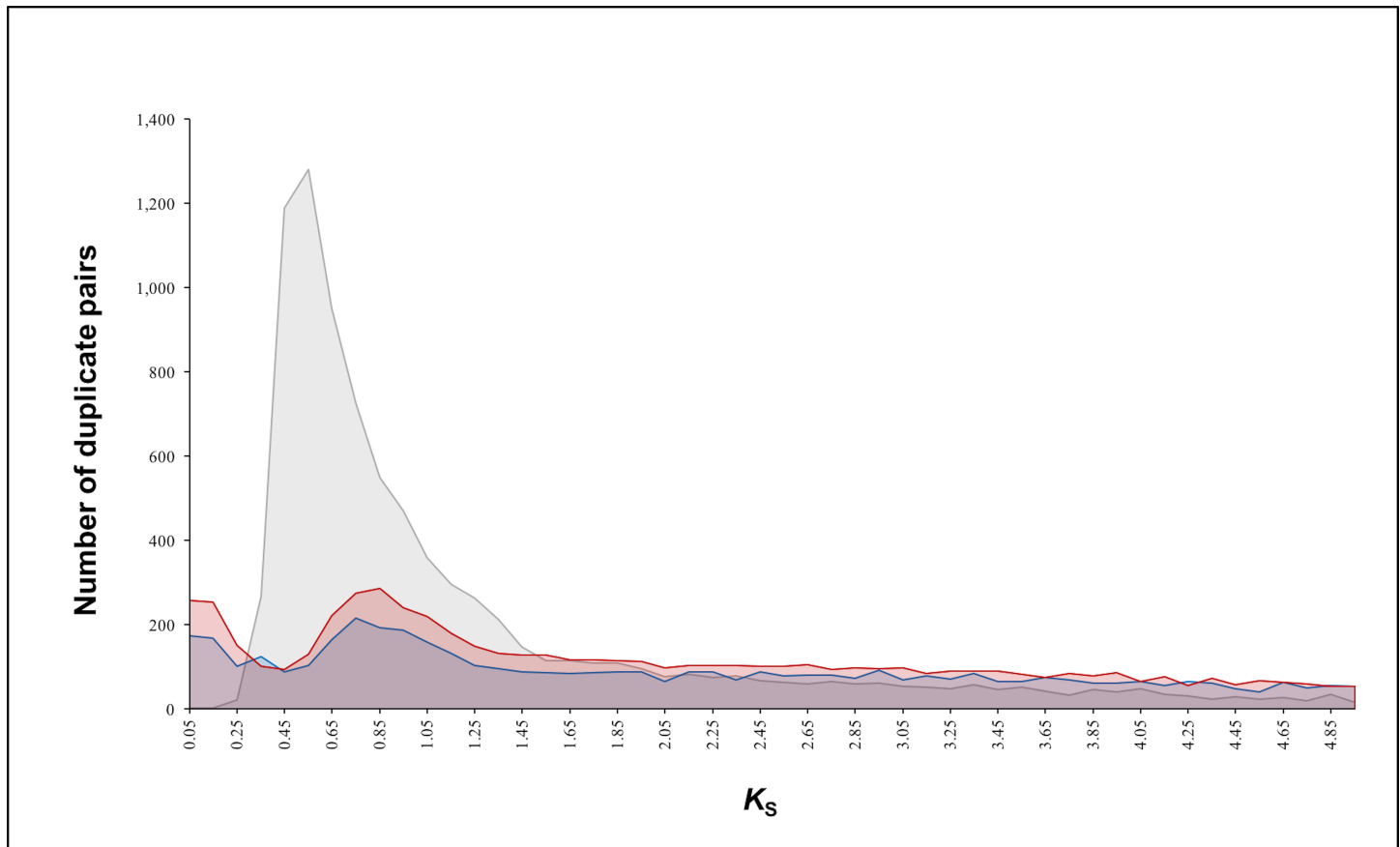


Fig 1. K_S age distributions for *Ginkgo biloba* and *Encephalartos natalensis* transcriptomes. The red graph represents the age distribution of duplicates (in the transcriptome) of *G. biloba*, while the blue graph represents the age distribution of duplicates of *E. natalensis*. The graph in grey denotes the K_S distribution for one-to-one orthologs of *G. biloba* and *E. natalensis*.

<https://doi.org/10.1371/journal.pone.0184454.g001>

earliest during gymnosperm evolution, with the ginkgo lineage following not long thereafter [16, 26, 31–33], others show cycads and ginkgo as having diverged from a common ancestor [17, 34–37] (Fig 2). In either case, the more parsimonious explanation would be to assume that the duplication peaks observed in both lineages represent a shared event, rather than independent ancient duplications that have occurred early in the evolution of both lineages. However, the correct phylogenetic placement of this WGD event is uncertain. The WGD might represent the ancient seed-plant-specific WGD assumed to have occurred ~340 Mya [1], but could also be gymnosperm-specific or nested within the gymnosperms (Fig 2) [38].

Due to the slow substitution rates of most of the gymnosperm taxa [13, 31, 39], these ancient WGD events in the lineages of cycad and ginkgo (Fig 1), and conifers [13] are still discernible as distinct peaks in their K_S distributions (Fig 1, Suppl. Fig S3 in [13]). Slow substitution rates result in gradual genomic change, conserving the remnants of ancient events in the genomes of these species. Because these lineages separated from one another hundreds of millions of years ago, absolute dating of these WGD events is challenging, but they likely indicate a shared event in a common ancestor (Fig 2).

It should be noted that recently a draft genome for *G. biloba* was presented, in which the authors claimed to have found evidence for two WGD events based on the detection of two different peaks in a K_S distribution of paralogs [40]. However, we think their K_S distributions

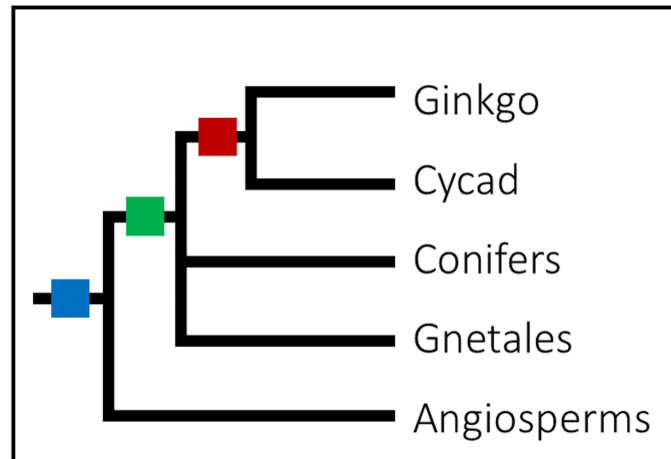


Fig 2. Possible locations for the whole genome duplication uncovered for *Encephalartos natalensis* and *Ginkgo biloba*. Cycad and ginkgo share a common ancestor and are sister to the ancestor of Gnetales and conifers [17, 30, 34, 35, 37, 39]: the inferred polyploidy event could have occurred in the common ancestor of the cycad and ginkgo lineages (red square), the gymnosperm ancestor (green square), or in the seed plant ancestor (blue square). Note that the position of Gnetales remains uncertain (see text for details).

<https://doi.org/10.1371/journal.pone.0184454.g002>

should be interpreted with caution, since the youngest peak, corresponding to a WGD estimated between 147 and 74 Mya [40], is likely artificial and the result of thresholds used to consider genes as duplicates rather than as identical. Furthermore, the inferred WGD dates as obtained by Guan *et al.* [40] should be considered unrealistic. For instance, the authors assume the older genome duplication to be between 735–515 Mya, which would predate the origin of land plants. As we have shown in the current study, the WGD in ginkgo is probably only slightly older (0.2 K_S) than the divergence between cycads and ginkgo (Fig 1).

If the WGD detected here did not occur in the ancestor of the cycads and ginkgo, but earlier in the gymnosperm evolution (green or blue squares in Fig 2), we would expect to see remnants of this event in the Gnetales. Yet the complete absence of an ancient peak in the Gnetales remains difficult to explain. One reason why such evidence is lacking might be the faster rate of evolution in Gnetales, compared to other gymnosperms [39, 41, 42], resulting in any traces of ancient polyploidies to be lost. On the other hand, the placement of the Gnetales within the gymnosperms remains elusive. While different molecular markers have placed them within the conifer lineage (the ‘Gnepine’ and ‘Gnecup’ hypotheses), sister to the conifers (the ‘Gnetifer’ hypothesis), or sister to all other gymnosperms (the ‘Gnetales-sister’ hypothesis) [30, 34, 36, 43–45], certain morphological traits even place them closer to the angiosperms, or as a basal seed plant [46, 47]. Therefore, if the WGD event detected here indeed represents an ancient gymnosperm or seed-plant WGD (see [38]), the lack of evidence for WGDs in the Gnetales, except the more recent one in *Welwitschia*, and the difficulty in resolving their exact phylogenetic placement, could suggest an evolutionary history that is different from the other gymnosperms. In conclusion, despite the fact that early seed plant evolution remains problematic with respect to both phylogenetic relationships and relative and absolute dating of WGD events, we here provide conclusive evidence that cycads have also undergone an ancient WGD.

Materials and methods

Materials from *Encephalartos natalensis* and *Ginkgo biloba* were obtained from plants grown at the Manie van der Schijff Botanical Garden at the University of Pretoria, South Africa. Leaflets

and rachis samples were collected separately from *E. natalensis*, while mature leaf and stem tissue were sampled from three male *G. biloba* trees. RNA was extracted using a standard CTAB RNA extraction method [48], followed by a clean-up step using the Qiagen RNeasy Plus Mini Kit. RNA was sent for sequencing at Novogene, China, generating 150-bp libraries.

RNA-seq libraries from these tissues were used to construct *de novo* transcriptome assemblies for *E. natalensis* (66,583,315 paired end reads; 20 Gbp) and *G. biloba* (98,762,732 paired end reads; 29.6 Gbp). The *de novo* assemblies were constructed with Velvet v1.2.10 [49] followed by Oases v0.2.08 [50] with a k-mer size of 101, retaining only contigs larger than 200 bases. The redundancy of the assemblies was removed with the Evidential gene pipeline and only the primary transcript of each loci was used in further analyses [51].

TransDecoder v.3.0.0 was used to predict coding regions in the transcriptomes of *E. natalensis* and *G. biloba* [52], after which the 19,991 and 23,845 longest coding and peptide sequence transcripts were selected for the species, respectively. K_S age distributions for *E. natalensis* and *G. biloba* were constructed as described previously [29]. Briefly, to construct the paranome an all-against-all BLASTP search was performed of all the longest transcripts with an E-value cutoff of 1×10^{-10} , followed by gene family construction and prediction using the mclblastline pipeline (v10-201, <http://micans.org/mcl>) [53]. Each gene family was aligned using MUSCLE (v3.8.31). To obtain K_S estimates for all pairwise comparisons in gene families, maximum likelihood estimation was performed using the CODEML program of the PAML package (v4.4c) [54, 55]. Gene families were then subdivided into subfamilies for which K_S estimates between members did not exceed a value of 5. To correct for the redundancy of K_S values (a gene family of n members produces $n(n-1)/2$ pairwise K_S estimates for $n-1$ retained duplication events), a phylogenetic tree was constructed for each subfamily using PhyML [56] under default settings. For each duplication node in the resulting phylogenetic tree, all m K_S estimates between the two child clades were added to the K_S distribution with a weight of $1/m$ (where m is the number of K_S estimates for a duplication event), so that the weights of all K_S estimates for a single duplication event sum up to one. One-to-one orthologous pairs between *E. natalensis* and *G. biloba* were created by performing a reciprocal best hit BLASTP analysis of the longest translated transcripts from one species against the other. Valid orthologous pairs were then identified as having at least 30% identity over 150 amino acids.

Accession numbers

Raw reads of both transcriptomes are available at the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>) under the submission number SUB2337915.

Acknowledgments

The authors thank Jason Samson, Philip Rousseau and the Manie van der Schijff Botanical Gardens of the University of Pretoria for plant material and assisting with sampling. EM acknowledges the UP Research Development Programme (RDP) and the Genomics Research Institute (GRI) for support. YVdP acknowledges the Multidisciplinary Research Partnership 'Bioinformatics: from nucleotides to networks' Project (no. 01MR0310W) of Ghent University, and funding from the European Union Seventh Framework Programme (FP7/2007-2013) under European Research Council Advanced Grant Agreement 322739 –DOUBLEUP.

In memory of Philip Rousseau.

Author Contributions

Conceptualization: Eshchar Mizrahi.

Data curation: Danielle Roodt, Rolf Lohaus, Riaan L. Swanepoel.

Formal analysis: Danielle Roodt.

Funding acquisition: Eshchar Mizrachi.

Investigation: Danielle Roodt, Rolf Lohaus, Yves Van de Peer, Eshchar Mizrachi.

Methodology: Danielle Roodt, Rolf Lohaus, Lieven Sterck, Riaan L. Swanepoel.

Project administration: Yves Van de Peer, Eshchar Mizrachi.

Resources: Yves Van de Peer, Eshchar Mizrachi.

Software: Rolf Lohaus, Lieven Sterck, Yves Van de Peer.

Supervision: Yves Van de Peer, Eshchar Mizrachi.

Validation: Rolf Lohaus.

Writing – original draft: Danielle Roodt.

Writing – review & editing: Danielle Roodt, Rolf Lohaus, Lieven Sterck, Riaan L. Swanepoel, Yves Van de Peer, Eshchar Mizrachi.

References

1. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. *Nature*. 2011; 473(7345):97–100. <https://doi.org/10.1038/nature09916> PMID: 21478875
2. De Bodt S, Maere S, Van de Peer Y. Genome duplication and the origin of angiosperms. *Trends Ecology Evolution*. 2005; 20(11):591–7. <https://doi.org/10.1016/j.tree.2005.07.008> PMID: 16701441.
3. Soltis DE, Bell CD, Kim S, Soltis PS. Origin and early evolution of angiosperms. *Annals of the New York Academy of Sciences*. 2008; 1133:3–25. <https://doi.org/10.1196/annals.1438.005> PMID: 18559813.
4. Mayrose I, Zhan SH, Rothfels CJ, Arrigo N, Barker MS, Rieseberg LH, et al. Methods for studying polyploid diversification and the dead end hypothesis: a reply to Soltis et al. (2014). *New Phytol*. 2015; 206(1):27–35. <https://doi.org/10.1111/nph.13192> PMID: 25472785.
5. Vanneste K, Maere S, Van de Peer Y. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Phil Trans R Soc B*. 2014; 369(1648):20130353. <https://doi.org/10.1098/rstb.2013.0353> PMID: 24958926
6. Arrigo N, Barker MS. Rarely successful polyploids and their legacy in plant genomes. *Curr Opin Plant Biol*. 2012; 15(2):140–6. <https://doi.org/10.1016/j.pbi.2012.03.010> PMID: 22480430
7. Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, et al. Recently formed polyploid plants diversify at lower rates. *Science*. 2011; 333(6047):1257–. <https://doi.org/10.1126/science.1207205> PMID: 21852456
8. Comai L. The advantages and disadvantages of being polyploid. *Nat Rev Genet*. 2005; 6(11):836–46. <https://doi.org/10.1038/nrg1711> PMID: 16304599
9. Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, et al. Polyploidy and angiosperm diversification. *American Journal of Botany*. 2009; 96(1):336–48. <https://doi.org/10.3732/ajb.0800079> PMID: 21628192.
10. Madlung A. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity (Edinb)*. 2013; 110(2):99–104. <https://doi.org/10.1038/hdy.2012.79> PMID: 23149459; PubMed Central PMCID: PMC3544449.
11. Van de Peer Y, Mizrachi E, Marchal K. The evolutionary conundrum of polyploidy. *Nat Rev Genet*. 2017. <https://doi.org/10.1038/nrg.2017.26> PMID: 28502977
12. Vanneste K, Baele G, Maere S, Van de Peer Y. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Research*. 2014; 24(8):1334–47. <https://doi.org/10.1101/gr.168997.113> PMID: 24835588
13. Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, et al. Early genome duplications in conifers and other seed plants. *Science Advances*. 2015; 1(10):e1501084. <https://doi.org/10.1126/sciadv.1501084> PMID: 26702445

14. Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, et al. Widespread genome duplications throughout the history of flowering plants. *Genome Research*. 2006; 16(6):738–49. <https://doi.org/10.1101/gr.4825606> PMID: 16702410; PubMed Central PMCID: PMCPMC1479859.
15. Rydin C, Pedersen KR, Friis EM. On the evolutionary history of *Ephedra*: Cretaceous fossils and extant molecules. *P Natl Acad Sci USA*. 2004; 101(47):16571–6.
16. Schneider H, Schuettpeiz E, Pryer KM, Cranfill R, Magallon S, Lupia R. Ferns diversified in the shadow of angiosperms. *Nature*. 2004; 428(6982):553–7. <https://doi.org/10.1038/nature02361> PMID: 15058303.
17. Burleigh JG, Barbazuk WB, Davis JM, Morse AM, Soltis PS. Exploring diversification and genome size evolution in extant gymnosperms through phylogenetic synthesis. *Journal of Botany*. 2012;2012.
18. Wu H, Ma Z, Wang MM, Qin AL, Ran JH, Wang XQ. A high frequency of allopolyploid speciation in the gymnospermous genus *Ephedra* and its possible association with some biological and ecological features. *Molecular Ecology*. 2016; 25(5):1192–210. <https://doi.org/10.1111/mec.13538> PMID: 26800145
19. Huang J, Giannasi DE, Price RA. Phylogenetic relationships in *Ephedra* (Ephedraceae) inferred from chloroplast and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*. 2005; 35(1):48–59. <https://doi.org/10.1016/j.ympev.2004.12.020> PMID: 15737581
20. Khoshoo T. Polyploidy in gymnosperms. *Evolution*. 1959:24–39.
21. Brenner ED, Stevenson DW, McCombie RW, Katari MS, Rudd SA, Mayer KF, et al. Expressed sequence tag analysis in *Cycas*, the most primitive living seed plant. *Genome Biology*. 2003; 4(12):R78.1. <https://doi.org/10.1186/gb-2003-4-12-r78> PMID: 14659015; PubMed Central PMCID: PMCPMC329417.
22. Nagalingum NS, Marshall CR, Quental TB, Rai HS, Little DP, Mathews S. Recent synchronous radiation of a living fossil. *Science*. 2011; 334(6057):796–9. <https://doi.org/10.1126/science.1209926> PMID: 22021670.
23. Jones D. *Cycads of the world: Ancient plant in today's landscape*. 2002. Epub 2.
24. Taylor EL, Taylor TN, Krings M. *Paleobotany: the biology and evolution of fossil plants*. 2009.
25. Calonje M, Stevenson DW, Stanberg L. *The World List of Cycads*. 2013. Epub Online edition [Internet].
26. Crisp MD, Cook LG. Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms. *New Phytol*. 2011; 192(4):997–1009. <https://doi.org/10.1111/j.1469-8137.2011.03862.x> PMID: 21895664
27. Condamine FL, Nagalingum NS, Marshall CR, Morlon H. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. *BMC Evolutionary Biology*. 2015; 15(1):65.
28. Vanneste K, Sterck L, Myburg Z, Van de Peer Y, Mizrahi E. Horsetails Are Ancient Polyploids: Evidence from *Equisetum giganteum*. *The Plant Cell*. 2015; 27(6):1567–78. <https://doi.org/10.1105/tpc.15.00157> PMID: 26002871
29. Vanneste K, Van de Peer Y, Maere S. Inference of genome duplications from age distributions revisited. *Mol Biol Evol*. 2013; 30(1):177–90. <https://doi.org/10.1093/molbev/mss214> PMID: 22936721.
30. Li Z, De la Torre AR, Sterck L, Canovas FM, Avila C, Sierra IM, et al. Single-copy genes as molecular markers for phylogenomic studies in seed plants. *Genome Biology and Evolution*. 2017; 9(5):1130–47. <https://doi.org/10.1093/gbe/evx070> PMID: 28460034
31. Magallón SA, Hilu KW, Quandt D. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany*. 2013; 100(3):556–73. <https://doi.org/10.3732/ajb.1200416> PMID: 23445823
32. de la Torre-Barcelona JE, Kolokotronis SO, Lee EK, Stevenson DW, Brenner ED, Katari MS, et al. The impact of outgroup choice and missing data on major seed plant phylogenetics using genome-wide EST data. *PLOS ONE*. 2009; 4(6):e5764. <https://doi.org/10.1371/journal.pone.0005764> PMID: 19503618; PubMed Central PMCID: PMCPMC2685480.
33. Chaw S, Parkinson CL, Cheng Y, Vincent TM, Palmer JD. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc Natl Acad Sci U S A*. 2000; 97(8):4086–91. PMID: 10760277
34. Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. *P Natl Acad Sci USA*. 2014; 111(45):E4859–E68. <https://doi.org/10.1073/pnas.1323926111> PMID: 25355905
35. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. From algae to angiosperms—inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evolutionary Biology*. 2014; 14:23. <https://doi.org/10.1186/1471-2148-14-23> PMID: 24533922; PubMed Central PMCID: PMCPMC3933183.

36. Xi ZX, Rest JS, Davis CC. Phylogenomics and Coalescent Analyses Resolve Extant Seed Plant Relationships. PLOS ONE. 2013; 8(11). <https://doi.org/10.1371/journal.pone.0080870> PMID: 24278335
37. Wu C-S, Chaw S-M, Huang Y-Y. Chloroplast phylogenomics indicates that *Ginkgo biloba* is sister to cycads. Genome Biology and Evolution. 2013; 5(1):243–54. <https://doi.org/10.1093/gbe/evt001> PMID: 23315384
38. Ruprecht C, Lohaus R, Vanneste K, Mutwil M, Nikoloski Z, Van de Peer Y, et al. Revisiting ancestral polyploidy in plants. Science Advances. 2017; 3(7):e1603195. <https://doi.org/10.1126/sciadv.1603195> PMID: 28695205
39. De La Torre AR, Li Z, Van de Peer Y, Ingvarsson PK. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. Mol Biol Evol. 2017; 34(6):1363–77. <https://doi.org/10.1093/molbev/msx069> PMID: 28333233
40. Guan R, Zhao Y, Zhang H, Fan G, Liu X, Zhou W, et al. Draft genome of the living fossil *Ginkgo biloba*. Gigascience. 2016; 5(49):1.
41. Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 2002; 19(1):101–9. PMID: 11752195.
42. Magallón SA, Sanderson MJ. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. Evolution. 2005; 59(8):1653–70. PMID: 16329238.
43. Bowe LM, Coat G, dePamphilis CW. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proceedings of the National Academy of Science of the United States of America. 2000; 97(8):4092–7. PMID: 10760278; PubMed Central PMCID: PMC18159.
44. Chaw S, Zharkikh A, Sung HM, Lau TC, Li WH. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol Biol Evol. 1997; 14(1):56–68. PMID: 9000754.
45. Lee EK, Cibrian-Jaramillo A, Kolokotronis SO, Katari MS, Stamatakis A, Ott M, et al. A functional phylogenomic view of the seed plants. PLOS Genet. 2011; 7(12):e1002411. <https://doi.org/10.1371/journal.pgen.1002411> PMID: 22194700; PubMed Central PMCID: PMC3240601.
46. Magallón SA, Sanderson MJ. Relationships among seed plants inferred from highly conserved genes: sorting conflicting phylogenetic signals among ancient lineages. American Journal of Botany. 2002; 89(12):1991–2006. <https://doi.org/10.3732/ajb.89.12.1991> PMID: 21665628
47. Doyle JA. Seed plant phylogeny and the relationships of Gnetales. International Journal of Plant Sciences. 1996:S3–S39.
48. Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. Plant Molecular Biology Reporter. 1993; 11(2):113–6.
49. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research. 2008; 18(5):821–9. <https://doi.org/10.1101/gr.074492.107> PMID: 18349386
50. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. 2012; 28(8):1086–92. <https://doi.org/10.1093/bioinformatics/bts094> PMID: 22368243; PubMed Central PMCID: PMC3324515.
51. Gilbert D. EvidentialGene: mRNA Transcript Assembly Software 7th annual arthropod genomics symposium. 2013:1. 10.7490/f1000research.1112594.1.
52. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols. 2013; 8(8):1494–512. <https://doi.org/10.1038/nprot.2013.084> PMID: 23845962; PubMed Central PMCID: PMC3875132.
53. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002; 30(7):1575–84. PMID: 11917018; PubMed Central PMCID: PMC101833.
54. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol. 1994; 11(5):725–36. PMID: 7968486.
55. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007; 24(8):1586–91. <https://doi.org/10.1093/molbev/msm088> PMID: 17483113.
56. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology. 2010; 59(3):307–21. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638.