Chloroplast Genome-Based Molecular Baraminology Analysis of Proteales

Matthew Cserhati

Key Words: chloroplast, molecular baraminology, Nelumbonaceae, Platanaceae, Proteaceae, Sabiaceae

Accepted for publication June 10, 2024

Abstract

Proteales (or proteans) is an order of eudicot angiosperm plants that have economic and medicinal importance. The chloroplast genome of 18 Proteales species, together with four outgroup species from the genus *Cucurbita* were downloaded, aligned, and clustered. According to the results, the Hopkins clustering statistic was 0.801, which reflects decently good clustering. Five putative groups were found: *Platanus, Sabia, Meliosma, Nelumbo,* and the family Proteaceae (*Macadamia+Grevillea+Helicia*). Similarities based on pollen morphology also lend support to this division of Proteales species into these five tentative holobaramins. Hybridization data within the genera *Platanus* and *Nelumbo* and between *Helicia* and *Macadamia* also imply that these groups may be putative holobaramins. Further evidence from mitochondrial genomes, nuclear genes, or more hybridization data would strengthen these clusters.

Simple Summary

Scientists are trying to understand how plants are grouped into "created kinds" (also called **holobaramins**), which are groups of species that can reproduce with each other but not with other groups. To do this for a group of plants called **Proteales** (which includes macadamia nut trees and lotus flowers), they studied the plants' **chloroplast genomes**. Chloroplasts are like tiny power factories in plant cells that produce sugar, and their DNA blueprint is very stable and similar within a kind, making it useful for identifying these groups. By looking at the chloroplast DNA of 18 Proteales species and four outgroup species from the *Cucurbita* genus (like gourds), the researchers found **five likely "kinds"** within Proteales: **Platanus** (plane trees), **Sabia** plants, **Nelumbo** (lotus flowers), and a group from the **Proteaceae family** that includes *Macadamia*, *Grevillea*, and *Helicia* species. For example, the *Platanus* trees show high genetic similarity and are known to hybridize, supporting them as a kind, and the *Nelumbo* lotuses also form a strong group with fertile hybrids. Although the Proteaceae group is promising due to its unique pollen, the scientists emphasized that the findings for some of these groups are **tentative** and need **more evidence** from other DNA sources (like mitochondrial or nuclear DNA) or more observations of natural hybridization to be completely sure of these "kind" boundaries.

Introduction

Baraminology is the study of the created kinds, as mentioned in the Genesis creation account. A kind, or holobaramin, can be described as a reproductive community consisting of species that are continuous with one another and discontinuous with all other species (Wood, 2006). Studies in baraminology aim to use various kinds of data (the Bible, hybridization data, DNA sequences, morphological, ecological, and behavioral data) to classify species into their appropriate baramins.

Proteales (proteans) is an order of eudicot angiosperm plants that live in North America, Southeast Europe, the Middle East, Australia, South Africa, Madagascar, and Southeast Asia. Taxonomists usually assign four families to this group, including Proteaceae, Platanaceae, Nelumbonaceae, and Sabiaceae (The Angiosperm Phylogeny Group, 2009). These families in turn consist of 85 genera and around 1,750 species. The order is dominated by the family Proteaceae with 80 genera and about 1,615 species.

Some of these species are used as medicinal plants (Wen et al., 2016),

such as *Sabia parviflora* Wall. ex Roxb. (Chen et al., 2021). Some Proteales species exhibit high antifungal activity (Pauw and Eloff, 2014), but not so much antibacterial activity. Other species are of economic importance due to their ornamental qualities, timber, and oil-lubricating qualities (Cheng et al., 2023).

These plants are morphologically diverse trees and shrubs, with small individual flowers that are combined into dense floral clusters called inflorescences. Well-known species such as the silky oak (*Grevillea robusta*), the macadamia nut tree (*Macadamia ternifolia*), the oriental plane tree (*Platanus orientalis*), and the sacred lotus (*Nelumbo nucifera*) belong to Proteales. Researchers have noted the gaps in gross morphology between the four families that belong to this order (Gobo et al., 2023).

Several hybrid species are known from these four families. For example, the London plane (*Platanus x hispanica*) is a hybrid of the American sycamore (*P. occidentalis*) and the Oriental plane tree (*P. orientalis*). There is also evidence that *P. rzedowskii* is a hybrid of *P. mexicana* and *P. occidentalis* (Grimm and Denk, 2008).

The chloroplast genome is highly conserved in terms of genome structure, GC [Guanine-Cytosine] content, gene content, and gene order (Wicke et al., 2011; Chen et al., 2022). This means that it is possible that all species within a given holobaramin all have very similar chloroplast genome structures. Since chloroplast genomes are independent of the nuclear genome, they can be used to identify species, and map phylogenetic relationships between various plant groups (Parks et al., 2009; Daniell et al., 2016).

Previous chloroplast genome studies include 163 species from the order Liliales (Cserhati, 2023) and 28 species from Cucurbitaceae (Cserhati, 2024, submitted). Cucurbit chloroplast genomes have also been analyzed (Brophy et al., 2023). A study of chloroplast genomes from 306 Myrtales species and the cytochrome oxidase I (COI) gene from 18 species was also analyzed by Di Martino et al. (2024).

Proteales was chosen in this study because it contains a much smaller number of species than other chloroplast genome-based studies (Cserhati, 2023, 2024; Di Martino et al., 2024). When a larger number of groups are analyzed, the discovered groups may be an artifact of a large number of species. In this study, 27 species of Proteales were analyzed, which is about ten times less than the approximately 300 species studied in Myrtales (Di Martino et al., 2024).

Biblical Analysis

Species from the order Proteales are not mentioned in the Bible. However, along with all other plants, they were created on Day 3 of Creation Week.

Materials and Methods

The chloroplast genome sequences of 18 species from Proteales were downloaded from NCBI at: https://www.ncbi.nlm.nih.gov/nucleotide. The species name, the accession, the corresponding family, GC%, and genome length are listed in Supplementary File 1. Four species from the genus *Cucurbita* were used as a control outgroup, since they are known to belong to a different baramin.

The genome sequences were aligned using MAFFT, version 7.505 (Katoh et al., 2002). Using a self-written R script, the sequence distance matrix was assembled using the 'dist.alignment' method from the 'seqinr' package. Figures 1 and 2 were created using ggplot2, and the heatmap in Figure 2 was created using the 'ward.D2' clustering algorithm in the heatmap command. See Cserhati (2023) for a more detailed description of the methods used in this analysis.

Silhouette analysis was performed in R using the 'silhouette' method. Silhouette plots were calculated using the 'fviz_silhouette' method. Two to eleven clusters were analyzed to calculate their average silhouette width, and the cluster number with the maximum value was chosen for further analysis.

Supplementary File 1, the supplementary figures and the R script used in this analysis can be found online on the Zenodo website at https://zenodo.org/records/11495534.

Results and Discussion

In Figure 1, we can see the 22 species in this study plotted for GC% and chloroplast genome length. While we cannot discern whether many species cluster into their own group, we do see that members of the genera *Sabia*, *Platanus*, and *Nelumbo* separate from one another (see Table I). These groups might be putative holobaramins. The outgroup *Cucurbita* also forms its own cluster in the lower left. It is puzzling that the two species of *Meliosma* (*M. aff. cuneifolia* and *M. oldhamii*) differ from one another somewhat significantly (although *M. aff. cuneifolia* does cluster near *Sabia* from the same family).

A total of seven putative clusters were found in our data, as shown in Figure 2. The Hopkins clustering statistic was 0.801, which indicates good clustering. Of these seven groups, only three (*Macademia+Grevillea+Helicia*, *Nelumbo*, and *Cucurbita*) were statistically significant at the 5% level. Two other groups had less than three species, therefore their statistics were not robust. These clusters and accompanying statistics can be seen in Supplementary File 1. Silhouette plot analysis revealed that the maximum average silhouette width was 0.66 for seven clusters. Two to eleven clusters were analyzed to estimate the optimal number of clusters present in the data set. (Figure 3).

The gene content and gene order are the same in the genera *Macadamia*, *Platanus*, and *Nelumbo*, which share 79 protein-coding genes, 30 tRNA, and 4 rRNA genes. Size differences are due to expansion of the inverted repeat (IR) regions, and reduction in the large and small single copy regions (LSC and SSC) in *Macadamia* relevant to *Platanus* (Nock et al., 2014).

Two genomes from the genus *Sabia* (*Sabia yunnaensis* and *Sabia parviflora*) form a clear cluster at the top right of the heatmap in Figure 2. The chloroplast genomes of *Sabia* are relatively conserved, with 85 coding genes, 37 tRNA genes, and eight rRNA genes. Terpenoids, alkaloids, and flavonoids have been extracted from *Sabia* to treat rheumatoid arthritis, trauma, hepatitis, and other illnesses. The GC% doesn't vary much and is between 38.56%–38.73%. No specific structural variation has been found in this genus (Chen et al., 2022). The clustering added *Meliosma aff. cuneifolia* to this group, but the p-value was insignificant (p=0.18). Therefore these two genera more likely belong to separate holobaramins.

Two species from the genus *Meliosma* (*Meliosma oldhamii* and *Meliosma aff. cuneifolia*, see top right, Figure 2) showed a large difference between one another, so that they were classified initially as members of two separate holobaramins. Their sequence similarity is 0.671. The sequence similarity between *M. oldhamii* and *Sabia yunnaenesis* is 0.669, not much different. It appears that just as *M. oldhamii* and *S. yunnaensis* belong to separate baramins, so would the two *Meliosma* species. The big question is what could cause such a large sequence difference between these two species? Both species have 85 protein-coding genes, 37 tRNA genes, and eight rRNA genes. Despite overall structural similarities, seven protein-coding genes (ndhB, rps12, rps7, rp12, rp123, rps19, and ycf2), as well as seven tRNAs (trnA-UGC, trnI-CAU, trnI-GAU, trnL-CAA, trnN-GUU, trnR-ACG, and trnV-GAC), and four rRNA genes (23S, 16S, 5S, and 4.5S) are duplicated in *M. oldhamii* (Cheng et al., 2023), which may explain the differences in the chloroplast genome

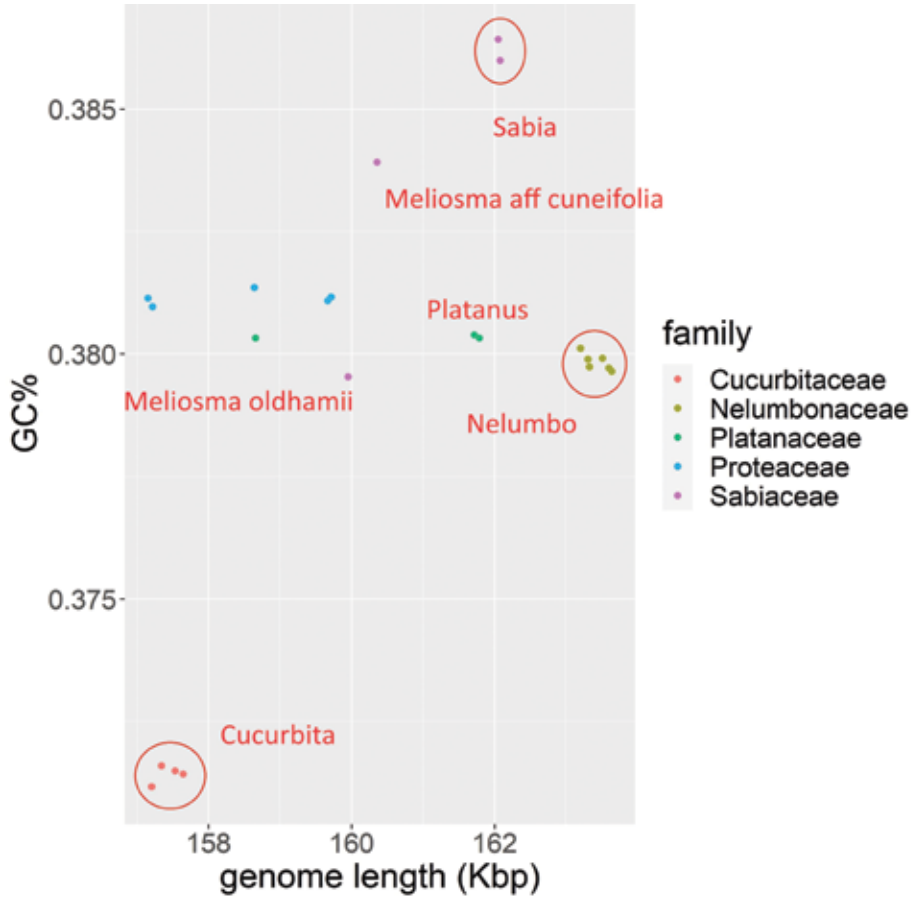


Figure 1. Chloroplast genome length (Kbp) and GC% [Guanine + Cytosine %] of the 18 Proteales species (plus the four outgroup *Cucurbita* species) included in this study. Species forming visible groups are circled.

Table I. Average genome length and GC% of several genera on the genome length/GC% plot.

Group/species	Genome length	GC%
Sabia	162,068±18 bp	38.62±0.031%
Nelumbo	163,432±160 bp	37.98±0.016%
Platanus	161,755±37 bp	38.01±0.003%
Cucurbita	157,431±196 bp	37.14±0.018%
Meliosma aff cuneofolia	160,358 bp	38.39%
Meliosma oldhamii	159,951 bp	37.95%

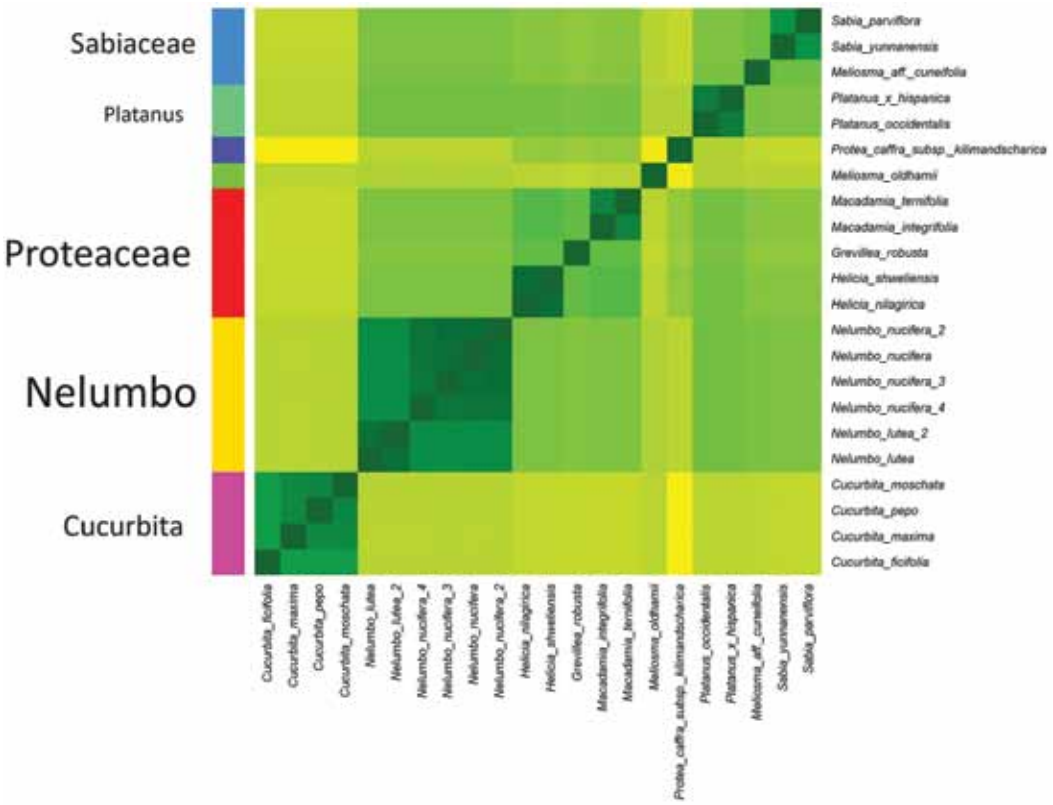
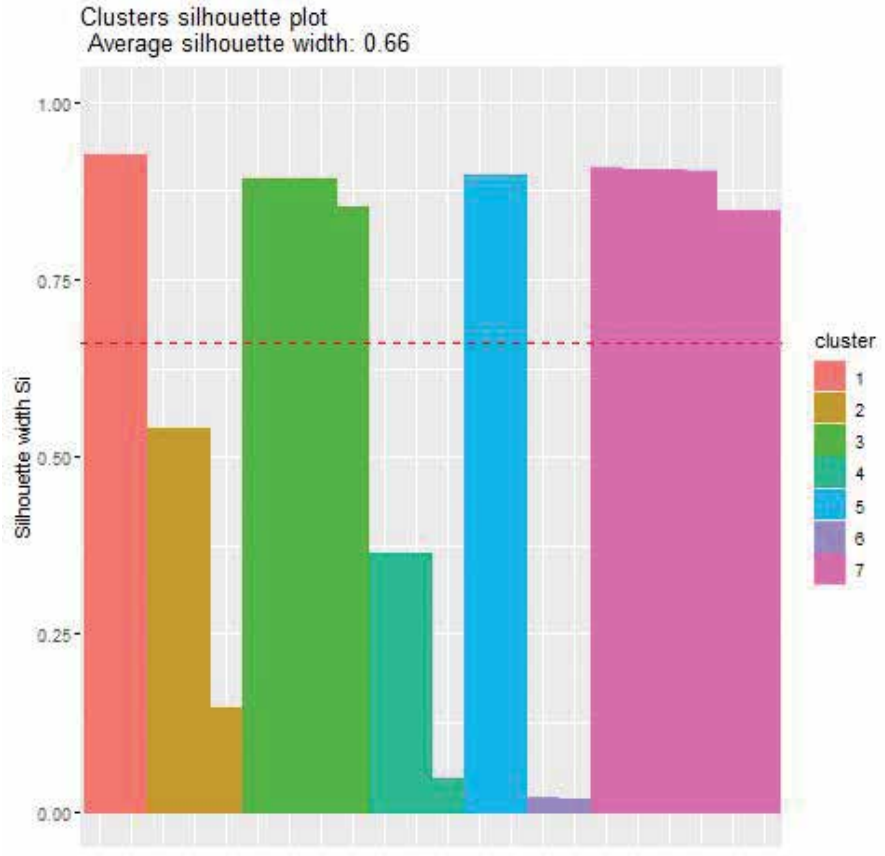


Figure 2 ***(right)*****. Heatmap showing the baraminic relationships of the 21 species in this study. Darker, greener colors indicate high sequence similarity between species, which belong to the same putative holobaramin. Lighter, yellower colors indicate lower sequence similarity values representing species from two different holobaramins.**

Figure 3 ***(lower right)*****. Silhouette plot showing the maximum average silhouette width value of 0.66 for an optimal number of seven clusters.**



sequence that contribute to such a large sequence distance. Taking this into account could mean that the two *Meliosma* species belong to the same holobaramin, after all.

The genus *Platanus* is made up of two species (second cluster from top right in Figure 2), *Platanus occidentalis* and *P. x hispanica*. The sequence similarity value between these two species is 0.966. The fossil record of *Platanus* is more diverse, in line with mass extinctions happening in the past, possibly during the Genesis Flood (Grimm and Denk, 2008). Some evidence based on sequence comparisons of the internal transcribed spacer (ITS) suggest widespread hybridization between various members of this genus (Grimm and Denk, 2008), suggesting that it is a holobaramin. The pollen in this genus

is generally uniform, especially with regards to the pollen size, the pattern of the reticulum, and the shape and size of the colpi. However, the pollen of *P. kerrii* differs from other species based on its high reticulum and folded shape, although this species was represented by only one sample in the study of Denk and Tekleva (2006).

The genus *Nelumbo* (group #4 from top right, Figure 2) consists of six genomes from two species, forming a well-defined cluster ($p=1.7x10^{-30}$). This genus is considered to be a living fossil, not having changed in morphology over its historical lifespan (Sanderson, 2001; Nock et al., 2014). The genus contains only two species, *N. lutea* (the American lotus) and *N. nucifera* (the sacred lotus), which have fertile hybrids (Xue et al., 2012). Several authors place *Nelumbo* into its own family because of its unique triaperturate pollen. Others think that any physical similarities in pollen between *Nelumbo* and water lilies from the order Nympheales are superficial (Kreunen and Osborn, 1999).

The two species from *Helicia*, *H. nilagirica* and *H. shweliensis*, form a significant cluster together with *Grevillea robusta*, *Macadamia integrifolia*, and *M. ternifolia* ($p=2.8x10^{-5}$) in Figure 2. All five of these species are members of the family Proteaceae. Niu and Liu (2020) also found that *H. nilagirica* and *M. integrifolia* cluster together very closely. They suggest that *H. nilagirica* may be used as rootstock or as a gene donor in macadamia breeding. This suggests that these two species may be able to hybridize, meaning that they might possibly be part of the same holobaramin. This grouping is highly tentative, since hybrids between *Helicia*, *Macadamia*, and *Grevillea* have not been observed. However, the porate structure of pollen from Proteaceae as well as the pattern of aperture positions on the meiotic tetrads of this family is unique among angiosperms (Sauquet and Cantrill, 2007). This may indicate the holobaraminic (and also monobaraminic) status of the family Proteaceae, although one must keep in mind that pollen may vary even among species within the same group. A monobaramin is a subset of a holobaramin, something like a lineage of species within the larger group.

Protea caffra subsp. kilimanscharica forms a cluster all by itself. In the lower right corner of Figure 2, the four outgroup species from the genus *Cucurbita* can be seen. These form a highly significant cluster ($p=6.7x10^{-9}$) and are separated from the Proteales by the lowest levels of similarity in the heatmap analysis.

Conclusion

The chloroplast genomes of 18 species of Proteales were analyzed. Several putative holobaramins were found at the level between the genus and the family. These include the genera *Platanus* (Platanaceae), *Sabia* (Sabiaceae), *Meliosma* (Sabiaceae) and *Nelumbo* (Nelumbonaceae), and the family Proteaceae (*Macadamia+Grevillea+Helicia*). Since two of these four genera belong to different families, and since the family is generally regarded as the level of the kind, this may indicate that these groups are holobaramins. The last group is very tentative, and more additive evidence are necessary to verify whether these three genera are either monobaramins within the same holobaramin or separate holobaramins altogether. The baraminic delineations of these groups may be solidified with more additive evidence, such as the examination of mitochondrial or hybridization data, or individual nuclear genes. Due to the small number of species and the limited amount of data, this is a preliminary study on the molecular baraminology of Proteales. Further analysis of mitochondrial and nuclear DNA would help further establish the present findings.

References

The Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161:105–121. https://doi.org/10.1111/j.1095–8339.2009.00996.x.

Brophy, Timothy R., Jack R. Gregory, and Brigitte Townsend. 2023. Hybridization and genetic distances suggest one large monobaramin in the gourd family (Cucurbitales: Cucurbitaceae) [poster]. In J.H. Whitmore (editor), *Proceedings of the Ninth International Conference on Creationism*, pp. 657–658. Cedarville University International Conference on Creationism, Cedarville, Ohio.

Chen, Q., C. Chen, B. Wang, Z. Wang, W. Xu, Y. Huang, and Q. Sun. 2022. Complete chloroplast genomes of 11 Sabia samples: Genomic features, comparative analysis, and phylogenetic relationship. *Frontiers in Plant Science* 13: 1052920. https://doi.org/10.3389/fpls.2022.1052920.

Chen, Q., W. Xu, C. Zhao, B. Wang, C. Chen, Q. Liu, Q. Sun, Q., and Y. Huang. 2021. Complete chloroplast genome of medicinal plant *Sabia parviflora* Wall. ex Roxb. (Sabiaceae). *Mitochondrial DNA. Part B, Resources*, 6(7): 1924–1925. https://doi.org/10.1080/23802359.2021.1935350.

Cheng, Y., X. Xu, L. Tong, L. Tian, C. Xia, and X. Yu. 2023. Complete chloroplast genome and phylogenetic analysis of *Meliosma oldhamii* Miq. ex Maxim. (Sabiaceae). *Mitochondrial DNA. Part B, Resources* 8(11):1306–1310. https://doi.org/10.1080/23802359.2023.2281034.

Cserhati, M. 2023. Chloroplast genome-based baraminology study of Liliales. *Creation Research Society Quarterly* 60(2):84–96.

Cserhati, M. 2024. Baraminology of Cucurbitaceae based on chloroplast genome analysis. *Answers Research Journal*, *submitted*.

Daniell, H., C.S. Lin, M. Yu, and W.J. Chang. 2016. Chloroplast genomes:

Diversity, evolution, and applications in genetic engineering. *Genome Biology* 17(1):134. https://doi.org/10.1186/s13059–016–1004–2.

Denk, T., and M.V. Tekleva. 2006. Comparative pollen morphology and ultrastructure of Platanus: Implications for phylogeny and evaluation of the fossil record. Grana 45(3):195–221. https://doi.org/10.1080/00173130600873901.

Gobo, W.V., L. Kunzmann, R. Iannuzzi, T.B. Dos Santos, D.M. da Conceição, D. Rodrigues do Nascimento Jr, W.F. da Silva Filho, J.B. Bachelier, and C. Coiffard. 2023. A new remarkable Early Cretaceous nelumbonaceous fossil bridges the gap between herbaceous aquatic and woody protealeans. *Scientific Reports* 13(1): 8978. https://doi.org/10.1038/s41598–023–33356-z.

Grimm, G.W., and T. Denk. 2008. Its evolution in Platanus (Platanaceae): Homoeologues, pseudogenes and ancient hybridization. *Annals of Botany* 101(3):403–419. https://doi.org/10.1093/aob/mcm305.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30(14):3059–3066. https://doi.org/10.1093/nar/gkf436.

Kreunen, S.S., and J.M. Osborn. 1999. Pollen and anther development in Nelumbo (Nelumbonaceae). *American Journal of Botany* 86(12):1662–1676.

Niu, Y.F., and J. Liu. 2019. Complete chloroplast genome of *Helicia nilagirica* Bedd. and its phylogenetic analysis. *Mitochondrial DNA. Part B, Resources* 5(1):342–343. https://doi.org/10.1080/23802359.2019.1703587.

Nock, C.J., A. Baten, and G.J. King. 2014. Complete chloroplast genome of Macadamia integrifolia confirms the position of the Gondwanan early-diverging eudicot family Proteaceae. *BMC Genomics* 15 Suppl 9(Suppl 9):S13. https://doi.org/10.1186/1471–2164–15-S9-S13.

Parks, M., R. Cronn, and A. Liston. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7:84. https://doi.org/10.1186/1741–7007–7-84.

Pauw, E., and J.N. Eloff. 2014. Which tree orders in southern Africa have the highest antimicrobial activity and selectivity against bacterial and fungal pathogens of animals? *BMC Complementary and Alternative Medicine* 14:317. https://doi.org/10.1186/1472–6882–14–317.

Sanderson, M.J., and J.A. Doyle. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. *American Journal of Botany* 88(8):1499–1516. https://doi.org/10.2307/3558458.

Sauquet, H., and D.J. Cantrill. 2007. Pollen diversity and evolution in Proteoideae (Proteales: Proteaceae). *Systematic Botany* 32(2):271–316. https://doi.org/10.1600/036364407781179743.

Wicke, S., G.M. Schneeweiss, C.W. dePamphilis, K.F. Müller, and D. Quandt. 2011. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology* 76(3–5):273–297. https://doi.org/10.1007/s11103–011–9762–4.

Wood, T.C. 2006. The current status of baraminology. *Creation Research Society Quarterly* 43(3):149–158.

Xue, J., S. Wang, and S.L. Zhou. 2012. Polymorphic chloroplast microsatellite loci in Nelumbo (Nelumbonaceae). *American Journal of Botany* 99(6):e240–e244. https://doi.org/10.3732/ajb.1100547.