

Minireview

Ancient flowering plants: DNA sequences and angiosperm classification

Mark W Chase and Michael F Fay

Address: Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK.

Correspondence: Mark W Chase. E-mail: m.chase@rbgkew.org.uk

Published: 22 March 2001

Genome Biology 2001, **2**(4):reviews1012.1–1012.4

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/4/reviews/1012>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

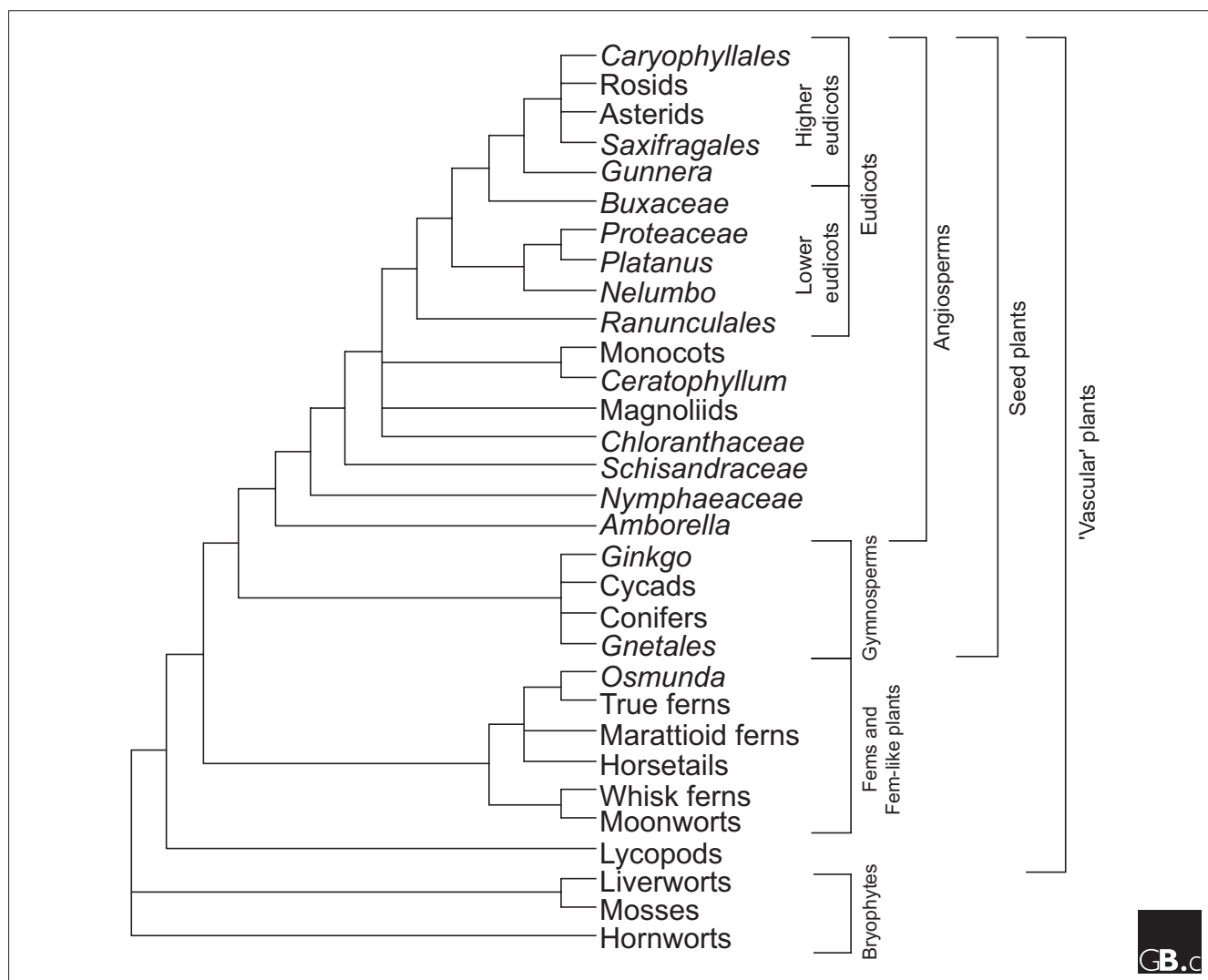
Phylogenetic analyses of gene sequences provide a clear pattern of which extant flowering plant genera diversified earliest. Combined with complete genomic sequences, these data will vastly improve understanding of the genetic basis of plant diversity.

Over the past ten years, botanists have produced a huge body of DNA sequences from genes in each of the three plant genomes - mitochondrial, nuclear, and plastidial. Some of the data sets are prodigious: 580 ribulose biphosphate carboxylase/oxygenase large subunit (*rbcL*) sequences for advanced dicotyledons [1], and 587 species covering all major lineages and families of plants for three genes (*rbcL*, ATP synthase β subunit (*atpB*) and 18S ribosomal DNA) [2,3]. Progress in sorting out major lineages has been both highly collaborative and rapid; the first paper to examine overall patterns with extensive sampling (500 *rbcL* sequences), which had 43 co-authors, was published as recently as 1993 [4]. In many respects, these studies are similar to the model-genome sequencing efforts, except that they encompass the breadth of plant diversity rather than examining a few species intensively. Similar work has focused on the relationships specifically among land plants, with equally noteworthy success [5].

The major accomplishments of this research fall into several categories. First of all, these studies [1-5] demonstrated that large phylogenetic analyses were themselves practical and sound [6,7], both conclusions that were previously thought unlikely [8,9]. Subsequent to publication of the empirical studies of flowering plant relationships [4,10], simulation studies reached the same conclusions [11,12]. In parallel, simulation and empirical studies have also demonstrated that existing software and personal computers are adequate

for these tasks; large analyses do not require powerful computers, elaborate software, or time-consuming analyses [7,13]. The reason for the apparent ease and simplicity of large phylogenetic analyses despite the dire prospects from the theoretical standpoint is that each of the genes used contains a relatively clear and congruent pattern, which, when the data are combined, immensely simplifies analysis [6,7]. On the basis of the results of these ground-breaking studies of plant phylogeny [1-5,10], large-scale phylogeny building, which is necessary for an understanding of broad patterns of biological diversity, no longer had to confront the problems previously expected to impede progress. The way was clear for major insights into patterns of flowering plant evolution once enough data were collected.

Although studies analyzing single genes were largely congruent in their general conclusions about the plants' relatedness [7], the placement of the root of the phylogenetic tree was not. The first study [4] using *rbcL* placed the root between an unusual aquatic genus, *Ceratophyllum*, and the rest of the flowering plants (angiosperms), whereas the second and third genes, *atpB* and 18S rDNA, located this point between *Amborella* and the rest (Figure 1) [3,14]. None of these results, however, withstood analysis with re-sampling techniques, such as the bootstrap and the jackknife [15,16], which are designed to demonstrate how clear a pattern is within a specified data matrix. When we added additional genes from the mitochondrial genome [17], however, this

**Figure 1**

Synopsis of relationships of the major groups of land plants, with more detail in the flowering plants [3,5]. Bryophytes have water-conducting tissues, although they are not the same as in other land plants, and hence we have put 'vascular' plants in quotes. Nodes with more than two branches indicate groups for which inter-relationships are not yet clear.

situation was remedied, and the rooting of the phylogenetic tree between *Amborella* and the rest of the angiosperms was well supported. Another analysis, using even more genes [18], also found a great deal of consistency and a similar rooting, but before tree construction they used a method of analysis that reduced the 'noise' caused by varying patterns of molecular evolution in each of the genes; it is unclear, however, how 'noise' should be defined or whether it is necessary to completely remove it from analyses. Nonetheless, the only major difference that the use of this method produced was that the water-lily family, *Nymphaea* and its relatives, joined *Amborella* on the first side branch, rather than this branch being occupied solely by *Amborella*. In the other analyses [2,3,17], *Nymphaea* was placed as the next lineage after *Amborella* to split off the ancient angiosperm stock

(Figure 1). In either scenario, most of the implications for angiosperm evolution would be similar, so such a finding is, overall, highly consistent with the other analyses using three or more genes.

Another approach to this problem was to use a pair of genes derived from a single gene that underwent duplication before any of extant angiosperms evolved but after they split from the gymnosperms; phylogenetic trees for each of the duplicated loci were then used to root the other [19]. This effort was, however, limited because some critical taxa were absent (*Ceratophyllum*, for example); only a single locus that did not clearly fall into one of the pair could be found in some plants. The potential of this method thus remains largely unevaluated, although it holds great promise.

Many of the patterns emerging from analyses of DNA sequences [3,4,10,17,18] are not particularly different from some parts of previous classifications. For example, families with fused petals (often previously classified as Asteridae, such as in the widely used system of Cronquist [20]) formed a group in the DNA results as expected; it would have been strange if all previous ideas about flowering plant classification and phylogenetic relationships were incorrect. Nevertheless, the patterns revealed by analyses of DNA sequences have produced a substantial number of greatly altered ideas about relationships, opening up a potential conflict between molecules and morphology. The differences could be the result of different underlying patterns in morphology and DNA sequences, but an alternative explanation is simply that the apparent discrepancies are, instead, the product of the different methods used. Phylogeneticists objectively give equal emphasis to each data type until clear evidence emerges that some parts are less reliable, whereas evolutionary taxonomists synthesize a large body of data but usually use intuitive weighting to determine which is the most reliable of the different categories of information. When we analyzed morphological data using the same techniques as were used for the DNA data, the results also differed from previous classifications using morphology and were much more similar to those produced with the gene sequences [21].

An independent classification of the families of angiosperms has been published that relies largely but not exclusively on DNA sequence data [22]; this makes angiosperms the first major group of organisms to be so treated. Like many of the previous DNA studies, this effort was highly collaborative and is thus cited as the Angiosperm Phylogeny Group Classification so that it will not be associated with the name of any particular researcher. This classification is, in effect, a work in progress and will be updated as more information emerges; the foundations of the Angiosperm Phylogeny Group Classification were laid on clear patterns consistent in all published studies, which are therefore unlikely to change in any substantial way. These patterns are both well supported by measures such as the bootstrap and well corroborated by many other kinds of studies. Some families (all small, many consisting of single species) remain poorly studied, however, and a few of the larger patterns remain unclear (see below); these are the foci of on-going research.

Although the general patterns of flowering plant relationships have been greatly clarified by studies employing multiple genes, the inter-relationships of some major groups remain unclear. For example, the three largest groups of eudicotyledons (or advanced dicotyledons), namely the asterids, rosids and caryophyllids, are each clearly defined, but their relationships to each other are not. It would appear that all three arose more or less simultaneously, perhaps in parallel with more advanced groups of pollinating insects about 100 million years ago [23], and their rapid appearance left little pattern in each gene to group any two of these

together. Until there are many more genes sequenced from a broad range of flowering plants (for example, roughly 600 species, as in previous studies [2,3]), these patterns will not be robustly addressed. Such work is underway.

Relationships of the angiosperms to other land plants are also now becoming clearer than ever before. On the basis of studies of morphological data, gymnosperms have long been thought to have given rise to the flowering plants, perhaps with the Gnetales as their closest extant relatives [24], but this view has now given way to one [5] in which the gymnosperms are monophyletic and are collectively the sister group to the angiosperms (Figure 1). The seed plants (angiosperms plus gymnosperms) are then related to the ferns and their relatives, which clearly include the horsetails and whisk ferns, groups that previously were of highly variable and speculative placements. Among the higher land plants (excluding the mosses, hornworts and liverworts), the lycopods occupy an isolated position outside the ferns (and their allies) and seed plants (Figure 1).

In spite of the work remaining, most of what is now known about relationships of the angiosperms is detailed and well founded for the first time. Plants are the basis of life on Earth, and knowing the patterns relating to their evolution is a great advantage because it permits research to be accurately focused and brings to bear an immense predictive power. All organisms are the products of both the constraints imposed by their evolutionary history and the action of natural selection. If the patterns of evolutionary descent can be estimated with confidence, then researchers have an enhanced potential to separate the action of selection from characteristics inherited from a common ancestor. Thus, botanists today are in the fortunate position of being able to combine an in-depth knowledge of genomic structure and content of several model organisms with a clear picture of how these model organisms are related to the rest of a hugely diverse group that provides us with food, fuel, medicines, and housing as well as bringing beauty to our lives. Such knowledge of phylogeny is not idle curiosity but is instead an important tool for comparative biology.

References

1. Savolainen V, Fay MF, Albach DC, Backlund A, van der Bank M, Cameron KM, Johnson SA, Lledó MD, Pintaud J-C, Powell M *et al.*: **Phylogeny of the eudicots: a nearly complete familial analysis based on *rbcl* gene sequences.** *Kew Bull* 2000, **55**:257-309.
2. Soltis PS, Soltis DE, Chase MW: **Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology.** *Nature* 1999, **402**:402-404.
3. Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF *et al.*: **Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcl*, and *atpB* sequences.** *Bot J Linn Soc* 2000, **133**:381-461.
4. Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L *et al.*: **Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcl*.** *Ann Missouri Bot Gard* 1993, **80**:528-580.
5. Pryer KM, Schneider H, Smith AM, Cranfill R, Wolf PG, Hunt JS, Sipes SD: **Horsetails and ferns are a monophyletic group and**

- the closest living relatives to the seed plants. *Nature* 2001, **409**:618-622.
6. Soltis DE, Soltis PS, Mort ME, Chase MW, Savolainen V, Hoot SB, Morton CM: **Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms.** *Syst Biol* 1998, **47**:32-42.
 7. Chase MW, Cox AV: **Gene sequences, collaboration, and analysis of large data sets.** *Austr Syst Bot* 1998, **11**:215-229.
 8. Felsenstein J: **The number of evolutionary trees.** *Syst Zool* 1978, **27**:27-33.
 9. Graur D, Duret L, Gouy M: **Phylogenetic position of the order Lagomorpha (rabbits, hares and allies).** *Nature* 1996, **379**:333-335.
 10. Soltis DE, Soltis PS, Nickrent DL, Johnson LA, Hahn WJ, Hoot SB, Sweere JA, Kuzoff RK, Kron KA, Chase MW: **Angiosperm phylogeny inferred from 18S ribosomal DNA sequences.** *Ann Missouri Bot Gard* 1997, **84**:1-49.
 11. Hillis DM: **Taxonomic sampling, phylogenetic accuracy, and investigator bias.** *Syst Biol* 1998, **47**:3-8.
 12. Graybeal A: **Is it better to add taxa or characters to a difficult phylogenetic problem?** *Syst Biol* 1998, **47**:9-17.
 13. Kei T, Nei M: **Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used.** *Mol Biol Evol* 2000, **17**:1251-1258.
 14. Savolainen V, Chase MW, Morton CM, Hoot SB, Soltis DE, Bayer C, Fay MF, de Bruijn A, Sullivan S, Qiu Y-L: **Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences.** *Syst Biol* 2000, **49**:306-362.
 15. Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG: **Parsimony jackknifing outperforms neighbor-joining.** *Cladistics* 1996, **12**:99-124.
 16. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39**:783-791.
 17. Qiu Y-L, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Chen Z, Savolainen V, Chase MW: **The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes.** *Nature* 1999, **402**:404-407.
 18. Barkman TJ, Chenery G, McNeal JR, Lyons-Weiler J, Ellisens WJ, Moore M, Wolfe AD, dePamphilis CW: **Independent and combined analyses of sequences from all three genome compartments converge on the root of flowering plant phylogeny.** *Proc Nat Acad Sci USA* 2000, **97**:13166-13171.
 19. Mathews S, Donoghue MJ: **The root of angiosperm phylogeny inferred from duplicate phytochrome genes.** *Science* 1999, **286**:947-950.
 20. Cronquist A: *An Integrated System of Classification of Flowering Plants.* New York: Columbia University Press; 1981.
 21. Nandi OI, Chase MW, Endress PK: **A combined cladistic analysis of angiosperms using *rbcL* and non-molecular data sets.** *Ann Missouri Bot Gard* 1998, **85**:137-212.
 22. Angiosperm Phylogeny Group: **An ordinal classification of the families of flowering plants.** *Ann Missouri Bot Gard* 1998, **85**:531-553.
 23. Chase MW, Fay MF, Savolainen V: **Higher-level classification in the angiosperms: new insights from the perspective of DNA sequence data.** *Taxon* 2000, **49**:685-704.
 24. Doyle JA, Donoghue MJ: **Seed plant phylogeny and the origin of the angiosperms: an experimental cladistic approach.** *Bot Rev* 1986, **52**:321-431.