BMC Biology

COMMENTARY **Open Access**

Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there

Charles C Davis[*], Zhenxiang Xi and Sarah Mathews

Abstract

A study in *BMC Evolutionary Biology* represents the most comprehensive effort to clarify the phylogeny of green plants using sequences from the plastid genome. This study highlights the strengths and limitations of plastome data for resolving the green plant phylogeny, and points toward an exciting future for plant phylogenetics, during which the vast and largely untapped territory of nuclear genomes will be explored.

See research article:
http://www.biomedcentral.com/1471-2148/14/23

Commentary

The plastid genome, or plastome, has so far been the most important source of data for plant phylogenetics in the era of comparative DNA sequencing. Its utility results from its relatively small size (between 75 and 250 kilobases), largely uniparental inheritance, conservation of gene content and order, and its high copy number in green plant cells. From the early use of a single plastid gene to infer the phylogeny of a broad sampling of seed plants [1], to the now common use of around 80 plastid genes to address finer-scale phylogenetic questions, this circular genome has been a mainstay for evolutionary botanists.

Efforts to understand green plant phylogeny from plastome data have now come full circle. In this issue, Ruhfel *et al.* [2] report results from their analyses of 78 plastid genes from 360 species, from green algae to angiosperms. Their results provide insights into this ongoing effort, adding support for some relationships and highlighting phylogenetic questions that require more data, especially from nuclear genomes.
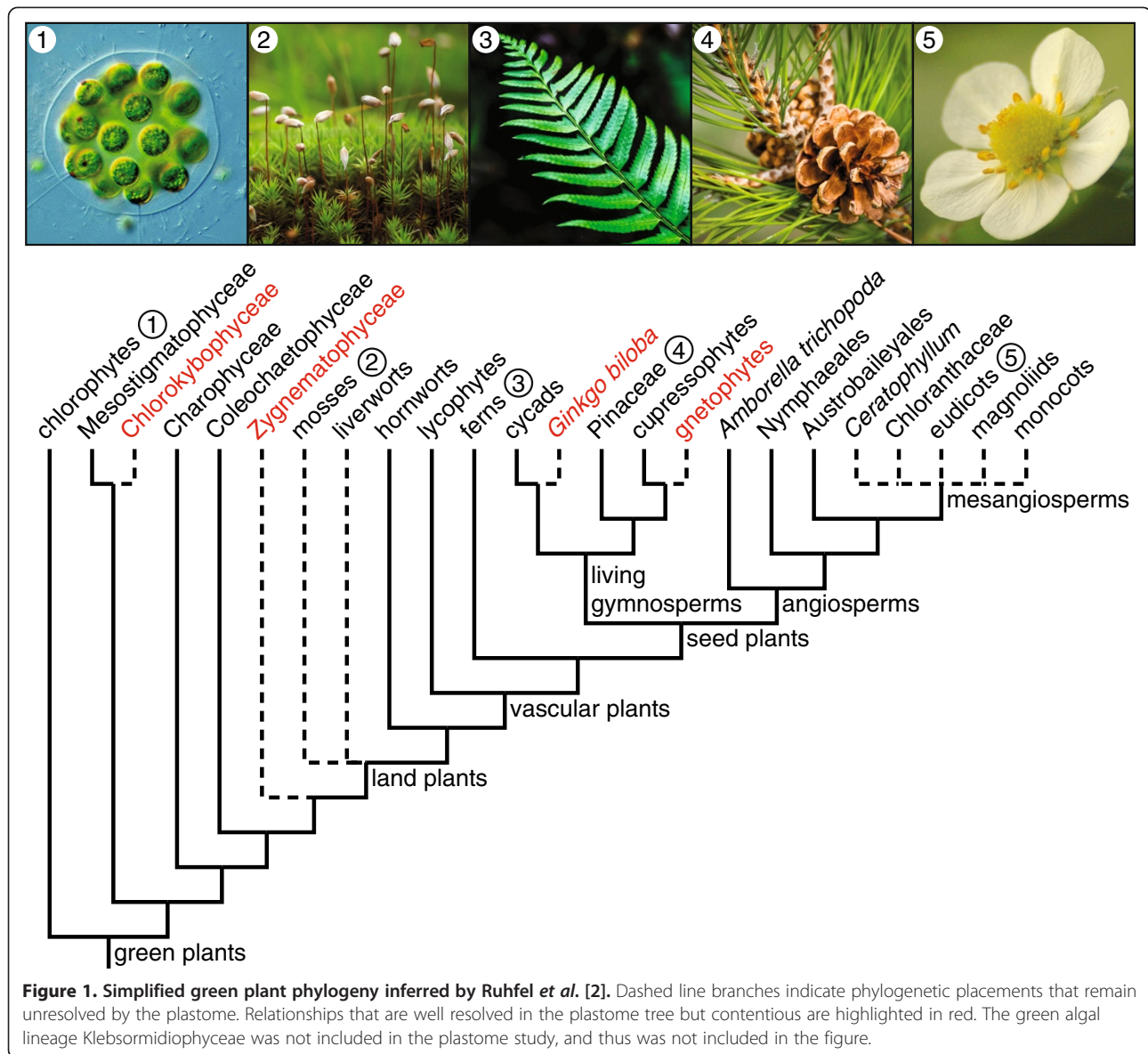
Congruence and conflict in plastid phylogenomics

Ruhfel *et al.* present a phylogeny that is well resolved at most nodes, and largely in agreement with previous studies, including at nodes that have been difficult to resolve (Figure 1). These include the splits between land plants and their algal sister clade [3,4], and between vascular plants and their non-vascular sister clade [5]. Here, Zygnematophyceae, a large clade of mostly freshwater algal species, is identified as sister to land plants. This suggests that shared components of auxin signaling and chloroplast movement likely were present in their common ancestor [3]. Their analyses also support the non-monophyly of bryophytes, or liverworts, mosses and hornworts. These land plants lack a well developed vascular system and have similar ecologies. Hornworts are sister to vascular plants in the plastid tree, consistent with evidence that their sporophytes may be at least partially free-living, unlike those of liverworts and mosses [5].

So how much closer does this new phylogeny bring us to a robust understanding of green plant evolution? This study, like many others, has difficulty resolving key relationships within green plants. This is most evident in the lack of resolution deep in the angiosperm phylogeny among the mesangiosperm clades, *Ceratophyllum*, Chloranthaceae, eudicots, magnoliids, and monocots. Branching order among non-vascular plants, especially involving liverworts and mosses, remains contentious. These problems persist due, in part, to the challenge of placing lineages that are species-poor and divergent in molecular trees, and due to the difficulty of assessing homologies among organisms with very diverse or reduced morphologies.

Despite these few persistent problems, the Rhufel *et al.* tree appears to address many if not all remaining questions. In some cases, however, high support for relationships should be interpreted cautiously because conflicting topologies are supported by other data. Key examples include the previously mentioned sister groups of land plants and vascular plants, but also relationships

* Correspondence: cdavis@oeb.harvard.edu
Department of Organismic and Evolutionary Biology, Harvard University, 22 Divinity Ave, Cambridge, MA 02138, USA

Figure 1. Simplified green plant phylogeny inferred by Ruhfel *et al.* [2]. Dashed line branches indicate phylogenetic placements that remain unresolved by the plastome. Relationships that are well resolved in the plastome tree but contentious are highlighted in red. The green algal lineage Klebsormidiophyceae was not included in the plastome study, and thus was not included in the figure.

among major seed plant clades. The latter involves the position of gnetophytes, and the close relationship of cycads and *Ginkgo biloba*. The gnetophytes especially represent a vexing problem in seed plant phylogenetics [6]. They form a small clade of approximately 90 species that are highly divergent from other seed plants in both morphology and molecules. Although plastid phylogenomic studies are converging on the 'Gnecup' topology, in which gnetophytes are united with cupressophyte conifers, recent nuclear phylogenomic analyses yield the alternative 'Gnepine' topology in which gnetophytes are united with Pinaceae conifers [7]. Even within individual plastid loci, different nucleotide sites have been shown to favor rival gnetophyte placements [6]. A similarly strong conflict in seed plants concerns the positions of cycads and *Ginkgo*

biloba, where their plastid trees strongly unite the two, but other studies place cycads alone as sister to extant gymnosperms.

Where do we go next?

It is well known that biases within molecular data may be exacerbated in large phylogenomic data sets, leading to erroneous but well supported results, especially when trying to resolve ancient splits. Biases may result from phenomena such as pattern heterogeneity and uneven base frequencies, and their exploration will help us to understand cases of incongruent relationships. For example, when Ruhfel *et al.* accounted for biased GC content, their data placed lycophytes as sister to ferns and seed plants, rather than as the lone sister to seed plants,

as in their total evidence tree. Additional approaches for mitigating biases in molecular data sets include increasing taxon sampling and better modeling of nucleotide evolution. Traditional data partitioning schemes try to account for variation in evolutionary rates by partitioning nucleotide sites based on their rate of evolution, most commonly, by gene or codon position. The approach developed by Xi *et al.* [8], in contrast, requires no *a priori* assumptions about evolutionary rates. Instead, the optimal number of partitions, and their contents, are identified using a Bayesian mixture model analysis, which is not influenced by preconceptions about nucleotide evolution. This approach improved resolution over analyses using traditional partitioning strategies and also reduced model complexity because the optimal number of partitions identified in the search was smaller than in commonly used schemes. Ruhfel *et al.* found this scheme to be computationally difficult to implement with their data, but improvements in the efficiency of Bayesian mixture model searches will help. Finally, despite the promise of these improvements to molecular phylogenetic studies, the evolution of green plants cannot be understood from molecular data alone. For example, 70% of seed plant lineages cannot be sampled for molecular datasets because they are extinct. Better integration of morphological evidence from living and fossil taxa are especially needed to reconstruct the evolutionary history of green plants [9].

The largest leap, however, is still ahead of us. Within the green plant species tree there is a 'cloud' of gene trees [10], of which the plastid genes comprise only a small fraction. An obvious next step is to understand the species tree more thoroughly by incorporating mitochondrial and nuclear data. Mitochondrial data have previously been neglected, but increasingly are being sampled for large-scale phylogenomic studies. However, their informativeness may be limited by slow nucleotide evolution and species relationships may be obscured by potentially rampant horizontal gene transfer involving mitochondrial DNA [11]. Nuclear genomic data, in contrast, have tremendous potential to improve phylogenetic resolution and illuminate the species tree. This source of data will likely reveal surprises when juxtaposed against our current understanding of relationships inferred from plastid data alone. Conflicts between plastid and species trees may result from introgression of the plastid from one species into another, and this may have gone undetected due to heavy reliance on phylogenetic data from uniparentally inherited plastomes. Recombination and gene conversion, which can occur in the plastome, as well as differential selective pressures acting on plastid genes, may also introduce biases and lead to incongruent gene and species trees. Along these lines, recent analyses already indicate that potential plastid-nuclear genome conflicts involve the gnetophytes, early diverging flowering plants, and the large flowering plant orders Lamiales, Malpighiales, and Myrtales. Evaluating the extent to which these incongruent placements demonstrate divergent genome histories requires further exploration, for which the nuclear genome will be a particularly valuable resource.

In addition to providing a wealth of new data for clarifying species trees, the nuclear genome will greatly improve our understanding of important innovations across green plants. Whole genome duplications (WGDs), for example, potentially enhance an organism's success. Consistent with this, recent analyses of transcriptomes from seed plants indicate that at least three major WGDs occurred very near to the origin of clades characterized by putative key innovations [12,13]. These include the origins of seeds, flowers, and pentamorous floral symmetry - the last of which characterizes more than approximately 70% of all angiosperms (the eudicots) and may be related to their coevolution with bees [14].

Nuclear genomic data also more directly facilitate our ability to connect unique phenotypes with their underlying genetic architectures. In an exemplar study of fungal relationships, Floudas *et al.* [15] investigated the origin of lignin decomposition in fungi - the ability of organisms to degrade lignin synthesized by green plants is a rare feature across the tree of life. This is especially relevant because the absence of lignin decomposition prior to the end of the Carboniferous era (approximately 300 million years ago) accounts for Earth's extensive stores of fossil fuels. The origin of lignin decomposition by fungi was implicated in the sharp decline in burial of organic carbon around this time. The authors tested this idea by investigating genes implicated in lignin degradation, thus discovering an association between key expansions of these genes coincident with the origins of fungal clades that can degrade lignin. These expansions broadly correspond with the disappearance of fossilized forests from the geological record. Such exemplar studies likely represent the tip of the iceberg, and highlight the tantalizing future research opportunities in plant nuclear genomics.

We are entering into a new and exciting era in plant phylogenetics. Plastid phylogenomics will continue to be a fast and inexpensive way to flesh out the green plant clade, but the next wave is to explore the uncharted terrain of the nuclear genome. It is already on the way, as evidenced by large-scale comparative transcriptome projects (for example [16]) and the growing number of genome sequencing projects focused on phylogenetically key species.

References

1. Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, et al: **Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*.** *Ann Missouri Bot Gard* 1993, **80**:528–580.
2. 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 Evol Biol* 2014, **14**:23.
3. Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, De Clerck O: **Phylogeny and molecular evolution of the green algae.** *Crit Rev Plant Sci* 2012, **31**:1–46.
4. Zhong B, Xi Z, Goremykin VV, Fong R, Mclenachan PA, Novis PM, Davis CC, Penny D: **Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes.** *Mol Biol Evol* 2014, **31**:177–183.
5. Qiu YL, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC: **The deepest divergences in land plants inferred from phylogenomic evidence.** *Proc Natl Acad Sci U S A* 2006, **103**:15511–15516.
6. Burleigh JG, Mathews S: **Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life.** *Am J Bot* 2004, **91**:1599–1613.
7. Xi Z, Rest JS, Davis CC: **Phylogenomics and coalescent analyses resolve extant seed plant relationships.** *PLoS One* 2013, **8**:e80870.
8. Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, Endress PK, Matthews ML, Stevens PF, Mathews S, Davis CC: **Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales.** *Proc Natl Acad Sci U S A* 2012, **109**:17519–17524.
9. Mathews S, Clements MD, Beilstein MA: **A duplicate gene rooting of seed plants and the phylogenetic position of flowering plants.** *Phil Trans R Soc B Biol Sci* 2010, **365**:383–395.
10. Maddison WP: **Gene trees in species trees.** *Syst Biol* 1997, **46**:523–536.
11. Xi Z, Wang Y, Bradley RK, Sugumaran M, Marx CJ, Rest JS, Davis CC: **Massive mitochondrial gene transfer in a parasitic flowering plant clade.** *PLoS Genet* 2013, **9**:e1003265.
12. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, de Pamphilis CW: **Ancestral polyploidy in seed plants and angiosperms.** *Nature* 2011, **473**:97–100.
13. Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ: **A genome triplication associated with early diversification of the core eudicots.** *Genome Biol* 2012, **13**:R3.
14. Cardinal S, Danforth BN: **Bees diversified in the age of eudicots.** *Proc R Soc Lond B* 2013, **280**:20122686.
15. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Górecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, et al: **The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes.** *Science* 2012, **336**:1715–1719.
16. **1000 Plants**. [http://onekp.com/].