

SPECIAL ISSUE INTRODUCTION

Using and navigating the plant tree of life

Douglas E. Soltis^{1,2}, Michael J. Moore^{3,5}, Emily B. Sessa², Stephen A. Smith⁴, and Pamela S. Soltis¹

Manuscript received 23 January 2018; revision accepted 7 February 2018.

¹ Florida Museum of Natural History, University of Florida, P. O. Box 117800, Gainesville, FL 32611, USA

² Department of Biology, University of Florida, P. O. Box 118525, Gainesville, FL 32611, USA

³ Department of Biology, Oberlin College, 119 Woodland Street, Oberlin, OH 44074, USA

⁴ Department of Ecology and Evolutionary Biology, University of Michigan, 830 North University, Ann Arbor, MI 48109, USA

⁵ Author for correspondence (e-mail: michael.moore@oberlin.edu)

Citation: Soltis D. E., Michael J. Moore, Emily B. Sessa, Stephen A. Smith, and Pamela S. Soltis. 2018. Using and navigating the plant tree of life. *American Journal of Botany* 105(3): 287–290.

doi:10.1002/ajb2.1071

The “tree of life” has become a metaphor for the interconnectivity and breadth of all life on Earth. It also has come to symbolize the broad investigation of biodiversity, including both the reconstruction of phylogeny and the numerous downstream analyses that are possible with a firm phylogenetic underpinning. Only a few decades ago, the construction of large phylogenetic trees of hundreds of taxa (or more) was considered an impossible task due to the immense computational challenges posed by analyses of large data sets. And no wonder—the number of possible trees that can describe the relationships of just 200 species exceeds the number of atoms in the universe (Hillis, 1996). As a result, building the tree of all named life, including the green plant branch (*Viridiplantae*)—a major clade with perhaps 500,000 species—has long been considered a grand challenge in biology. However, a perfect storm of algorithm development, increases in computational power, and DNA sequencing improvements over the last decade has not only made the construction of large trees more feasible, but also allowed us to attain some far-reaching goals—a noteworthy example being the recent publication of a first draft tree of all life (Hinchcliff et al., 2015).

In plant biology, the frequent reconstruction of large phylogenetic trees has had an immense impact on the field. Large trees have helped to resolve deep-level relationships and resulted in the revision of classifications, including some of the most profound changes in our view of plant relationships over the past 200 years (e.g., reviewed in part by Gitzendanner et al., 2018, in this issue). Large trees have also ushered in a renaissance in the study of conservation, ecology, methods development, crop improvement, genome evolution,

and much more. Building the plant tree of life has come to represent the biodiversity equivalent of the human genome project, with numerous and often unanticipated downstream outcomes.

Accompanying these exciting advances are equally significant challenges that remain for the construction of a better and more complete picture of the evolution of plant lineages. In addition to the computational challenges of larger data sets, these include conceptual and methodological barriers. For example, where it was once thought that simply increasing DNA sequence data would increase resolution of relationships, we now understand that increasing data leads to increasing analytical complexity. Furthermore, this complexity is not due solely to limitations in computational power and methodology, but in part reflects the underlying complexity of the evolutionary process and its impact on genomes. Nevertheless, current conceptual and computational limitations present fantastic opportunities for transformative developments in our understanding of plant evolution. In this special issue, we explore many of the uses and challenges of big trees and big data in plant biology. Diverse papers provide overviews of the current status of the green plant tree of life and describe some of the myriad applications of the knowledge of phylogenetic relationships as well as some of the challenges inherent in handling plant phylogenomic data.

PROGRESS IN PLANT PHYLOGENOMICS

Since the last *American Journal of Botany* Tree of Life issue in 2004 (Palmer et al., 2004), the rise of high-throughput sequencing has

both revolutionized our understanding of the processes that have generated plant diversity and, to a great extent, confounded our attempts to reconstruct phylogeny by unlocking the evolutionary history of the plant nuclear genome. More than a decade later, papers in this issue demonstrate the use of numerous nuclear loci to examine plant evolution, with a particular focus on evaluating three important, and often interrelated, phenomena: (1) conflict among gene trees, (2) gene/genome duplication, and (3) hybridization/introgression. Each of these can and does confound estimates of phylogeny, but also provides important insights into processes that have generated plant diversity. For example, using 300 transcriptomes, Walker et al. (2018) found that whole-genome duplication (WGD) has been common throughout the evolutionary history of Caryophyllales and demonstrate that despite rampant gene tree conflict, estimates of deep-level phylogeny in the order are now largely consistent with and better supported than in previous studies. McManus et al. (2018) used organellar phylogenomics to clarify relationships and identify conserved plastome gene order within the chlorophytic algal family Hydrodictyaceae, noting that this conservation contrasts with the dynamic plastome evolution characteristic of Sphaeropleales. Outside of *Viridiplantae*, Parks et al. (2018) investigated the history of WGD in diatoms, using multiple data sources and analyses to conduct a thorough survey of the history of polyploidy in this group. They conclude that WGD has also played an important and ongoing role in diatom evolution.

A theme shared by several papers in this compilation is exploration of new methods for examining the tree of life and understanding processes that lead to conflict among gene trees. The first of these papers is a timely review by Folk et al. (2018a) of the challenges and prospects for detecting hybridization across the plant tree of life. Knowles et al. (2018) discuss a recently developed method for identifying incomplete lineage sorting (ILS) and horizontal gene transfer (HGT) across data partitions in phylogenomic data sets. They also explore how the timing of these events can affect the accuracy of their detection. Pease et al. (2018) show that quartet sampling can be used to distinguish lack of support for a topology from conflicting support in the plant tree of life, and Lam et al. (2018) explore the challenges of reconstructing relationships in extreme groups, such as mycoheterotrophic lineages. They show that phylogenomic analyses can succeed even when genomes are greatly reduced in gene content and have accelerated rates of evolution. The longstanding issue of the impact of multiple haplotypes in analyses of nuclear genes was tackled by Kates et al. (2018). They review various methods for assessing haplotype diversity within individuals and propose a new method for assembling haplotype sequences that better accounts for heterozygosity in phylogenetic analyses.

PLANT PHYLOGENOMICS AT DEEPER LEVELS

Despite decades of molecular analysis, often based on a relatively small number of genes, major questions of evolutionary relationships remain in green plants. Several papers undertake large-scale analyses that update our understanding of green plant phylogeny as a whole, including insights from multiple genomic compartments. Gitzendanner et al. (2018) analyzed near-complete plastid genomes to provide an overview of the current framework of green plant phylogeny, while Smith and Brown (2018) have taken a major step by providing a dated green plant tree of life that includes all the seed plant taxa in the Open Tree of Life taxonomy. They also detail some of

the significant gaps in our knowledge and discuss challenges that will need to be addressed to further improve our understanding of large-scale plant phylogenetics. Beaulieu and O'Meara (2018) demonstrate that analysis of large sparse matrices can lead to reliable phylogenies, illustrating this with a large phylogeny of campanulids, and yet they caution that such data sets may require significant curation.

Other papers focus on specific significant clades. Two papers examine Apocynaceae at both deeper levels and the species level, respectively, to examine the evolution of growth form, flowers, and defense mechanisms (Fishbein et al., 2018a, b). Similarly, Walker et al. (2018) used phylogenomic inference to provide novel insights into the evolution of Caryophyllales, a clade with highly diverse growth forms ranging from cacti to carnivorous plants. Edger et al. (2018) used plastid genes to address evolution, paleopolyploidy, and the origin of chemical defenses in Brassicales. Stull et al. (2018) reconstructed an updated tree for *Asteridae* and estimated morphological character evolution in this huge clade of flowering plants, demonstrating that some of the key traits that have traditionally been used to define this clade—most notably sympetaly—have a more complicated evolutionary history than previously appreciated.

APPLICATIONS TO MAJOR QUESTIONS IN GREEN PLANT EVOLUTION

Many of the papers in this issue demonstrate the utility of big trees for addressing major questions in green plant evolution and illustrate the diverse uses of phylogenies as tools for understanding trait evolution, diversification, and the assembly of plant communities. Landis et al. (2018) used a large phylogenetic tree for angiosperms to assess the association between numerous WGD events and diversification, showing that although clades with WGDs have greater species richness than their sister clades, only approximately 40% of the 106 WGD events examined are significantly associated with increased rates of diversification. Testo and Sundue (2018) similarly investigated lineage diversification rates in ferns and how these may be related to body size evolution; they conclude that diversification and leaf size evolution are largely decoupled across this clade. Using morphological data, Stoughton et al. (2018) conducted integrated analyses of leaf shape and ecology in a phylogenomic context to understand species boundaries among recently diverged lineages of *Claytonia*, while Sessa et al. (2018) used a phylogenetic framework to explore patterns of community assembly, focusing on the ferns of Florida. Integrating phylogeny and functional traits for the oaks of North America, Cavender-Bares et al. (2018) found a shift from phylogenetic overdispersion to clustering with increasing scale and convergent evolution in leaf traits with clustering at nearly all spatial scales. Heyduk et al. (2018) used phylogenetics and transcriptomics to investigate key questions surrounding the evolution of CAM photosynthesis, while Goolsby et al. (2018) used sequence capture methods to examine the evolution of 83 gene families at transitions of C_4 and CAM photosynthesis.

MOVING FORWARD

Despite significant progress, the enormous size of the green plant clade of life continues to impede overall synthesis—and researchers outside of the botanical community often overestimate the extent of what we have achieved vs. what we still do not know. By

some estimates, fewer than 30% of green plant species currently have sequence data available (see Hinchliff et al., 2015; Folk et al., 2018b), and, in addition to the lack of molecular data necessary for reconstructing the tree itself, there remain significant gaps in other sources of data which thwart our ability to fully use emerging plant phylogenetic frameworks. For example, well-curated specimen occurrence and morphometric databases are growing but still do not include a majority of plant taxa, thus hindering analyses of global patterns of plant diversity and evolution. In this issue, Folk et al. (2018b) illustrate the challenges of data limitation in the massive rosid clade (90,000 species), and Stull et al. (2018) show that great gaps remain in our understanding of morphological character states in several phylogenetically key asterid lineages. It will require immense investment of time and resources to expand these data sources and integrate them into our improving phylogenetic framework, but these initiatives are well worth the effort, especially given the central importance of autotrophs as primary drivers of terrestrial and aquatic ecosystems. As one example, Eiserhardt et al. (2018) provide a vision for the future of green plant phylogenetics that will lead to a global phylogenetic synthesis and an integrated, open, and continually updated green plant tree of life; we enthusiastically endorse their recommendations.

The overall goal of this compilation of papers is to illustrate the incredible progress and promise of big trees and big data in the study of the green plant tree of life, highlighting outstanding challenges that will be fruitful avenues for future research. The advancements that have been made in this field have been substantive and rapid, and the future affords enormous new opportunities for plant researchers. What an exciting time to be a plant evolutionary biologist! Importantly, much of this tremendous progress continues to be attributable to the collaborative spirit of this community, a trend that started with the first attempts decades ago to reconstruct green plant phylogeny on a major scale (Chase et al., 1993; Palmer et al., 2004). It is a trend we hope will continue.

LITERATURE CITED

- Beaulieu, J. M., and B. C. O'Meara. 2018. Can we build it? Yes we can, but should we use it?: Assessing the quality and value of a very large phylogeny of campanulid angiosperms. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1020>.
- Cavender-Bares, J., S. Kothari, J. E. Meireles, M. A. Kaproth, P. S. Manos, and A. L. Hipp. 2018. The role of diversification in the continental scale community assembly of the American oaks (*Quercus* L.). *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1049>.
- Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, et al. 1993. Phylogenetic relationships among seed plants based on *rbcL* sequence data. *Annals of the Missouri Botanical Garden* 80: 528–580.
- Edger, P. P., J. C. Hall, A. Harkess, M. Tang, J. Coombs, S. Mohammadin, M. E. Schranz, et al. 2018. Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1040>.
- Eiserhardt, W. L., A. Antonelli, D. J. Bennett, L. R. Botigué, J. G. Burleigh, S. Dodsworth, B. J. Enquist, et al. 2018. A roadmap for global synthesis of the plant tree of life. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1041>.
- Fishbein, M., T. Livshultz, S. C. K. Straub, A. O. Simões, J. Boutte, A. McDonnell, and A. Foote. 2018a. Evolution on the backbone: Apocynaceae phylogenomics and new perspectives on growth forms, flowers, and fruits. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1067>.
- Fishbein, M., S. C. K. Straub, J. Boutte, K. Hansen, R. C. Cronn, and A. Liston. 2018b. Evolution at the tips: *Asclepias* phylogenomics and new perspectives on leaf surfaces. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1062>.
- Folk, R. A., P. S. Soltis, D. E. Soltis, and R. Guralnick. 2018a. New prospects in the detection and comparative analysis of hybridization in the Tree of Life. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1018>.
- Folk, R. A., M. Sun, P. S. Soltis, S. A. Smith, D. E. Soltis, and R. P. Guralnick. 2018b. Challenges of comprehensive taxon sampling in comparative biology: Wrestling with rosids. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1059>.
- Gitzendanner, M. A., P. S. Soltis, G. K.-S. Wong, B. R. Ruhfel, and D. E. Soltis. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1048>.
- Goolsby, E. W., A. J. Moore, L. P. Hancock, J. M. De Vos, and E. J. Edwards. 2018. Molecular evolution of photosynthesis genes during transitions to C₄ and CAM metabolism. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1051>.
- Heyduk, K., J. N. Ray, S. Ayyampalayam, and J. Leebens-Mack. 2018. Shifting gene expression profiles are associated with weak and strong Crassulacean acid metabolism. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1017>.
- Hillis, D. M. 1996. Inferring complex phylogenies. *Nature* 383: 130–131.
- Hinchliff, C. E., S. A. Smith, J. F. Allman, J. G. Burleigh, R. Chaudhary, L. M. Coghill, K. A. Crandall, et al. 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences, USA* 112: 12764–12769.
- Kates, H. R., M. G. Johnson, E. M. Gardner, N. J. C. Zerega, and N. J. Wickett. 2018. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus* (Moraceae). *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1068>.
- Knowles, L., S. A. Smith, H. Huang, and J. Sukumaran. 2018. A matter of phylogenetic scale: Distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1064>.
- Lam, V., H. Darby, V. Merckx, G. Lim, T. Yukawa, K. Neubig, J. R. Abbott, et al. 2018. Phylogenomic inference in extremis: A case study with mycoheterotroph plastomes. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1070>.
- Landis, J. B., D. E. Soltis, Z. Li, H. E. Marx, M. S. Barker, D. C. Tank, and P. S. Soltis. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1060>.
- McManus, H. A., K. Fučíková, P. O. Lewis, L. A. Lewis, and K. G. Karol. 2018. Organellar phylogenomics inform systematics in the green algal family Hydrodictyaceae (Chlorophyceae) and provide clues to the complex evolutionary history of plastid genomes in the green algal tree of life. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1066>.
- Palmer, J. D., D. E. Soltis, and M. W. Chase. 2004. The plant tree of life: an overview and some points of view. *American Journal of Botany* 91: 1437–1445.
- Parks, M., T. Nakov, E. Ruck, N. Wickett, and A. Alverson. 2018. Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1056>.
- Pease, J. B., J. W. Brown, J. F. Walker, C. E. Hinchliff, and S. A. Smith. 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1016>.
- Sessa, E. B., S. M. Chambers, D. Li, L. Trotta, C. L. Endara, J. G. Burleigh, and B. Baiser. 2018. Community assembly of the ferns of Florida. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1073>.
- Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1019>.

- Stoughton, T., R. Kriebel, D. D. Jolles, and R. L. O'Quinn. 2018. Next generation lineage discovery: A case study of tuberous *Claytonia* L. (Montiaceae). *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1061>.
- Stull, G. W., M. Schori, D. E. Soltis, and P. S. Soltis. 2018. Character evolution and missing (morphological) data across *Asteridae*. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1050>.
- Testo, W. L., and M. Sundue. 2018. Are rates of species diversification and body size evolution coupled in the ferns? *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1044>.
- Walker, J. F., Y. Yang, T. Feng, A. Timoneda, J. Mikenas, V. Hutchinson, C. Edwards, et al. 2018. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight to the evolution of Caryophyllales. *American Journal of Botany*. <https://doi.org/10.1002/ajb2.1069>.