American Journal of Botany

RESEARCH ARTICLE

INTRODUCTION

For the Special Issue: Exploring Angiosperms353: a Universal Toolkit for Flowering Plant Phylogenomics

Exploring Angiosperms353: An open, community toolkit for collaborative phylogenomic research on flowering plants

William J. Baker[1,10], Steven Dodsworth[2], Félix Forest[1], Sean W. Graham[3], Matthew G. Johnson[4], Angela McDonnell[5], Lisa Pokorny[1,6], Jennifer A. Tate[7], Susann Wicke[8,9], and Norman J. Wickett[5]

Manuscript received 12 April 2021; revision accepted 14 May 2021.

[1] Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, UK

[2] School of Life Sciences, University of Bedfordshire, University Square, Luton LU1 3JU, UK

[3] Department of Botany, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada

[4] Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA

[5] Plant Science and Conservation, Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, IL 60022, USA

[6] Present address: Centre for Plant Biotechnology and Genomics (CBGP) UPM-INIA, Pozuelo de Alarcón (Madrid) 28223, Spain

[7] School of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand

[8] Plant Evolutionary Biology, Institute for Evolution and Biodiversity, University of Münster, Münster, Germany

[9] Plant Systematics and Biodiversity, Institute for Biology, Humboldt-Universität zu Berlin, Berlin, Germany

[10] Author for correspondence (e-mail: w.baker@kew.org)

Citation: Baker, W. J., S. Dodsworth, F. Forest, S. W. Graham, M. G. Johnson, A. McDonnell, L. Pokorny, J. A. Tate, et al. 2021. Exploring Angiosperms353: an open, community toolkit for collaborative phylogenomic research on flowering plants. *American Journal of Botany* 108(7): 1059–1065.

doi:10.1002/ajb2.1703

The unveiling of the angiosperm (flowering plant) tree of life over the past three decades has been one of the great success stories of modern plant biology. Flowering plants underpin most terrestrial biomes: they fix vast amounts of terrestrial carbon, in turn producing a substantial fraction of planetary oxygen, and drive major biogeochemical cycles. The bulk of human calories are derived either directly (crops) or indirectly (fodder) from angiosperms, as are many medicines, fuel, dyes, beverages, timber, fibers, and other materials. Countless indispensable and mundane items that impact human existence find their origins in flowering plants, and without them, life would be decidedly drearier—imagine a world without herbs, spices, or garden flowers, for example. In this context, the importance of a comprehensive understanding of the angiosperm tree of life cannot be overstated. The tree of life is the fundamental, biological roadmap to the evolution and properties of plants (e.g., Wong et al., 2020). For evolutionary biologists, phylogenies allow us to better understand the spectacular rise of the flowering plants to dominance over the past 140 million or so years (e.g., Lutzoni et al., 2018; Ramírez-Barahona et al., 2020). Information about angiosperm phylogenetic relationships also underpins modern angiosperm classification (e.g., APG IV, 2016), and helps us to better understand species origins and boundaries (e.g., Fazekas et al., 2009). Today, tree of life research is undergoing a renaissance due to the development of powerful, new phylogenomic methods (Dodsworth et al., 2019). In this special issue of the *American Journal of Botany*, together with a companion issue of *Applications in Plant Sciences*, we gather a set of papers that focus on a new, common phylogenomic toolkit, the Angiosperms353 probe set (Johnson et al., 2019), and illustrate its potential for evolutionary synthesis by promoting open collaboration across our community.

Flowering-plant systematists and evolutionary biologists worked cooperatively from the earliest days of molecular systematics, making major discoveries that were fuelled by community-wide agreement on common sets of DNA markers (e.g., Chase et al., 1993; CBOL Plant Working Group, 2009). Until fairly recently, most researchers focused their efforts on sequencing a handful of common genes from the plastid genome and the nuclear ribosomal cistron, as these were readily retrievable using standard laboratory

protocols, namely, the polymerase chain reaction (PCR) and Sanger sequencing. This community-wide approach facilitated data sharing and phylogenetic syntheses at extraordinary scales. For example, data for two of the most commonly sequenced plastid markers, *rbcL* and *matK*, are publicly available for almost 20% of ~350,000 vascular plant species (RBG Kew, 2016). By converging on tractable, standard markers and using well-established and straightforward molecular techniques, the angiosperm systematics community organically established a remarkably effective distributed data set for plant comparative research.

With these advances came an improved understanding of the intractable parts of the angiosperm tree of life (Wickett et al., 2014), reflecting data limitations and/or molecular evolutionary phenomena such as incomplete lineage sorting (ILS; e.g., Maddison, 1997). Fortunately, technological innovations in genomics have created new opportunities to address these and other issues, effectively destabilizing the status quo in plant phylogenetics. Our community has rapidly transitioned toward high-throughput sequencing (HTS) methods, which offer vastly greater volumes of DNA sequence data per taxon, at ever-decreasing cost per base pair. For example, it is now relatively straightforward to recover plastid genomes using HTS methods (e.g., Cronn et al., 2008). However, the greatest rewards from HTS are to be found in the nuclear genome, a much larger source of phylogenetic evidence that was scarcely tapped in the Sanger sequencing era due to the challenges of retrieving single-copy genes (e.g., Small et al., 2004; Mason-Gamer et al., 1998; Oh and Potter, 2003). Today, the development of methods and data that leverage the nuclear genome is the primary focus of the plant phylogenomic community.

The One Thousand Plant Transcriptomes (1KP) initiative (Wickett et al., 2014; One Thousand Plant Transcriptomes Initiative, 2019) has been especially influential in addressing phylogenetic questions with nuclear genome-scale data, as it has placed hundreds of green plant transcriptomes in the public domain (Matasci et al., 2014; Carpenter et al., 2019), providing deep insights into the evolution of green plants. However, transcriptome-based data sets are generated at a scale that exceeds the requirements of current phylogenetic approaches. Transcriptome sequencing also remains relatively costly and has exacting requirements for sample preparation from fresh tissue, which can be limiting for many phylogenetic projects. In contrast, target sequence capture is an HTS technique that subsamples the nuclear genome to a more manageable scale, yielding hundreds of genes per sample at an affordable cost (Hale et al., 2020). Its use is increasingly popular in plants (e.g., Andermann et al., 2020; Dodsworth et al., 2019), which often have very large genomes (Pellicer et al., 2018) that must be rendered tractable for phylogenetic inference. Target sequence capture uses RNA or DNA probes to isolate a specific set of genes from genomic DNA, which are then sequenced using HTS approaches. The method has the added advantage of being highly effective with degraded DNA, opening up avenues for sequencing even centuries-old herbarium specimens (e.g., Hart et al., 2016; Brewer et al., 2019). Customized target-capture probe sets are now routinely designed for specific plant groups, with potentially immense benefits for phylogenetic research in these groups. However, the gene overlap between these custom probe sets is generally minimal, limiting opportunities for data sharing and re-use across studies, and creating obstacles for large-scale data synthesis. A universal, open access toolkit for target sequence capture in any angiosperm group was therefore sorely needed.

ANGIOSPERMS353: A UNIVERSAL TOOLKIT

Angiosperms353 is a target sequence capture probe set that is designed to work across all angiosperm families (Johnson et al., 2019). The toolkit was inspired by analyses of 1KP transcriptomes, which identified 410 low-copy, protein-coding, nuclear genes across all green plants for phylogeny reconstruction (One Thousand Plant Transcriptomes Initiative, 2019). Using a clustering approach, Johnson et al. (2019) defined the minimum set of variants of these genes present in the 1KP data that, when converted to target capture probes, would have a reasonable prospect of capturing each gene in any angiosperm family. Their refinements resulted in a reduction to 353 target genes, hence the toolkit's name. Evidence for the universal effectiveness of Angiosperms353 is rapidly growing. For example, preliminary tests by Johnson et al. (2019), based on 42 samples, showed the kit to be highly effective across angiosperm orders, recovering a median of 137 kb of target sequence out of a theoretical maximum of 261 kb. In a more extensive study involving 2374 samples, a median of 161 kb of target sequence was successfully recovered across 292 (70%) angiosperm families (Baker et al., 2021), from DNA templates of varying quality. In addition to a high degree of universality, the kit offers additional advantages. It can be used off-the-shelf, without costly start-up investments (avoiding production of precursor genomic data). It also alleviates the need for bioinformatic expertise in probe design, levelling the playing field for those seeking to make use of targeted sequence capture approaches. In addition, early evidence suggests that Angiosperms353 genes, if analyzed appropriately, can be phylogenetically informative across multiple taxonomic scales (including at the population level; Beck et al., 2021 [Preprint]; Slimp et al., 2021; Wenzell et al., 2021). Broad uptake of Angiosperms353 is now being reported at conferences (e.g., Lagomarsino and Jabaily, 2020), but to date, published empirical studies in which applications of Angiosperms353 have been fully explored remain relatively few in number (Brewer et al., 2019; Van Andel et al., 2019; Gaynor et al., 2020; Howard et al., 2020 [Preprint]; Larridon et al., 2020, 2021a, 2021b; Murphy et al., 2020; Shee et al., 2020; Baker et al., 2021; Beck et al., 2021 [Preprint]; Starr et al., 2021). This special issue, and its companion in *Applications in Plant Sciences*, aims to address this gap, and to thoroughly explore the potential and pitfalls of this exciting new toolkit.

Higher-level angiosperm phylogeny: from genus to order levels

The initial driver for the design of Angiosperms353 arose in the early stages of the Plant and Fungal Trees of Life project (PAFTOL; www.paftol.org), which aimed to build a complete genus-level angiosperm tree of life using genomic approaches. "First pass" analyses of Angiosperms353 loci across 2333 (17%) angiosperm genera (Baker et al., 2021) resulted in a highly resolved and widely supported phylogenetic tree that included a number of familial and ordinal relationships that challenge established concepts (APG IV, 2016). Much deeper analyses are underway to fully assess the signal and conflict in these data. However, an increasing body of evidence—including 10 studies across monocots, rosids, and asterids presented in this issue—shows that Angiosperms353 data are indeed highly informative for resolving relationships among families within orders (in this special issue: Antonelli et al., 2021; Lee et al., 2021; Maurin et al., 2021; Thomas et al., 2021b; Zuntini et al., 2021), and among genera within families (Gaynor et al., 2020; Howard et al., 2020 [Preprint]; Larridon et al., 2021a, b; Starr et al., 2021; and in this special issue:

Buerki et al., 2021; Clarkson et al., 2021; Pérez-Escobar et al., 2021; Pillon et al., 2021; Shah et al., 2021). Within each of the groups covered here, Angiosperms353 is resetting the phylogenetic baseline with far larger data quantities than have previously been utilized, including the sampling of many genera for which no DNA sequence data are currently available in public repositories (e.g., Buerki et al., 2021; Clarkson et al., 2021). In some families (e.g., Commelinaceae: Zuntini et al., 2021; Cunoniaceae: Pillon et al., 2021; Cyperaceae: Larridon et al., 2021b, Starr et al., 2021; Sapindaceae: Buerki et al., 2021) classifications are being revised as a result.

Applications at the species level and below

There is mounting evidence that the sequence data recovered by the Angiosperms353 probe set are sufficiently variable to reconstruct relationships at the species level, especially when noncoding sequences serendipitously captured from regions flanking the target genes (the so-called "splash zone") are also taken into account. A number of studies exploring the suitability of Angiosperms353 for species-level phylogenetic inference have already been published (e.g., Gaynor et al., 2020; Howard et al., 2020 [Preprint]; Larridon et al., 2020, 2021a, 2021b; Murphy et al., 2020; Shee et al., 2020; Starr et al., 2021). In this issue, A. Thomas et al. (2021a) specifically investigated species-level relationships among recently radiated lineages in *Veronica* sect. *Hebe* (Plantaginaceae), similar to Larridon et al. (2020) and Shee et al. (2020), who also dealt with young genera. Other studies have also provided insights into the utility of Angiosperms353 at the species level (e.g., *Cornus*, Cornaceae: S. Thomas et al., 2021b; *Combretum*, Combretaceae: Maurin et al., 2021; *Lonicera*, Caprifoliaceae; and *Viburnum*, Adoxaceae: Lee et al., 2021).

Targeted enrichment approaches have generally not been the method of choice for population genomics studies, which usually favor techniques such as RADseq or genotyping-by-sequencing (e.g., Schley et al., 2020; Hunt et al., 2021). Taxon-specific probe sets have been shown to be suitable at various taxonomic levels (e.g., Nicholls et al., 2015; Villaverde et al., 2018; Soto Gomez et al., 2019; Christe et al., 2021) and might be expected to perform somewhat better at resolving relationships at lower taxonomic levels than a universal one (but see Larridon et al., 2020; Siniscalchi et al., 2021; Ufimov et al., 2021). In this special issue, two studies pay particular attention to the potential of Angiosperms353 for deciphering relationships below the species level. Wenzell et al. (2021) found limited genetic variation in the Angiosperms353 genes among recently diverged species of *Castilleja* (Orobanchaceae), which are otherwise well characterized by floral color diversity. Ottenlips et al. (2021) attempted to shed new light on the *Lomatium triternatum* (Apiaceae) species complex, which Sanger-based sequencing data have so far failed to resolve with confidence. On the basis of their Angiosperms353 data, they postulated that strong ILS effects are at play and inferred that this group may comprise a number of cryptic species, some of which relate to previous taxonomic circumscriptions. Slimp et al. (2021) provide one of the most in-depth investigations to date on the suitability of Angiosperms353 for population genomics. They examined a regional data set rather than focusing on a particular flowering-plant group and show that Angiosperms353 have excellent promise for conservation genomics and environmental DNA-based approaches. This approach should be useful for considering all angiosperm species in an ecosystem without the need for species-specific marker design, as would be required with other population genomics techniques. This paves the way for community-level Angiosperms353 studies. The studies noted above indicate that the potential of Angiosperms353 could be considerable for species- and population-level applications and has potential as a novel tool for DNA barcoding for the molecular identification of modern, historical, ancient, and mixed environmental samples (Kistler et al., 2020; Folk et al., 2021).

Going off-target: organellar genomes and high-copy nuclear elements

Plastid DNA sequence (and even whole plastid genomes) can be recovered from target capture sequence data, including those enriched for Angiosperms353 loci, by capitalizing on the off-target sequence reads from the unenriched background DNA (Weitemier et al., 2014; Antonelli et al., 2021; Lee et al., 2021; Zuntini et al., 2021). The elevated stoichiometry of the plastid genome in plant cells makes this possible as there are generally many plastids per cell, and multiple genome copies per plastid. In a recent study of 2394 samples, a median of ~20% of the plastid genome was recovered from off-target reads in Angiosperms353-enriched libraries (Baker et al., 2021). Plastid genome recovery per library can also be enhanced by addition of unenriched library template to a target-enriched one prior to sequencing (Dodsworth et al., 2019). Though often incomplete, the plastid genome data recovered as by-catch from target sequence capture libraries usually includes at least partial recovery of the ~78 protein-coding genes of the plastid genome, including genes such as *rbcL* or *matK* that have been widely used for decades in plant molecular phylogenetics.

Angiosperms353 data sets have several considerable advantages over plastid genome data sets: they tend to be larger, cover a greater number of more variable genes, and sample a much wider breadth of genomic space. They can also be analyzed using multispecies coalescent (MSC) approaches (e.g., ASTRAL-III, Zhang et al., 2018), which assume that genes are unlinked, whereas plastid genomes are generally considered to be single-linkage groups (e.g., Birky, 2001). However, analyzing plastid data in parallel with Angiosperms353 data is useful as it permits integration with existing large-scale data sets (e.g., Zuntini et al., 2021), creating opportunities for comparison and contrast. The vast back-catalog of plastid data produced primarily in the pre-genomic era remains extremely relevant as the foundation of much current phylogenetic thinking (e.g., APG IV, 2016, and its predecessors, which relied extensively on plastid gene data sets), and because plastid phylogenomic data continue to be actively collected (e.g., Gitzendanner et al., 2018; Lam et al., 2018).

The mitochondrial genome has been largely ignored for phylogenomic inference in flowering plants because of its generally extremely slow substitution rates (Wolfe et al., 1987) and very high rates of genome structural rearrangement (Gualberto and Newton, 2017); the latter makes assembly of whole genomes difficult. However, the ~40 protein-coding mitochondrial genes considered together can perform well in higher-level phylogenomic inference (e.g., Bell et al., 2020; Soto Gomez et al., 2020; Sousa et al., 2020). More effort could therefore be made to investigate the mitochondrion as an additional usable by-product of sequence data from target sequence capture studies (e.g., Liu et al., 2019), although mitochondrial genome recovery rates are expected to be lower (approximately 10-fold lower copy number per cell, relative to plastid genomes; Gualberto and Newton, 2017).

In addition to organellar off-target markers, it is possible to recover high-copy nuclear elements, such as the nuclear ribosomal cistron, including the nuclear ribosomal internal transcribed spacers (nrITS),

which has been a mainstay of plant phylogenetics for decades. These can then be integrated with other large-scale data sets (e.g., Shee et al., 2020). Baker et al. (2021) were able to recover a median of 85% of the nuclear ribosomal cistron from off-target reads from 2349 Angiosperms353-enriched DNA libraries. It is also possible to retrieve genomic repeats, including satellite repeats, DNA transposons, and retroelements from off-target reads, which have been shown to have usable signal for phylogenetic inference (Dodsworth et al., 2015; Vitales et al., 2020).

CHALLENGES

While targeted sequencing methods are revolutionizing plant molecular phylogenetic studies across taxonomic levels, a number of analytical and biological challenges remain (McKain et al., 2018). Here, we highlight a few areas relevant for users of the Angiosperms353 probe set: (1) gene duplication, resulting from gene family expansions or whole-genome duplications (WGD), often followed by genome contraction (i.e., diploidization; Wendel, 2015); (2) missing data; and (3) gene-tree conflict. Among the most difficult current challenges is the analytical handling of gene duplications (and losses) to achieve proper identification and separation of paralogs or homeologs (Fernández et al., 2020). Paralogs result from gene duplication within a genome, whereas homeologs are duplicates that are reunited in an allopolyploid (or homoploid) after hybridization between species (Glover et al., 2016). Although the Angiosperms353 probe set was designed specifically to target putative single-copy genes, the prevalence of WGD and polyploidy in flowering plants means that duplications and losses, even of these genes, are inevitable in many flowering plant lineages. At present, most users remove paralogs following bioinformatic "warnings" (Johnson et al., 2016), as they can result in discordant gene trees and increase phylogenetic noise (e.g., Soto Gomez et al., 2019). However, excluding these paralogs might mean missing out on informative markers (Gardner et al., 2020), particularly for understanding and reconstructing hybrid and polyploid species origins (Morales-Briones et al., 2018, 2021; Nauheimer et al., 2021). For example, Lee et al. (2021) identified paralogs in their data set that were mostly restricted to a particular clade. Visual inspection of the "paralog warnings" (Johnson et al., 2016) in a phylogenetic context allowed these authors (Lee et al., 2021) to rescue a large number of markers, which would have otherwise been discarded, for downstream phylogenomic inference. The overall impact of paralogs can vary as well and depend on the group under study. For example, duplicates identified by Soto Gomez et al. (2020) were found to have little effect on species-tree inference. A. Thomas et al. (2021b) report that few paralog warnings were identified in *Veronica* sect. *Hebe*, even though the group is known to contain high polyploids (12*x* and higher). The authors attempted to use available phasing programs to separate duplicated gene copies, but these proved inefficient, possibly because the programs were designed for diploid organisms. Overall, we expect development of new analytical tools will continue to improve our ability to use Angiosperms353 and related data types to reconstruct difficult reticulate evolutionary histories (e.g., Freyman et al., 2020 [Preprint]; Rothfels, 2021). New methods, such as ASTRAL-Pro (Zhang et al., 2020), should allow improved handling of paralogs in phylogenomic analyses.

Another challenge of the Angiosperms353 probe set is that gene recovery rates can be moderately to substantially patchy (e.g., Baker et al., 2021), resulting in missing data (partial gene recovery, missing genes). Various lines of evidence from simulated and empirical data downplay the negative effects of missing data on phylogenetic inference in simultaneous analyses of concatenated data (e.g., Wiens, 2006; Wiens and Tiu, 2012) and analyses within a MSC framework (e.g., Hosner et al., 2016; Molloy and Warnow, 2018; Nute et al., 2018). In this special issue, Shah et al. (2021) tested the impact of missing data using different levels of filtering within their Ochnaceae Angiosperms353 data set. By testing different thresholds of missing data included in the alignments (in this case at individual sites), Shah et al. (2021) showed that certain thresholds decreased phylogenetic noise and increased the number of informative characters, resulting in more robust relationships. Somewhat elevated levels of missing data are an expected trade-off of a universal probe set, versus custom probes designed for greater fidelity to a specific lineage. Nevertheless, the patchiness of Angiosperms353 data sets appears not to inhibit the recovery of robust and well-supported phylogenetic trees (e.g., Baker et al., 2021). Bioinformatic improvements also promise to enhance gene recovery, further mitigating the impacts of missing data (McLay et al., 2021).

Quantifying and characterizing gene-tree conflict is a challenge that is not unique to Angiosperms353 data and is explicitly considered as part of species-tree inference when accommodating conflicting gene trees (e.g., Maddison, 1997). In addition to the paralogy issues discussed above, gene trees reconstructed from individual nuclear loci can experience dissimilar phylogenetic histories resulting from several biological processes, including incomplete lineage sorting (ILS) and horizontal gene transfer (HGT) mediated by introgression and other mechanisms. Both ILS and HGT conflicts can be well accommodated by using MSC methods (e.g., Davidson et al., 2015; Zhang et al., 2018). These approaches also allow derivation of quartet-based support measures summarizing main and conflicting relationships found among the underlying gene trees. Several of the papers in this special issue also use MSC approaches to analyze plastome data sets. This practice may be problematic, as neither ILS nor HGT are expected to result in disagreements among individual plastid gene trees, which are in general thought to be fully linked (e.g., Birky, 2001). In this sense, MSC frameworks are inappropriate for analyzing plastid data generated using Angiosperms353 or by other means. Intriguingly, however, tree disagreement among plastid genes estimated in quartet-based inference can approach levels seen for nuclear data sets (Antonelli et al., 2021; Pérez-Escobar et al., 2021; Zuntini et al., 2021). These conflicts likely reflect gene-tree estimation error, or other sources of conflict among plastid loci such as long-branch attraction (rather than ILS or HGT) and serve as a useful reminder that these other sources of gene-tree conflict can affect nuclear data too. Further work could be done to validate the use of MSC approaches for plastid data from Angiosperms353 studies or elsewhere and to better understand what the uncovered gene-tree conflicts mean in such analyses.

PROSPECTS AND OPPORTUNITIES

With tree of life research transitioning inexorably toward a genomic future, the Angiosperms353 probe set offers an open and accessible mechanism for the angiosperm phylogenomics community to embrace new methods in an integrated way, building on the collaborative spirit of the Sanger sequencing era, from which plant science continues to reap so many benefits. As this special issue shows, applications of the toolkit are wide-ranging, and these studies represent just the first wave of many in the pipeline. Target sequence capture has become the method of choice for genomic characterization of herbarium specimens, opening the door to species-level phylogenetic research at ever finer scale and a wealth of new research avenues that leverage the

world's preserved collections (Buerki and Baker, 2016; Hart et al., 2016; Brewer et al., 2019). Importantly, the Angiosperms353 toolkit presents the opportunity to achieve this in a coordinated manner, via the standardization of the target gene set, maximizing benefits across the community through data sharing and reuse. It is even being adopted for continental scale biodiversity genomic programmes, such as the Genomics for Australian Plants project (https://www.genomicsforaustralianplants.com/), which is coordinating efforts with the PAFTOL project to sequence all Australian genera and focal radiations, drawing extensively from the specimens held in the network of Australian herbaria. The open data approach underlying the development of Angiosperms353 (Johnson et al., 2019; Baker et al., 2021) is ensuring that a substantial foundational body of data comprising thousands of samples is already available in the public domain, for example, via the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) or the Kew Tree of Life Explorer (https://treeoflife.kew.org/). Moreover, a compatible probe set targeting many of the same loci has been developed for nonflowering land plants (Breinholt et al., 2021), presenting the enticing prospect of data integration across all land plants.

Researchers who have developed custom probe sets for their group of choice may now feel that they face an invidious choice between the taxon-specific focus of their kit and integrative benefits of Angiosperms353. In practice, this dichotomy no longer applies—different probe kits can be combined in a single hybridization reaction to retrieve multiple target gene sets simultaneously, as has recently been demonstrated by Hendriks et al. (2021). An even more efficient solution that is gaining popularity is the inclusion of Angiosperms353 genes in the design of custom probe sets (e.g., Jantzen et al., 2020; Christe et al., 2021; Eserman et al., 2021; Ogutcen et al., 2021). Looking forward, however, we can envisage a time in the not-too-distant future when sequencing costs decrease so far that target capture via hybridization in vitro will itself become redundant, with target genes being retrieved bioinformatically from deeply sequenced samples. A fuller understanding of the nature of the genes themselves will be an important step toward substantiating their potential as standard markers for the phylogenomic era. As the race to sequence the genomes of life on Earth gathers pace (Cheng et al., 2018; Lewin et al., 2018), tools such as Angiosperms353 have a central role to play in bringing us ever closer to our ultimate collective goal of a complete, unified tree of life for all species.

ACKNOWLEDGMENTS

We thank all authors for their contributions to this special issue. We are also grateful to all reviewers for thoughtful evaluations of the manuscripts. Special thanks go to the BSA publications team, especially to Amy McPherson and Pamela Diggle, for supporting the guest editors throughout with care, diligence, and good humor.

LITERATURE CITED

Andermann, T., M. F. Torres Jiménez, P. Matos-Maraví, R. Batista, J. L. Blanco-Pastor, A. L. S. Gustafsson, L. Kistler, et al. 2020. A guide to carrying out a phylogenomic target sequence capture project. *Frontiers in Genetics* 10: 1407.

Antonelli, A., J. J. Clarkson, K. Kainulainen, O. Maurin, G. E. Brewer, A. P. Davis, N. Epitawalage, et al. 2021. Settling a family feud: a high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. *American Journal of Botany* 108: 1143–1165.

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.

Baker, W. J., P. Bailey, V. Barber, A. Barker, S. Bellot, D. Bishop, L. R. Botigué, et al. 2021. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology*: syab035. https://doi.org/10.1093/sysbio/syab035.

Beck, J. B., M. L. Markley, M. G. Zielke, J. R. Thomas, H. J. Hale, L. D. Williams, and M. G. Johnson. 2021. Is Palmer's elm leaf goldenrod real? The angiosperms353 kit provides within-species signal in *Solidago ulmifolia* s.l. *bioRxiv* 2021.01.07.425781 [Preprint].

Bell, D., Q. Lin, W. K. Gerelle, S. Joya, Y. Chang, Z. N. Taylor, C. J. Rothfels, et al. 2020. Organellomic data sets confirm a cryptic consensus on (unrooted) land–plant relationships and provide new insights into bryophyte molecular evolution. *American Journal of Botany* 107: 91–115.

Birky, C. W. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics* 35: 125–148.

Breinholt, J. W., S. B. Carey, G. P. Tiley, E. C. Davis, L. Endara, S. F. McDaniel, L. G. Neves, et al. 2021. A target enrichment probe set for resolving the flagellate land plant tree of life. *Applications in Plant Sciences* 9: e11406.

Brewer, G. E., J. J. Clarkson, O. Maurin, A. R. Zuntini, V. Barber, S. Bellot, N. Biggs, et al. 2019. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science* 10: 1102.

Buerki, S., and W. J. Baker. 2016. Collections-based research in the genomic era. *Biological Journal of the Linnean Society* 117: 5–10.

Buerki, S., M. W. Callmander, P. Acevedo-Rodriguez, P. P. Lowry II, J. Munzinger, P. Bailey, O. Maurin, et al. 2021. An updated infra-familial classification of Sapindaceae based on targeted enrichment data. *American Journal of Botany* 108: 1234–1251.

Carpenter, E. J., N. Matasci, S. Ayyampalayam, S. Wu, J. Sun, J. Yu, F. R. Jimenez Vieira, et al. 2019. Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *GigaScience* 8: giz126.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA* 106: 12794–12797.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden* 80: 528–580.

Cheng, S., M. Melkonian, S. A. Smith, S. Brockington, J. M. Archibald, P. M. Delaux, F. W. Li, et al. 2018. 10KP: A phylodiverse genome sequencing plan. *GigaScience* 7: giy013.

Christe, C., C. G. Boluda, D. Koubínová, L. Gautier, and Y. Naciri. 2021. New genetic markers for Sapotaceae phylogenomics: more than 600 nuclear genes applicable from family to population levels. *Molecular Phylogenetics and Evolution* 160: 107123.

Clarkson, J. J., A. R. Zuntini, O. Maurin, S. R. Downie, G. M. Plunkett, A. N. Nicolas, J. F. Smith, et al. 2021. A higher-level nuclear phylogenomic study of the carrot family (Apiaceae). *American Journal of Botany* 108: 1252–1269.

Cronn, R., A. Liston, M. Parks, D. S. Gernandt, R. Shen, and T. Mockler. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research* 36: e122.

Davidson, R., P. Vachaspati, S. Mirarab, and T. Warnow. 2015. Phylogenomic species tree estimation in the presence of incomplete lineage sorting and horizontal gene transfer. *BMC Genomics* 16: S1.

Dodsworth, S., M. W. Chase, L. J. Kelly, I. J. Leitch, J. Macas, P. Novák, M. Piednoël, et al. 2015. Genomic repeat abundances contain phylogenetic signal. *Systematic Biology* 64: 112–126.

Dodsworth, S., L. Pokorny, M. G. Johnson, J. T. Kim, O. Maurin, N. J. Wickett, F. Forest, et al. 2019. Hyb-seq for flowering plant systematics. *Trends in Plant Science* 24: 887–891.

Eserman, L. A., S. K. Thomas, E. E. D. Coffey, and J. H. Leebens-Mack. 2021. Target sequence capture in orchids: Developing a kit to sequence hundreds of single-copy loci. *Applications in Plant Sciences* 9: e11416.

Fazekas, A. J., P. R. Kesanakurti, K. S. Burgess, D. M. Percy, S. W. Graham, S. C. H. Barrett, S. G. Newmaster, et al. 2009. Are plant species inherently harder to

discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources* 9: 130–139.

Fernández, R., T. Gabaldon, and C. Dessimoz. 2020. Orthology: definitions, prediction, and impact on species phylogeny inference. *In* C. Scornavacca, F. Delsuc, and N. Galtier [eds.], Phylogenetics in the genomic era, 2.4:1–2.4:14. [no commercial publisher, authors open access book].

Folk, R. A., H. R. Kates, R. LaFrance, D. E. Soltis, P. S. Soltis, and R. P. Guralnick. 2021. High-throughput methods for efficiently building massive phylogenies from natural history collections. *Applications in Plant Sciences* 9: e11410.

Freyman, W. A., M. G. Johnson, and C. J. Rothfels. 2020. homologizer: Phylogenetic phasing of gene copies into polyploid subgenomes. *bioRxiv* 2020.10.22.351486 [Preprint].

Gardner, E. M., M. G. Johnson, J. T. Pereira, A. S. Ahmad Puad, D. Arifiani, N. J. Wickett, N. J. C. Zerega, et al. 2020. Paralogs and off-target sequences improve phylogenetic resolution in a densely sampled study of the breadfruit genus (*Artocarpus*, Moraceae). *Systematic Biology* 70: 558–575.

Gaynor, M. L., C.-N. Fu, L.-M. Gao, L.-M. Lu, D. E. Soltis, and P. S. Soltis. 2020. Biogeography and ecological niche evolution in Diapensiaceae inferred from phylogenetic analysis. *Journal of Systematics and Evolution* 58: 646–662.

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* 105: 291–301.

Glover, N. M., H. Redestig, and C. Dessimoz. 2016. Homoeologs: What are they and how do we infer them? *Trends in Plant Science* 21: 609–621.

Gualberto, J. M., and K. J. Newton. 2017. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annual Review of Plant Biology* 68: 225–252.

Hale, H., E. M. Gardner, J. Viruel, L. Pokorny, and M. G. Johnson. 2020. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in Plant Sciences* 8: e11337.

Hart, M. L., L. L. Forrest, J. A. Nicholls, and C. A. Kidner. 2016. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* 65: 1081–1092.

Hendriks, K. P., T. Mandáková, N. M. Hay, E. Ly, A. Hooft van Huysduynen, R. Tamrakar, S. K. Thomas, et al. 2021. The best of both worlds: combining lineage specific and universal bait sets in target enrichment hybridization reactions. *Applications in Plant Sciences* 9: e11438.

Hosner, P. A., B. C. Faircloth, T. C. Glenn, E. L. Braun, and R. T. Kimball. 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution* 33: 1110–1125.

Howard, C. C., A. A. Crowl, T. S. Harvey, and N. Cellinese. 2020. Peeling back the layers: the complex dynamics shaping the evolution of the Ledebouriinae (Scilloideae, Asparagaceae). *bioRxiv* 2020.11.02.365718 [Preprint].

Hunt, H. V., N. A. S. Przelomska, M. G. Campana, J. Cockram, H. F. J. Bligh, C. J. Kneale, O. I. Romanova, et al. 2021. Population genomic structure of Eurasian and African foxtail millet landrace accessions inferred from genotyping-by-sequencing. *Plant Genome* 14: e20081.

Jantzen, J. R., P. Amarasinghe, R. A. Folk, M. Reginato, F. A. Michelangeli, D. E. Soltis, N. Cellinese, et al. 2020. A two-tier bioinformatic pipeline to develop probes for target capture of nuclear loci with applications in Melastomataceae. *Applications in Plant Sciences* 8: e11345.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. Zerega, et al. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: 1600016.

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eiserhardt, et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68: 594–606.

Kistler, L., V. C. Bieker, M. D. Martin, M. W. Pedersen, J. Ramos Madrigal, and N. Wales. 2020. Ancient plant genomics in archaeology, herbaria, and the environment. *Annual Review of Plant Biology* 71: 605–629.

Lagomarsino, L. P., and R. S. Jabaily. 2020. Virtual Botany Conference 2020 Symposium – Angiosperms353: a new essential tool for plant systematics. Website: http://2020.botanyconference.org/engine/search/index.php?func=detail&aid=941 [accessed 31 March 2021].

Lam, V. K. Y., H. Darby, V. S. F. T. Merckx, G. Lim, T. Yukawa, K. M. Neubig, J. R. Abbott, et al. 2018. Phylogenomic inference *in extremis*: A case study with mycoheterotroph plastomes. *American Journal of Botany* 105: 480–494.

Larridon, I., T. Villaverde, A. R. Zuntini, L. Pokorny, G. E. Brewer, N. Epitawalage, I. Fairlie, et al. 2020. Tackling rapid radiations with targeted sequencing. *Frontiers in Plant Science* 10: 1655.

Larridon, I., A. R. Zuntini, R. L. Barrett, K. L. Wilson, J. J. Bruhl, P. Goetghebeur, W. J. Baker, et al. 2021a. Resolving generic limits in Cyperaceae tribe Abildgaardieae using targeted sequencing. *Botanical Journal of the Linnean Society* 196: 163–187.

Larridon, I., A. R. Zuntini, É. Léveillé-Bourret, R. L. Barrett, J. R. Starr, M. Muasya, T. Villaverde, et al. 2021b. A new classification of Cyperaceae (Poales) supported by phylogenomic data. *Journal of Systematics and Evolution*. In press.

Lee, A. K., I. S. Gilman, M. Srivastav, A. D. Lerner, M. J. Donoghue, and W. L. Clement. 2021. Reconstructing Dipsacales phylogeny using Angiosperms353: issues and insights. *American Journal of Botany* 108: 1122–1142.

Lewin, H. A., G. E. Robinson, W. J. Kress, W. J. Baker, J. Coddington, K. A. Crandall, R. Durbin, et al. 2018. Earth biogenome project: sequencing life for the future of life. *Proceedings of the National Academy of Sciences, USA* 115: 4325–4333.

Liu, Y., M. G. Johnson, C. J. Cox, R. Medina, N. Devos, A. Vanderpoorten, L. Hedenäs, et al. 2019. Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nature Communications* 10: 1–11.

Lutzoni, F., M. D. Nowak, M. E. Alfaro, V. Reeb, J. Miadlikowska, M. Krug, A. E. Arnold, et al. 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications* 9: 5451.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

Mason-Gamer, R. J., C. F. Weil, and E. A. Kellogg. 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. *Molecular Biology and Evolution* 15: 1658–1673.

Matasci, N., L. H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3: 2047-217X-3-17.

Maurin, O., A. Anest, S. Bellot, E. Biffin, G. Brewer, T. Charles-Dominique, R. S. Cowan, et al. 2021. A nuclear phylogenomic study of the angiosperm order Myrtales, exploring the potential and limitations of the universal Angiosperms353 probe set. *American Journal of Botany* 108: 1087–1111.

McKain, M. R., M. G. Johnson, S. Uribe-Convers, D. Eaton, and Y. Yang. 2018. Practical considerations for plant phylogenomics. *Applications in Plant Sciences* 6: e1038.

McLay, T. G. B., J. L. Birch, B. F. Gunn, W. Ning, J. A. Tate, L. Nauheimer, E. M. Joyce, L. Simpson, et al. 2021. New targets acquired: improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* 9: e11420.

Molloy, E. K., and T. Warnow. 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. *Systematic Biology* 67: 285–303.

Morales-Briones, D. F., G. Kadereit, D. T. Tefarikis, M. J. Moore, S. A. Smith, S. F. Brockington, A. Timoneda, et al. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. *Systematic Biology* 70: 219–235.

Morales-Briones, D. F., A. Liston, and D. C. Tank. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist* 218: 1668–1684.

Murphy, B., F. Forest, T. Barraclough, J. Rosindell, S. Bellot, R. Cowan, M. Golos, et al. 2020. A phylogenomic analysis of *Nepenthes* (Nepenthaceae). *Molecular Phylogenetics and Evolution* 144: 106668.

Nauheimer, L., N. Weigner, E. Joyce, D. Crayn, C. Clarke, and K. Nargar. 2021. HybPhaser: a workflow for the detection and phasing of hybrids in target capture datasets. *Applications in Plant Sciences* 9: e11441.

Nicholls, J., R. Pennington, E. Koenen, C. Hughes, J. Hearn, L. Bunnefeld, K. Dexter, et al. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science* 6: 710.

Nute, M., J. Chou, E. K. Molloy, and T. Warnow. 2018. The performance of coalescent-based species tree estimation methods under models of missing data. *BMC Genomics* 19: 286.

Oh, S.-H., and D. Potter. 2003. Phylogenetic utility of the second intron of *LEAFY* in *Neillia* and *Stephanandra* (Rosaceae) and implications for the origin of *Stephanandra*. *Molecular Phylogenetics and Evolution* 29: 203–215.

Ogutcen, E., C. Christe, K. Nishii, N. Salamin, M. Möller, and M. Perret. 2021. Phylogenomics of Gesneriaceae using targeted capture of nuclear genes. *Molecular Phylogenetics and Evolution* 157: 107068.

One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679–685.

Ottenlips, M. V., D. H. Mansfield, S. Buerki, M. A. E. Feist, S. R. Downie, S. Dodsworth, F. Forest, et al. 2021. Resolving species boundaries in a recent radiation with the Angiosperms353 probe set: The *Lomatium packardiae/L. anomalum* clade of the *L. triternatum* (Apiaceae) complex. *American Journal of Botany* 108: 1217–1233.

Pellicer, J., O. Hidalgo, S. Dodsworth, and I. J. Leitch. 2018. Genome size diversity and its impact on the evolution of land plants. *Genes* 9: 88.

Pérez-Escobar, Ó. A., S. Dodsworth, D. Bogarín, S. Bellot, J. A. Balbuena, R. J. Schley, I. A. Kikuchi, et al. 2021. Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *American Journal of Botany* 108: 1166–1180.

Pillon, Y., H. C. F. Hopkins, O. Maurin, N. Epitawalage, J. Bradford, Z. S. Rogers, W. J. Baker, and F. Forest. 2021. Phylogenomics and biogeography of Cunoniaceae (Oxalidales) with complete generic sampling and taxonomic realignments. *American Journal of Botany* 108: 1181–1200.

Ramírez-Barahona, S., H. Sauquet, and S. Magallón. 2020. The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology and Evolution* 4: 1232–1238.

RBG Kew. 2016. The state of the world's plants report – 2016. Royal Botanic Gardens, Kew, UK.

Rothfels, C. J. 2021. Polyploid phylogenetics. *New Phytologist* 230: 66–72.

Schley, R. J., R. T. Pennington, O. A. Pérez-Escobar, A. J. Helmstetter, M. de la Estrella, I. Larridon, I. A. B. S. Kikuchi, et al. 2020. Introgression across evolutionary scales suggests reticulation contributes to Amazonian tree diversity. *Molecular Ecology* 29: 4170–4185.

Shee, Z. Q., D. G. Frodin, R. Cámara-Leret, and L. Pokorny. 2020. Reconstructing the complex evolutionary history of the Papuasian *Schefflera* radiation through herbariomics. *Frontiers in Plant Science* 11: 258.

Shah, T., J. V. Schneider, G. Zizka, O. Maurin, W. Baker, F. Forest, G. E. Brewer, V. Savolainen, et al. 2021. Joining forces in Ochnaceae phylogenomics: A tale of two targeted sequencing probe kits. *American Journal of Botany* 108: 1201–1216.

Siniscalchi, C. M., O. Hidalgo, L. Palazzesi, J. Pellicer, L. Pokorny, O. Maurin, I. J. Leitch, et al. 2021. Lineage-specific vs. universal: comparison of the Compositae1061 and Angiosperms353 enrichment panels in the sunflower family. *Applications in Plant Sciences* 9: e11422.

Slimp, M., L. D. Williams, H. Hale, and M. G. Johnson. 2021. On the potential of Angiosperms353 for population genomic studies. *Applications in Plant Sciences* 9: e11419.

Small, R. L., R. Cronn, and F. J. Wendel. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17: 145–170.

Soto Gomez, M., Q. Lin, E. da Silva Leal, T. J. Gallaher, D. Scherberich, C. B. Mennes, S. Y. Smith, et al. 2020. A bi-organellar phylogenomic study of Pandanales: inference of higher-order relationships and unusual rate-variation patterns. *Cladistics* 36: 481–504.

Soto Gomez, M., L. Pokorny, M. B. Kantar, F. Forest, I. J. Leitch, B. Gravendeel, P. Wilkin, et al. 2019. A customized nuclear target enrichment approach for developing a phylogenomic baseline for *Dioscorea* yams (Dioscoreaceae). *Applications in Plant Sciences* 7: e11254.

Sousa, F., P. Cíváň, J. Brazão, P. G. Foster, and C. J. Cox. 2020. The mitochondrial phylogeny of land plants shows support for Setaphyta under composition-heterogeneous substitution models. *PeerJ* 8: e8995.

Starr, J. R., P. Jiménez-Mejías, A. R. Zuntini, É. Léveillé-Bourret, I. Semmouri, M. Muasya, W. J. Baker, et al. 2021. Targeted sequencing supports morphology and embryo features in resolving the classification of Cyperaceae tribe Fuireneae s.l. *Journal of Systematics and Evolution* (in press). https://doi.org/10.1111/jse.12721.

Thomas, A. E., J. Igea, H. M. Meudt, D. C. Albach, W. G. Lee, and A. J. Tanentzap. 2021a. Using target sequence capture to improve the phylogenetic resolution of a rapid radiation in New Zealand *Veronica*. *American Journal of Botany* 108: 1289–1306.

Thomas, S. K., X. Liu, Z.-Y. Du, Y. Dong, A. Cummings, L. Pokorny, Q.-Y. Xiang, and J. H. Leebens-Mack. 2021b. Comprehending Cornales: phylogenetic reconstruction of the order using the Angiosperms353 probe set. *American Journal of Botany* 108: 1112–1121.

Ufimov, R., V. Zeisek, S. Píšová, W. J. Baker, T. Fér, M. van Loo, C. Dobeš, and R. Schmickl. 2021. Relative performance of customized and universal probe sets in target enrichment: a case study in subtribe Malinae (Rosaceae). *Applications in Plant Sciences* 9 (in press).

Van Andel, T., M. A. Veltman, A. Bertin, H. Maat, T. Polime, D. Hille Ris Lambers, J. Tjoe Awie, et al. 2019. Hidden rice diversity in the Guianas. *Frontiers in Plant Science* 10: 1161.

Villaverde, T., L. Pokorny, S. Olsson, M. Rincón-Barrado, M. G. Johnson, E. M. Gardner, N. J. Wickett, et al. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* 220: 636–650.

Vitales, D., S. Garcia, and S. Dodsworth. 2020. Reconstructing phylogenetic relationships based on repeat sequence similarities. *Molecular Phylogenetics and Evolution* 147: 106766.

Weitemier, K., S. C. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: 1400042.

Wendel, J. F. 2015. The wondrous cycles of polyploidy in plants. *American Journal of Botany* 102: 1753–1756.

Wenzell, K. E., A. J. McDonnell, N. J. Wickett, J. B. Fant, and K. A. Skogen. 2021. Incomplete reproductive isolation and low genetic differentiation despite floral divergence across varying geographic scales in *Castilleja*. *American Journal of Botany* 108: 1270–1288.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalaya, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* 111: E4859–E4868.

Wiens, J. J. 2006. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* 39: 34–42.

Wiens, J. J., and J. Tiu. 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. *PLoS One* 7: e42925.

Wolfe, K. H., W. H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058.

Wong, G. K.-S., D. E. Soltis, J. Leebens-Mack, N. J. Wickett, M. S. Barker, Y. Van de Peer, S. W. Graham, et al. 2020. Sequencing and analyzing the transcriptomes of a thousand species across the Tree of Life for green plants. *Annual Review of Plant Biology* 71: 741–765.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.

Zhang, C., C. Scornavacca, E. K. Molloy, and S. Mirarab. 2020. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution* 37: 3292–3307.

Zuntini, A. R., L. P. Frankel, L. Pokorny, F. Forest, and W. J. Baker. 2021. A comprehensive phylogenomic study of the monocot order Commelinales, with a new classification of Commelinaceae. *American Journal of Botany* 108: 1066–1086.