











INTRODUCTION

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

Exploring Angiosperms353: Developing and applying a universal toolkit for flowering plant phylogenomics

Angela J. McDonnell^{1,10} , William J. Baker² , Steven Dodsworth³ , Félix Forest² , Sean W. Graham⁴ , Matthew G. Johnson⁵ , Lisa Pokorny^{2,6} , Jennifer Tate⁷ , Susann Wicke^{8,9} , and Norman J. Wickett¹ 

Manuscript received 11 June 2021; revision accepted 25 June 2021.

¹Negaunee Institute for Plant Conservation Science and Action, Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, Illinois 60022, USA

²Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, United Kingdom

³School of Life Sciences, University of Bedfordshire, University Square, Luton LU1 3JU, United Kingdom

⁴Department of Botany, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada

⁵Department of Biological Sciences, Texas Tech University, Lubbock, Texas 79409, USA

⁶Centre for Plant Biotechnology and Genomics (CBGP) UPM-INIA-CSIC, 28223 Pozuelo de Alarcón (Madrid), Spain

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

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

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

¹⁰Author for correspondence: amcdonnell@chicagobotanic.org

Citation: McDonnell, A. J., W. J. Baker, S. Dodsworth, F. Forest, S. W. Graham, M. G. Johnson, L. Pokorny, J. Tate, et al. 2021. Exploring Angiosperms353: Developing and applying a universal toolkit for flowering plant phylogenomics. *Applications in Plant Sciences* 9(7): e11443.

doi:10.1002/aps3.11443

Target enrichment represents a useful, cost-effective method for researchers working on the phylogenomics of non-model organisms (e.g., Cronn et al., 2012; Hale et al., 2020). The ability to sequence a customizable predefined genomic subset for several dozens or even hundreds of taxa allows in-depth analyses and the testing of phylogenetic hypotheses in ways that were not previously possible (reviewed in McKain et al., 2018). The most popular methods for targeted sequencing of genomic loci in phylogenomics include (long-)amplicon sequencing (Rothfels et al., 2017) and hybridization capture (Mandel et al., 2014; Weitemier et al., 2014). Targeted amplicon sequencing is based on single-fragment PCR amplification or by using multiplexing methods such as a microfluidic PCR-based amplification of multiple pre-selected genomic regions (e.g., Zhang and Ozdemir, 2009; Ho et al., 2014), which can then be pooled and sequenced. Massively parallel amplicon sequencing was first used in medical diagnostics (Turner et al., 2009) and was later applied to metazoan phylogenetics (Bybee et al., 2011; O'Neill et al., 2013). Microfluidic PCR and long-amplicon sequencing were subsequently applied in

plant systematics (Uribe-Convers et al., 2014, 2016; Gostel et al., 2015). Amplicon-based methods can be time consuming as they require careful optimization and validation of primers. These methods are also susceptible to many of the common problems in PCR (such as nonspecific products, inability to amplify large loci in their entirety, or simply no products). Recently, amplicon approaches have been largely supplanted by hybridization-based targeted enrichment, which allows for relatively rapid probe design with reference to a few related transcriptomes or genomes, and allows simultaneous and efficient recovery of many hundreds of genes.

Target enrichment via hybridization-based sequence capture can use customized or universal probes (short nucleotide fragments between 80 and 120 bp in length, also called baits). This versatile and powerful approach relies on probes to capture complementary sets of genomic targets from whole genome DNA or cDNA samples using solid-phase or in-solution hybridization (Gnirke et al., 2009; Mamanova et al., 2010; Cronn et al., 2012). To account for frequent (ancient) whole genome duplication events, plant systematists can

identify sets of consistently low-to-single-copy markers that balance two common issues: ensuring recovery across a broad range of taxa, while containing sufficient phylogenetic informativeness. Target regions can be customized to address specific research questions, including targeting genes common to a metabolic pathway (Gardner et al., 2016; Folk et al., 2021) or genes of agronomic interest (Witek et al., 2016; Soto Gomez et al., 2019). The targets are often complete genes or exonic sequences, but non-coding DNA such as introns or untranslated regions can also be enriched. In-solution capture of the probe targets depends on the thermodynamics of hybridization. This process is governed, in part, by the DNA complexity of the bait and the probe length. Longer probes tolerate more mismatches, making them more suitable for capture across different species. Higher tiling density (the degree to which probes overlap, e.g., 75%) can compensate for failures of specific probes to hybridize or for low-quality input material, such as those stemming from historical herbarium specimens, so-called herbariomic approaches (Brewer et al., 2019). While overlapping probes increase the chance of capturing the target of interest via neighboring probes, there is a trade-off, as a higher tiling density tends to reduce the number of loci that can be examined per capture kit.

Targeted enrichment of nuclear genes also overcomes several potential drawbacks of estimating phylogenies from organelle data alone. The single-locus behavior of organellar DNAs means that only a single history is provided: the entire plastid genome is expected to function as a single linkage group (Birky, 2001; but see Zhang et al., 2020). Organellar data are also prone to high rates of genetic drift in small population sizes, which could lead to a less accurate estimate of lineage differentiation, compared with large sets of nuclear genes (reviewed in Wicke and Schneeweiss, 2015). Although target capture data provide many more gene histories, analysis of these data requires specific bioinformatic approaches, including assembly of gene sequences without genomic references and detection of local gene duplication events that violate assumptions of orthology. In this regard, the plant systematics community has been innovative in addressing the related computational challenges, resulting in the wide adoption of targeted sequencing for plant systematics in recent years.

Angiosperms353 is a tool for enabling phylogenomic-level analysis of any angiosperm group at any scale; the targeted genes and associated probes (Johnson et al., 2019) were developed from a set of low-copy nuclear genes used to infer the phylogeny of all green plants (One Thousand Plant Transcriptomes Initiative et al., 2019) to facilitate the collection of phylogenomic data in a relatively cost-effective and repeatable way. When considering Angiosperms353, it is important to distinguish between the low-copy orthologous genes found in all angiosperms and the 120-bp probe sequences used to enrich genomic libraries. The genes were selected based on copy number (ideally, genes that are single copy across more than 1400 available plant transcriptomes) and contain useful phylogenetic signal at deep and recent time scales. The probe sequences were designed from ~600 angiosperm transcriptomes and genomes, and were selected to maximize sequence similarity to other angiosperms, while minimizing the number of sequences used for probe design. Distinguishing the genes from the probes is important when considering the advantages, limitations, and future directions for Angiosperms353 in phylogenetic systematics, as discussed below.

Although initial reports of the utility of Angiosperms353 genes and probes have been promising (Dodsworth et al., 2019), a few key

questions remain: (1) Can Angiosperms353 probes truly be used to reliably recover orthologous genes across all groups of flowering plants?; (2) Is an angiosperm-wide probe design the most efficient for all targeted sequencing projects?; (3) What new bioinformatics tools are needed to apply Angiosperms353 genes to previously intractable research questions?; and (4) Are Angiosperms353 genes variable enough within species to extend the utility of the probes to population genetic studies? In this special issue and its companion issue in the *American Journal of Botany* (Baker et al., 2021b), the papers collected address these questions and demonstrate a number of ways in which research employing Angiosperms353 will benefit future angiosperm phylogenomic studies. This issue of *Applications in Plant Sciences* is focused on a variety of methodological issues, including comparisons between the Angiosperms353 probe set and taxon-specific probe sets, the development of new tools for specific lineages, a new lab technique to enrich libraries using Angiosperms353 and taxon-specific probe sets simultaneously, and bioinformatic approaches to facilitate careful assembly and analysis of data obtained using the Angiosperms353 kit.

PROBE SET COMPARISONS AND DEVELOPMENT OF NEW TOOLS

Low-to-single-copy nuclear markers have been an important tool in evolutionary studies of angiosperms (e.g., Small et al., 2004; Wu et al., 2006; Duarte et al., 2010). The identification of a set of hundreds of nuclear markers shared across angiosperms has facilitated the design of probes for targeted enrichment and subsequent high-throughput sequencing (Johnson et al., 2019). In this issue, three papers address the application and expansion of the Angiosperms353 loci. These papers build on existing resources to design new lineage-specific probes that contain the Angiosperms353 loci (Eserman et al., 2021; Ufimov et al., 2021), generate careful comparisons of the performance of an existing lineage-specific probe set with the Angiosperms353 probe set (Siniscalchi et al., 2021), and investigate the efficacy of simultaneous enrichment of libraries using universal and lineage-specific probes in a single hybridization reaction (Hendriks et al., 2021).

Eserman et al. (2021) describe the development of Orchidaceae963, a new bait set that incorporates 254 of the Angiosperms353 loci, which they tested on members of three subfamilies of Orchidaceae. This resource will facilitate systematic work in the family and is also expected to be useful in population and conservation genomic studies. The incorporation of the Angiosperms353 loci along with custom marker genes is becoming more common (Yardeni et al., 2019; Jantzen et al., 2020; Christe et al., 2021; Ogutcen et al., 2021), and will allow collaboration between research teams and foster data sharing across studies. Comparative analyses of the custom sets and the Angiosperms353 set will naturally follow (e.g., see Larridon et al., 2020; Shah et al., 2021). For example, Ufimov et al. (2021) found a similar amount of parsimony informative sites using the Angiosperms353 probe set and a custom probe set in their study of Rosaceae subtribe Malinae. They also showed that the custom probes allowed for improved locus recovery rates across samples and lower levels of gene tree discordance (but see Yardeni et al., 2019). Siniscalchi et al. (2021) compared their lineage-specific Compositae1061 probes with the Angiosperms353 probes in eight members of the sunflower family. Although they were able to recover more loci on average with the lineage-specific probes, they found lower levels of paralogy in the gene trees based

on the Angiosperms353 loci, which may simplify downstream analyses of the latter versus the former. Hendriks et al. (2021) performed a test case for combining both lineage-specific (i.e., Brassicaceae) and universal probes in a single hybridization reaction to simultaneously enrich libraries for both sets of loci. They found high levels of enrichment and locus recovery for both sets of targets across 26 Brassicaceae samples that included 16 samples from a single tribe and 10 samples reflecting broader sampling across the family. Their study shows that it is possible to generate data for multiple probe sets with little extra cost or work per sample.

Taken together, these papers highlight that researchers have several options when choosing loci for target capture in flowering plants. The Angiosperms353 genes provide a ready set of loci that are low copy and reliably recovered using the universal Angiosperms353 probes, or by using custom probes designed with orthologs of the Angiosperms353 genes from the target taxa. The feasibility of combining probe sets during hybridization suggests additional flexibility. There is a trade-off, as custom probe designs will have higher per-sample costs and require existing sequence data in the focal group. However, far fewer probes will be needed for the Angiosperms353 genes in a custom design, allowing for additional loci for a given number of probes. Whichever choice is made, by including Angiosperms353 in a target capture project, researchers will benefit the entire phylogenetics community by enabling meta-analysis that parallels the broad adoption of single-gene markers in past decades.

NEW INNOVATIONS: BIOINFORMATICS

The development of Angiosperms353 into a common tool for molecular phylogenetics and population genetics has been paralleled by the need for new analytical workflows. One of the most active areas of tool development is in using extensions to existing tools to facilitate analysis of Angiosperms353 data. In this issue, three tools specifically tie in with HybPiper (Johnson et al., 2016), a popular tool for assembly and recovery of genes and flanking regions from target-capture sequencing reads. A common issue identified with Angiosperms353 data assembled using HybPiper is relatively low sequence recovery (e.g., Gaynor et al., 2020), but it is often unclear whether the issue lies with the molecular tools (i.e., poor hybridization between probes and target DNA) or bioinformatic tools (i.e., poor recovery in HybPiper or other tools). To test this, McLay et al. (2021) developed new target files to improve the mapping frequency of reads to Angiosperms353 target genes, a result independently confirmed by Slimp et al. (2021) and Lee et al. (2021). McLay et al. (2021) also provide a workflow that uses a hidden Markov model to choose clade-specific target sequences from a database of potential Angiosperms353 sequences. Their results indicate that the sequence divergence between the Angiosperms353 probe sequences and the target DNA of angiosperms during nucleic acid hybridization and subsequent sequencing is more forgiving than the effect of the same divergence on gene recovery during bioinformatic analysis. For researchers struggling with poor target enrichment efficiency using Angiosperms353, McLay et al. provide a potential bioinformatic solution that may reduce the need to re-enrich libraries. Their approach is to improve the target file used to recover genes in HybPiper by selecting more appropriate orthologs of the Angiosperms353 genes.

Nauheimer et al. (2021) describe HybPhaser, a method for detecting hybrid taxa using HybPiper assemblies. Using an Angiosperms353 data set collected for *Nepenthes*, they employ an

innovative two-step strategy in which an initial phylogeny is first used to identify representatives of different clades that have low heterozygosity (and are thus presumably not hybrids). In the second step, reads from putative hybrids are mapped to these chosen clade representatives. Differential mapping information is then used to phase hybrid sequences, which can then be placed on a multiply-labeled tree to identify hybrid origins. The HybPhaser method also has promise for addressing a common issue in target-capture phylogenetics: the identification of paralogous genes. This builds upon previous work using clade references to identify paralogs (Gardner et al., 2021), but the development of a more generalizable workflow in HybPhaser will be of use to many researchers unsure about whether target-captured genes are single copy in their data.

Slimp et al. (2021) also describe new methods for downstream analysis of HybPiper output, for use in population genomic studies and the calculation of within-species demographic parameters. Their method adapts previously described workflows (Kates et al., 2018) to use HybPiper “supercontigs” (targeted coding regions and flanking non-coding regions) as reference sequences within species. Within-species variants can then be assessed jointly and used with standard population genomics programs (also see Beck et al., 2021 and Wenzell et al., 2021). Slimp et al. further demonstrate that the Angiosperms353 loci are variable within populations of several species, which suggests that genome-scale estimation of population parameters is now feasible without the need to develop taxon-specific methods.

THE FUTURE OF TARGETED ENRICHMENT IN ANGIOSPERM PHYLOGENOMICS

Since its initial publication, there has been a rapid and substantial uptake of Angiosperms353 to address systematics questions. The papers in this issue and its companion *American Journal of Botany* issue further illustrate that the utility of data generated through target enrichment extends beyond phylogenomics, with promising applications into population and conservation genomics. Of course, with the adaptability of data comes a requirement for flexible and innovative methods to process and analyze these data. While new methods and novel applications are already emerging, we recognize that there are some challenges and improvements that lie ahead.

The need to evaluate the efficiency of recovering the targeted loci, both in vitro and in silico, is reflected in several papers in this issue, highlighting the distinction between targeted Angiosperms353 genes and the probes used to recover them. The papers presented here and in the companion issue build further evidence that the Angiosperms353 genes can be reliably recovered across the diversity of flowering plant species, with strong phylogenetic signal at scales ranging from deep time to the population level. Several papers in this issue describe innovations in how these target genes are recovered. Custom probe sets can still provide additional phylogenetic resolution within focal taxa by capturing more genes with the same number of probes, but the inclusion of Angiosperms353 loci along with custom probe sets will enable meta-analyses to compare species across flowering plants. In silico, the use of enhanced target files offers potential improvements for the assembly of enriched sequence libraries and illustrates that newly available genomic data can be used to complement the original Angiosperms353 target file and probe set. In this regard, the availability of an ever-increasing number of

complete flowering plant genome and transcriptome sequences (e.g., One Thousand Plant Transcriptomes Initiative et al., 2019), and the massive increase in available sequence data facilitated by Angiosperms353, should allow for future refinement of the probes and choice of target sequences to better represent the diversity of angiosperms and maximize the length of sequences recovered across a broader set of taxa, including those with extremely altered lifestyles that affect the rate of sequence evolution, such as in heterotrophic taxa (e.g., Lam et al., 2018). Of course, the utility of Angiosperms353 data, or any other set of target enrichment data, relies on the accessibility of these genomic resources.

Currently, there is no central data repository specifically tailored to access target enrichment data in a way that is not taxon specific. The Kew Tree of Life Explorer (<https://treeoflife.kew.org/>) was recently launched (Baker et al., 2021a) to make freely available pre- and post-publication Angiosperms353 data for thousands of taxa. The continued development of accessible tools to store and distribute these data will accelerate the progress that has been made thus far. For example, the ability to download a specific set of genes for a specific set of taxa from a central repository, rather than having to download and then filter multiple data sets from publication-specific repositories, would significantly improve the ability to incorporate existing data for addressing new research questions. Similarly, a centralized repository for uploading assembled and annotated target enrichment data generated for any taxon across the entire tree of life facilitates standardization and accessibility. A streamlined set of tools for identifying Angiosperms353 homologs for building custom probe sets would also increase the efficiency of data collection, much like how MarkerMiner (Chamala et al., 2015) provides a way to identify orthologs from a starting data set. The adoption of Angiosperms353 and target enrichment more generally in angiosperm systematics should cultivate further growth in data and in methods, as we have seen in this special issue. As more people use these data and methods, the opportunities for positive developments will only increase further, and improvements to existing probe sets, the development of new probe sets, opportunities to create solutions to challenges such as the resolution of paralogs, and the establishment of companion databases are all likely to be addressed head-on and embraced by the angiosperm systematics community.

ACKNOWLEDGMENTS

We thank all authors and reviewers for their valuable contributions to this special issue of *Applications in Plant Sciences*. We also acknowledge the tireless work of the APPS publications team, especially Managing Editor Beth Parada and Editor-in-Chief Theresa Culley.

LITERATURE CITED

- Baker, W. J., P. Bailey, V. Barber, A. Barker, S. Bellot, D. Bishop, L. R. Botigué, et al. 2021a. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* syab035. <https://doi.org/10.1093/sysbio/syab035>.
- Baker, W. J., S. Dodsworth, F. Forest, S. W. Graham, M. G. Johnson, A. McDonnell, L. Pokorny, et al. 2021b. Exploring Angiosperms353: An open, community toolkit for collaborative phylogenomic research on flowering plants. *American Journal of Botany* 108(7): 1059–1065.
- 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] [posted 8 January 2021] Available at <https://doi.org/10.1101/2021.01.07.425781> [accessed 28 June 2021].
- Birky, C. W., Jr. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics* 35: 125–148.
- 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.
- Bybee, S. M., H. Bracken-Grissom, B. D. Haynes, R. A. Hermansen, R. L. Byers, M. J. Clement, J. A. Udall, et al. 2011. Targeted amplicon sequencing (TAS): A scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution* 3: 1312–1323.
- Chamala, S., N. Garcia, G. T. Godden, V. Krishnakumar, I. E. Jordon-Thaden, R. De Smet, W. B. Barbazuk, et al. 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3(4): 1400115.
- Christe, C., C. G. Boluda, D. Koubinová, 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: 1055–7903.
- Cronn, R., B. J. Knaus, A. Liston, P. J. Maughan, M. Parks, J. V. Syring, and J. Udall. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99: 291–311.
- Dodsworth, S., L. Pokorny, M. G. Johnson, J. T. Kim, O. Maurin, N. J. Wickett, F. Forest, and W. J. Baker. 2019. Hyb-Seq for flowering plant systematics. *Trends in Plant Science* 24: 887–891.
- Duarte, J. M., P. K. Wall, P. P. Edger, L. L. Landherr, H. Ma, J. C. Pires, J. Leebens-Mack, and C. W. dePamphilis. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* 10: 61.
- 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(7): e11416.
- 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(2): e11410.
- Gardner, E. M., M. G. Johnson, D. Ragone, N. J. Wickett, and N. J. C. Zerega. 2016. Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery. *Applications in Plant Sciences* 4(7): 1600017.
- Gardner, E. M., M. G. Johnson, J. T. Pereira, A. S. A. Puad, D. Arifiani, N. J. Wickett, and N. J. C. Zerega. 2021. 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. Fu, L. Gao, L. 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.
- Gnrirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182–189.
- Gostel, M. R., K. A. Coy, and A. Weeks. 2015. Microfluidic PCR-based target enrichment: A case study in two rapid radiations of *Commiphora* (Burseraceae) from Madagascar. *Journal of Systematics and Evolution* 53: 411–431.
- 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(4): e11337.
- Hendriks, K. P., T. Mandáková, N. M. Hay, E. Ly, A. Hoof 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(7): e11438.
- Ho, T., L. Cardle, X. Xu, M. Bayer, K. S. J. Prince, R. N. Mutava, D. F. Marshall, and N. Syed. 2014. Genome-tagged amplification (GTA): A PCR-based method

- to prepare sample-tagged amplicons from hundreds of individuals for next generation sequencing. *Molecular Breeding* 34: 977–988.
- Jantzen, J. R., P. Amarasinghe, R. A. Folk, M. Reginato, F. A. Michelangeli, D. E. Soltis, N. Cellinese, and P. S. Soltis. 2020. A two-tier bioinformatic pipeline to develop probes for target capture of nuclear loci with applications in Melastomataceae. *Applications in Plant Sciences* 8(5): e11345.
- Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4(7): 1600016.
- Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eisehardt, 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.
- 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*. *American Journal of Botany* 105: 404–416.
- Lam, V. K. Y., H. Darby, V. S. F. T. Merckx, G. Lim, T. Yukawa, K. M. Neubig, J. R. Abbott, G. E. Beatty, 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. Epiawalage, I. Fairlie, et al. 2020. Tackling rapid radiations with targeted sequencing. *Frontiers in Plant Science* 10: 1655.
- Lee, A. K., I. Gilman, M. J. Donoghue, and W. L. Clement. 2021. Reconstructing Dipsacales phylogeny using Angiosperms353: Issues and insights. *American Journal of Botany* 108(7): 1122–1142.
- Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, et al. 2010. Target-enrichment strategies for next-generation sequencing. *Nature Methods* 7: 111–118.
- Mandel, J. R., R. B. Dikow, V. A. Funk, R. R. Masalia, S. E. Staton, A. Kozik, L. Rieseberg, and J. M. Burke. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences* 2(2): 1300085.
- 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(3): e1038.
- McLay, T. G. B., J. L. Birch, B. F. Gunn, W. Ning, J. A. Tate, L. Nauheimer, E. M. Joyce, et al. 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* 9(7): e11420.
- 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 data sets. *Applications in Plant Sciences* 9(7): e11441.
- 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.
- O'Neill, E. M., R. Schwartz, C. T. Bullock, J. S. Williams, H. B. Shaffer, X. Aguilar-Miguel, G. Parra-Olea, and D. W. Weisrock. 2013. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Molecular Ecology* 22: 111–129.
- One Thousand Plant Transcriptomes Initiative, J. H. Leebens-Mack, M. S. Barker, E. J. Carpenter, M. K. Deyholos, M. A. Gitzendanner, S. W. Graham, et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679–685.
- Rothfels, C. J., K. M. Pryer, and F.-W. Li. 2017. Next-generation polyploid phylogenetics: Rapid resolution of hybrid polyploid complexes using PacBio single-molecule sequencing. *New Phytologist* 213: 413–429.
- Shah, T., J. V. Schneider, G. Zizka, O. Maourin, W. J. Baker, F. Forest, G. E. Brewer, et al. 2021. Joining forces in Ochnaceae phylogenomics: A tale of two targeted sequencing probe kits. *American Journal of Botany* 108(7): 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: A comparison of the Compositae1061 and Angiosperms353 enrichment panels in the sunflower family. *Applications in Plant Sciences* 9(7): 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(7): e11419.
- Small, R. L., R. C. Cronn, and J. F. Wendel. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17: 145–170.
- 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(6): e11254.
- Turner, E. H., S. B. Ng, D. A. Nickerson, and J. Shendure. 2009. Methods for genomic partitioning. *Annual Review of Genomics and Human Genetics* 10: 263–284.
- Ufimov, R., V. Zeisek, S. Pišová, 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. *Applications in Plant Sciences* 9(7): e11442.
- Uribe-Convers, S., J. R. Duke, M. J. Moore, and D. C. Tank. 2014. A long PCR-based approach for DNA enrichment prior to next-generation sequencing for systematic studies. *Applications in Plant Sciences* 2(1): 1300063.
- Uribe-Convers, S., M. L. Settles, and D. C. Tank. 2016. A phylogenomic approach based on PCR target enrichment and high throughput sequencing: Resolving the diversity within the South American species of *Bartsia* L. (Orobanchaceae). *PLoS ONE* 11: e0148203.
- Weitemier, K., S. C. K. 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(9): 1400042.
- 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(7): 1270–1288.
- Wicke, S., and G. M. Schneeweiss. 2015. Next-generation organellar genomics: Potentials and pitfalls of high-throughput technologies for molecular evolutionary studies and plant systematics. In E. Hörandl and M. Appelhans [eds.], *Next-generation sequencing in plant systematics*. Koeltz Scientific Books, Königstein, Germany.
- Witek, K., F. Jupe, A. I. Witek, D. Baker, M. D. Clark, and J. D. G. Jones. 2016. Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nature Biotechnology* 34: 656–660.
- Wu, F., L. A. Mueller, D. Crouzillat, V. Pétiard, and S. D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: A test case in the euasterid plant clade. *Genetics* 174: 1407–1420.
- Yardeni, G., J. Viruel, M. Paris, J. Hess, C. G. Crego, M. de La Harpe, N. Rivera, et al. 2019. Taxon-specific or universal? Using target capture to study the evolutionary history of a rapid radiation. *bioRxiv* 2021.05.20.444989 [Preprint] [posted 22 May 2021]. Available at <https://doi.org/10.1101/2021.05.20.444989> [accessed 25 June 2021].
- Zhang, R., Y.-H. Wang, J.-J. Jin, G. W. Stull, A. Bruneau, D. Cardoso, L. P. De Queiroz, et al. 2020. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Systematic Biology* 69: 613–622.
- Zhang, Y., and P. Ozdemir. 2009. Microfluidic DNA amplification—A review. *Analytica Chimica Acta* 638: 115–125.

Exploring Angiosperms353: Developing and applying a universal toolkit for flowering plant phylogenomics

McDonnell A

2021-07-26
