Received: 6 November 2023 | Accepted: 14 May 2024

DOI: 10.1002/aps3.11615

GENOMIC RESOURCE ARTICLE

Applications in Plant Sciences

Truly the best of both worlds: Merging lineage-specific and universal probe kits to maximize phylogenomic inference

Luiz Henrique M. Fonseca[1] | **Pieter Asselman**[1] | **Katherine R. Goodrich**[2] | **Francis J. Nge**[3] | **Vincent Soulé**[3] | **Kathryn Mercier**[4,5] | **Thomas L. P. Couvreur**[3] | **Lars W. Chatrou**[1]

[1]Systematic and Evolutionary Botany Laboratory, Department of Biology, Ghent University, Ghent, Belgium

[2]Department of Biology, Widener University, Chester, Pennsylvania, USA

[3]Institute of Research for Development (IRD), UMR DIADE, Université de Montpellier, Montpellier, France

[4]Department of Biology, City College of New York, New York, New York, USA

[5]The Graduate Center of the City University of New York, New York, New York, USA

Correspondence

Luiz Henrique M. Fonseca, Systematic and Evolutionary Botany Laboratory, Department of Biology, Ghent University, Ghent, Belgium.
Email: luizhmf@gmail.com

Lars W. Chatrou, Systematic and Evolutionary Botany Laboratory, Department of Biology, Ghent University, Ghent, Belgium.
Email: lars.chatrou@ugent.be

Abstract

Premise: Hybridization capture kits are now commonly used for reduced representation approaches in genomic sequencing, with both universal and clade-specific kits available. Here, we present a probe kit targeting 799 low-copy genes for the plant family Annonaceae.

Methods: This new version of the kit combines the original 469 genes from the previous Annonaceae kit with 334 genes from the universal Angiosperms353 kit. We also compare the results obtained using the original Angiosperms353 kit with our custom approach using a subset of specimens. Parsimony-informative sites and the results of maximum likelihood phylogenetic inference were assessed for combined matrices using the genera *Asimina* and *Deeringothamnus*.

Results: The Annonaceae799 genes derived from the Angiosperms353 kit have extremely high recovery rates. Off-target reads were also detected. When evaluating size, the proportion of on- and off-target regions, and the number of parsimony-informative sites, the genes incorporated from the Angiosperms353 panel generally outperformed the genes from the original Annonaceae probe kit.

Discussion: We demonstrated that the new sequences from the Angiosperms353 probe set are variable and relevant for future studies on species-level phylogenomics and within-species studies in the Annonaceae. The integration of kits also establishes a connection between projects and makes new genes available for phylogenetic and population studies.

KEYWORDS

Angiosperms353, Annonaceae, *Asimina*, bait set, *Deeringothamnus*, phylogenetics

Target enrichment methods allow for the sequencing of hundreds of independent low-copy loci from the nuclear genome and cytoplasmic regions (Mamanova et al., 2010; Cronn et al., 2012; Weitemier et al., 2014). This wealth of molecular data can be applied in phylogenetic and population studies of non-model organisms and has opened a new era for plant systematics and species delimitation, among other fields (Soltis et al., 2013; McKain et al., 2018; Andermann et al., 2020; Zuntini et al., 2024). Strategies of genome reduction that incorporate hundreds of thousands to millions of base pairs are allowing the inference of robust phylogenetic hypotheses while taking into account the effects of diverse processes, such as incomplete lineage sorting (Pamilo and Nei, 1988; Maddison, 1997) or introgression, or accounting for the simultaneous effects of both (Nakhleh, 2013). Independent loci such as those from the nuclear genome are critical for species tree inference under a multispecies coalescent model or a multispecies network coalescent (Wen and Nakhleh, 2018). The use of target enrichment methods can also be extended to population

studies with the possibility of orthologous gene variants to be separated (phased) and single-nucleotide polymorphisms to be assessed (Nauheimer et al, 2021; Jiménez-Mena et al., 2022).

Gene capture techniques have been quite successfully applied in multiple plant studies at diverse levels. Universal kits targeting all angiosperms are available (Johnson et al., 2019; Waycott et al., 2021), as well as customized kits designed for plant families such as the Apocynaceae (Weitemier et al., 2014), Asteraceae (Mandel et al., 2014; Moore-Pollard et al., 2024), Bignoniaceae (Fonseca et al., 2023), Gesneriaceae (Ogutcen et al., 2021), Orchidaceae (Eserman et al., 2021), and Rubiaceae (Ball et al., 2023), or for less inclusive clades such as *Dioscorea* L. (Dioscoreaceae) (Soto Gomez et al., 2019) and *Euphorbia* L. (Euphorbiaceae) (Villaverde et al., 2018). Even species-specific sets have been designed for population genetics and breeding (e.g., for barley [*Hordeum vulgare* L.]; Hill et al., 2019). Attempts to compare customized and universal hybridization capture kits at shallow phylogenetic levels often conclude that the performances of both types of kit are similar in phylogenetic studies (e.g., Larridon et al., 2020). Rather than a comparison, a more interesting approach is the merging of both universal and specific probe sets into a single kit (Hendriks et al., 2021), combining the versatility of the former and the efficiency of the latter to assemble complete or nearly complete genes in a single framework (Eserman et al., 2021; Fonseca et al., 2023).

The Annonaceae family consists of lianas, shrubs, and trees that are common components of tropical forests. The phylogenetic resolution within the Annonaceae was hindered by low variability when using Sanger sequencing data (Chatrou et al., 2012). To address this, a probe kit was specifically designed for this family (Couvreur et al., 2019), improving the phylogenetic resolution by targeting up to 469 exons (genes) and a total of 364,653 bp. The probe kit was designed with consideration of the molecular variability across the whole family and was successfully tested at both the family level (65 different genera) and the species level in the Piptostigmateae tribe (29 species and multiple individuals) (Couvreur et al., 2019). Several recent studies have confirmed the utility of the kit for inferring species-level relationships in the Annonaceae (Dagallier et al., 2023; Martínez-Velarde et al., 2023; Lopes et al., 2024); however, hundreds of genes were flagged as paralogous using this kit (Couvreur et al., 2019; Lopes et al., 2024), a recurrent problem in custom probe kits (e.g., Ogutcen et al., 2021; Ball et al., 2023; Fonseca et al., 2023). Additionally, many of the selected genes lack introns (Couvreur et al., 2019), reducing the power to infer species-level and/or population-level phylogenies due to the low molecular variability of exons. The lack of overlap with the widely used Angiosperms353 target capture panel (Johnson et al., 2019) is another limitation of the customized Annonaceae probe set (Couvreur et al., 2019), limiting the reuse of the data generated for the Annonaceae and its contribution to ongoing efforts to assemble the plant tree of life (Baker et al., 2021).

In this study, we address the problems mentioned above by generating an Annonaceae-specific probe kit that combines the genes selected by Couvreur et al. (2019) with single-copy genes from the Angiosperms353 panel (Johnson et al., 2019). We built a set of 799 putatively single-copy nuclear genes, which we believe will allow for more robust estimates of relationships with stronger support. This new probe kit, named Annonaceae799, was tested using the Annonaceae genera *Asimina* Adans. and *Deeringothamnus* Small, the only non-tropical genera in the family. These genera form a clade within which the phylogenetic relationships are not known or well supported (Mercer et al., 2016; Li et al., 2017). The phylogenetic relationships within and among *Asimina* and *Deeringothamnus* have also been blurred by putative gene flow and introgression between species (Kral, 1960; Mercer et al., 2016), presenting a formidable challenge to any phylogeneticist and a rigorous test for probe kits designed for phylogenetic analyses.

To validate the new Annonaceae799 bait kit, we compared the results obtained using the probes designed using Annonaceae references for the Angiosperms353 panel (referred to here as "custom Angiosperms353") with the results obtained using the probes from the original Annonaceae probe kit (referred to here as "old Annonaceae") in terms of variability and utility for phylogenetic inference. We also evaluated the molecular variability in multiple individuals of *Asimina triloba* (L.) Dunal, demonstrating the effectiveness of the probe kit within species. Furthermore, we compared the results for the *Asimina* and *Deeringothamnus* species obtained using the custom Angiosperms353 probe set developed in this study with those of the original Angiosperms353 probe kit (referred to here as "original Angiosperms353"; Johnson et al., 2019). This comparison provides a valuable opportunity to test enrichments using the same species, gene set, and different probes; however, these results should be interpreted carefully as they are not formal tests between kits with controlled variables as we will further discuss.

METHODS

We generated data for 16 specimens of *Asimina*, two *Deeringothamnus*, and four outgroups. The data used for the comparison of the custom and original Angiosperms353 kits were, respectively, generated here and obtained from a parallel initiative to sequence the Annonaceae. The laboratory procedures and sequencing used to generate the original Angiosperms353 data followed those of Couvreur et al. (2019). Sequence analyses, alignments, statistics, and phylogenetic inference were performed for all samples.

Gene selection and probe design

The original alignments from the Angiosperms353 panel (available from https://github.com/mossmatters/Angiosperms353 [accessed 14 July 2024]) were used in

this study as reference sequences for the probe design. The Annonaceae transcriptome sequences from the 1KP project (Leebens-Mack et al., 2019) were included (i.e., *Annona muricata* L. [YZRI, original acronym from the 1KP project] and *Uvaria macrocarpa* (DC.) Hook. f. & Benth. [PSJT]), while species outside the family were excluded. For 312 genes, at least one sequence of Annonaceae was available and used as a reference. For the remaining 22 genes, we used *Eupomatia bennettii* F. Muell. (DHPO) as the reference, because the Annonaceae and Eupomatiaceae are sister families (Chatrou et al., 2012). In total, the exons of 803 genes were initially combined. These included the 469 genes selected by Couvreur et al. (2019) plus 334 single-copy genes from the Angiosperms353 panel (Johnson et al., 2019). All genes were evaluated to determine their number of copies with BLAT (Kent, 2002), using *Liriodendron chinense* (Hemsl.) Sarg. (Magnoliaceae) (Chen et al., 2019; GenBank: PRJNA418360) as the closest genome available. Duplicates were identified in a search using the program CD-HIT-EST 4.5 (with –c 0.91) (Fu et al., 2012). Of the original 803 genes, four were flagged by CD-HIT-EST for having similar copies to others, in which case the longest copy available was kept. The probes were designed using 799 genes and covered 698,154 bp. Using the original alignments from the 1KP project (Johnson et al., 2019; Leebens-Mack et al., 2019) and keeping only sequences from the Annonaceae or closely related species, we were able to save thousands of probes in our final set, enabling the merger of the custom Angiosperms353 and the old Annonaceae probe sets.

Sampling

We sampled 21 specimens of *Asimina*, *Deeringothamnus*, and outgroups, including all 11 known species within the focal genera. As taxonomic references for *Asimina* and *Deeringothamnus*, we used Kral (1960) and the Flora of North America (http://floranorthamerica.org/). For *Asimina triloba*, seven specimens from different localities were included in our study. As outgroups, we selected the closely related *Annona glabra* L., *Annona scandens* Diels, *Diclinanona calycina* R. E. Fr., and *Goniothalamus laoticus* (Finet & Gagnep.) Bân (Tables 1 and S1; see Supporting Information) (Erkens et al., 2014). For most of the species, the leaves were collected in the field in 2006, 2021, and 2022 and dried in silica gel. To complement the sampling, herbarium material was also used as a source of leaf tissue (herbaria FLAS and L [herbarium acronyms per Index Herbariorum (Thiers, 2024)]; Table S1). The *Asimina* and *Deeringothamnus* samples used for the original Angiosperms353 assay were from leaves collected in the field (Table S2).

DNA extraction and library construction

Total DNA was extracted using the 2× cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). The cell lysis step was carried out overnight at 60°C. The integrity of the DNA extractions was evaluated using agarose gel electrophoresis, their purity was determined using a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, Massachusetts, USA), and the concentration was measured using a Qubit 2.0 (Thermo Fisher Scientific). To prepare the Illumina (San Diego, California, USA) sequencing libraries, we used the NEBNext Ultra FS II Library Prep kit (New England Biolabs, Ipswich, Massachusetts, USA) with default parameters. Approximately 500 ng of genomic DNA (gDNA) was used as the input for each sample. Libraries of around 350 bp were generated after the enzymatic digestion of samples with high DNA quality. Herbarium samples were not digested. The DNA quality after library preparation was evaluated using a BioAnalyzer 2100 (Agilent Technologies, Santa Clara, California, USA), and the concentration was determined using a Qubit 2.0. The enrichment step used probes specifically designed for the Annonaceae by Daicel Arbor Biosciences (Ann Arbor, Michigan, USA), merging the original Annonaceae kit (Couvreur et al., 2019) with genes selected from the Angiosperms353 panel (Johnson et al., 2019). With this strategy, we targeted exons from 799 nuclear genes. Samples were enriched in sets of 8–10 samples, with groupings determined based on the type of material (silica dried or herbarium samples) and the concentration of the sample. The samples were incubated for 24 h at 60°C, within the interval between 65°C and 60°C suggested in the manufacturer's instructions (Daicel Arbor Biosciences, 2022). Using this temperature, we could ensure the probes would capture genes from all the samples. Initial trials using 62°C for 16 h and 20 h recovered extremely low concentrations of DNA, leading us to update the parameters. The final enriched sets of libraries were sequenced by Macrogen (Amsterdam, the Netherlands) using an Illumina NovaSeq 6000 platform. This platform generated around 1.4 Gbp of data for each sample and paired-end reads of 150 bp.

Sequence analyses

The Illumina reads were demultiplexed at the sequencing facility, and low-quality reads were trimmed using fastp version 0.23.4 (Chen, 2023), with a Phred quality score threshold of 30 (-q 30). The read quality was assessed using FastQC version 0.12.1 (Andrews, 2010) and MultiQC version 1.21 (Ewels et al., 2016). The HybPiper pipeline (version 2.1.5; Johnson et al., 2016) was used to assemble the genes at the Flemish Supercomputer Center (Vlaams Supercomputer Centrum; VSC-HPC, Ghent, Belgium). As an input for the pipeline, we used the 799 genes selected as references (-t_dna), paired-end reads (-r), unpaired reads (--unpaired), the BWA version 07.17 (Li and Durbin, 2009) aligner (--bwa), and ran the intronerate script (--run_intronerate) for up to 1200 seconds for each gene (--timeout_exonerate_contigs). The remaining parameters of HybPiper were left as default. We retrieved FASTA files containing the exon-only and supercontig (exons + introns/intergenic regions) sequences for all samples. The following HybPiper commands were run to obtain summary statistics: all assembled genes (retrieve_sequences) and paralog

TABLE 1 Summary statistics of sequencing success, including the number of raw paired-end reads obtained, percentage of on-target reads, number of loci obtained, percentage of gene recovery, and number of loci retained after paralogs removed.

Species[a]	No. of paired-end reads	No. of mapped reads	Genes assembled	Percentage of on-target reads	No. of loci at 50% (nuclear)	No. of loci at 50% (plastid)
Asimina incana22	10,928,202	4,563,719	799	41.8	774	77
Asimina longifolia24	22,213,929	15,580,155	801	70.1	775	77
Asimina manasota26	9,025,321	3,847,949	802	42.6	769	77
Asimina obovata27	17,434,662	7,318,892	797	42.0	771	74
Asimina parviflora55	25,938,311	13,672,264	802	52.7	783	77
Deeringothamnus pulchellus36	18,048,328	12,567,453	801	69.6	773	76
Asimina pygmaea37	20,569,514	8,120,781	798	39.5	782	75
Asimina reticulata39	27,276,263	10,540,115	802	38.6	782	77
Deeringothamnus rugelii42	17,776,967	8,250,058	801	46.4	784	77
Asimina tetramera45	12,015,857	3,779,394	799	31.5	780	77
Asimina triloba47	24,546,914	9,130,145	791	37.2	787	77
Asimina triloba48	24,380,205	8,826,222	800	36.2	787	77
Asimina triloba49	23,772,110	11,66,0892	798	49.1	788	77
Asimina triloba51	9,299,096	5,663,854	800	60.9	782	69
Asimina triloba53	17,785,852	10,985,132	799	61.8	784	75
Asimina triloba54	10,356,918	6,170,999	797	59.6	782	72
Asimina triloba71	14,164,353	8,455,548	800	59.7	782	62
Annona glabra	23,699,026	10,340,480	798	43.6	762	69
Annona scandens	9,790,566	5,092,223	800	52.0	763	54
Diclinanona calycina	20,969,178	12,109,113	799	57.7	776	77
Goniothalamus laoticus	4,754,971	2,268,171	783	47.7	748	15
Mean	17,368,883	8,364,133	798.4	49.5	776.9	70.9

[a]Numbers after species names were used to identify the individuals sampled in our study.

statistics (paralog_retriever). Two genes completely failed, and two were excluded due to low representation (i.e., they appeared in fewer than 80% of the samples). To evaluate the recovery of plastid genes, we used HybPiper version 2.1.5 (Johnson et al., 2016) with the same parameters as described previously, and a reference file with sequences of 78 protein-coding genes from the plastome of *Annona muricata* (Niu et al., 2020; GenBank: MT742546.1). Attempts to assemble complete or nearly complete plastomes using the pipeline GetOrganelle version 1.7.4.1 (Jin et al., 2020) resulted in fragmented and incomplete results, so here we will focus on the assemblies obtained using protein-coding genes as references.

Sequence alignments, statistics, and phylogenetic inference

Each step detailed here was repeated for the exon-only, supercontig, and plastome data sets. Gene sequences were aligned using MAFFT version 7.450 (Katoh and Standley, 2013) with an automatic selection of alignment mode and a maximum of 1000 interactions. Poorly aligned regions and bases for which more than 50% of the samples were missing data were removed using ClipKIT version 1.3 (Steenwyk et al., 2020). The alignments were concatenated using AMAS (Borowiec, 2016) to generate a supermatrix with all molecular data combined. The statistical properties of each alignment, including the combined alignments, were obtained using R version 4.3.1 (R Core Team, 2023) package ape version 5.7 (Paradis and Schliep, 2019). Parsimony-informative sites (PIS) were identified using AMAS for the "complete" (including outgroups) and "ingroup-only" alignments (excluding outgroups). The genetic distance between samples of *Asimina triloba* was also evaluated using the package ape, function *dist.dna*, and the K80 evolutionary model. The gene matrices of all genes retained after removing the paralogous genes were combined in the exon-only, supercontig, and plastome data sets and were used as inputs for phylogenetic reconstructions using the maximum likelihood (ML) criterion in IQ-TREE version 2.1.3 (Nguyen et al., 2015) with model

selection (Kalyaanamoorthy et al., 2017) using the option -m TESTMERGE and 1000 ultrafast bootstrap replicates (Hoang et al., 2018). Trees were also inferred using the exon-only and supercontig data sets of the custom and original Angiosperms353 panel. The same parameters described above were used for both panels and data sets.

Comparison of the custom and original Angiosperms353 probe sets

To further evaluate the efficiency of the custom Angiosperms353 probe kit, we compared the results obtained with data generated for 10 species of *Asimina* using the original Angiosperms353 kit during the enrichment step. Of the 17 specimens of *Asimina* and *Deeringothamnus* sampled here, three were shared between the experiments (*A. tetramera* Small, *A. triloba*, and *D. rugelii* Small), and specimens for the remaining species were from the same or nearby populations (Table S2). The laboratory procedures and sequencing followed the descriptions outlined by Couvreur et al. (2019). In short, DNA was extracted from silica-dried samples using the mixed alkyltrimethylammonium bromide (MATAB) and chloroform separation methods (Mariac et al., 2014). Sequencing libraries were constructed following a modified protocol of Rohland and Reich (2012) using 6-bp barcodes and Illumina indexes to allow multiplexing. After cleaning and quantification, libraries were combined in sets of eight. Targeted regions were enriched using the designed probes for the Annonaceae and amplified using 14–16 cycles. Libraries were sequenced on an Illumina HiSeq version 3 platform using paired-end reads of 150 bp at CIRAD facilities (Montpellier, France). The gene assembly was performed as described above.

RESULTS

Target sequences

Using the complete Annonaceae799 probe kit, we recovered an average of 17,368,883 paired-end reads after cleaning with fastp. The maximum number of reads was 27,276,263 (*Asimina reticulata* Shuttlew. ex Chapm.), and the minimum number of reads was 4,754,971 (*G. laoticus*). An average of 49.5% (31.5–70.1%) of reads were on target (Table 1). We obtained sequences for an average of 798.4 genes per sample (99.9%), ranging from a minimum of 783 genes (*G. laoticus*) to a maximum of 802 genes (*Asimina manasota26* DeLaney, *Asimina parviflora55* (Michx.) Dunal, and *Asimina reticulata39*). We assembled genes covering at least 50% of the full sequence for 776.9 genes on average, ranging from a minimum of 748 genes (*G. laoticus*) to a maximum of 788 genes (*Asimina triloba49*) (Figure 1, Table 1). Four genes completely failed or were present in fewer than half of the specimens and were excluded (Figure 1). Assemblies shorter than 200 bp were recovered for 10 genes, which were excluded as well. A total of 56 genes were identified in a single specimen, and 70 genes were identified in fewer than five samples. Paralogs were flagged for 90 genes, of which 16 were from the Angiosperms353 and 74 from the old Annonaceae set. After removing genes with putative paralogs, the final set to be used for phylogenetic studies with *Asimina*, *Deeringothamnus*, and outgroups comprised 701 genes.

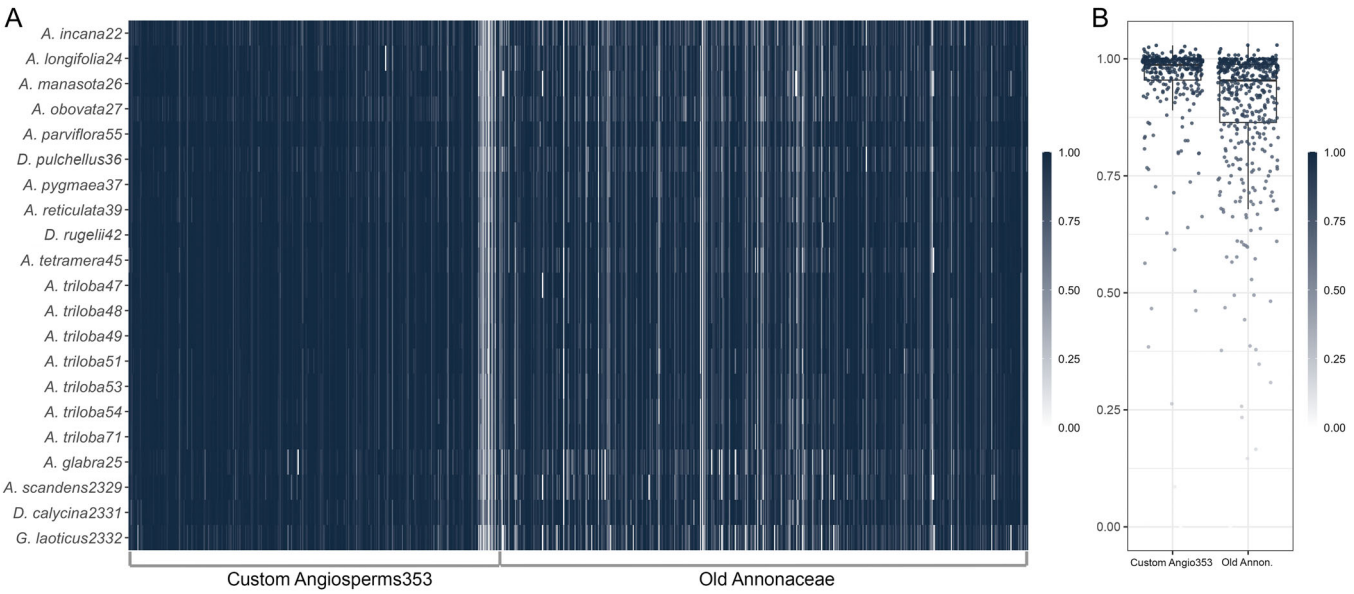

FIGURE 1 Proportion of gene recovery. (A) Heatmap of the 799 genes targeted. (B) Box plot of the proportional values of gene assembly for genes from the Angiosperms353 (Custom Angio353) panel and genes from the Annonaceae (Old Annon.) panel. Boxes show the upper and lower quartiles; whiskers indicate the extent of variability observed outside the boxes.

The length of the per-gene alignments ranged from 201 bp to 5652 bp, with a mean of 862.9 bp. Alignments from the genes incorporated from the Angiosperms353 set ranged from 240 bp to 4653 bp, with a mean size of 984.5 bp. The extreme values were obtained for the genes from the old Annonaceae set, with a mean value of 763.2 bp. The number of PIS considering the whole gene set for "complete" (Figure 2A) and "ingroup-only" (Figure 2B) alignments ranged from 0/0 to 360/122, with a mean value of 49.1/13.5 sites, respectively. The number of PIS for genes from the custom Angiosperms353 set ranged from 0/0 to 232/122, with a mean value of 56.0/15.8 sites, respectively (Figure 2A, B). The number of PIS for the old Annonaceae set ranged from 1/0 to 360/96, with a mean of 43.5/11.5 sites (Figure 2A, B). The total length of all per-gene alignments combined (exon-only) was 604,924 bp, of which 34,438 were PIS when outgroups were considered, and 9337 were PIS when only ingroup species were considered (Table 2). The alignment combining genes from the custom Angiosperms353 set was 311,095 bp long, of which 17,684 were PIS when outgroups were considered, and 4987 were PIS when only species of the ingroup were considered. The combined alignment of the old Annonaceae set was 293,829 bp, of which 16,754 were PIS when outgroups were considered, and 4350 were PIS for "ingroup-only" (Table 2). The pairwise genetic distance between samples of *Asimina triloba* ranged from 0.13% to 0.326% (Table 3).

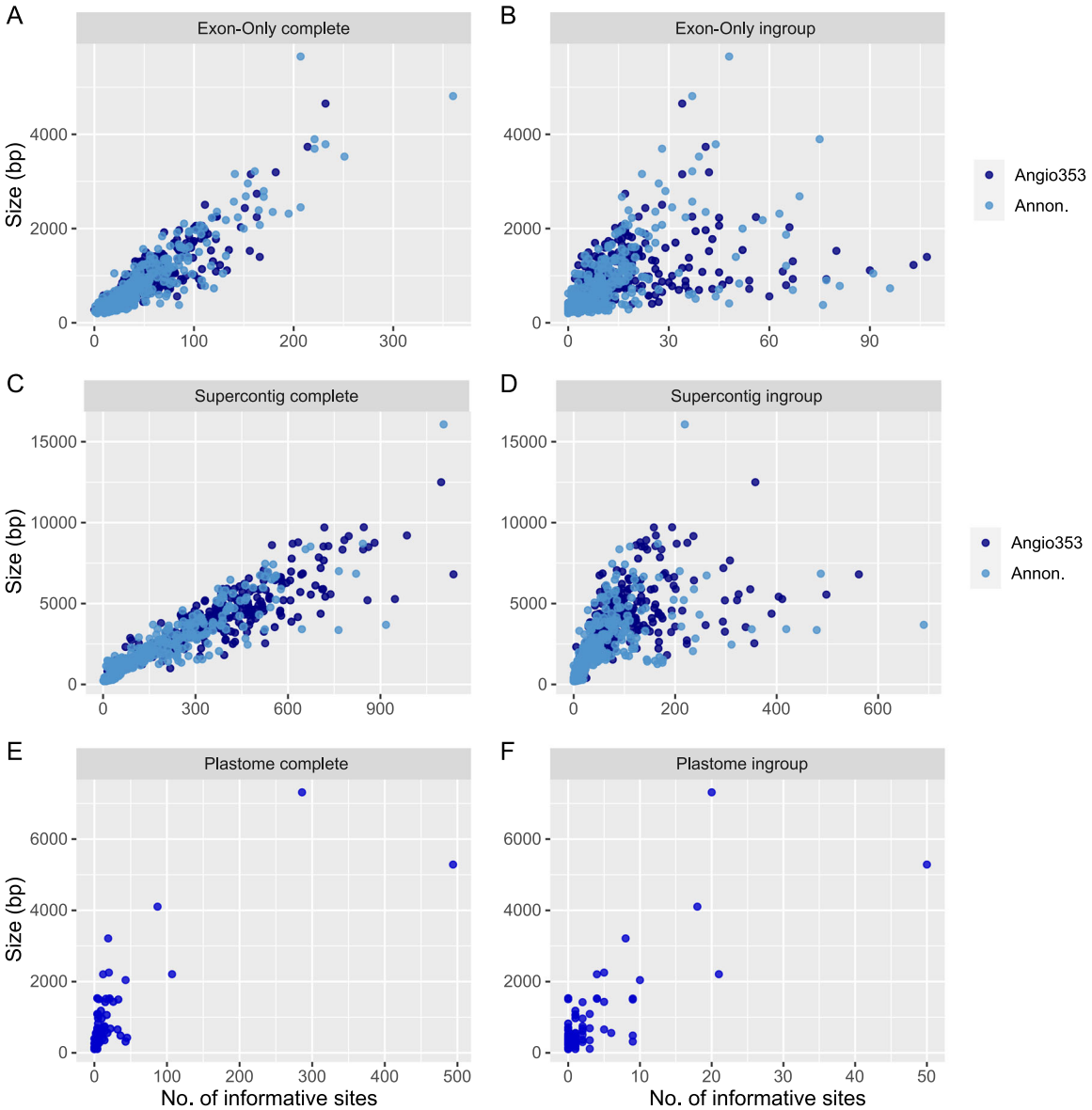

FIGURE 2 Size and number of parsimony-informative sites for each of the 701 genes compared. (A, B) Genes from the exon-only data set, with (A) and without (B) outgroups. (C, D) Genes from the supercontig data set, with (C) and without (D) outgroups. (E, F) Genes from the plastome data set, with (E) and without (F) outgroups.

Nuclear non-targeted sequences

Off-target regions from introns and intergenic regions were obtained for *Asimina*, *Deeringothamnus*, and outgroups. In total, off-target regions were obtained for 499 genes, 277 genes from the Angiosperms353 set and 222 from the Annonaceae set. The length of the alignments ranged from 656 bp to 16,069 bp, with a mean of 3677 bp. Alignments from the Angiosperms353 set ranged from 824 bp to 12,498 bp, with a mean size of 4222 bp. The extreme values were again obtained from the Annonaceae set, with a mean value of 2997.1 bp. The number of PIS considering the whole gene set for "complete" (Figure 2C) and "ingroup-only" (Figure 2D) alignments ranged from 10/3 to 1137/690, with a mean of 331.4/120 sites, respectively. The number of PIS for the genes incorporated from the Angiosperms353 set ranged from 14/5 to 1137/562, with a mean of 392.1/142.4 sites (Figure 2C, D). The number of PIS from the old Annonaceae set ranged from 10/3 to 1105/690, with a mean of 255.5/92.1 sites (Figure 2C, D). When all the alignments were combined (supercontig), the length was 1,939,083 bp. When outgroups were considered, the data set contained 171,052 PIS, with 47,208 PIS in the "ingroup-only" alignment (Table 2). When the different data sets were evaluated, the custom Angiosperms353 alignment was 1,198,434 bp long, of which 110,194 sites were PIS when outgroups were considered, and 29,432 were PIS when only species of *Asimina* were considered. The Annonaceae alignment was 740,649 bp long, of which 60,858 sites were PIS when outgroups were considered, and 17,776 were PIS for "ingroup-only" (Table 2). The pairwise genetic distance between samples of *Asimina triloba* ranged from 0.26% to 0.52% (Table 3).

TABLE 2 Number of aligned sequences and parsimony-informative sites for each data set.

Probe set	No. of aligned sites	No. of parsimony-informative sites	
		Complete set	Ingroup-only set
Angiosperms353	311,095	17,684	4987
Annonaceae	293,829	16,754	4350
Exon-only	604,924	34,438	9337
Angiosperms353	1,198,434	110,194	29,432
Annonaceae	740,649	60,858	17,776
Supercontig	1,939,083	171,052	47,208
Plastome	66,237	1643	241

Plastome non-targeted sequences

Off-target regions from the plastome were obtained from ingroup and outgroup specimens, and for all 78 genes used as references (Figure S1). *pet*L, *pet*N, and *rpo*A were excluded because their sequences were obtained for less than half of the individuals or had sequences shorter than 100 bp. The retained sequence alignments ranged from 105 bp to 7314 bp, with a mean of 883.2 bp (Figure 1, Table 1). The number of PIS for "complete" (Figure 2E) and "ingroup-only" (Figure 2F) alignments ranged from 0/0 to 494/50, with a mean of 21.9/3.2 sites. The length of all the alignments combined (plastome) was 66,237 bp, of which 1643 bp were PIS when outgroups were considered and 241 bp were PIS when only species of the ingroup were considered (Table 2).

Comparison between Angiosperms353 probe kits

The custom Angiosperms353 probe kit targeted 334 genes, 332 of which could be assembled, while the original kit targeted 353 genes, 350 of which could be assembled for *Asimina* species (Table 4). The mean gene recovery value for the custom kit was 0.94, while the value for the original probe kit was 0.79 (Figure 3, Table 4). The custom genes had a longer combined alignment of 314,337 bp, compared with that of the original genes (240,833 bp), and the former had more variable sites (2935 vs. 1887 bp). For supercontig

TABLE 3 Comparison of within-species molecular variation for eight samples of *Asimina triloba* using the K80 evolutionary model. Values above the diagonal 0 values are pairwise comparisons for the exon-only data set, and values below the diagonal are pairwise comparisons for the supercontig data set.

	A. triloba47	*A. triloba48*	*A. triloba49*	*A. triloba51*	*A. triloba53*	*A. triloba54*	*A. triloba71*
A. triloba47	0	0.002349	0.002201	0.002770	0.003157	0.003240	0.003101
A. triloba48	0.003098	0	0.001966	0.002430	0.003084	0.003257	0.003014
A. triloba49	0.003176	0.003402	0	0.002507	0.003084	0.002186	0.002719
A. triloba51	0.004655	0.004379	0.004868	0	0.003055	0.003265	0.002970
A. triloba53	0.003815	0.004009	0.002729	0.004796	0	0.001301	0.002181
A. triloba54	0.004147	0.004227	0.002922	0.005269	0.002607	0	0.002213
A. triloba71	0.004048	0.003792	0.003799	0.004580	0.003552	0.004067	0

TABLE 4 Comparison between the custom and original Angiosperms353 probe kits using one individual of each species of *Asimina*.

Probe kit comparison		Genes			Alignments	
		Targeted	Assembled	Mean recovery	No. of aligned sites	No. of parsimony-informative sites
Exon-only	Custom	334	332	0.94	314,337	2935
	Original	353	350	0.79	240,833	1887
Supercontig	Custom	334	285	NA	1,138,782	13,824
	Original	353	216	NA	870,032	11,817

Note: NA, not applicable.

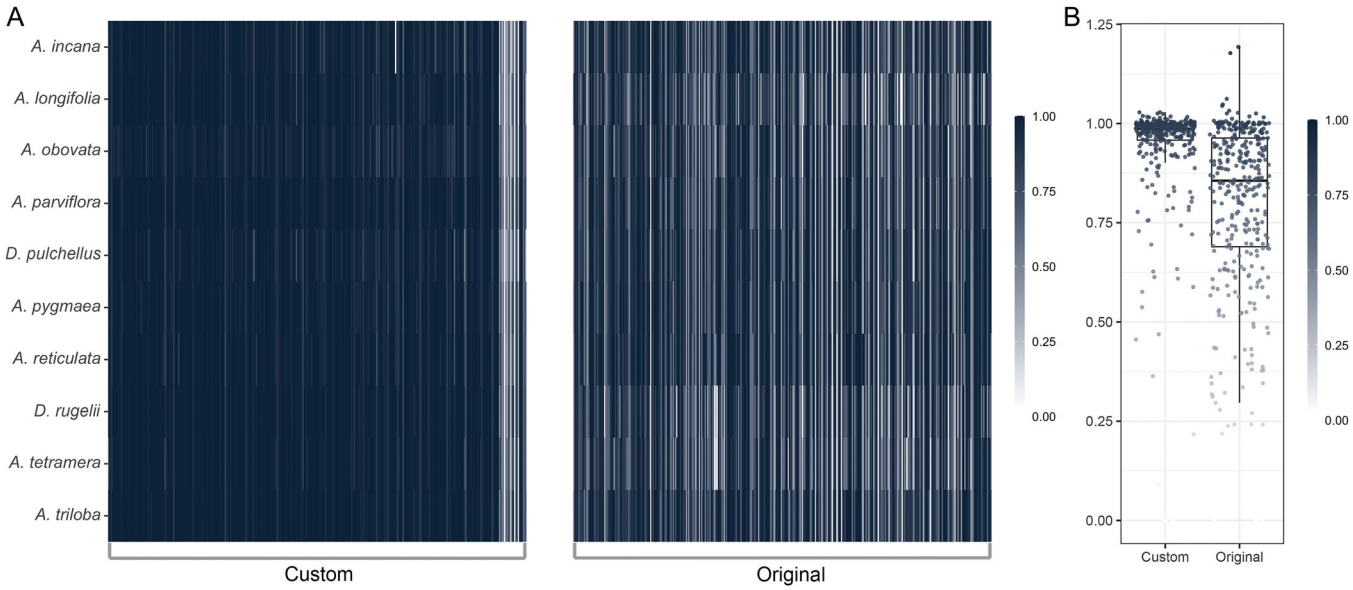

FIGURE 3 Proportion of gene recovery. (A) Heatmap comparing the results of the custom and original Angiosperms353 gene set. (B) Box plot of the proportional values of gene assembly for genes from the custom and original Angiosperms353 sets. Boxes show the upper and lower quartiles; whiskers indicate the extent of variability observed outside the boxes.

regions, the custom kit recovered 285 genes, while the original kit recovered 216 genes. The assembled custom genes had a longer combined alignment (1,138,782 bp) than that of the original genes (870,032 bp). The custom set recovered 13,824 bp of PIS, compared with 11,817 bp for the original kit (Table 4).

Phylogenetic reconstructions

The single ML tree derived from the analysis of the concatenated exon-only data set with 701 genes and 604,924 bp had high ultra-fast bootstrap (UFBoot) values (>90%) across most branches (Figure 4), with a mean of 90.7% for the entire tree. The tree was rooted with *G. laoticus* following previous phylogenetic reconstructions of the Annonaceae (Chatrou et al., 2012). The genus *Asimina* emerged as monophyletic, with maximum UFBoot support. The eight branches inside the genus received support values below 100% (Figure 4), including three within *Asimina triloba*. The ML tree derived from the analysis of the concatenated supercontig data set with 701 genes and 1,939,083 bp had high UFBoot values (>90%) across most branches (Figure 4), with a mean of 98.8% for the entire tree. *Asimina* emerged as monophyletic with maximum UFBoot support. Four branches inside the genus received support values below 100% (Figure 4), one of which was within *Asimina triloba*. The ML tree derived from the analysis of the concatenated plastome data set with 75 genes and 66,237 bp had high UFBoot values (>90%) across most branches (Figure 4), with a mean value of 88.1%. *Asimina* emerged as monophyletic with maximum UFBoot support. Six branches inside the genus received support values below 100% (Figure 4), three of which were within *Asimina triloba*.

Trees for the Angiosperms353 custom and original data sets were also inferred using both the exon-only and supercontig data sets (Figure S2). The custom trees recovered one branch with a UFBoot support value below 100%, while

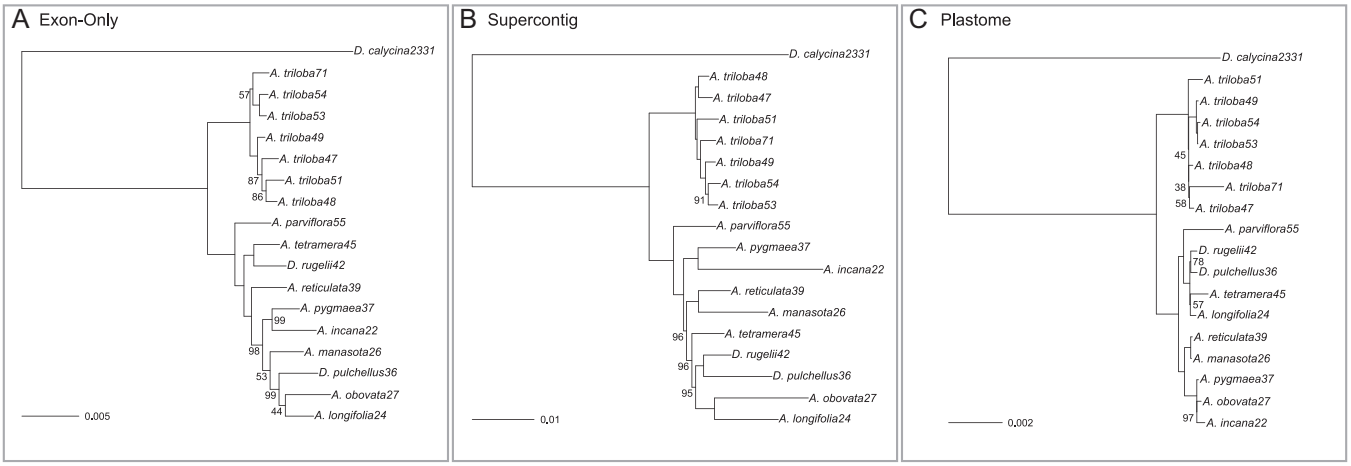

FIGURE 4 Maximum-likelihood phylogenetic trees obtained using the different data sets. The outgroups *Goniothalamus laoticus*, *Annona glabra*, and *A. scandens* were omitted to facilitate visualization. Values on the branches are Ultra-Fast Bootstrap results. Maximum values were omitted. The phylogenies were obtained with (A) the exon-only data set, (B) the supercontig data set, and (C) the plastome data set.

trees from the original data set recovered all internal branches with maximum support (Figure S2).

DISCUSSION

Target enrichment in plant studies

In order to make the original Annonaceae probe kit more universal, reduce the proportion of paralogs assembled, and increase molecular variability, we generated an augmented version of the old Annonaceae probe kit (Couvreur et al., 2019) by including a portion of the Angiosperms353 exons. This new kit allows the sequencing of 799 single- to low-copy nuclear genes, instead of the original 469 exons from the original kit. The new version of the kit does not require an increase in costs as it uses the maximum number of probes available in Daicel Arbor Bioscience entry-level kits (i.e., 20,000). The advent of hybrid kits, including probes designed using genes selected from universal kits (i.e., Angiosperms353) and family-specific kits, allows the interconnection of data generated at broad taxonomic scales with projects focusing on lower taxonomic levels of the phylogeny (Baker et al., 2021). One limitation of the universal kits is that the number of genes assembled is restricted by the number of probes in the kit, and it is challenging to embrace all the molecular variation present within the flowering plants. The solution to maximize molecular variation while keeping the number of different probes reasonable reached its limitation for some species and clades, however. This was clear in the original publication of the Angiosperms353 panel, with the probe kit extensively failing to capture genes for some of the species tested (i.e., *Upuna borneensis* Symington and *Nicotiana heterantha* Symon & Kenneally; Johnson et al., 2019). The maintenance of the Angiosperms353 gene set coupled with clade-specific probes is a step forward, assuring high recovery values for all the genes in most cases (e.g., Ogutcen et al., 2021; Fonseca et al., 2023). The recovery rate of the new version of the Annonaceae probe kit is extremely high (Figure 1), with a mean value of 798.4 genes or 99.9% of those included, and up to 50% of the gene sequence assembled for 776.9 genes across the species (Table 1), echoing previous results obtained for the Bignoniaceae, for which 98% of genes had an extremely high recovery rate (Fonseca et al., 2023).

The merger of gene sets provides the best of both worlds: the connection between projects with different taxonomic scopes and the inclusion of highly variable genes to resolve shallow relationships in the tree of life (Hendriks et al., 2021). Positive results were recently obtained for diverse plant clades that implemented a merged approach (Ogutcen et al., 2021; Fonseca et al., 2023), but the second assertion was not always met. Some genes from the Angiosperms353 were as good as genes from lineage-specific hybridization capture kits (e.g., Larridon et al., 2020), in some cases even outperforming the latter in terms of variation and number of PIS (Fonseca et al., 2023). This is also true for the Annonaceae799 kit, with the variation and PIS for the genes from the custom Angiosperms353 panel outperforming the genes from the old Annonaceae kit (Figure 2A–D, Table 2). The genes from the custom Angiosperms353 panel were also longer and recovered more flanking regions than those in the original Annonaceae probe kit. This is due to the gene selection processes used for these kits (Johnson et al., 2019; Leebens-Mack et al., 2019).

We found that 49.5% of reads were on-target (Table 1), a value comparable with those obtained using other clade-specific kits (e.g., 31.6% in *Dioscorea* [Soto Gomez et al., 2019] and 48.6% in *Euphorbia* [Villaverde et al., 2018]). This relatively high number of off-target reads allows the assembly of plastome genes through the

HybPiper pipeline (Johnson et al., 2016). It is not clear how to implement the necessary parameters for the enrichment step to ensure high recovery rates for off-target regions (Weitemier et al., 2014). In fact, some studies failed to assemble complete or nearly complete plastomes or plastome genes for most of the samples (e.g., Villaverde et al., 2018), highlighting the excellent results obtained here with plastome data (Figure S1). Here, we followed the protocol (Daicel Arbor Biosciences, 2022), with an enrichment step lasting for 24 h at 60°C. Increments of non-enriched libraries were not included to boost the results for off-target regions. The temperature was notably lower than the 65°C suggested by Daicel Arbor Biosciences. For the Annonaceae799 probe kit, this lower temperature guarantees extremely high rates of recovery for on-target genes while allowing the assembly of off-target regions (Figures 1 and 3) without the addition of extra volumes of non-enrichment libraries before sequencing, such as the volume ratio of 1:1 used by Hendriks et al. (2023).

Annonaceae799 probe kit

The new version of the Annonaceae target capture kit will allow users to infer phylogenies, study species complexes, and infer population histories using exons from up to 799 genes. Considering the *Asimina* and *Deeringothamnus* framework with four outgroups (*Annona glabra*, *Annona scandens*, *Diclinanona calycina*, and *Goniothalamus laoticus*), 90 genes were flagged as containing paralogs, including 16 genes from the custom Angiosperms353 panel and 74 from the old Annonaceae set. Considering previous studies (e.g., of the Bignoniaceae, for which 202 out of 711 genes were flagged as paralogs [Fonseca et al., 2023], and of the Gesneriaceae, with 219 out of 830 genes flagged [Ogutcen et al., 2021]), the number of genes with putative paralogs obtained here is relatively low. The paralogs are a minor problem if we consider that 56 genes had a single specimen flagged, and 70 genes had fewer than five samples flagged. Further analyses on putative paralogs can also increase the number of useful genes (Yang and Smith, 2014; Morales-Briones et al., 2021), and even provide new and hidden orthologous sets among the current reference genes (Yang and Smith, 2014).

The old Annonaceae probe kit was designed to be used from the large-scale family level all the way down to within-population phylogeographies, and has been successfully applied in many studies (Helmstetter et al., 2020a, 2020b; Dagallier et al., 2023; Martínez-Velarde et al., 2023; Lopes et al., 2024). The new set designed here was first intended to tackle taxonomically complex taxa, as well as population studies using the maximum amount of molecular data possible through target enrichment. The first application of the Annonaceae799 probe kit will be in a phylogenomic and population study of *Asimina* and *Deeringothamnus* (Fonseca et al., in prep.). The molecular variation within *Asimina* and *Deeringothamnus* is low compared with the results including outgroups (Figure 2A–D, Table 2); however, thousands of bases are available for species-level phylogenetic relationships (Figure 2A–D, Table 2), and hundreds of bases are informative when different individuals of *Asimina triloba* are compared (Table 3). A positive surprise is the performance of the custom Angiosperms353 probe set compared with the old Annonaceae probe kit in terms of molecular variation (Figure 2A–D, Table 2). The new genes certainly improved the original probe kit beyond the increase in the number of bases and genes, providing more variable bases and PIS for phylogenetic and population studies. The new version of the Annonaceae probe kit will also allow for data reuse and will contribute to ongoing efforts to reconstruct the plant tree of life using the Angiosperms353 kit (Baker et al., 2021) as a source of molecular data.

Angiosperms353 gene sets

Here we compared two different probe kits targeting the same set of genes. The original Angiosperms353 kit was designed to target orthologous genes for all flowering plants and included a comprehensive set of more than 75,000 different probes (Johnson et al., 2019). To reduce the number of probes necessary to achieve the same objective in a smaller clade of the flowering plants, we used the 410 alignments from Leebens-Mack et al. (2019) and kept only the genes from the Angiosperms353 set (Johnson et al., 2019) and the species of Annonaceae (*Annona muricata* and *Uvaria macrocarpa*). Using this strategy, we reduced the number of necessary probes to target a subset of the genes (334 out of the 353 genes) using a small fraction of probes (just 9.3%). Despite targeting fewer genes, the custom probe kit recovered longer combined alignments for both exon-only and supercontig data sets when compared with those obtained using the original Angiosperms353 set. The number of PIS was also higher for both the exon-only and supercontig data sets (Table 4). The combined supercontig alignment is almost 50% larger and includes 84% more PIS, providing more informative positions. Highlights of the comparison between strategies to obtain sequences for the same genes are: (1) the higher efficiency of the custom kit during gene assembly compared with the original Angiosperms353 set, as already suggested by Baker et al. (2021); and (2) the higher number of PIS for the custom kit, allowing more positions to be used during the tree search at shallow levels of the phylogeny. The latter observation is extremely important for individuals and research groups designing custom target capture kits for shallow levels of the angiosperm tree of life, or in intra-specific studies designed to use the Angiosperms353 gene set.

These results should be read with caution, however. The comparison between the custom and original Angiosperms353 bait sets is not based on standardized procedures, and differences during library assembly, target enrichment, and sequencing could be responsible for the different results obtained. Nevertheless, we consider the comparison between kits provided here to be informative, due to the huge differences in gene length and PIS between the results

obtained from the two sets. It is unlikely that these distances are driven by methodological differences alone. As shown here, combining kits is possible and could be technically superior. Whether or not to include the Angiosperms353 gene set in the researcher's custom kit or to use different kits separately may also depend on other factors, such as technical or financial limitations. Today, it is typically financially convenient to include the Angiosperms353 gene set in a custom kit instead of using it separately; however, the pricing policies of the companies responsible for synthesizing the probes and making available the enrichment kits are fluid and this situation could change quickly.

Towards a well-resolved phylogeny of *Asimina*

Here we present the first phylogenetic results obtained for *Asimina* and *Deeringothamnus* using a supermatrix approach. The clade emerged as monophyletic in all analyses, although the topology varied between data sets (Figure 4). Considering the morphological variation within *Asimina*, the topology derived from the supercontig data set aligns with expected patterns, such as the emergence of a well-supported clade containing the *Deeringothamnus* species *D. pulchellus* Small and *D. rugelii*. The supercontig data set also yielded a topology with less conflict and higher UFBoot support values (Figure 4). When custom and original probes were tested using the exon-only and supercontig data sets, different topologies were obtained, highlighting that the complete Annonaceae799 set may be necessary for improved resolution within the clade (Figure S2).

The trees generated for *Asimina* and *Deeringothamnus* in this study are preliminary, as they do not account for other evolutionary processes such as incomplete lineage sorting (Pamilo and Nei, 1988; Maddison, 1997) and introgression (Nakhleh, 2013). A more comprehensive phylogenetic study of the genus using more samples and different phylogenetic methods is currently in progress (Fonseca et al., in prep.); however, the initial findings presented here show the relevance of the Annonaceae799 probe kit for phylogenetic purposes in *Asimina*, due to the low molecular variability observed (Tables 2 and 4). The custom Angiosperms353 probes were designed using two different transcriptomes from various clades within the family and one from the Eupomatiaceae, suggesting that the successful gene assembly for *Asimina*, *Deeringothamnus*, and outgroups can be replicated in other contexts, whether for population studies within species, other genera, or broader clades within the Annonaceae.

AUTHOR CONTRIBUTIONS

L.H.M.F., T.L.P.C., and L.W.C. conceived the study; L.H.M.F. and L.W.C. wrote the manuscript; L.H.M.F., V.S., and F.J.N. obtained and analyzed the data; P.A. and V.S. assisted with laboratory work, including DNA extractions, library preparations, and the enrichment step, and provided feedback on the manuscript; K.R.G. and K.M. provided samples and feedback on the manuscript. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

L.H.M.F. and L.W.C. would like to thank Ghent University and the Foundation Arboretum Wespelaar for funding the project. This project also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant no. 865787). K.M. would like to thank the Maxwell-Hanrahan Foundation for supporting the fieldwork. The authors are grateful to Dr. Anne Cox (Florida International University) and Dr. Élaine Norman (Stetson University) for help during fieldwork and extremely valuable information about species locality; Dr. Robert Raguso (Cornell University) and Dr. Deise Gonçalves (University of Michigan) for sharing silica-dried samples; and Dr. Lucas Majure and Kent Perkins for sharing silica-dried and herbarium-sample leaf fragments from specimens housed in the University of Florida Herbarium (FLAS). We are also thankful to the associate editor and two anonymous reviewers for suggestions that improved our manuscript. We would also like to thank the Vlaams Supercomputer Center at Ghent University for providing computational infrastructure, guidance, and support.

DATA AVAILABILITY STATEMENT

Raw sequencing reads are available via the National Center for Biotechnology Information (NCBI; BioProject: PRJNA1086860 and PRJNA508895). Details of the read accessions are available in Table S1. Alignments and custom scripts are available at GitHub (https://github.com/luizhhziul/Annonaceae799).

ORCID

Luiz Henrique M. Fonseca http://orcid.org/0000-0002-8140-3244
Katherine R. Goodrich https://orcid.org/0000-0002-3014-0683
Francis J. Nge https://orcid.org/0000-0002-0361-8709
Vincent Soulé https://orcid.org/0009-0001-3910-2317
Kathryn Mercier https://orcid.org/0000-0001-7898-1353
Thomas L. P. Couvreur https://orcid.org/0000-0002-8509-6587
Lars W. Chatrou https://orcid.org/0000-0003-0131-0302

REFERENCES

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.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Website: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [accessed 12 July 2024].

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

Ball, L. D., A. M. Bedoya, C. M. Taylor, and L. P. Lagomarsino. 2023. A target enrichment probe set for resolving phylogenetic relationships in the coffee family, Rubiaceae. *Applications in Plant Sciences* 11: e11554.

Borowiec, M. L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660.

Chatrou, L. W., M. D. Pirie, R. H. J. Erkens, T. L. P. Couvreur, K. M. Neubig, J. R. Abbott, J. B. Mols, et al. 2012. A new subfamilial and tribal classification of the pantropical flowering plant family Annonaceae informed by molecular phylogenetics. *Botanical Journal of the Linnean Society* 169: 5–40.

Chen, J., Z. Hao, X. Guang, C. Zhao, P. Wang, L. Xue, Q. Zhu, et al. 2019. *Liriodendron* genome sheds light on angiosperm phylogeny and species–pair differentiation. *Nature Plants* 5: 18–25.

Chen, S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* 2: e107.

Couvreur, T. L. P., A. J. Helmstetter, E. J. M. Koenen, K. Bethune, R. D. Brandão, S. A. Little, H. Sauquet, and R. H. J. Erkens. 2019. Phylogenomics of the major tropical plant family Annonaceae using targeted enrichment of nuclear genes. *Frontiers in Plant Science* 9: 1941.

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.

Dagallier, L. P. M. J., F. M. Mbago, M. Couderc, M. Gaudeul, A. Grall, C. Loup, J. J. Wieringa, et al. 2023. Phylogenomic inference of the African tribe Monodoreae (Annonaceae) and taxonomic revision of *Dennettia*, *Uvariodendron*, and *Uvariopsis*. *Phytokeys* 233: 1–200.

Daicel Arbor Biosciences. 2022. MyBaits: Hybridization capture for targeted NGS, version 5.02, p. 8. Daicel Arbor Biosciences, Ann Arbor, Michigan, USA. Available at: https://arborbiosci.com/wp-content/uploads/2022/03/myBaits_v5.02_Manual.pdf [accessed 8 August 2024].

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Erkens, R. H. J., L. W. Chatrou, T. Chaowasku, L. Y. T. Westra, J. W. Maas, and P. J. M. Maas. 2014. A decade of uncertainty: Resolving the phylogenetic position of *Diclinanona* (Annonaceae), including taxonomic notes and a key to the species. *Taxon* 63: 1244–1252.

Eserman, L. A., S. K. Thomas, E. E. 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.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32: 3047–3048.

Fonseca, L. H. M., M. M. Carlsen, P. V. Fine, and L. G. Lohmann. 2023. A nuclear target sequence capture probe set for phylogeny reconstruction of the charismatic plant family Bignoniaceae. *Frontiers in Genetics* 13: 1085692.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150–3152.

Helmstetter, A. J., B. E. N. Amoussou, K. Béthune, N. G. Kamden, R. G. Kakaï, B. Sonké, and T. L. P. Couvreur. 2020a. Phylogenomic approaches reveal how climate shapes patterns of genetic diversity in an African rain forest tree species. *Molecular Ecology* 29: 3560–3573.

Helmstetter, A. J., K. Béthune, N. G. Kamden, B. Sonké, and T. L. P. Couvreur. 2020b. Individualistic evolutionary responses of Central African rain forest plant to Pleistocene climatic fluctuations. *Proceedings of the National Academy of Sciences, USA* 117: 32509–32518.

Hendriks, K. P., C. Kiefer, I. A. Al-Shehbaz, C. D. Bailey, A. Hooft van Huysduynen, L. A. Nikolov, L. Nauheimer et al. 2023. Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. *Current Biology* 33: 4052–4068.

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.

Hill, C. B., D. Wong, J. Tibbits, K. Forrest, M. Hayden, X. Q. Zhang, S. Westcott, et al. 2019. Targeted enrichment by solution-based hybrid capture to identify genetic sequence variants in barley. *Scientific Data* 6: 12.

Hoang, D. T., O. Chernomor, A. van Haeseler, B. Q. Minh, and L. S. Vinh. 2018. UFBoot2: Improving the Ultrafast Bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.

Jiménez-Mena, B., H. Flávio, R. Henriques, A. Manuzzi, M. Ramos, D. Meldrup, J. Edson et al. 2022. Fishing for DNA? Designing baits for population genetics in target enrichment experiments: Guidelines, considerations and the new tool supeRbaits. *Molecular Ecology Resources* 22: 2105–2119.

Jin, J. J., W. B Yu., J. B. Yang, Y. Song, C. W. DePamphilis, T. S. Yi, and D. Z. Li. 2020. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21: 241.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. Zerega, and N. J. Wickett. 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.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Kent, W. J. 2002. BLAT—the BLAST-like alignment tool. *Genome Research* 12: 656–664.

Kral, R. 1960. A revision of *Asimina* and *Deeringothamnus* (Annonaceae). *Brittonia* 12: 233–278.

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.

Leebens-Mack, J. H., M. S. Barker, E. J. Carpenter, M. K. Deyholos, M. A. Gitzendanner, S. W. Graham, I. Grosse, et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679–685.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.

Li, P. S., D. C. Thomas, and R. M. K. Saunders. 2017. Historical biogeography and ecological niche modelling of the *Asimina-Disepalum* clade (Annonaceae): Role of ecological differentiation in Neotropical-Asian disjunctions and diversification in Asia. *BMC Evolutionary Biology* 17: 188.

Lopes, J. C., L. H. M. Fonseca, D. M. Johnson, F. Luebert, N. Murray, F. J. Nge, C. Rodrigues-Vaz, et al. 2024. Dispersal from Africa to the Neotropics was followed by multiple transitions across Neotropical biomes facilitated by frugivores. *Annals of Botany* 133: 659–676,

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

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, R. W. Michelmore, et al. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences* 2: e1300085.

Mariac, C., N. Scarcelli, J. Pouzadou, A. Barnaud, C. Billot, A. Faye, A. Kougbeadjo, et al. 2014. Cost effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. *Molecular Ecology and Resources* 14: 1103–1113.

Martínez-Velarde, M. F., C. Rodriguez-Vaz, V. Soulé, F. J. Nge, G. E. Schatz, T. L. P. Couvreur, and A. E. Ortiz-Rodriguez. 2023. *Desmopsis terriflora*, an extraordinary new species of Annonaceae with flagelliflory. *Phytokeys* 227: 181–198.

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.

Mercer, E., B. Griffin, J. Steele, K. R. Goodrich, and C. M. Bush. 2016. Phylogenetic relationships of *Asimina* and *Deeringothamnus* (Annonaceae) based on morphology, floral scent chemistry, and Inter-Simple Sequence Repeat data. *Journal of the Torrey Botanical Society* 143: 58–68.

Moore-Pollard, E. R., D. S. Jones, and J. R. Mandel. 2024. Composite-ParaLoss-1272: A complementary sunflower-specific probe set reduces paralogs in phylogenomic analyses of complex systems. *Applications in Plant Sciences* 12: e11568.

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 sl. *Systematic Biology* 70: 219–235.

Nakhleh, L. 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends in Ecology and Evolution* 28: 719–728.

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: e11441.

Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Niu, Y. F., K. X. Li, and J. Liu. 2020. The complete chloroplast genome of *Annona muricata* L.: A tropical fruit with important medicinal properties. *Mitochondrial DNA Part B* 5: 3330–3332.

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.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5: 568–583.

Paradis, E., and K. Schliep. 2019. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.

R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: https://www.R-project.org/ [accessed 12 July 2024].

Rohland, N., and D. Reich. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* 22: 939–946.

Soltis, D. E., M. A. Gitzendanner, G. Stull, M. Chester, A. Chanderbali, S. Chamala, I. Jordon-Thaden, et al. 2013. The potential of genomics in plant systematics. *Taxon* 62: 886–898.

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.

Steenwyk, J. L., T. J. Buida III, Y. Li, X. X. Shen, and A. Rokas. 2020. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biology* 18: e3001007.

Thiers, B. 2024 (continuously updated). Index Herbariorum. Website: http://sweetgum.nybg.org/science/ih/ [accessed 14 July 2024].

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.

Waycott, M., K. van Dijk, and E. Biffin. 2021. A hybrid capture RNA bait set for resolving genetic and evolutionary relationships in angiosperms from deep phylogeny to intraspecific lineage hybridization. *bioRxiv* 456727 [Preprint]. Available at: https://doi.org/10.1101/2021.09.06.456727 [posted 7 September 2021; accessed 29 July 2024].

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: e1400042.

Wen, D., and L. Nakhleh. 2018. Coestimating reticulate phylogenies and gene trees from multilocus sequence data. *Systematic Biology* 67: 439–457.

Yang, Y., and S. A. Smith. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution* 31: 3081–3092.

Zuntini, A. R., T. Carruthers, O. Maurin, P. C. Bailey, K. Leempoel, G. E. Brewer, N. Epitawalage, et al. 2024. Phylogenomics and the rise of the angiosperms. *Nature* 629: 843–850.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Table S1. Voucher information for the 21 specimens used in the enrichment reactions with the Annonaceae799 probe kit.

Table S2. Voucher information for the 10 specimens used in the enrichment reactions with the custom Angiosperms353 probe kit.

Figure S1. Heatmap of the 78 genes from the plastome.

Figure S2. Phylogenies obtained using the custom and original Angiosperms353 data sets. The trees were rooted using *Asimina triloba*. Values on the branches are Ultra-Fast Bootstrap results. Maximum values were omitted. (A) Phylogeny obtained with the custom exon-only data set. (B) Phylogeny obtained with the custom supercontig data set. (C) Phylogeny obtained with the original exon-only data set. (D) Phylogeny obtained with the original supercontig data set.

How to cite this article: Fonseca, L. H. M., P. Asselman, K. R. Goodrich, F. J. Nge, V. Soulé, K. Mercier, T. L. P. Couvreur, and L. W. Chatrou. 2024. Truly the best of both worlds: Merging lineage-specific and universal probe kits to maximize phylogenomic inference. *Applications in Plant Sciences* 12(6): e11615. https://doi.org/10.1002/aps3.11615