Received: 30 October 2023 | Accepted: 19 February 2024

DOI: 10.1002/aps3.11597

APPLICATION ARTICLE

Applications in Plant Sciences

Developing Asparagaceae1726: An Asparagaceae-specific probe set targeting 1726 loci for Hyb-Seq and phylogenomics in the family

Philip C. Bentz | **Jim Leebens-Mack**

Department of Plant Biology, University of Georgia, 120 Carlton St., Athens, Georgia 30605, USA

Correspondence

Philip C. Bentz, Jim Leebens-Mack, Department of Plant Biology, University of Georgia, 120 Carlton St., Athens, Georgia 30605, USA.
Email: philip.bentz@uga.edu
and jleebensmack@uga.edu

Abstract

Premise: Target sequence capture (Hyb-Seq) is a cost-effective sequencing strategy that employs RNA probes to enrich for specific genomic sequences. By targeting conserved low-copy orthologs, Hyb-Seq enables efficient phylogenomic investigations. Here, we present Asparagaceae1726—a Hyb-Seq probe set targeting 1726 low-copy nuclear genes for phylogenomics in the angiosperm family Asparagaceae—which will aid the often-challenging delineation and resolution of evolutionary relationships within Asparagaceae.

Methods: Here we describe and validate the Asparagaceae1726 probe set (https://github.com/bentzpc/Asparagaceae1726) in six of the seven subfamilies of Asparagaceae. We perform phylogenomic analyses with these 1726 loci and evaluate how inclusion of paralogs and bycatch plastome sequences can enhance phylogenomic inference with target-enriched data sets.

Results: We recovered at least 82% of target orthologs from all sampled taxa, and phylogenomic analyses resulted in strong support for all subfamilial relationships. Additionally, topology and branch support were congruent between analyses with and without inclusion of target paralogs, suggesting that paralogs had limited effect on phylogenomic inference.

Discussion: Asparagaceae1726 is effective across the family and enables the generation of robust data sets for phylogenomics of any Asparagaceae taxon. Asparagaceae1726 establishes a standardized set of loci for phylogenomic analysis in Asparagaceae, which we hope will be widely used for extensible and reproducible investigations of diversification in the family.

KEYWORDS

Asparagaceae, Hyb-Seq, lineage-specific target sequence capture probe set, phylogenomics

The combination of target enrichment and genome skimming, otherwise known as Hyb-Seq (Weitemier et al., 2014), is a popular method to generate robust data sets of hundreds to thousands of genomic loci for phylogenomic inference (e.g., Grover et al., 2012; Heyduk et al., 2016a), with much lower costs and computational challenges compared to whole genome sequencing (Hale et al., 2020). Organellar genome sequences (e.g., plastomes) can also be obtained for phylogenomics by extracting reads after combining target-enriched libraries with a small aliquot of unenriched shotgun sequencing libraries for each sample (Talavera et al., 2023), or simply mining off-target Hyb-Seq reads for those mapping to the organellar genomes (Granados Mendoza et al., 2020; Schneider et al., 2021; Baldwin et al., 2023). The former approach may avoid potential bias in sequence sampling of organellar sequences that may result from the targeted hybridization/enrichment process.

Universal probe sets that incorporate conserved, low-copy loci spanning taxonomic orders are especially useful in non-model systems lacking reference genomes (e.g., Lemmon et al., 2012; Johnson et al., 2019). However, universal probe

sets may not provide sufficient sequence information to accurately elucidate patterns of lineage diversification at shallower taxonomic scales, especially in clades with incomplete lineage sorting (ILS), historical introgression, or low sequence divergence among lineages (Yardeni et al., 2022). On the other hand, lineage-specific probe sets designed for use across narrower taxonomic scales, with an increased number of target loci, can provide phylogenomic data sets that yield robust estimates of relationships even in the face of rapid radiation (e.g., Mandel et al., 2014; Eserman et al., 2021).

In this study, we validate a newly developed Hyb-Seq probe set designed specifically for phylogenomics in the angiosperm family Asparagaceae (Asparagales). Asparagaceae consists of seven subfamilies (Chase et al., 2009) containing approximately 121 genera (POWO, 2023) and 2595 species (Table 1) that extend across every geographical continent except Antarctica (Stevens, 2001), including the economic and culturally important *Asparagus officinalis* L. (garden asparagus) and *Agave tequilana* F.A.C.Weber (tequila agave). The estimated stem age of Asparagaceae is just ~51.8 million years ago (Ma) (Bentz et al., 2024). Phylogenetic work based on plastome sequences suggests that a rapid radiation of lineages spawned all the Asparagaceae subfamilies (Steele et al., 2012), making ancient ILS a complicating factor for phylogenomic analysis (see Table 1 for age estimates and taxon diversity of each subfamily). Ancient and contemporary hybridization has also occurred within Asparagaceae subfamilies (Good-Avila et al., 2006; McKain et al., 2012, 2016; Heyduk et al., 2021; Meng et al., 2021; Howard et al., 2022, 2023), leading to complicated evolutionary histories that are difficult to elucidate using either phylogenetic analysis of few nuclear or plastome sequences, or morphology. Additionally, the relatively recent origins of species-rich genera in Asparagaceae have resulted in low interspecies sequence divergence; for example, the crown group ages for *Agave* L. (287 species) and *Asparagus* Tourn. ex L. (215 species) are estimated ~6.2 Ma (Jiménez-Barron et al., 2020) and ~5.1 Ma (Bentz et al., 2024), respectively. Lastly, whole genome duplications (WGDs) have occurred multiple times within Asparagaceae subfamilies (Hanson et al., 2003; McKain et al., 2012; Šmarda et al., 2014; Harkess et al., 2017; Gunn et al., 2020; Nath et al., 2023), resulting in paralogy issues that are also difficult to overcome in phylogenetic analysis with limited locus sampling. Altogether, these phenomena make it exceedingly difficult to conduct systematic, taxonomic, and macro-evolutionary studies in Asparagaceae without large multi-locus data sets and an accurate phylogenomic framework.

Many studies have attempted to resolve relationships at various taxonomic scales across Asparagaceae; most of these have used either a few plastid or nuclear loci (e.g., Gándara et al., 2014; Norup et al., 2015; Gutiérrez et al., 2017; Takawira-Nyenya et al., 2018; Flores-Abreu et al., 2019; Meng et al., 2021) or whole plastomes for tree inference (McKain et al., 2016; Gunn et al., 2020; Lu et al., 2022; Ji et al., 2023). Fewer studies have employed a Hyb-Seq approach for phylogenomics in Asparagaceae, likely due to limited genomic resources outside of *Asparagus* (Harkess et al., 2017, 2020; Li et al., 2020). However, the universal probe set designed for angiosperms (i.e., Angiosperms353 by Johnson et al., 2019) was tested in Scilloideae and resulted in strong support for major clades in the subfamily, but failed to robustly infer many of the shallower, species-level relationships (Howard et al., 2022). Other studies have used transcriptome assemblies to design lineage-specific probes from <300 nuclear genes for targeted sequencing of Agavoideae taxa (Heyduk et al., 2016b; Yoo et al., 2021). Both of those studies resolved relationships with strong or moderate support for some clades to very weak support for others. Lastly, a more recent study employed a Hyb-Seq approach with custom RNA probes, corresponding to 515 nuclear genes, to investigate relationships within *Polygonatum* Mill. (Nolinoideae) (Qin et al., 2024). Coalescent-based species tree analyses of those 515 genes resulted in extensive gene tree discordance across much of the genus *Polygonatum* (see figure 2 in Qin et al., 2024). Whether the lack of resolution in previous phylogenomic analyses is a consequence of insufficient data or represents true polytomies—due to radiations resulting in ancestral nodes with more than two daughter branches—needs to be tested with a multi-locus phylogenomic data set including significantly more information content.

In this study, we use genome sequences from Asparagaceae taxa to develop Asparagaceae1726—a lineage-specific Hyb-Seq probe set targeting coding regions from 1726 low-copy nuclear loci conserved across the family. This probe set was designed as part of a larger phylogenomics project aiming to test species relationships across the genus *Asparagus* (Asparagoideae). However, we designed Asparagaceae1726 probes for universal applicability across all Asparagaceae taxa,

TABLE 1 Taxonomy, age, and diversity of extant subfamilies of Asparagaceae.[a]

Clade	Authority	Genera	Species[b]	Age (Ma)
Agavoideae	Herb	23	637	43.5
Aphyllanthoideae	Lindl.	1	1	47.6
Asparagoideae	Burmeist.	2	175–305	40.1
Brodiaeoideae	Traub	12	62	39.5
Lomandroideae	Thorne & Reveal	12	186 (+15)	56.6
Nolinoideae	Burnett	26	590	40.1
Scilloideae	Burnett	41–70	800–1025	39.5

Note: Ma = million years ago.

[a]Taxonomy of subfamilies based on Chase et al. (2009). Age estimates for subfamilies represent stem ages and are based on the following: Agavoideae, Asparagoideae, Brodiaeoideae, Nolinoideae, and Scilloideae = mean age estimates from Bentz et al. (2024); Aphyllanthoideae = mean age estimates for clade of Agavoideae + Aphyllanthoideae from Givnish et al. (2018); Lomandroideae = median age estimates from Gunn et al. (2020). The number of total genera and species were compiled from Stevens (2001; accessed 9 October 2023).

[b]Ranges represent uncertainty in the number of accepted species, and (+) indicates that more species are to be named.

by testing for copy number and sequence similarity of target orthologs in genome assemblies for Asparagoideae and Agavoideae species. Considering the ancestral WGDs that have been documented in Asparagaceae evolution (Hanson et al., 2003; McKain et al., 2012; Šmarda et al., 2014; Harkess et al., 2017; Gunn et al., 2020; Nath et al., 2023), in addition to the prevalence of gene duplications in plant genomes more generally (Panchy et al., 2016), we expect Asparagaceae1726 data sets to yield some proportion of paralogs. We, therefore, assess evidence for multiple gene copies and compare the results of analyses that include single, putative orthologs with analyses that include paralogous gene copies when they are assembled.

Asparagaceae1726 is a publicly available resource (https://github.com/bentzpc/Asparagaceae1726) and represents the first attempt at establishing a standardized set of loci for phylogenomic analysis in Asparagaceae, which will enable data sharing, consistency, and cross-study reproducibility. Here, we describe the development of this new Asparagaceae-specific probe set and report results from Hyb-Seq and phylogenomic experiments with multiple samples from six of the seven subfamilies of Asparagaceae. Aphyllanthoideae samples were not available to us when data were generated to assess the utility of the Asparagaceae1726 bait set. Additionally, we explore the phylogenetic utility of plastomic sequences mined from off-target Hyb-Seq reads and how those may compare to plastome assemblies generated through genome skimming of unenriched shotgun sequence library spike-ins (Weitemier et al., 2014).

METHODS

Asparagaceae1726 probe design

To construct custom probes from loci conserved in low copy numbers across Asparagaceae, we assessed orthology and gene copy numbers across three Asparagaceae genomes (see below) and 11 monocot outgroups using Orthofinder v.2.5.4 (Emms and Kelly, 2015, 2019) with default options. Protein annotations for these assemblies were obtained from the National Center for Biotechnology Information's (NCBI) GenBank (https://www.ncbi.nlm.nih.gov/genbank/) or JGI Phytozome (https://phytozome-next.jgi.doe.gov): *Asparagus setaceus* v.1 (GenBank: GCA_012295165.1; Li et al., 2020), *Acorus americanus* v.1.1 (Phytozome: 586; DOE-JGI, 2020a), *Ananas comosus* v.3 (Phytozome: 321; Ming et al., 2015), *Asparagus officinalis* v.1.1 (Phytozome: 498; Harkess et al., 2017), *Brachypodium distachyon* v.3.2 (Phytozome: 556; The International Brachypodium Initiative, 2010), *Joinvillea ascendens* v.1.1 (Phytozome: 587; DOE-JGI, 2020b), *Musa acuminata* v.1 (Phytozome: 321; D'hont et al., 2012), *Oryza sativa* v.7 (Phytozome: 323; Ouyang et al., 2007), *Phoenix dactylifera* v.1 (GenBank: GCA_009389715.1; Hazzouri et al., 2019), *Phalaenopsis equestris* (GenBank: GCF_001263595.1; Cai et al., 2015), *Sorghum bicolor* v.3.1.1 (Phytozome: 454; McCormick et al., 2018), *Setaria italica* v.2.2 (GenBank: GCF_000263155.2; Bennetzen et al., 2012), and *Yucca aloifolia* v.2.1 (Phytozome: 839; DOE-JGI, 2023). Loci were inferred as conserved and low copy across Asparagaceae if ≤5 ortholog copies were detected for a shared locus among *A. officinalis*, *A. setaceus* (Kunth) Jessop, and *Y. aloifolia* L. We extracted these candidate loci from the assembly for *A. setaceus* using BEDTools getfasta (Quinlan and Hall, 2010), then re-tested sequence similarity of those exons with gene models from the *A. officinalis* and *Y. aloifolia* assemblies via a local BLAST (Altschul et al., 1990) search with a minimum word size of 7, reporting only the best alignment for each exon.

We also used BLAST (*E*-value ≤ 1E−30) to test whether any of these candidate loci are orthologous with targets from Angiosperms353. To include additional targets orthologous with Angiosperms353, we performed a separate BLAST analysis with the same parameters but compared Angiosperms353 targets with all *A. officinalis* gene models, filtering out those with multi-hit exons and alignment lengths <80 nucleotides. Exon sequences that passed filtering were sent to Arbor Biosciences (Ann Arbor, Michigan, USA) and used to design 90-nucleotide-long probes with 2× tiling. Only exon sequences were used in probe design to prevent probes from extending into more variable intronic regions. To filter these probes, target sequences were soft masked for simple repeats and those in the monocot repeat database. Probes were removed if they were composed of ≥25% soft masked sequence or if they failed moderate BLAST filtering that accounted for probe hybridization melting temperature (T_m) (i.e., temperature at which 50% of molecules are hybridized) (Appendix S1; see Supporting Information with this article). In the final probe set, exons that were ≥80 but <90 nucleotides long were padded to 90 nucleotides with thymine bases (T's). Exons shorter than 80 nucleotides were not included in probe design to avoid overpadding probes with ≥10 T's.

DNA isolation, library preparation, hybridization, and sequencing

DNA was extracted from silica-dried, fresh, flash-frozen, or herbarium specimen leaf tissue using the DNeasy Plant Mini Kit (Qiagen, Venlo, the Netherlands) or Quick-DNA Plant/Seed Miniprep Kit (Zymo Research, Irvine, California, USA) (Appendix S2) and quantified with the Qubit 2.0 fluorometer broad-range assay with software version 3.1 (Thermo Fisher Scientific, Waltham, Massachusetts, USA). DNA sequencing libraries were constructed using the KAPA HyperPlus Kit (Roche, Basel, Switzerland) for Illumina with universal Y-yoke stub adapters and dual-indexed iTru primers from Adapterama I (Glenn et al., 2019), provided by the University of Georgia Environmental Health Science's (EHS) DNA Laboratory (Athens, Georgia, USA). Using the KAPA HyperPlus Kit, total genomic DNA was fragmented for 8 min, aiming for an average fragment size of 350–450 bp.

The Y-yoke stub adapters were then ligated to A-tailed molecules during an incubation period of 3 h at 20°C and then overnight at 4°C. Adapter-ligated molecules were then amplified using six PCR cycles with KAPA HiFi HotStart ReadyMix (Roche) and the dual-indexed primers. Libraries were checked for quality using the Bioanalyzer High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, California, USA) and quantified using real-time PCR (qPCR) with the KAPA Library Quantification Kit and KAPA SYBR Fast qPCR Master Mix (Roche). Prior to pooling, absolute quantification of each library was size corrected (mean nM × [452/mean fragment size]) and converted from nM to ng/μL ([(660 × mean fragment size) × size-corrected nM]/1,000,000). Additional DNA isolation and library preparation details can be found in Appendix S2.

To hybridize library molecules with Asparagaceae1726 RNA probes, we equally pooled 16 different libraries, aiming for 200 ng input from each library, for each separate hybridization reaction. With these pools (each containing 16 individual libraries), we then performed a 0.8× bead cleanup using KAPA HyperPure Beads (Roche). After eluting the cleaned pools in 100 μL of H_2O, we concentrated each to a total of 7 μL (i.e., the protocol-recommended input volume for hybridization reactions) using a Savant Speed-Vac SC110 (Thermo Fisher Scientific). We then performed separate hybridization reactions with Asparagaceae1726 probes on each concentrated pool, using the myBaits Custom DNA-Seq kit v.5.03 (Arbor Biosciences). We incubated each hybridization reaction at 60°C for 24 h, then amplified each for 14 PCR cycles to enrich for Asparagaceae1726 targets. PCR reactions were performed as before, then purified using the Monarch PCR & DNA Cleanup Kit (New England Biolabs, Ipswich, Massachusetts, USA). The final target-enriched pools were assessed for quality and quantity as before, equimolarly pooled at 10 nM, then sequenced with the Illumina platform NovaSeq X Plus (300 cycles) 2.5B (paired-end 150-bp reads) at SeqCenter (Pittsburgh, Pennsylvania, USA).

Taxon sampling

In this study, we tested Asparagaceae1726 probes in multiple species from all Asparagaceae subfamilies (Table 1) except Aphyllanthoideae, for which we lacked tissue samples. We included *Dianella tasmanica* Hook.f. (Asphodelaceae, Asparagales) as a close outgroup taxon (Chase et al., 2009) for phylogenomic analyses (see Table 2 for detailed sample information).

Target ortholog and paralog sequence assembly

We removed lingering adapter sequences from Hyb-Seq reads, corrected mismatched base pairs, and filtered out reads shorter than 21 bases using fastp v.0.23.2 (Chen et al., 2018). We then used the HybPiper v.2.1.6 pipeline (Johnson et al., 2016) to (i) map the filtered reads to target nucleotide sequences with BWA-MEM (Li, 2013), (ii) assemble mapped reads into contiguous sequences with SPAdes (Bankevich et al., 2012), (iii) extend gene assemblies into flanking intronic regions using HybPiper's intronerate function, and (iv) test for paralogs with default HybPiper parameters. We constructed the reference target sequence file for read mapping in HybPiper by concatenating exon sequences for each target locus, allowing for alternative mapping to more similar orthologs. For this, we used ortholog reference sequences from *A. setaceus*, *A. officinalis*, and *Y. aloifolia* for 1587 loci, representing those reported as low copy by Orthofinder. An additional 139 target loci from *A. officinalis* were also included as reference sequences in this file, representing orthologs of Angiosperms353 targets. Considering only concatenated exon lengths from one ortholog per locus (139 loci from *A. officinalis* and 1587 from *A. setaceus*), the sum length of all 1726 targets was 2,432,740 nucleotides (Appendix S3). The reference target sequence file used in HybPiper in this study can be downloaded from GitHub (https://github.com/bentzpc/Asparagaceae1726; see Data Availability Statement). For the HybPiper run, we increased the timeout parameter for Exonerate (Slater and Birney, 2005) to 300 s for samples requiring additional time at this step. Potential paralogs were scored for each locus using default HybPiper run parameters, which report paralogs based on the coverage and number of SPAdes contig assemblies mapping to a target sequence with inferred paralog scoring based on "depth" (>75% of target length covered by >2 shorter contigs in addition to the putative ortholog assembly) and contiguous "length" (>1 SPAdes contig mapping to >75% of the reference target sequence) of non-primary assemblies for each target. Lastly, we tested whether target recovery could be increased with the addition of more closely related orthologs in the target reference file used for mapping. To do this, we re-ran HybPiper using two samples exhibiting the lowest target recovery, with additional ortholog exon assemblies (from the first HybPiper run) added to the reference file from a sample with the highest target recovery in the same subfamily.

Phylogenomic analysis

Multiple sequence alignments (MSAs) for each target locus were produced using MAFFT v.7.487 with the flag *--auto*, instructing MAFFT to test for the best alignment strategy (Katoh and Standley, 2013), and poorly aligned sequences were trimmed using trimAl v.1.4.1 (Capella-Gutiérrez et al., 2009) and the flag *-automated1*. We then used IQ-TREE v.1.6.12 (Nguyen et al., 2015) for maximum likelihood (ML) gene tree inference with 1000 ultrafast bootstraps (BS) and the *MFP* option, which allows ModelFinder (Kalyaanamoorthy et al., 2017) to determine the best-fit substitution model for each gene using Bayesian information criterion. The resulting collection of ML unrooted gene trees were analyzed with ASTRAL-III v.5.7.8 (Zhang et al., 2018), which uses a gene tree summary method for a species tree

TABLE 2 Taxa sampled for Hyb-Seq experiments with Asparagaceae1726 probes.[a]

Subfamily	Genus	Species	Authority	Coll. number	Year	Herbarium
Agavoideae	*Agave*	*virginica*	L.	P.C. Bentz 126	2022	GA
Agavoideae	*Behnia*	*reticulata*	(Thunb.) Didr.	Burrows & Burrows 15952	2019	J
Agavoideae	*Chlorophytum*	*comosum*	(Thunb.) Jacques	P.C. Bentz 127	2023	GA
Agavoideae	*Yucca*	*filamentosa*	L.	P.C. Bentz 131	2023	GA
Asparagoideae	*Asparagus*	*bayeri*	(Oberm.) Fellingham & N.L.Mey.	Burrows & Burrows 11919	2010	J
Asparagoideae	*Asparagus*	*burchellii*	Baker	Burrows & Burrows 8209	2003	J
Asparagoideae	*Asparagus*	*capensis*	L.	Burrows & Burrows 14691	2015	J
Asparagoideae	*Asparagus*	*concinnus*	(Baker) Kies	Burrows & Burrows 16277	2020	J
Asparagoideae	*Asparagus*	*concinnus*	(Baker) Kies	Burrows & Burrows 15358	2019	J
Asparagoideae	*Asparagus*	*dauricus*	Fisch. ex Link	P.C. Bentz 40	2021	GA
Asparagoideae	*Asparagus*	*divaricatus*	(Oberm.) Fellingham & N.L.Mey.	Crouch 1088	2008	J
Asparagoideae	*Asparagus*	*exuvialis*	Burch.	Burrows & Burrows 15913	2019	J
Asparagoideae	*Asparagus*	*gobicus*	N.A.Ivanova ex Grubov	P.C. Bentz 75	2023	GA
Asparagoideae	*Asparagus*	*intricatus*	(Oberm.) Fellingham & N.L.Mey.	Burrows & Burrows 15593	2019	J
Asparagoideae	*Asparagus*	*officinalis*	L.	P.C. Bentz 81	2023	GA
Asparagoideae	*Asparagus*	*officinalis*	L.	P.C. Bentz 19	2021	GA
Asparagoideae	*Asparagus*	*schoberioides*	Kunth	P.C. Bentz 13	2021	GA
Asparagoideae	*Asparagus*	*setaceus*	(Kunth) Jessop	Burrows & Burrows 16255	2020	J
Asparagoideae	*Asparagus*	*suaveolens*	Burch.	Burrows & Burrows 16446	2021	J
Asparagoideae	*Hemiphylacus*	*hintoniorum*	L.Hern.	P.C. Bentz 55	2021	GA
Brodiaeoideae	*Dichelostemma*	*congestum*	(Sm.) Kunth	D. Hee 08	1998	GA
Brodiaeoideae	*Triteleia*	*laxa*	Benth.	R. Halse 8737	2013	GA
Lomandroideae	*Acanthocarpus*	sp.	Genus: Lehm.	K. Thiele s.n.	2008	–
Lomandroideae	*Dichopogon*	*preissii*	(Endl.) Brittan	K. Thiele RRT3701	2008	–
Lomandroideae	*Eustrephus*	*latifolius*	R.Br.	W. Zomlefer 2326	2011	GA
Lomandroideae	*Thysanotus*	sp.	Genus: R.Br.	K. Thiele s.n.	2008	–
Nolinoideae	*Aspidistra*	*elatior*	Blume	P.C. Bentz 130	2023	GA
Nolinoideae	*Liriope*	*muscari*	(Decne.) L.H.Bailey	P.C. Bentz 129	2023	GA
Nolinoideae	*Ophiopogon*	*japonicus*	(Thunb.) Ker Gawl.	P.C. Bentz 128	2023	GA
Scilloideae	*Albuca*	cf. *recurva*	(Oberm) J.C.Manning & Goldblatt	C.C. Howard 205	2023	GA
Scilloideae	*Bowiea*	*volubilis*	Harv. ex T.Moore & Mast.	C.C. Howard 156	2023	GA
Scilloideae	*Dipcadi*	sp.	Genus: Medik.	C.C. Howard 222	2023	GA
Scilloideae	*Drimia*	*intricata*	(Baker) J.C.Manning & Goldblatt	C.C. Howard HBG90233	2023	HBG
Scilloideae	*Drimiopsis*	*botryoides* subsp. *botryoides*	Baker	C.C. Howard 154	2023	GA
Scilloideae	*Ledebouria*	*scabrida*	Jessop	C.C. Howard s.n.	2023	GA
Asphodelaceae family (outgroup)	*Dianella*	*tasmanica*	Hook.f.	T. Givnish s.n.	2008	–

[a]Collection (Coll.) numbers are associated with each voucher specimen located at various herbaria as indicated by the herbarium code in accordance with Index Herbariorum (https://sweetgum.nybg.org/science/ih/). Year indicates time of collection. Taxon names correspond to those accepted by POWO (2023).

estimate that is statistically consistent under the multi-species coalescent (MSC) model. Before running ASTRAL, branches in gene trees with <10% BS support were collapsed to improve the accuracy of species tree inference (Zhang et al., 2018). The resulting unrooted MSC species tree was rooted with *Dianella tasmanica* as the outgroup (Table 2).

Lastly, we used all paralogs recovered and assembled by HybPiper in a second phylogenomic analysis with ASTRAL-Pro2 v.1.15.2.4, which is an extention of ASTRAL that accounts for data sets with multiple gene copies for one or more taxa represented in a gene tree, thus accounting for paralogs and orthologs in a data set (Zhang et al., 2020; Zhang and Mirarab, 2022). MSAs and multi-copy gene trees were estimated using the methods described above. Due to the nature of paralog detection performed by HybPiper (Johnson et al., 2016), only trees based on concatenated exon sequence were used for the ASTRAL-Pro2 analysis. MSC-based, rather than concatenation, methods were used to analyze nuclear genes because concatenation implicitly assumes an absence of recombination and ILS (i.e., a single history is shared among all genes) and can result in statistical inconsistencies in multi-locus data sets with high levels of gene tree discordance (Kubatko and Degnan, 2007; Edwards et al., 2016); whereas coalescent approaches assume free recombination between genes, accounting for gene tree discordance due to ILS (Zhang et al., 2018).

Mining plastome sequences from off-target reads for phylogenetic analysis

Using the filtered Hyb-Seq reads, we de novo assembled plastomes using GetOrganelle v.1.7.5.2 (Jin et al., 2020) with default parameters for plastome assembly and annotated with Geneious Prime v.2021.2.2 (http://www.geneious.com) based on 80% sequence similarity of annotations from the following GenBank reference plastomes: *Asparagus officinalis* (NC_034777.1), *Yucca filamentosa* (NC_032712.1), or *Eustrephus latifolius* (NC_025305.1). Allowing for variation in plastome size, we arbitrarily defined plastomes as complete when ≥90% of the total reference plastome length assembled in a sample, using the reference NC_034777, which exhibits a plastome 156,699 nucleotides long. We then tested for differences in plastome sequencing coverage/depth between plastomic reads mined from off-target Hyb-Seq reads with reads obtained from typical low-coverage, whole genome sequencing (WGS) (i.e., genome skimming) experiments, by mapping reads from both sources to an *A. officinalis* reference plastome (NC_034777) using BWA-MEM and the SAMtools (v.1.16) coverage function (Li et al., 2009). After filtering alignments that aligned to >10 plastome regions, we compared mapping results across samples for which we had genome skimming data and off-target Hyb-Seq reads. For the samples with off-target Hyb-Seq reads, we compared those that yielded full plastome assemblies with those that only yielded partial plastome assemblies. Prior to mapping, sequencing adapters, PCR duplicates, and reads shorter than 21 nucleotides were removed from the WGS reads using fastp, then down-sampled to resemble low-coverage levels typical of genome skimming experiments (~10× average coverage) using BBTools reformat.sh (Bushnell, 2018). Read mapping depth was plotted using karyoploteR (Gel and Serra, 2017).

To test for phylogenetic utility of plastome sequences mined from target-enriched libraries, first we separately aligned all plastome gene sequences using the same methods described above for nuclear genes. We then concatenated all MSAs into a supermatrix with defined gene partitions using SequenceMatrix v.1.9 (Vaidya et al., 2011) and analyzed the matrix using ML with IQ-TREE v.1.6.12 and 1000 ultrafast BS. Genes were partitioned and allowed to evolve at independent rates and substitution models, as determined by ModelFinder in IQ-TREE (with option *MFP + MERGE*). To assess the impact of missing data across samples, we performed two independent IQ-TREE analyses: one including samples with an arbitrary minimum gene cutoff of >5 and another >30. The resulting ML trees were compared to trees inferred with ASTRAL under assumptions of the MSC model. We performed a concatenated analysis of plastome genes because they share a single history and do not recombine in a phylogenetically relevant way (Doyle, 2022), both of which violate assumptions of the MSC model (Edwards et al., 2016).

RESULTS

Asparagaceae1726 probe design

Following the Orthofinder run and subsequent filtering steps, 1587 loci remained as candidate low-copy loci for the final probe design. A local BLAST search comparing these candidate loci with gene models from *A. officinalis* and *Y. aloifolia* showed an average sequence similarity of 96.2% (range: 71.4–100%) and 86.6% (range: 70.2–100%), respectively (Figure 1). Of these 1587 candidate loci, 67 were found as orthologous to Angiosperms353 targets, then another 139 Angiosperms353 orthologs from *A. officinalis* were added to the candidate list based on a separate BLAST search. After additional filtering performed by the Arbor Biosciences team, the final probe set was constructed from 1726 loci (8743 exons), of which 206 overlap with Angiosperms353. The raw probe set consisted of 40,116 probes, and the filtered/final set was composed of 38,581 probes. A list of all loci used to construct the final probes and their corresponding orthologs from related genomes is provided in Appendix S4, and a list of the Angiosperms353 orthologs included in Asparagaceae1726 is provided in Appendix S5.

DNA isolation, library preparation, probe hybridization, and sequencing

DNA isolations yielded >200 ng of genomic DNA from each sample. Mean library fragment sizes ranged from

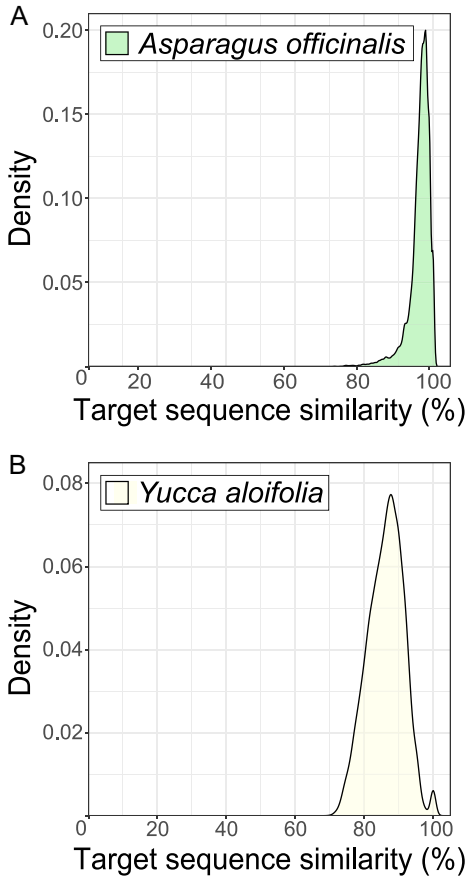

FIGURE 1 Histograms of percent sequence similarity between exons used in Asparagaceae1726 probe design and protein-coding sequences from genome assemblies for (A) *Asparagus officinalis* (Harkess et al., 2017) and (B) *Yucca aloifolia* (DOE-JGI, 2023). Similarity is based on percent identity from a local BLAST search. Histogram bin size = 100.

355–716 bp, and size-corrected concentrations ranged from 36.5–196.4 nM before normalization. The total number of trimmed reads for all samples ranged from 6,401,892 to 189,277,592 (mean = 43,347,135; median = 36,570,137) (Table 3). The percentage of on-target reads was variable within and between Asparagaceae subfamilies, overall ranging from 8.4% to 61.7% (mean = 29.5%; median = 27.2%) (Figure 2B). Percent on-target reads is calculated based on the fraction of reads that map to the reference target sequences. Asparagoideae samples had the highest percentage of on-target reads, with a mean of 43.2%; the next highest mean recovery rate was 31.4%, from Lomandroideae. The remaining subfamilies resulted in a mean of <20% of on-target reads: Agavoideae = 19.1%, Brodiaeoideae = 17.5%, Nolinoideae = 14.3%, and Scilloideae = 11.3% (see Table 3 for additional details and Appendix S6 for all HybPiper output statistics). Across all samples, the mean read coverage and sequencing depth per captured target was approximately 77% (avg. median = 81%) and 389× (avg. median = 271×), respectively. The approximate mean read coverage and sequencing depth per captured targets for each subfamily was: Agavoideae = 85% and 261×; Asparagoideae = 93% and 701×; Brodiaeoideae = 57% and 66×; Lomandroideae = 58% and 225×; Nolinoideae = 69% and 62×; and Scilloideae = 60% and 63×, respectively. The outgroup *Dianella* sample yielded a mean of 21.7% on-target reads, as well as 54% mean read coverage and 173× sequencing depth per captured target.

Target recovery

The *Asparagus capensis* (Asparagoideae) sample yielded the most assembled targets, recovering 100% of targets (Table 3). The *Drimiopsis* Lindl. & Paxton and *Ledebouria* Roth samples (both from Scilloideae) recovered fewer targets than did the outgroup, with 1416 and 1454 (82% and 84% of total targets), respectively, compared to 1538 targets recovered in the Asphodelaceae sample (Table 3). We tested whether target recovery could be increased in these samples by re-running HybPiper with added ortholog reference sequences from the *Dipcadi* Medik. sample, which recovered the most targets (1545) out of all Scilloideae samples (Table 3). This resulted in slightly increased target recovery in both *Drimiopsis* and *Ledebouria* samples, with an additional 25 and 28 targets recovered for each, respectively; their percent of on-target reads also increased to 9.2% and 11.4%, respectively. All other relevant target recovery and assembly statistics for all samples can be found in Table 3.

Paralog recovery

Using the "length" detection criterion of HybPiper, 502 target loci yielded at least one paralog assembly across all samples (Appendix S7). *Hemiphylacus hintoniorum* L.Hern. (Asparagoideae) yielded the most paralog assemblies (225 total) followed by three Agavoideae samples: *Behnia* Didrichsen (185), *Chlorophytum* Ker Gawl. (166), and *Agave* (163) (Table 3). The remaining samples yielded <100 target genes with paralogs, including three samples with <10: *Asparagus officinalis* (coll. P.C. Bentz 81; Asparagoideae) (9), *Dichelostemma* Kunth (Brodiaeoideae) (8), and *Drimiopsis* (Scilloideae) (7) (Table 3). On average, Agavoideae samples yielded the most paralog assemblies (mean = 138), followed by Lomandroideae (mean = 58), whereas all other sampled subfamilies yielded <50 paralogs/sample on average: Asparagoideae (37), Brodiaeoideae (12), Nolinoideae (48), and Scilloideae (12). Without paralogs from *Hemiphylacus* S.Watson, *Asparagus* (Asparagoideae) samples yielded a mean of 25 paralogs. Compared to the above results, approximately 0.5–8× more paralog warnings were reported by HybPiper's "depth" paralog detection scheme (Appendix S6).

Phylogenomic analysis using only putative orthologs (ASTRAL-III)

In the analysis of one-to-one target orthologs, we generated 1726 ML gene trees using IQ-TREE and estimated a species tree using a coalescent-based approach with ASTRAL-III. Summary statistics of multiple sequence alignments, substitution models, and gene tree analyses are provided in Appendix S8 (Figures S1–S4) and Appendix S9. In the ASTRAL-III tree, all branches were strongly supported, with

TABLE 3 Results from Hyb-Seq experiments with Asparagaceae1726 probes in six Asparagaceae subfamilies.

Subfamily	Species	Total reads[a]	On-target reads[b]	Total targets[c]	Mean target cov.[d]	Mean target depth[e]	Paralogs[f]	Plastome length[g]	Mean plastome depth[h]
Agavoideae	*A. virginica*	36,547,346	23%	1673	92%	263×	163	96%	159×
Agavoideae	*B. reticulata*	16,600,390	14%	1576	84%	74×	185	100%	78×
Agavoideae	*C. comosum*	189,277,592	12%	1593	72%	253×	38	9%	782×
Agavoideae	*Y. filamentosa*	49,100,336	27%	1698	94%	452×	166	100%	88×
Asparagoideae	*A. bayeri*	28,003,278	48%	1723	92%	463×	18	19%	111×
Asparagoideae	*A. burchellii*	27,672,596	52%	1723	93%	442×	14	100%	349×
Asparagoideae	*A. capensis*	122,262,910	62%	1726	99%	2425×	32	100%	330×
Asparagoideae	*A. concinnus*[i]	14,969,886	36%	1720	92%	179×	17	81%	85×
Asparagoideae	*A. concinnus*[j]	62,973,884	53%	1722	91%	1233×	17	28%	150×
Asparagoideae	*A. dauricus*	19,432,486	16%	1685	89%	106×	13	100%	104×
Asparagoideae	*A. divaricatus*	111,657,624	62%	1724	99%	2225×	37	100%	251×
Asparagoideae	*A. exuvialis*	70,563,140	59%	1722	96%	1118×	16	26%	79×
Asparagoideae	*A. gobicus*	19,178,694	30%	1712	91%	142×	97	100%	510×
Asparagoideae	*A. intricatus*	64,343,788	45%	1725	93%	1209×	41	17%	134×
Asparagoideae	*A. officinalis*[k]	14,242,278	33%	1717	92%	160×	9	100%	791×
Asparagoideae	*A. officinalis*[l]	8,774,448	32%	1715	90%	97×	11	100%	576×
Asparagoideae	*A. schoberioides*	28,269,132	33%	1692	91%	328×	12	10%	19×
Asparagoideae	*A. setaceus*	6,401,892	40%	1718	93%	103×	17	80%	47×
Asparagoideae	*A. suaveolens*	12,701,300	49%	1721	90%	191×	19	67%	115×
Asparagoideae	*H. hintoniorum*	69,794,294	43%	1710	91%	787×	225	100%	472×
Brodiaeoideae	*D. congestum*	22,122,098	21%	1508	54%	64×	8	82%	33×
Brodiaeoideae	*T. laxa*	39,880,246	14%	1501	60%	68×	16	97%	74×
Lomandroideae	*Acanthocarpus*	36,592,928	28%	1551	54%	180×	17	80%	600×
Lomandroideae	*D. preissii*	52,454,994	36%	1604	61%	329×	152	80%	1166×
Lomandroideae	*E. latifolius*	34,217,530	38%	1606	62%	222×	15	15%	48×
Lomandroideae	*Thysanotus*	42,000,920	25%	1468	53%	170×	47	73%	390×
Nolinoideae	*A. elatior*	38,730,368	9%	1521	66%	34×	32	100%	39×
Nolinoideae	*L. muscari*	24,566,332	18%	1596	69%	81×	50	100%	62×
Nolinoideae	*O. japonicus*	26,316,166	15%	1581	72%	72×	61	100%	39×
Scilloideae	*A.* cf. *recurva*	38,329,100	12%	1534	61%	79×	20	100%	216×
Scilloideae	*B. volubilis*	40,672,914	10%	1489	58%	48×	10	99%	70×
Scilloideae	*Dipcadi*	38,519,436	13%	1545	62%	94×	20	100%	254×
Scilloideae	*D. intricata*	42,873,774	13%	1532	61%	83×	15	100%	151×

TABLE 3 (Continued)

Subfamily	Species	Total reads[a]	On-target reads[b]	Total targets[c]	Mean target cov.[d]	Mean target depth[e]	Paralogs[f]	Plastome length[g]	Mean plastome depth[h]
Scilloideae	*D. botryoides*	32,256,226	8%	1416	58%	33×	7	100%	107×
Scilloideae	*L. scabrida*	24,569,116	11%	1454	60%	39×	13	100%	171×
Asphodelaceae family (outgroup)	*D. tasmanica*	53,627,432	22%	1538	54%	172×	4	100%	373×

[a]Total sequencing reads after adapter trimming and quality filtering.
[b]Percentage of reads that mapped to targets.
[c]Total number of target sequences in which coding sequence successfully assembled.
[d]Mean percentage per target sequence covered by mapped reads.
[e]Mean read mapping depth per target sequence.
[f]Total number of paralogs recovered per sample.
[g]Percentage of total plastome length recovered from off-target reads, based on the reference plastome length for *Asparagus officinalis* (GenBank: NC_034777.1).
[h]Mean read mapping depth for each of the 82 protein-coding genes from the reference NC_034777.
[i]Collection number (coll.) Burrows & Burrows 16277. (See Table 1 to match coll. with samples.)
[j]Coll. Burrows & Burrows 15358.
[k]Coll. P.C. Bentz 81.
[l]Coll. P.C. Bentz 19.

local posterior probabilities (LPP) equal to 1.0, except for one branch representing the most recent common ancestor of *A. setaceus* (LPP = 0.60) (Figure 3A). Furthermore, comparisons of quartet frequencies supporting the main topology (*q1*) versus the two possible alternatives (*q2* and *q3*) revealed minimal gene tree discordance for branches supporting monophyly of each subfamily, each exhibiting >70% support for *q1* (Figure 3A). Although a Lomandroideae + Nolinoideae + Asparagoideae clade was highly supported by LPP, *q1* for this clade was <0.50 (*q1* = 0.46), with skewed quartet frequencies for alternative topologies (*q2* = 0.20; *q3* = 0.34) (Figure 3A). Furthermore, *q1* support for a Lomandroideae + Asparagoideae clade sister to Nolinoideae was <0.50 (*q1* = 0.40) but exhibited less variance for alternative resolutions (*q2* = 0.26; *q3* = 0.34). The Agavoideae + Brodiaeoideae + Scilloideae clade was supported with a *q1* frequency of 0.59 (*q2* = 0.21; *q3* = 0.20) (Figure 3A), while intersubfamilial relationships within that clade exhibited even higher *q1* frequencies: Brodiaeoideae + Scilloideae clade (*q1* = 0.83) sister to Agavoideae (Figure 3A). Quartet frequencies <0.50 supporting *q1* were also observed within the two most-sampled subfamilies: one bifurcation within Scilloideae and several in *Asparagus* (Asparagoideae) (Figure 3A). According to a polytomy test with ASTRAL-III, we can reject the possibility of polytomies at all of these nodes (*P* value = 0) except for the stem node of *A. setaceus* (*P* value = 0.4).

Phylogenomic analysis with paralog assemblies (ASTRAL-Pro2)

All assembled orthologs and paralogs were included in multi-copy gene trees and analysis with ASTRAL-Pro2. The resulting species tree resulted in strong support (LPP = 1) for all branches except the stem leading to *A. setaceus* (LPP = 0.64) (Figure S8 in Appendix S8). Aside from this branch, the analysis with ASTRAL-Pro2 (Figure S8 in Appendix S8) resulted in an identical topology to that inferred with ASTRAL-III (Figure 3A), which did not incorporate paralogs. Like in the ASTRAL-III analysis, >70% of quartets supported *q1* and the monophyly of each subfamily (Figure S8 in Appendix S8), although frequencies varied between analyses. Notably, in the ASTRAL-Pro2 tree, the Asparagoideae + Lomandroideae + Nolinoideae clade had the same quartet support for *q1* (=0.46) but even more skewed support for alternative topologies (*q2* = 0.19; *q3* = 0.35) (Figure S8 in Appendix S8) compared to the ASTRAL-III tree (Figure 3A). We also observed increased quartet support for *q1*, along with increasingly skewed support for alternatives for a Lomandroideae + Asparagoideae clade sister to Nolinoideae (*q1* = 0.43; *q2* = 0.35; *q3* = 0.22) (Figure S8 in Appendix S8). Skewed support for *q2* and *q3* was also evident in the resolution of an *Asparagus gobicus* + *Asparagus dauricus* clade (*q1* = 0.56; *q2* = 0.11; *q3* = 0.33) as well as a *Behnia reticulata* + *Yucca filamentosa* + *Agave virginica* clade (*q1* = 0.51; *q2* = 0.38; *q3* = 0.11) (Figure S8 in Appendix S8).

Mining plastomic sequences from Hyb-Seq reads

Out of 36 total samples, we generated 21 complete plastome assemblies from mined off-target Hyb-Seq reads (Table 3), while the remaining samples yielded plastome assemblies representing 15–89% (mean = 84%; median = 100%) of the reference plastome length (Table 3, Figure S5 in Appendix S8). Plastome read mapping depth across the reference protein-coding genes varied between 19–1166× (mean =

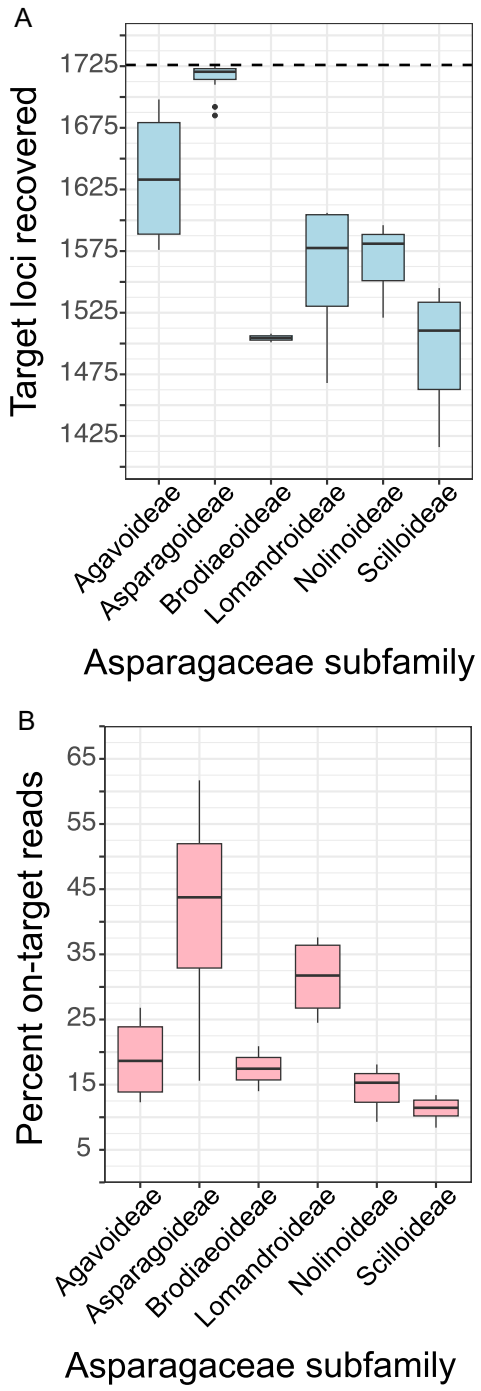

FIGURE 2 Boxplots showing (A) total number of Asparagaceae1726 target loci recovered and (B) percent of on-target reads (i.e., those that mapped to reference target sequences) summarized for each subfamily tested in this study. A minimum of 1416 targets (~82%) were recovered in all samples (black dotted line in A represents maximum targets: 1726). Asparagoideae samples yielded the highest percentage of recovered targets and on-target reads, whereas Scilloideae samples yielded the least in both categories.

251×; median = 142×) per sample (Table 3, Figure S6 in Appendix S8). Three samples yielded <5 plastome genes; seven samples yielded <10, including *Asparagus schoberioides* Kunth (coll. P.C. Bentz 13), which yielded six genes (Figure 4B); and 10 samples resulted in <30 genes (Appendix S10). In contrast to the highly variable plastome read coverage from mined off-target reads for *A. schoberioides*, WGS skimming data from the same sample yielded relatively uniform read coverage when mapped to the reference plastome (Figure 4A), as was also seen in a similar comparison with a sample that yielded complete plastomes from both data types (i.e., *Asparagus gobicus* N.A.Ivanova ex Grubov) (Figure S7 in Appendix S8). *Triteleia laxa* Benth. recovered the most plastid genes, with 104 annotations, but exhibited uneven yet overall deeper read coverage across the plastome (Figure 4C).

Plastome phylogenetics

We used mined plastomic gene sequences in a partitioned ML analysis of concatenated alignments using IQ-TREE. In the analysis excluding samples that yielded <30 plastome genes, we found strong support (BS = 100%) for the monophyly of each subfamily (Figure 3B) and two major clades in the family, each encompassing at least three subfamilies (not including Aphyllanthoideae, which was not sampled here): (i) an Agavoideae + Brodiaeoideae + Scilloideae clade and (ii) an Asparagoideae + Nolinoideae + Lomandroideae clade (Figure 3B). This was also the case in a similar analysis only excluding samples with <5 plastome genes recovered (Figure S9 in Appendix S8). Across both ML analyses, all branches had 100% BS support, except for several weakly supported branches among *Asparagus* samples (Figure 3B, Figure S9 in Appendix S8). In contrast to both ASTRAL trees, the plastome trees strongly support a Nolinoideae + Asparagoideae clade sister to Lomandroideae (Figure 3B, Figure S9 in Appendix S8).

DISCUSSION

Hyb-Seq with Asparagaceae1726

In this study, we describe the development of the Asparagaceae1726 probe set and report test results from Hyb-Seq trials performed on six of the seven subfamilies of Asparagaceae. We show that these probes can effectively generate robust data sets for phylogenomic analysis in and across all of these Asparagaceae subfamilies. A minimum of 1416 target loci were captured in all samples (Table 3), suggesting that at least that number of targets are recoverable across all Asparagaceae lineages. We suspected that the lower capture efficiency seen in some subfamilies may be partially due to including only *Asparagus* and *Yucca* L. sequences as reference targets for read mapping with BWA and HybPiper. We tested this by adding orthologs from Scilloideae to the reference target sequence file and found an increase in targets recovered in lesser-performing Scilloideae samples, increasing the minimum number of recovered targets from 1416 to 1441 loci. Although an

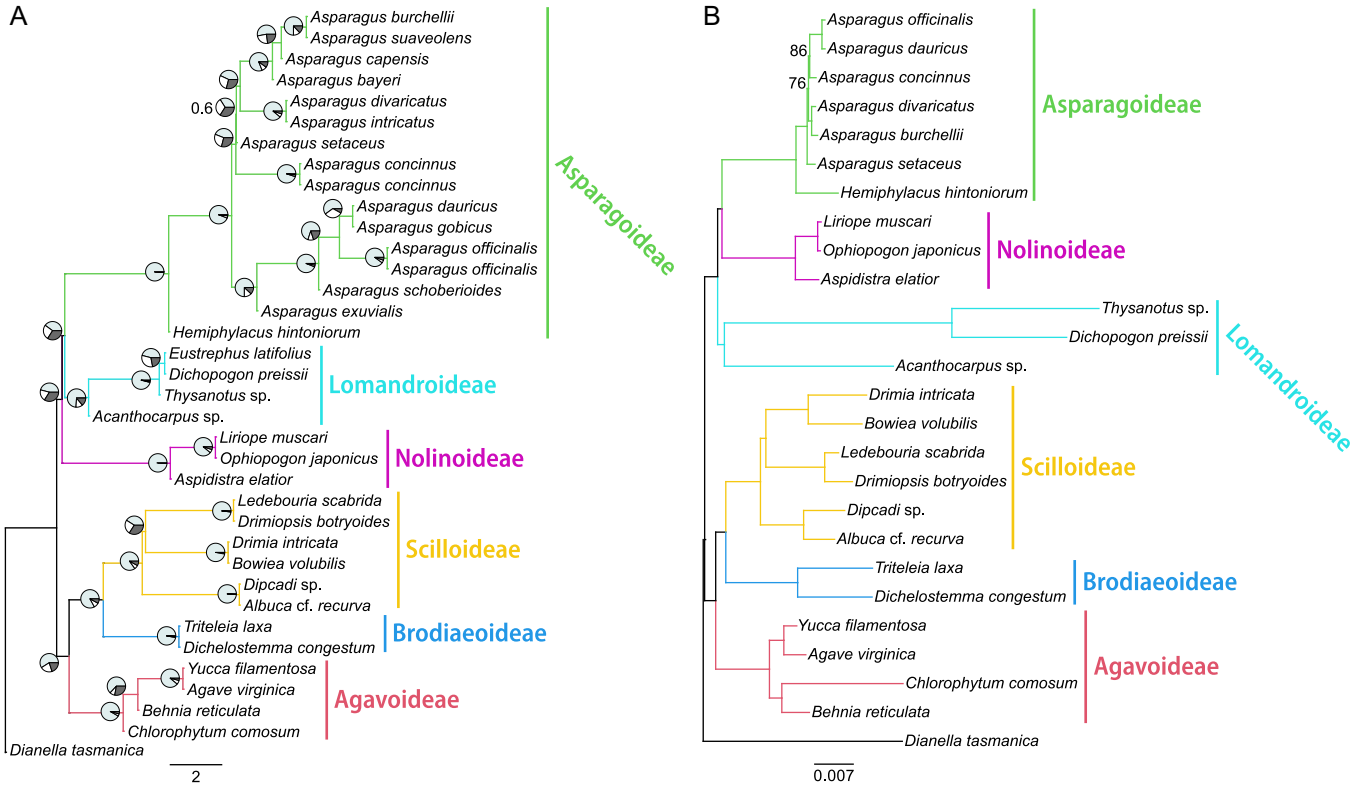

FIGURE 3 Contrasting tree topologies inferred from different data sets generated from Asparagaceae1726 Hyb-Seq reads: (A) multi-species coalescent summary species tree inferred from 1726 orthologs with ASTRAL-III (Zhang et al., 2018); (B) maximum likelihood tree inferred from concatenated alignments of plastid genes mined from off-target Hyb-Seq reads, only including samples with >30 genes recovered. Notably, the coalescent-based species tree (A) shows strong support for Asparagoideae + Lomandroideae sister to Nolinoideae, whereas the plastome tree (B) shows strong support for Asparagoideae + Nolinoideae sister to Lomandroideae. Pies in (A) represent gene tree quartet frequencies supporting the main topology (light blue) as illustrated here and two alternatives (white and gray). (A) Branch support = local posterior probability shown only when <1.0; branch lengths = coalescent units. (B) Total plastid genes per sample ranged from 32–107; branch support = ultrafast bootstrap approximations from IQ-TREE (Nguyen et al., 2015) shown only when <100; branch lengths = substitution rates.

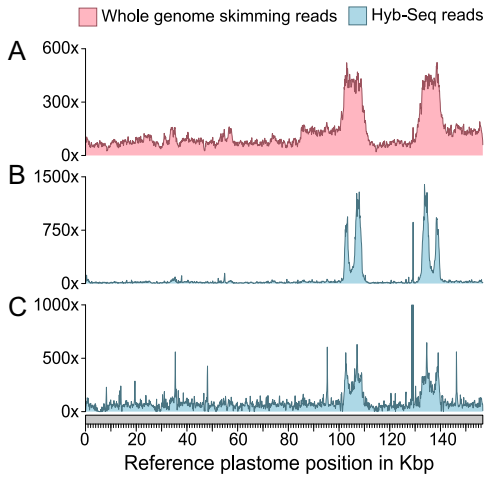

FIGURE 4 Plastome read mapping depth compared between low-coverage whole genome sequencing (i.e., genome skimming) (A; pink) and Hyb-Seq reads (B and C; blue), which yielded either complete (A and C) or incomplete (B) plastome assemblies. A and B are reads from independent sequencing runs for an individual sample of *Asparagus schoberioides* (coll. P.C. Bentz 13). C is from the sample *Tritelia laxa* (coll. Halse 8737), which yielded the most recovered plastome genes in this study. Reference plastome = *Asparagus officinalis* (NCBI GenBank: NC_034777.1). Plots show read mapping depth in 100-nucleotide sliding windows.

increase of 25 targets may not be significant in an analysis of >1700 loci, these results indicate that there are likely "uncaptured" target orthologs in the Hyb-Seq reads that did not map sufficiently well to reference sequences to enable locus assembly. Additionally, lower target read coverage for some samples (e.g., ~60% in Scilloideae samples; Table 3) could have contributed to lower assembly numbers; therefore, read coverage, and thus target assemblies, could be extended for some targets if more similar orthologs were included in the reference sequence file, as shown in previous studies using universal bait sets (McLay et al., 2021). As reference genome assemblies become available for each of the Asparagaceae subfamilies, they could be used to improve the HybPiper target file. Even without additional reference genomes, as more Hyb-Seq experiments are conducted with Asparagaceae1726 and different taxa, new target orthologs should be added to read mapping analyses to increase target recovery. This includes those from this study, which represent orthologs from six of the seven subfamilies of Asparagaceae. Asparagaceae1726 probes should also be tested in the phylogenetically elusive *Aphyllanthes monspeliensis* L.—the monotypic taxon of Aphyllanthoideae—to better understand its relationship among the other six subfamilies of Asparagaceae.

Asparagaceae1726 probes were constructed using *Asparagus* sequences, and therefore the highest capture and on-target efficiency was expected in Asparagoideae compared to the other subfamilies (Figure 2B). We also predicted high target recovery rates in Agavoideae samples because we verified presence and sequence similarity of Asparagaceae1726 targets in *Yucca* (Figure 1B). Outside of Asparagoideae, the *Yucca filamentosa* L. sample yielded the most recovered targets with 1698 (Table 3), which was likely influenced by the presence of *Yucca* sequences in the reference target file used for read mapping. Lomandroideae samples yielded the second-highest on-target efficiency (averaging ~32% on-target reads), resulting in 1468–1606 recovered targets, which are comparable results to the other subfamilies tested in this study (Table 3). Increased capture efficiency in Lomandroideae makes sense due to its close relationship to Asparagoideae (Figure 3A) and the overwhelming presence of *Asparagus* sequences in the reference target file.

Phylogenomics with Asparagaceae1726

Our coalescent-based analyses with ASTRAL showed limited effects of paralogs present in Asparagaceae1726 data sets, as evident by nearly 100% topological congruence between the species tree analysis inferred from strict orthologs (Figure 3A) and the other incorporating paralogs (Figure S8 in Appendix S8). The only incongruency between our two ASTRAL analyses involved the poorly supported placement of *A. setaceus* in both (Figure 3A, Figure S8 in Appendix S8). All phylogenomic analyses in this study showed strong support for two major clades in Asparagaceae, each consisting of at least three subfamilies (Aphyllanthoideae not sampled here): (i) an Agavoideae + Brodiaeoideae + Scilloideae clade and (ii) an Asparagoideae + Lomandroideae + Nolinoideae clade (Figure 3; Figures S8, S9 in Appendix S8). These results support previous findings based on plastome sequences (Steele et al., 2012; Chen et al., 2013; Lu et al., 2022; Ji et al., 2023; Bentz et al., 2024) and hundreds of nuclear genes (One Thousand Plant Transcriptomes Initiative, 2019; Timilsena et al., 2022). However, in those previous plastome trees, Lomandroideae is placed sister to a Nolinoideae + Asparagoideae clade with strong support, which agrees with our plastome phylogeny (Figure 3B) but disagrees with our ASTRAL trees inferred from 1726 nuclear genes (Figure 3A). Our ASTRAL trees show strong support (LPP = 1) for Nolinoideae sister to a Lomandroideae + Asparagoideae clade, with at least 40% support from gene tree quartets in both (Figure 3A, Figure S8 in Appendix S8). Concatenation-based analyses of 410 nuclear genes by the OneKP Initiative (i.e., One Thousand Plant Transcriptomes Initiative, 2019, see their supplementary figure 2) and 602 nuclear genes by Timilsena et al. (2022) also strongly support the relationships inferred in our ASTRAL analyses (Figure 3A, Figure S8 in Appendix S8). However, ASTRAL analysis of the same genes in both of those studies either supported the plastome topology shown in Figure 3B (Timilsena et al., 2022) or could not resolve relationships among Asparagoideae, Lomandroideae, and Nolinoideae (supplementary figure 1 in One Thousand Plant Transcriptomes Initiative, 2019). The incongruence of ASTRAL results between this and previous studies (i.e., One Thousand Plant Transcriptomes Initiative, 2019; Timilsena et al., 2022) could be explained by limited sampling of Lomandroideae in the earlier analyses. Whereas our study included four Lomandroideae samples, the previous studies included only one representative for the subfamily. Additionally, ASTRAL analyses in this study were based on over three times as many nuclear loci as in the previous studies, providing more robust tree inference.

Regardless, the extensive gene tree discordance in our ASTRAL analyses, as evident by the skewed or otherwise mixed quartet frequencies supporting alternative resolutions across the species trees (Figure 3A), is an indication of the complex evolutionary history of Asparagaceae. For example, alternative quartet frequencies supporting resolution of Nolinoideae as sister to Lomandroideae + Asparagoideae in both coalescent-based trees (Figure 3A, Figure S8 in Appendix S8) were relatively skewed and may indicate historical introgression (Pease et al., 2018). Along with the short branches across the backbone of Asparagaceae, quartet frequency profiles seen in both ASTRAL trees (Figure 3A, Figure S8 in Appendix S8) support previous studies positing that a rapid radiation of lineages (likely accompanied by ILS and introgression) marked the origin and early divergence of extant Asparagaceae subfamilies (Steele et al., 2012). The discordance between ASTRAL and plastome tree relationships among Asparagoideae, Lomandroideae, and Nolinoideae may be explained by chloroplast capture and/or genome-wide introgression of nuclear loci, or by ILS alone. Regardless, whereas the ASTRAL trees are estimates of species relationships, the plastome tree is an estimate of the history of the plastid genome (i.e., comparable to any single nuclear gene tree), not species relationships (Doyle, 2022).

To test for inter- and intra-subfamilial relationships, as well as the many interesting evolutionary patterns in Asparagaceae, a much denser sampling scheme is required. Given a sufficient taxon sampling scheme, our analyses demonstrate that Asparagaceae1726 will enable robust phylogenomic analyses of the family. The Asparagaceae1726 probe set represents a toolkit that can be used to explore the dynamic evolution of the family and to facilitate better understanding of where rapid radiations have occurred across the tree, as well as the roles that ILS, hybridization, and genome duplications may have played in the origin and evolution of Asparagaceae lineages.

Paralogs and their utility in phylogenomics

In our data set, we recovered at least one paralog for 502 target loci. Overall, Agavoideae samples yielded the most of

such paralogs, with a mean of 138 per sample. The next highest number of paralogs was found in Lomandroideae, which averaged 58 paralogs per sample, although most of those resulted from a single sample (i.e., *Dichopogon preissii* (Endl.) Brittan yielded 152 paralogs). Interestingly, the *Dianella* (Asphodelaceae) outgroup species yielded the fewest paralogs with four total (Table 3), followed by Brodiaeoideae and Scilloideae, both of which averaged 12 paralogs per sample. Asparagoideae averaged 37 paralogs per sample, with even fewer found in *Asparagus* (24 on average). The relatively small proportions of target paralogs found in Asparagoideae, Brodiaeoideae, Lomandroideae, Nolinoideae, and Scilloideae suggests that paralogy complications pose little threat to phylogenomic analyses in these groups. Even in Agavoideae samples, only approximately 8% of the 1726 possible targets yielded paralogs. Based on congruent results between ASTRAL analyses with and without paralogs, the paralogs present in Asparagaceae1726 data sets seem to pose little threat to robust species tree inference; however, paralogs should always be evaluated and considered for phylogenomic analysis, especially in Agavoideae taxa.

The presence of multi-copy genes in this data set, and others generated from Asparagaceae1726, is not surprising given the multiple WGD events across the Asparagaceae phylogeny (Hanson et al., 2003; McKain et al., 2012; Šmarda et al., 2014; Harkess et al., 2017; Gunn et al., 2020; Nath et al., 2023). It is likely that additional paralogs exist in various Asparagaceae lineages; further WGS and comparative genomics work across the family is likely to yield new insights into the impacts of polyploidization and genome fractionation. At the same time, MSC summary approaches have been developed to account for orthology–paralogy relationships in multi-copy gene families (e.g., with ASTRAL-Pro2).

In a larger context, paralogs should be expected in most angiosperm phylogenomic data sets composed of hundreds to thousands of nuclear loci, due to the repeated gains and losses associated with polyploidy across the phylogeny. Single-copy loci may be most ideal for phylogenomic analysis using the currently available tools, but given the ubiquity of polyploidy events across the angiosperm phylogeny, it is likely that homoeologous (paralogous) gene lineages have been retained even in single-copy gene families. In these cases, single homologs are seen in complete genome assemblies and annotations (Duarte et al., 2010; One Thousand Plant Transcriptomes Initiative, 2019), as well as loci targeted in Asparagaceae1726 and other targeted gene-capture bait sets (e.g., Johnson et al., 2019). Much like incomplete sorting of allelic variation between speciation events, the persistence of duplicated genes between speciation events can contribute to gene tree–species tree discordance and may mislead gene duplication estimates and species tree estimation when not accounted for (Bansal et al., 2012; Rasmussen and Kellis, 2012).

Mining plastomic sequences from Hyb-Seq reads

Here we show that partial to whole plastomes can be assembled from mining off-target Hyb-Seq reads, which is repeatedly shown in other studies (e.g., Granados Mendoza et al., 2020; Schneider et al., 2021; Baldwin et al., 2023; Bratzel et al., 2023). However, comparisons between plastomic read coverage/depth from separate WGS skimming and Hyb-Seq experiments revealed that the former results in more uniform coverage across different regions of the plastome (Figure 4A). Although mining off-target Hyb-Seq reads can result in completely circularized plastome assemblies, we caution against inference of gene presence or absence based on these data because they often result in partial plastome assemblies (Table 3). Nonetheless, in samples for which whole plastomes cannot be assembled, those successfully assembled genes hold ample phylogenetic information content. The phylogenetic analysis of mined plastome genes in this study (Figure 3B, Figure S9 in Appendix S8) strongly supports the same clade-level relationships found in previous plastome-based studies (Steele et al., 2012; Chen et al., 2013; Lu et al., 2022; Ji et al., 2023; Bentz et al., 2024). The weaker-supported branches found among *Asparagus* samples in our plastome trees are likely explained by the short branches and/or low sequence divergence among samples (Figure 3B), as shown in previous analyses of nearly complete plastomes (Bentz et al., 2024) and few nuclear markers (Norup et al., 2015).

Sequencing recommendations

While testing Asparagaceae1726 in this study, we aimed for 4–5 Gbp of sequence data per sample. This level of sequencing depth is unnecessary for typical Hyb-Seq experiments, thus future studies should aim for less sequencing depth to minimize flow cell space on sequencing runs (allowing space for more samples) and maximize target recovery. Our data indicate that Asparagaceae1726 Hyb-Seq experiments should aim for at least 8× sequencing depth per target, based on the minimum read mapping depth for the top 95% of successfully assembled targets in this study. Solely based on the *Asparagus* reference target sequences (totaling 2,432,740 nucleotides), this equates to ~19.5 Mbp of sequencing data with 100% capture efficiency. When estimating the optimal read depth for recovery of maximum targets from a pool of libraries, researchers should also consider that (1) target orthologs from other taxa may vary in length, (2) capture efficiency will vary across libraries, and (3) increased depth is required to recover paralogs. Unless an updated version of Asparagaceae1726 is developed that incorporates probes with orthologous sequence from additional subfamilies, given the lower capture efficiency observed in samples outside of Asparagoideae, we recommend that researchers favor higher-throughput sequencing platforms (e.g., Illumina NovaSeq) over lower-output platforms like

Illumina MiSeq. The presence of target paralogs for some target loci is unknown in advance, therefore we recommend increasing the sequencing depth for untested taxa to enable phasing of ortholog and paralog assemblies. Assembled paralogs can then be included in phylogenomic analysis, thus adding more information to an increasingly robust data set. Lastly, off-target sequences may be decreased by increasing hybridization temperatures in vitro, forcing stronger probe–target specificity; however, this allows for less probe–target sequence divergence and should be tested in taxa outside of Asparagoideae.

Future directions

The Asparagaceae1726 probe set described here is a first attempt at establishing a standardized set of loci for phylogenomic analysis in Asparagaceae. We imagine multiple future iterations aimed to improve the consistency of these probes across all Asparagaceae lineages. We invite researchers to explore alternative loci or expand upon this probe set. For example, to improve capture efficiency of these probes in taxa outside of Asparagoideae, target assemblies from this study could be used to design orthologous probes with increased sequence similarity to more distant relatives. Either way, target orthologs generated in this and future studies should be included in an updated target reference sequence file for read mapping in taxa that are not currently represented. As more genomic resources are generated for Asparagaceae lineages, gene copy numbers for these targets could be re-tested and filtered or accounted for accordingly. As taxon sampling across the family increases, meta-analyses could assess lineage-specific changes in gene copy number. Nonetheless, we encourage researchers to settle on a standardized set of loci for phylogenomic analysis in this group, which will enable data sharing, consistency across studies, and cross-study reproducibility.

AUTHOR CONTRIBUTIONS

P.C.B. and J.L-M. conceived the study, developed experimental designs, performed early data analysis, and contributed to manuscript edits. P.C.B. conducted experiments, analyzed and curated data, and wrote the manuscript. Both authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank Karolina Heyduk (University of Connecticut) and the Joint Genome Institute (JGI-DOE) for sharing the unpublished genome assembly for *Yucca aloifolia*, and JGI-DOE for sharing *Joinvillea ascendens* and *Acorus americanus* assemblies. We also thank Cody C. Howard (Oklahoma State University) for providing Scilloideae samples, Steven Hughes (University of Georgia Herbarium) for helping sample herbarium tissue, John and Sandra Burrows (Buffelskloof Herbarium) for providing *Asparagus* and *Behnia* samples, and Arbor Biosciences for helping with probe design.

DATA AVAILABILITY STATEMENT

Target sequence files, individual probe sequences, and ordering information for Asparagaceae1726 can be found on GitHub (https://github.com/bentzpc/Asparagaceae1726), along with the target orthologs and paralogs from each sample in this study. Relevant data, results, and scripts from this study are archived on Zenodo (https://doi.org/10.5281/zenodo.10442712). Sequencing reads from this study are available in the NCBI Short Read Archive (https://www.ncbi.nlm.nih.gov/sra) under BioProject PRJNA1034624.

ORCID

Philip C. Bentz http://orcid.org/0000-0003-2232-7488
Jim Leebens-Mack http://orcid.org/0000-0003-4811-2231

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Baldwin, E., M. McNair, and J. Leebens-Mack. 2023. Rampant chloroplast capture in *Sarracenia* revealed by plastome phylogeny. *Frontiers in Plant Science* 14: 1237749.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.

Bansal, K. S., E. J. Alm, and M. Kellis. 2012. Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss. *Bioinformatics* 28: i283–i291.

Bennetzen, J. L., J. Schmutz, H. Wang, R. Percifield, J. Hawkins, A. C. Pontaroli, M. Estep, et al. 2012. Reference genome sequence of the model plant *Setaria*. *Nature Biotechnology* 30: 555–561.

Bentz, P. C., Z. Liu, J.-B. Yang, L. Zhang, S. Burrows, J. Burrows, Akira Kanno, et al. 2024. Young evolutionary origins of dioecy in the genus *Asparagus*. *American Journal of Botany* 111(2): e16276.

Bratzel, F., J. Paule, J. Leebens-Mack, E. M. Leme, R. C. Forzza, M. A. Koch, S. Heller, et al. 2023. Target-enrichment sequencing reveals for the first time a well-resolved phylogeny of the core Bromelioideae (family Bromeliaceae). *Taxon* 72: 47–63.

Bushnell, B. 2018. BBTools: A suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. Website: https://sourceforge.net/projects/bbmap/ [accessed 30 September 2023].

Cai, J., X. Liu, K. Vanneste, S. Proost, W.-C. Tsai, K.-W. Liu, L.-J. Chen, et al. 2015. The genome sequence of the orchid *Phalaenopsis equestris*. *Nature Genetics* 47: 65–72.

Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.

Chase, M. W., J. L. Reveal, and M. F. Fay. 2009. A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. *Botanical Journal of the Linnean Society* 161: 132–136.

Chen, S., D.-K. Kim, M. W. Chase, and J.-H. Kim. 2013. Networks in a large-scale phylogenetic analysis: reconstructing evolutionary history of Asparagales (Lilianae) based on four plastid genes. *PLoS ONE* 8: e59472.

Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34: i884–i890.

D'hont, A., F. Denoeud, J.-M. Aury, F.-C. Baurens, F. Carreel, O. Garsmeur, B. Noel, et al. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488: 213–217.

DOE-JGI. 2020a. *Acorus americanus* v1.1 [online]. Website: http://phytozome-next.jgi.doe.gov/ [accessed 6 October 2020].

DOE-JGI. 2020b. *Joinvillea ascendens* v1.1 [online]. Website: http://phytozome-next.jgi.doe.gov/ [accessed 6 October 2020].

DOE-JGI. 2023. *Yucca aloifolia* Ya24Inoko v2.1 [online]. Website: https://phytozome.jgi.doe.gov/info/YaloifoliaYa24Inoko_v2_1/ [accessed 18 December 2023].

Doyle, J. J. 2022. Defining coalescent genes: Theory meets practice in organelle phylogenomics. *Systematic Biology* 71: 476-489.

Duarte, J. M., P. K. Wall, P. P. Edger, L. L. Landherr, H. Ma, P. K. 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.

Edwards, S. V., Z. Xi, A. Janke, B. C. Faircloth, J. E. McCormack, T. C. Glenn, B. Zhong, et al. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution* 94: 447–462.

Emms, D. M., and S. Kelly. 2015. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16: 115.

Emms, D. M., and S. Kelly. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* 20: 238.

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.

Flores-Abreu, I. N., R. E. Trejo-Salazar, L. L. Sanchez-Reyes, S. V. Good, S. Magallon, A. Garcia-Mendoza, and L. E. Eguiarte. 2019. Tempo and mode in coevolution of *Agave sensu lato* (Agavoideae, Asparagaceae) and its bat pollinators, Glossophaginae (Phyllostomidae). *Molecular Phylogenetics and Evolution* 133: 176–188.

Gándara, E., C. D. Specht, and V. Sosa. 2014. Origin and diversification of the Milla clade (Brodiaeoideae, Asparagaceae): A Neotropical group of six geophytic genera. *Molecular Phylogenetics and Evolution* 75: 118–125.

Gel, B., and E. Serra. 2017. karyoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33: 3088–3090.

Givnish, T. J., A. Zuluaga, D. Spalink, M. Soto Gomez, V. K. Y. Lam, J. M. Saarela, C. Sass, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *American Journal of Botany* 105: 1888–1910.

Glenn, T. C., R. A. Nilsen, T. J. Kieran, J. G. Sanders, N. J. Bayona-Vásquez, J. W. Finger, T. W. Pierson, et al. 2019. Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ* 7: e7755.

Good-Avila, S. V., V. Souza, B. S. Gaut, and L. E. Eguiarte. 2006. Timing and rate of speciation in *Agave* (Agavaceae). *Proceedings of the National Academy of Sciences, USA* 103: 9124–9129.

Granados Mendoza, C., M. Jost, E. Hágsater, S. Magallón, C. van den Berg, E. M. Lemmon, A. R. Lemmon, et al. 2020. Target nuclear and off-target plastid hybrid enrichment data inform a range of evolutionary depths in the orchid genus *Epidendrum*. *Frontiers in Plant Science* 10: 1761.

Grover, C. E., A. Salmon, and J. F. Wendel. 2012. Targeted sequence capture as a powerful tool for evolutionary analysis. *American Journal of Botany* 99: 312–319.

Gunn, B. F., D. J. Murphy, N. G. Walsh, J. G. Conran, J. C. Pires, T. D. Macfarlane, and J. L. Birch. 2020. Evolution of Lomandroideae: Multiple origins of polyploidy and biome occupancy in Australia. *Molecular Phylogenetics and Evolution* 149: 106836.

Gutiérrez, J., T. Terrazas, I. Luna-Vega, and G. A. Salazar. 2017. Phylogenetic analyses of the *Milla* complex (Brodiaeoideae: Asparagaceae), with an emphasis on *Milla*. *Botanical Journal of the Linnean Society* 185: 445–462.

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

Hanson, L., R. L. Brown, A. Boyd, M. A. Johnson, and M. D. Bennett. 2003. First nuclear DNA C-values for 28 angiosperm genera. *Annals of Botany* 91: 31–38.

Harkess, A., J. Zhou, C. Xu, J. E. Bowers, R. van der Hulst, S. Ayyampalayam, F. Mercati, et al. 2017. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nature Communications* 8: 1279.

Harkess, A., K. Huang, R. van der Hulst, B. Tissen, J. L. Caplan, A. Koppula, M. Batish, B. C. Meyers, et al. 2020. Sex determination by two Y-linked genes in garden asparagus. *The Plant Cell* 32: 1790–1796.

Hazzouri, K. M., M. Gros-Balthazard, J. M. Flowers, D. Copetti, A. Lemansour, M. Lebrun, K. Masmoudi, et al. 2019. Genome-wide association mapping of date palm fruit traits. *Nature Communications* 10: 4680.

Heyduk, K., D. W. Trapnell, C. F. Barrett, and J. Leebens-Mack. 2016a. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society* 117: 106–120.

Heyduk, K., M. R. McKain, F. Lalani, and J. Leebens-Mack. 2016b. Evolution of a CAM anatomy predates the origins of Crassulacean acid metabolism in the Agavoideae (Asparagaceae). *Molecular Phylogenetics and Evolution* 105: 102–113.

Heyduk, K., E. V. McAssey, J. Grimwood, S. Shu, J. Schmutz, M. R. McKain, and J. Leebens-Mack. 2021. Hybridization history and repetitive element content in the genome of a homoploid hybrid, *Yucca gloriosa* (Asparagaceae). *Frontiers in Plant Science* 11: 573767.

Howard, C. C., A. A. Crowl, T. S. Harvey, and N. Cellinese. 2022. Peeling back the layers: First phylogenomic insights into the Ledebouriinae (Scilloideae, Asparagaceae). *Molecular Phylogenetics and Evolution* 169: 107430.

Howard, C. C., L. Nanyeni, N. Mollel, D. Chuba, A. R. Zuntini, P. Malakasi, T. S. Harvey, and N. Cellinese. 2023. From southern Africa and beyond: Historical biogeography of a monocotyledonous bulbous geophyte. *Journal of Biogeography* 50: 1623–1638.

Ji, Y., J. B. Landis, J. Yang, S. Wang, N. Zhou, Y. Luo, and H. Liu. 2023. Phylogeny and evolution of Asparagaceae subfamily Nolinoideae: New insights from plastid phylogenomics. *Annals of Botany* 131: 301–312.

Jiménez-Barron, O., R. García-Sandoval, S. Magallón, A. García-Mendoza, J. Nieto-Sotelo, E. Aguirre-Planter, and L. E. Eguiarte. 2020. Phylogeny, diversification rate, and divergence time of *Agave* sensu lato (Asparagaceae), a group of recent origin in the process of diversification. *Frontiers in Plant Science* 11: 536135.

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. 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.

Kubatko, L. S., and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* 56: 17–24.

Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61: 727–744.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [preprint]. Available at: https://doi.org/10.48550/arXiv.1303.3997 [posted 16 March 2013; accessed 8 May 2024].

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25: 2078-2079.

Li, S.-F., J. Wang, R. Dong, H.-W. Zhu, L.-N. Lan, Y.-L. Zhang, N. Li, et al. 2020. Chromosome-level genome assembly, annotation and evolutionary analysis of the ornamental plant *Asparagus setaceus*. *Horticulture Research* 7: 48.

Lu, Q.-X., X. Chang, J. Gao, X. Wu, J. Wu, Z.-C. Qi, R.-H. Wang, et al. 2022. Evolutionary comparison of the complete chloroplast genomes in *Convallaria* species and phylogenetic study of Asparagaceae. *Genes* 13: 1724.

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: 1300085.

McCormick, R. F., S. K. Truong, A. Sreedasyam, J. Jenkins, S. Shu, D. Sims, M. Kennedy, et al. 2018. The *Sorghum bicolor* reference genome: Improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *The Plant Journal* 93: 338–354.

McKain, M. R., N. Wickett, Y. Zhang, S. Ayyampalayam, W. R. McCombie, M. W. Chase, J. C. Pires, et al. 2012. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). *American Journal of Botany* 99: 397–406.

McKain, M. R., J. R. McNeal, P. R. Kellar, L. E. Eguiarte, J. C. Pires, and J. Leebens-Mack. 2016. Timing of rapid diversification and convergent origins of active pollination within Agavoideae (Asparagaceae). *American Journal of Botany* 103: 1717–1729.

McLay, T. G., 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: 11420.

Meng, R., Y. Meng, Y.-P. Yang, and Z.-L. Nie. 2021. Phylogeny and biogeography of *Maianthemum* (Asparagaceae: Nolinoideae) revisited with emphasis on its divergence pattern in SW China. *Plant Diversity* 43: 93–101.

Ming, R., R. VanBuren, C. M. Wai, H. Tang, M. C. Schatz, J. E. Bowers, E. Lyons, et al. 2015. The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics* 47: 1435–1442.

Nath, S., S. Sarkar, S. D. Patil, P. S. Saha, M. M. Lekhak, S. Ray, S. R. Rao, et al. 2023. Cytogenetic diversity in Scilloideae (Asparagaceae): A comprehensive recollection and exploration of karyo-evolutionary trends. *The Botanical Review* 89: 158–200.

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.

Norup, M. F., G. Petersen, S. Burrows, Y. Bouchenak-Khelladi, J. Leebens-Mack, J. C. Pires, H. P. Linder, and O. Seberg. 2015. Evolution of *Asparagus* L. (Asparagaceae): Out-of-South-Africa and multiple origins of sexual dimorphism. *Molecular Phylogenetics and Evolution* 92: 25–44.

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

Ouyang, S., W. Zhu, J. Hamilton, H. Lin, M. Campbell, K. Childs, F. Thibaud-Nissen, et al. 2007. The TIGR rice genome annotation resource: Improvements and new features. *Nucleic Acids Research* 35: D883–D887.

Panchy, N., M. Lehti-Shiu, and S.-H. Shiu. 2016. Evolution of gene duplication in plants. *Plant Physiology* 171: 2294–2316.

Pease, J. B., J. W. Brown, J. F. Walker, C. E. Hinchliff, and S. A. Smith. 2018. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany* 105: 385–403.

POWO. 2023. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Website: https://powo.science.kew.org [accessed 15 October 2023].

Qin, Y. Q., M. H. Zhang, C. Y. Yang, Z. L. Nie, J. Wen, and Y. Meng. 2024. Phylogenomics and divergence pattern of *Polygonatum* (Asparagaceae: Polygonateae) in the north temperate region. *Molecular Phylogenetics and Evolution* 190: 107962.

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.

Rasmussen, D., and M. Kellis. 2012. Unified modeling of gene duplication, loss, and coalescence using a locus tree. *Genome Research* 22: 755–765.

Schneider, J. V., J. Paule, T. Jungcurt, D. Cardoso, A. M. Amorim, T. Berberich, and G. Zizka. 2021. Resolving recalcitrant clades in the pantropical Ochnaceae: Insights from comparative phylogenomics of plastome and nuclear genomic data derived from targeted sequencing. *Frontiers in Plant Science* 12: 638650.

Slater, G., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31.

Šmarda, P., P. Bureš, L. Horová, I. J. Leitch, L. Mucina, E. Pacini, L. Tichý, et al. 2014. Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proceedings of the National Academy of Sciences, USA* 111: E4096–E4102.

Steele, P. R., K. L. Hertweck, D. Mayfield, M. R. McKain, J. Leebens-Mack, and J. C. Pires. 2012. Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* 99: 330–348.

Stevens, P. F. 2001 onwards. Angiosperm Phylogeny Website, Version 14. Website: https://www.mobot.org/mobot/research/apweb/ [accessed 15 October 2023].

Takawira-Nyenya, R., L. Mucina, W. M. Cardinal-Mcteague, and K. R. Thiele. 2018. *Sansevieria* (Asparagaceae, Nolinoideae) is a herbaceous clade within *Dracaena*: Inference from non-coding plastid and nuclear DNA sequence data. *Phytotaxa* 376: 254–276.

Talavera, A., Z. L. Nie, Z. Y. Ma, G. Johnson, S. M. Ickert-Bond, E. A. Zimmer, and J. Wen. 2023. Phylogenomic analyses using a new 1013-gene Vitaceae bait-set support major groups of North American *Vitis*. *Molecular Phylogenetics and Evolution* 186: 107866.

The International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463: 763–768.

Timilsena, P. R., E. K. Wafula, C. F. Barrett, S. Ayyampalayam, J. R. McNeal, J. D. Rentsch, M. R. McKain, et al. 2022. Phylogenomic resolution of order- and family-level monocot relationships using 602 single-copy nuclear genes and 1375 BUSCO genes. *Frontiers in Plant Science* 13: 876779.

Vaidya, G., D. J. Lohman, and R. Meier. 2011. SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27: 171–180.

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

Yardeni, G., J. Viruel, M. Paris, J. Hess, C. Groot Crego, M. de La Harpe, N. Rivera, et al. 2022. Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Molecular Ecology Resources* 22: 927–945.

Yoo, M.-J., B.-Y. Lee, S. Kim, and C. E. Lim. 2021. Phylogenomics with Hyb-Seq unravels Korean *Hosta* evolution. *Frontiers in Plant Science* 12: 645735.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 15–30.

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

Zhang, C., and S. Mirarab. 2022. ASTRAL-Pro 2: Ultrafast species tree reconstruction from multi-copy gene family trees. *Bioinformatics* 38: 4949–4950.

SUPPORTING INFORMATION

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

Appendix S1. Sequence/probe filtering performed by Arbor Biosciences.

Appendix S2. DNA extraction and sequencing library preparation.

Appendix S3. Target locus lengths from *Asparagus* genomes.

Appendix S4. Target locus labels and orthologs found in Asparagaceae genomes.

Appendix S5. List of target orthologs shared with Angiosperms353.

Appendix S6. Output statistics from HybPiper.

Appendix S7. List of targets with paralog warnings.

Appendix S8. Supplemental figures.

Appendix S9. Multiple sequence alignment and gene tree estimation for Asparagaceae1726 single-copy gene trees.

Appendix S10. Plastome sequence assembly and alignment results.

How to cite this article: Bentz, P. C., and J. Leebens-Mack. 2024. Developing Asparagaceae1726: An Asparagaceae-specific probe set targeting 1726 loci for Hyb-Seq and phylogenomics in the family. *Applications in Plant Sciences* 12(5): e11597. https://doi.org/10.1002/aps3.11597