Brittonia (2025) 77:375–389
https://doi.org/10.1007/s12228-025-09854-x

ORIGINAL RESEARCH

Phylogenomics of *Lithospermum* (Boraginaceae) based on genome skimming

James Cohen[1]

Received: 9 June 2025 / Revised: 13 September 2025 / Accepted: 3 October 2025 / Published online: 3 November 2025 / Associate Editor: Stephen D. Gottschalk
© The Author(s) 2025

Abstract
Lithospermum (Boraginaceae) is a cosmopolitan genus of ca. 80 species native to both the Old and New Worlds. The genus has been the subject of multiple phylogenetic studies during the past 20 years, with different evolutionary relationships recovered depending on the taxon-locus combinations. Using whole plastomes and nuclear ribosomal cistrons, the phylogenetics, biogeographic patterns, and plastome selection of *Lithospermum* and its relatives were examined. *Lithospermum* is here resolved as non-monophyletic, with *Ulugbekia tschimganica* forming the sister group to *L. hancockianum*. Therefore, an expanded circumscription of *Lithospermum* is tentatively proposed, which aligns with other, prior studies. Based on multiple phylogenetic methods, evolutionary relationships inferred by the plastid genome are incongruent with those from the nuclear ribosomal cistron as well as prior analyses based on other nuclear DNA loci. This incongruence is likely due to incomplete lineage sorting within the nuclear genome and between nuclear and plastid markers during early diversification of the genus in the New World. *Lithospermum* is resolved to have originated in the Old World, with independent dispersal events to southern Africa and to North America, and a subsequent a long-distance dispersal event to the northern Andes mountain range in South America. Only six genes of the plastid genome are under positive selection, and based on prior studies, these are similar to those across Lithospermeae but differ from those among other tribes in the family.

Keywords Biogeography · Ka/Ks · Lithospermeae · Plastome

Introduction

Lithospermum L. (Boraginaceae), a cosmopolitan genus of ca. 80 species, is the second largest genus in subfamily Boraginoideae and the only one native to both the Old and New Worlds (Cohen & Davis, 2009; Chacón et al., 2016). Phylogenetic relationships in the genus have been the subject of multiple studies over the past 20 years, the results of which had encouraged a revised taxonomy for the genus. Specifically, Cohen & Davis (2009) resolved six segregate New World genera of Lithospermeae to be included in *Lithospermum*, and Weigend et al. (2009) recovered the monotypic Asian genus *Ulugbekia* Zakirov as a close relative of *Lithospermum*. While Weigend et al. (2009) advocated returning *Ulugbekia* to *Lithospermum*, as Boraginaceae expert Johnston (1952) had once done, this taxonomic change has not been completely adopted (POWO, 2025), which could be due to *Ulugbekia* often being placed as sister to *Lithospermum*, rather than nested within it. Later studies by Cohen (2011, 2012) and Chacón et al. (2019) recovered results similar to the earlier ones. All of these studies have included multiple plastid DNA (cpDNA) regions, and some have also used parts of the nuclear ribosomal (nrDNA) cistron (i.e., the internal transcribed spacer [ITS]). Given the various combinations of DNA regions, different evolutionary relationships have often been recovered, and this has, to some extent, made it challenging to effectively compare phylogenies or to use them for other purposes, such as to infer the biogeographic history of the group. More recently, Cohen (2022) employed 298 nuclear DNA (nDNA) regions to reconstruct the phylogeny of *Lithospermum*, and the phylogeny differed in multiple ways from prior evolutionary trees, particularly in terms of greater geographic structure as well as fewer origins of distyly and other derived floral morphologies. Cohen (2022) hypothesized that the incongruence

✉ James Cohen
jamescohen@weber.edu

Department of Botany and Plant Ecology, Weber State University, 1415 Edvalson St., Dept. 2504, Ogden, UT 84408-2504, USA

among phylogenies resolved from different genomes (and within the nuclear genome) was due to incomplete lineage sorting during early diversification of the genus in the New World. However, in comparing phylogenies from different genomes, an alternative explanation is possible, namely that the plastid phylogenies were based on only a fraction of that genome (Saarela et al., 2018). By comparing phylogenetic results to a sampling based on the entire genome, it is possible that the incongruence observed may be solely an artifact of incomplete sampling of the genome rather than differences in the evolutionary histories of the genomes. Indeed, Noroozi et al. (2022) subsampled the plastid genome and found that while a small fraction of the genome can recover a phylogeny similar to one with the entire genome, it is more likely to resolve one that differs, sometimes to a large extent.

Over the past decade, it has become more common to include entire plastid genomes in phylogenetics (e.g., Jost & Wanke, 2024). This is due in large part to sufficient reads obtained via low-coverage sequencing from second- and third-generation sequencing technologies, an approach known as genome skimming (Straub et al., 2012). Researchers have taken advantage of genome skimming for multiple studies across tribes, subtribes, and genera of Boraginaceae — Lithospermeae (Noroozi et al., 2022), Amsinckiinae (Ripma et al., 2014; Simpson et al., 2017a), and *Arnebia* Forssk. (Park et al., 2020; Sun et al., 2022) — resulting in a greater understanding of the evolutionary relationships in these groups, including increased resolution in challenging areas of the tree. Despite these phylogenomic advances, the plastid genome is limited in that it only provides one evolutionary history given its maternal pattern of inheritance in angiosperms (Doyle, 2022). At the same time, complete plastomes can allow for a better understanding of phylogenetics compared to using only bits and pieces of this genome, and using entire plastomes can complement hybridization-enrichment sequencing efforts that oftentimes employ hundreds of nuclear genes for phylogenomics (Weitemier et al., 2014; Straub et al., 2020; Cohen, 2025). Furthermore, assembly of the entire plastid genome allows for interspecific comparisons to explore patterns of evolution and selection for plastid genes (Saarela et al., 2018; Ye et al., 2018; Noroozi et al., 2022), which are involved in building photosynthetic protein complexes.

The objectives of the present study are: (1) to utilize complete plastid genomes and nrDNA cistrons to reconstruct the phylogeny and biogeographic history of *Lithospermum* and relatives, and (2) to investigate patterns of selection for genes of the plastid genome. I hypothesize that the plastid genome and nrDNA cistron will resolve phylogenies incongruent to each other as well as with the relationships recovered by Cohen (2022).

Materials & Methods

Taxon Sampling

A total of 57 species were included in the present study (Table 1). Forty-seven species were newly sequenced, while nine were previously published from the study of Noroozi et al. (2022), and a tenth was published by Cecchi et al. (2014). Forty species of *Lithospermum* were include in the study, which is the largest sampling of the genus to date and represents both the taxonomic and geographic diversity of the genus; however, one species, *L. hancockianum* Oliv., included only the limited nrDNA ITS sequence data available from a study on *Buglossoides* Moench (Cecchi et al., 2014). Additionally, seven species from six closely related genera, *Aegonychon* Gray, *Ancistrocarya* Maxim., *Buglossoides*, *Glandora* D.C.Thomas, Weigend & Hilger, *Lithodora* Griseb., and *Ulugbekia*, were also included in the study as well as three other genera from Lithospermeae (Table 1). While the sampling is concentrated in Boraginoideae (primarily Lithospermeae), two species of each Echiochiloideae and Cynoglossoideae were also included. Most species were collected from wild populations, but some samples were obtained from botanical gardens or herbaria. Fresh-collected leaf material was dried and preserved in silica gel and stored at 4° C. Voucher specimens were deposited at BH, CIIDIR, MICH, or WSCO (Thiers, updated continuously). Specimen information is presented in Table 1. Species starting with GenBank accession number SAMN48 were newly sequenced for the present study (47 # of sequences), while the remaining sequences were derived from prior studies (10 # of sequences).

DNA Sequencing

DNA was isolated from leaf material via homogenization with the method of Alexander et al. (2007), using the Norgen Plant/Fungi DNA isolation kit (Thorold, Ontario, Canada). Isolated DNA was sent to SNPsaurus (Eugene, Oregon, USA) for low-coverage sequencing (genome skimming), using an Illumina HiSeq 2500, to generate 150 base-pair (bp) paired-end (PE) reads. Raw reads were deposited at GenBank (PRJNA1269078, SAMN48776634-SAMN48776680), and alignments are available as Suppl. Material 1.

Low-quality reads (average Phred < 20) were removed from raw sequences, using *TrimGalore* (Krueger, 2015). Statistics on reads were calculated with *FastQC* (Andrews, 2010) and *MultiQC* (Ewels et al., 2016). The plastome and nuclear-ribosomal cistron were assembled with

Table 1 Species, collection information, and sequencing statistics for species included in study

Species	Collection	GenBank Accession	Herbarium	Unique reads	Duplicate reads	CP mean coverage	nrDNA mean coverage	CP genome length (bp)
Aegonychon purpurocaeruleum	19460250—RBGE	SAMN48776640	WSCO	1679299	206447	168	281	151739
Aegonychon zollingeri	K.Yonekura & T. Faukuda 95504	SAMN40923860	MO					
Alkanna orientalis	851242-DBG	SAMN48776634	WSCO	2444869	357591	220	639	151058
Anchusa capensis	DBG022032	SAMN48776635	WSCO	1162002	121548	96	62	150721
Ancistrocarya japonica	s.n	SAMN40923866	MIE					
Antiphytum floribundum	Cohen 227	SAMN48776637	BH	3241138	933194	326	312	150863
Antiphytum hintorium	Cohen 425	SAMN48776636	WSCO	2088615	346347	142	245	152698
Buglossoides arvensis	Cohen 355X1	SAMN48776639	MICH	2332193	399987	326	727	150868
Buglossoides tenuiflora	Nafisi, s.n	SAMN40923871	T					
Glandora rosmarinifolia	Plume 73	SAMN48776669	BH	622169	143407	32	220	150290
Hyunhia pulchria	Cohen 761	SAMN48776638	WSCO	1994683	376985	332	763	151453
Lithodora fruticosa	C. C. H. Jongkind & R. M. J. Nieuwenhuis 1169	SAMN40923882	MO					
Lithodora zahnii	Cohen 767	SAMN48776668	WSCO	2162971	451593	342	1332	150762
Lithospermum afromontanum	P. Lovett & C.J. Kayombo, 81	SAMN40923883	MO					
Lithospermum azuayense	C. Ulloa, D. Minga, D. fernadez & S. Sheth 1435	SAMN40923884	MO					
Lithospermum calcicola	Cohen 194	SAMN48776643	BH	1904960	406412	245	764	150388
Lithospermum californicum	Cohen 475	SAMN48776644	WSCO	1878058	297804	93	677	150049
Lithospermum calycosum	Cohen 197	SAMN48776645	BH	3710961	777895	314	1184	150620
Lithospermum canescens	Cohen 167	SAMN48776646	BH	2910058	833998	372	827	149862
Lithospermum caroliniense	Cohen 442	SAMN48776647	WSCO	5837544	1037258	184	1458	150426
Lithospermum cobrense	Cohen 457	SAMN48776648	WSCO	4422980	842366	282	1144	150537
Lithospermum discolor	Cohen 246	SAMN48776649	BH	1843864	316766	76	252	149150
Lithospermum distichum	Cohen 224	SAMN48776650	BH	1268268	218454	137	246	150499
Lithospermum erythrorhizon	Cohen 173	SAMN48776651	BH	4748852	1020186	399	2466	150618
Lithospermum exsertum	Cohen 244	SAMN48776670	BH	6692291	1969131	514	11152	150587
Lithospermum flavum	Cohen 226	SAMN48776673	BH	5420469	1463551	218	2937	150047
Lithospermum guatemalense	D. E. Breedlove 51263	SAMN40923888	MO					

Table 1 (continued)

Species	Collection	GenBank Accession	Herbarium	Unique reads	Duplicate reads	CP mean coverage	nrDNA mean coverage	CP genome length (bp)
Lithospermum hancockianum	*Handel-Mazzetti 6058*	KJ394990.1	KUN					
Lithospermum helleri	Cohen 132	SAMN48776652	BH	2507351	547161	217	788	149846
Lithospermum incisum	Cohen 371	SAMN48776653	WSCO	910234	169770	131	196	149220
Lithospermum johnstonii	Cohen 218	SAMN48776671	BH	5704694	2238594	488	3712	150462
Lithospermum latifolium	Cohen 87	SAMN48776654	BH	5876431	2569725	297	2338	150561
Lithospermum leonotis	Cohen 195	SAMN48776672	BH	6162814	2490606	791	1178	150535
Lithospermum macbridei	M. Weigend, K. Weigned, M. Binder & E. Rodriguz 5073	SAMN40923890	MO					
Lithospermum matamorense	Cohen and Straub 68	SAMN48776655	BH	1371722	199262	131	294	150545
Lithospermum mirabile	Cohen 180	SAMN48776656	BH	2241013	617427	422	553	150591
Lithospermum muelleri	R. Fernández N. & J. Rzedowski 3347	SAMN48776657	MICH	883724	130530	65	216	150592
Lithospermum multiflorum	Cohen 81	SAMN48776658	BH	2089737	572209	130	746	150608
Lithospermum nelsonii	Cohen 184	SAMN48776659	BH	836381	90685	33	380	150633
Lithospermum notatum	Cohen 188	SAMN48776660	BH	2120310	405906	140	1010	150641
Lithospermum oblongifolium	Cohen 201	SAMN48776661	BH	664341	111503	55	305	150503
Lithospermum obovatum	Cohen 208	SAMN48776662	BH	3474777	1146317	154	1669	150472
Lithospermum occidentale	Cohen 122	SAMN48776678	BH	2055627	436383	173	956	149844
Lithospermum officinale	Cohen 171	SAMN48776663	BH	2733246	729748	498	1285	150618
Lithospermum revolutum	Cohen 199	SAMN48776679	BH	1349103	321349	159	142	150647
Lithospermum rosei	Cohen 207	SAMN48776675	BH	10145463	3123515	1084	5396	150465
Lithospermum ruderale	Cohen 31	SAMN48776664	BH	1119246	282634	217	349	149438
Lithospermum strictum	Cohen 225	SAMN48776665	BH	1556258	317640	164	580	150535
Lithospermum thurberi	Cohen 468	SAMN48776680	WSCO	4739283	1349335	390	2138	150578
Lithospermum trinverium	Cohen 220	SAMN48776642	BH	1708424	275192	82	632	150266
Lithospermum tuberosum	Cohen 108	SAMN48776641	BH	5868244	2408976	370	3979	150514
Lithospermum tubuliflorum	M. Gonzalez 4001	SAMN48776666	CIIDIR	3843142	780476	182	1729	150501

Table 1 (continued)

Species	Collection	GenBank Accession	Herbarium	Unique reads	Duplicate reads	CP mean coverage	nrDNA mean coverage	CP genome length (bp)
Lithospermum viride	Cohen 451	SAMN48776667	WSCO	3098997	1207225	251	1052	150629
Mertensia ciliata	Cohen 261	SAMN48776674	BH	1998076	568812	480	271	151018
Oncaglossum pringlei	Cohen 219	SAMN48776676	BH	2126594	385742	173	623	149476
Onosma alborosea	Cohen 763	SAMN48776677	WSCO	4490495	660591	144	441	150792
Ulugbekia tschimganica	A. Orazova & V. Fissjun 5785	SAMN40923909	MO					

Newly sequenced specimens start with GenBank Access number SAMN48 and include information on reads, coverage, and length. Herbarium acronyms follow Thiers (updated continuously). *bp* is base pairs, *CP* is plastid, and *nrDNA* is nuclear ribosomal DNA

GetOrganelle (Jin et al., 2020) using default settings. These analyses were conducted with the Kettering University High Performance Cluster (KUHPC). Coverage for these regions was calculated using *BWA* for mapping (Vasimuddin et al., 2019), *samtools* for conversion of sam to bam files (Danecek et al., 2021), and *plotcoverage* in *deeptools* (Ramírez et al., 2016). Plastid genomes were annotated using *GeSeq* (Tillich et al., 2017) and the inverted repeat junctions were visualized with *IRplus* (Díez Menéndez et al., 2023). Individual genome regions (e.g., genes, introns, etc.) were extracted using *PhyloSuite* (Zhang et al., 2020).

Phylogenetic Analyses

The plastid genome was initially aligned using *MAFFT* (Katoh & Standley, 2013), with the iterative refinement method (*FFT-NS-i*), followed by the extraction and realignment of the long-single copy (LSC), short-single copy (SSC), and one inverted repeat (IR). The nrDNA cistron was aligned with *MAFFT* using the global homology algorithm (*G-INS-i*). Alignment statistics were calculated with *AMAS* (Borowiec, 2016). The outgroups for the present study included all taxa sampled that are not in *Lithospermum*: *Aegonychon*, *Alkanna* Tausch, *Anchusa* L., *Ancistrocarya*, *Antiphytum* DC. ex Meisn., *Buglossoides*, *Glandora*, *Huynhia Greuter*, *Lithodora*, *Mertensia* Rocth, *Oncaglossum* Sutorý, *Onosma* L., and *Ulugbekia*.

I conducted maximum parsimony (MP) and maximum likelihood (ML) analyses using the LSC, SSC, IR and nrDNA cistron as separate partitions, as well as with five combinations — IR + LSC, IR + SSC, LSC + SSC, IR + LSC + SSC, nrDNA + IR + LSC + SSC — each of which were concatenated using *2matrix* (Salinas & Little, 2014). Maximum parsimony analyses were undertaken with *TNT* (Goloboff & Catalano, 2016) using the following parameters: a random seed, 1 million trees held in memory, 1,000 ratchets with a 10% probability of upweighting and of downweighting, 1,000 tree-drift iterations, 50 sectorial searches, and 100 rounds of tree fusing; this search was repeated 10 times to identify the MP tree. Additionally, 1,000 jackknife and bootstrap replicates were run. *IQ-Tree* (Minh et al., 2020) was used for ML analyses. With *IQ-Tree*, *ModelFinder* (Kalyaanamoorthy et al., 2017) was employed to identify the optimal model, and with partition analyses, the best model was identified with and without the FreeRate heterogeneity models examined (i.e., MFP + MERGE and TESTMERGE). For all approaches, the Akaike Information Criterion (AIC), the corrected Akaike Information Criterion (AICc), and the Bayesian Information Criterion (BIC) were each considered to determine if model choice impacted phylogeny reconstruction. Along with searches for the optimal tree, 1,000 ultrafast bootstrap replicates were undertaken with the -bnni option to minimize potential issues with model violations. Quartet Sampling (Pease et al., 2018) was examined with concatenated trees, and quartet concordance, differential, and informativeness scores were calculated. Results were visualized with the QS_visualization script (https://github.com/ShuiyinLIU/QS_visualization).

Bayesian analyses were conducted, with *BEAST* 2 v2.6 (Bouckaert et al., 2019), for two matrices using a reduced set of 47 species that includes *Lithospermum* and close relatives: (1) the three plastid genome regions, and (2) the nrDNA cistron and plastid genome. For each analysis, the following parameters were employed: one billion Markov Chain Monte Carlo (MCMC) replicates with trees samples every 10,000 replicates, a GTR + I + G model, a relaxed log normal clock model, and a Yule birth–death model, with the fossil *Lithospermum dakotense* M.L.Gabel (Gabel, 1987) included at the base of the clade of New World species of *Lithospermum* with a log normal distribution, a mean of 1.25 and a standard deviation of 1, and an offset of 6.6 million years, which is consistent with Cohen (2022). In order to ensure appropriate rooting, the New World species of

Lithospermum were treated as a monophyletic group. To resolve consensus trees, *TreeAnnotator* was used with an 80% burnin (Bouckaert et al., 2019).

Patterns of Selection & Biogeography

Using *PAML* 4 (Yang, 2007), patterns of selection were investigated for 78 plastid genes. For these analyses, genes were aligned using *TranslatorX* (Abascal et al., 2010) to ensure that the alignment was based on codons, and the ML tree from the three plastid regions, using the optimal model from BIC, was employed. For *PAML* 4 analyses, the codon frequency was set to the average frequency at the positions, no clock was used, amino acid distances were equal, one Ka/Ks (omega) value was recognized for the entire tree, and a kappa of 2, an omega of 0.4, six gamma categories, and all six site models were used. Ka/Ks values for each model were determined for each gene. The effective number of codons (ENC) was calculated for each plastid gene, using *CodonW* (Peden, 2000), and relative synonymous codon usage (RSCU) was calculated for each gene with *cubar* (Zhang et al., 2024). Pairwise T-tests were conducted in R (R Developement Core Team, 2009) to examine differences among RSCU for various codons.

Patterns of historical biogeography were explored with *BioGeoBEARS* (Matzke, 2014) using the dispersal-extinction-cladogenesis (DEC), BayArea-like, and dispersal-vicariance (DIVA) models, with and without the J parameter, and the optimal model identified with AIC and AICc. For these analyses, the concatenated nrDNA and plastome phylogeny, from *BEAST* 2, was employed, and eight biogeographic areas were determined based on native geographic ranges for species of *Lithospermum* and relatives (Eurasia, southern Africa, eastern North America, western North America, eastern Mexico, central and southern Mexico and Central America, western Mexico, and South America). Most of these regions correspond to those from Cohen (2022), with new ones added to accommodate broader biogeographic sampling in the present study. All phylogenetic analyses were undertaken with resources at the University of Utah's Center for High Performance Computing.

Results

Taxon Sampling

Of the 47 newly sequenced species, the number of unique reads ranged from 622,169 in *Glandora rosmarinifolia* (Ten.) D.C.Thomas to 10,145,463 in *Lithospermum rosei* (I.M.Johnst.) J.I.Cohen, with a mean of 2,979,616. The number of duplicate reads ranged from 90,685 in *L. nelsonii* Greenm. to 3,123,515 in *L. rosei*, with a mean of 779,962. Plastid genome coverage ranged from 32× in *Glandora rosmarinifolia* to 1,084× in *L. rosei*, with a mean of 260×. Nuclear ribosomal cistron coverage ranged from 62× in *Anchusa capensis* Thunb. to 11,152× in *Lithospermum exsertum* (D.Don) J.I.Cohen, with a mean of 1,290×. Coverage statistics are in Table 1.

Assembled plastid genomes range in length from 149,150 bp in *Lithospermum discolor* M.Martens & Galeotti to 152,698 bp in *Antiphytum hintoniorum* L.C.Higgins & B.L.Turner. The plastid genome across the studied species of Boraginaceae have the typical quadripartite structure (Wicke et al., 2011) and include 78 protein genes, 42 tRNA genes, and four rRNA genes. The genes at the inverted-repeat boundaries generally are consistent across the family (Fig. 1), but there are some minimal differences regarding the position of *ndhF* and *ycf1* across the SSC-IRa junction. Also, in Lithospermeae *rpl22* is at the IRa-LSC junction, but in the few species of Borageineae and Echiochiloideae, *rps3* and *rps19* are, respectively, in this position.

Phylogenetic Analyses

The one included IR copy has an aligned length of 26,846 bp, and the aligned lengths of the LSC and SSC are 91,677 bp and 18,336 bp, respectively, for a total aligned length of 136,859 bp for the analyzed plastome (i.e., IR + LS + SSC regions). The LSC and SSC had more parsimony informative characters than the IR (Table 2). The aligned nrDNA cistron was 5,963 bp. While this region has the fewest number of variable sites, 8% of the characters are parsimony informative, which is a greater proportion than either the IR or LSC. Alignment statistics are provided in Table 2.

Overall, phylogenetic results were comparable among the various methodologies and models, and while outgroup relationships were similar between genomes, relationships within *Lithospermum* differed between the plastid genome and nrDNA cistron (Fig. 2). Phylogenies from the concatenated nrDNA + cpDNA matrix were, generally, similar to those from the cpDNA matrix (Fig. 3, Suppl. Material 2).

All outgroup nodes are strongly supported (100% ML bootstrap [BS], MP jackknife [JK]), with the exception of the clade that unites *Huynhia* + *Onosma* as sister to *Ancistrocrya*, *Lithospermum*, and relatives (65% BS, 99% JK). Lithospermeae is resolved as monophyletic (100% BS, MP JK), but *Lithospermum* is not due to the placement of *Ulugebekia*, which is, depending on the dataset, either sister to *L. hancockianum* (92% BS, 52% JK) or included in *Lithospermum*, with *L hancockianum* as sister to it and the rest of the genus, albeit with low support in this particular position (concatenated 4% JK). The clade of *Ulugbekia* and *Lithospermum* is monophyletic (100% BS, 93–95% JK). Sister to this clade is *G. rosmarinifolia* in cpDNA or concatenated cpDNA + nrDNA analyses (100% BS, 98% JK), a clade of

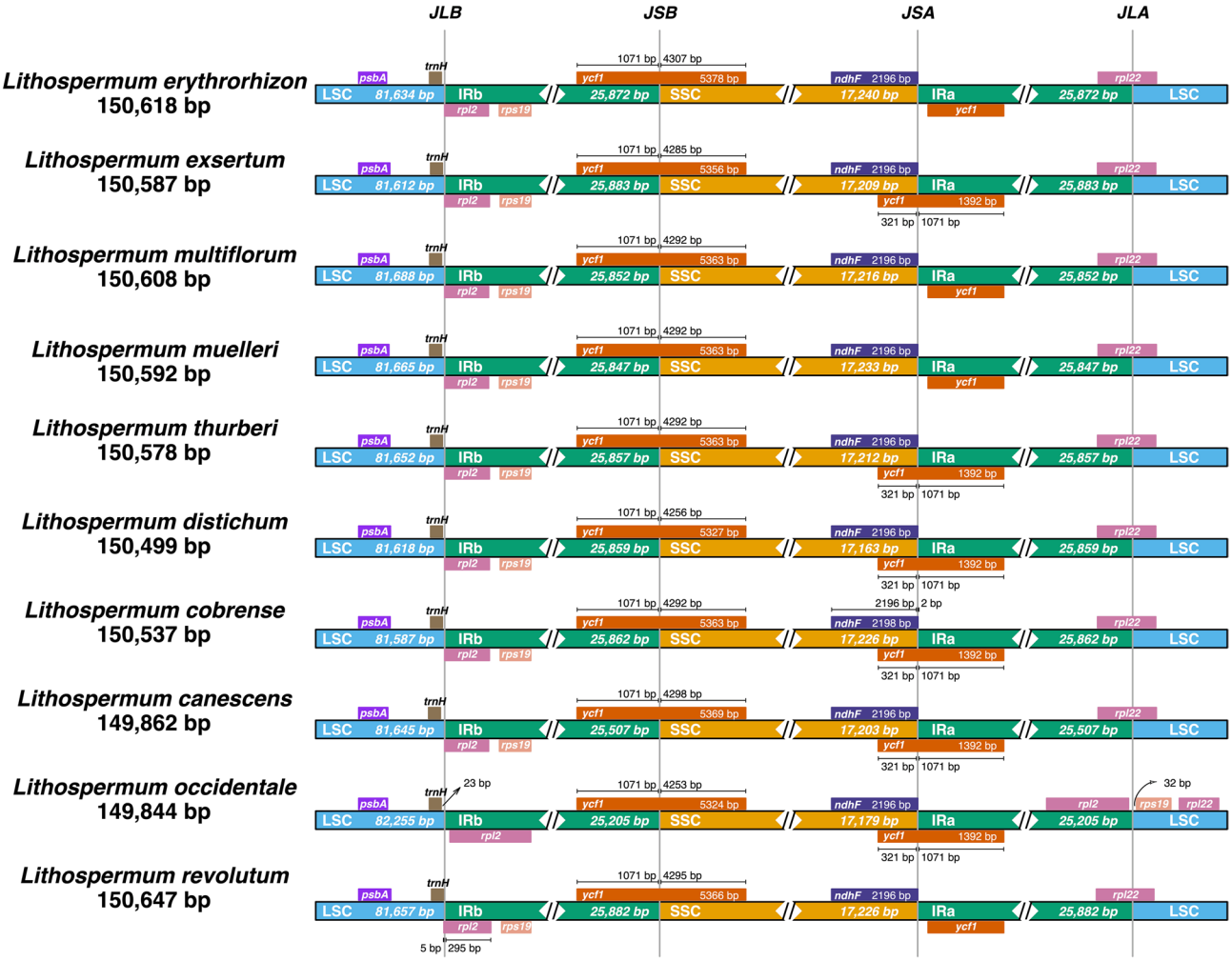


Fig. 1. Gene presence and length at inverted repeat junction boundaries for species of *Lithospermum*, including plastid genome length. LSC is long single copy, SSS is short single copy, IR is inverted repeat, and bp is base pairs.

Table 2 Alignment statistics for three plastid DNA regions and nuclear ribosomal cistron. bp is base pairs

DNA Region	Alignment length (bp)	Percent missing data	Variable sites (bp)	Percent variable sites	Parsimony informative sites	Percent parsimony informative
Inverted repeat	26,846	5.7%	1616	6.0%	412	1.5%
Long single copy	91,677	12.8%	14630	16.0%	6244	6.8%
Short single copy	18,336	7.8%	3858	21.0%	1696	9.2%
nrDNA cistron	5,963	1.0%	724	12.1%	477	8.0%

Ancistrocarya and two species of *Buglossoides* (30% BS) in nrDNA analyses, or a polytomy of multiple clades from across Lithospermeae in MP nrDNA analyses (Figs. 2, 3).

Within cpDNA + nrDNA analyses of *Lithospermum*, most nodes have > 75% BS and > 80% JK, but 13 branches in the MP analyses and eight branches in the ML analyses have only 45–60% BS and 0–66% JK (Fig. 3). These branches with lower support are primarily along the backbone of the tree for the genus. In most analyses, the Old-World species of *Lithospermum* diverge prior to the New-World ones. In the cpDNA and concatenated analyses, *L. afromontanum* Weim. is sister to the rest of the genus (100% BS, JK), apart from *L. hancockianum*, but in nrDNA analyses, *L. afromontanum* is in a clade (82–97% BS) with two other Old-World species,

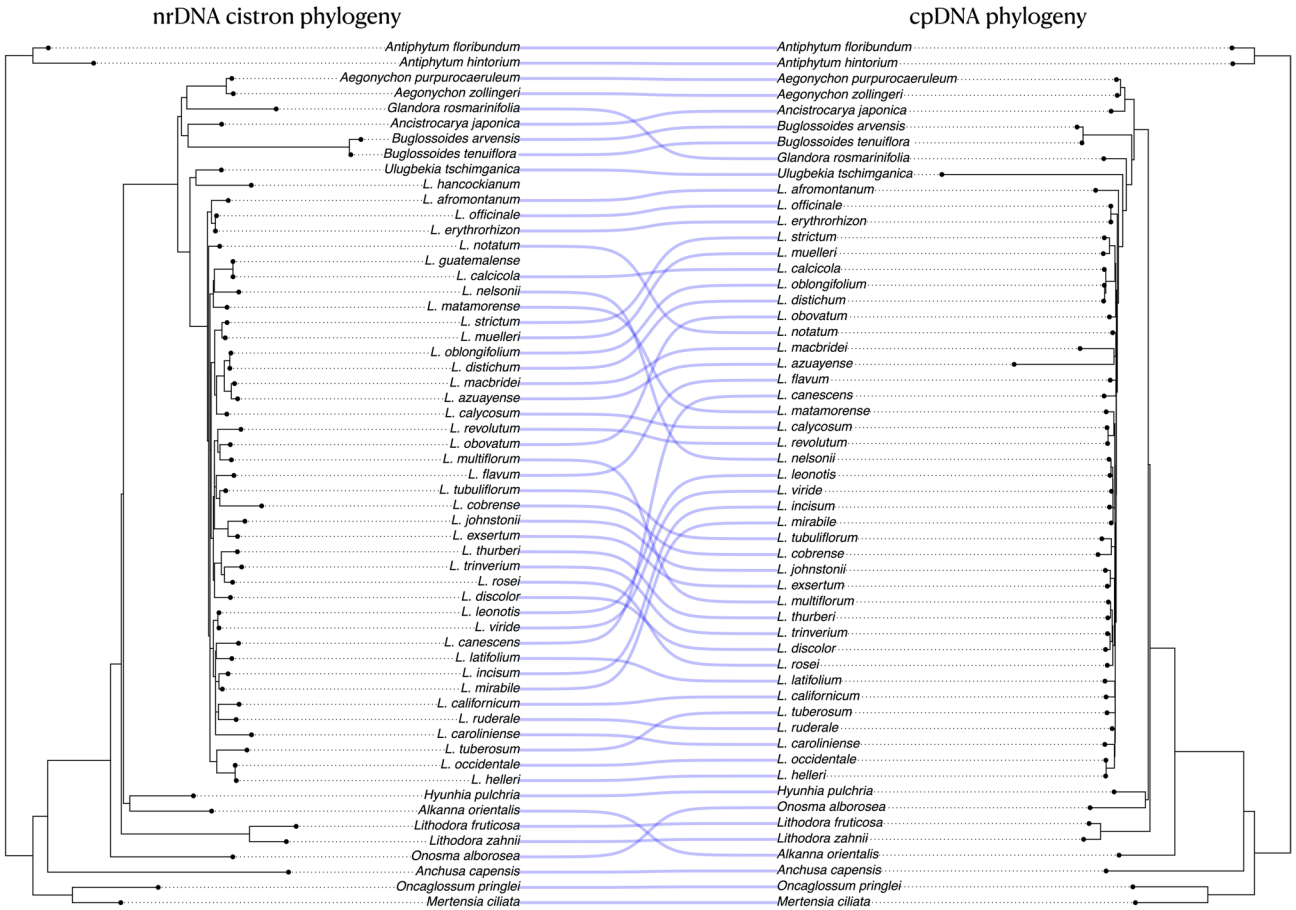


Fig. 2. Paired phylogenies from plastid and nuclear ribosomal DNA alignments. Lines between phylogenies show position of each species in trees. The generic abbreviation *L.* represents *Lithospermum.*

L. erythrorhizon Siebold & Zucc. and *L. officinale* L., that is sister to the rest of the genus (33–35% BS, 65% JK), apart from *L. hancockianum*. The exceptions to these two topologies are analyses with solely IR and with IR + SSC using *TESTMERGE*, and in these analyses, *L. azuayense* Weigend & Nürk is sister to *Ulugbekia*, two species that have long branches. Within *Lithospermum*, branch lengths for the backbone of the genus are quite short in ML analyses, and BS values generally are less than 60% (Fig. 3, thin branches). Additionally, Quartet Concordance scores generally range from < 0.2 to negative, but this is not the case for all branches along the backbone.

In the concatenated tree (Fig. 3, Suppl. Material 2), the earliest-diverging clade in the New World is one composed of species in eastern Mexico in cpDNA and concatenated analyses (100% BS, JK), but in ML nrDNA analyses, *L. notatum* (I.M.Johnst.) J.I.Cohen is the New-World member sister to the rest of the genus (39% BS). Species that used be included in *Macromeria* D.Don are resolved across the phylogeny in all analyses, and species formerly in *Onosmodium* Michx. are monophyletic, with *L. tuberosum* Rugel ex A.DC. (nrDNA, 61–76% BS, 52% JK; concatenated MP 2% JK), *L. californicum* A.Gray (e.g., cpDNA and concatenated ML, 38–51% BS), or *L. caroliniense* (J.F.Gmel.) MacMill. (e.g., cpDNA, 14–57% JK) as sister. The two South-American species of the genus included are resolved as sisters (98–100% BS, 99–100% JK), and in analyses of cpDNA and the concatenated matrix, *L. notatum* is sister to this clade (33–83% BS, 22–39% JK); however, in nrDNA analyses, a clade of *L. distichum* Ortega + *L. oblongifolium* Greenm. is sister to the South American species (99% BS, 98% JK).

Analyses with *BEAST 2* yielded similar topologies and support values to those with MP and ML datasets that had more outgroup sampling (Figs. 3, 4). Results of the *BEAST 2* analyses resolve the stem and crown age of *Lithospermum* as ca. 12.7–13.1 million years ago (MYA) and 11.3 MYA, respectively. The genus colonized the New World ca. 8.8–9 MYA and migrated to South America from North America ca. 6.7–6.9 MYA.

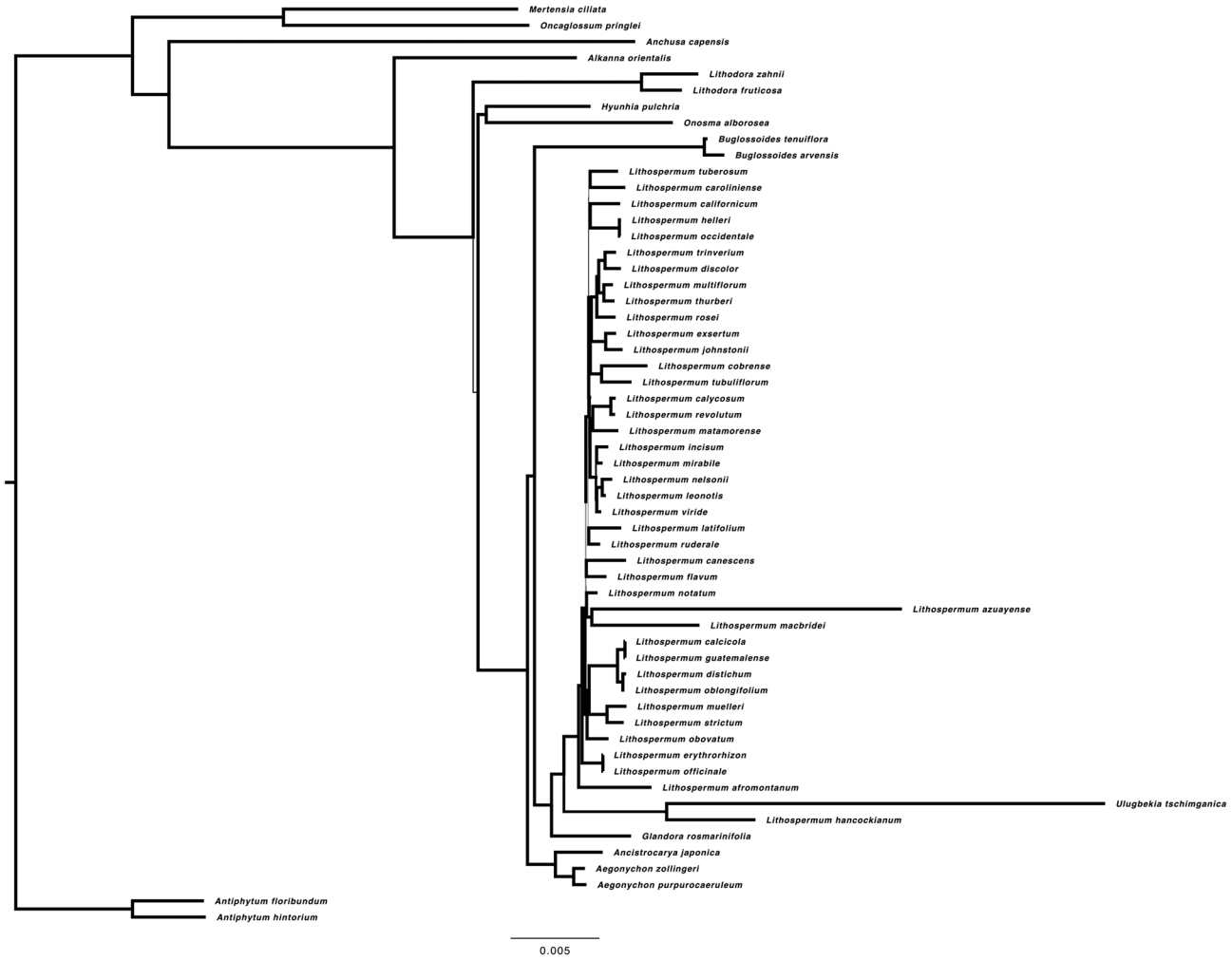


Fig. 3. Maximum likelihood phylogeny of concatenated plastid and nuclear ribosomal DNA data. Thick lines represented bootstrap values > 70%, but generally more than 90%, and thin branches denote bootstrap values 45–60%.

Patterns of Selection & Biogeography

Based on *PAML* analyses, five genes, *pbf1*, *petL*, *petN*, *psbF*, and *psbJ*, have omega (Ka/Ks) values greater than 1 among four of six models, and one gene, *rpl20*, has an omega value greater than 1 for three models (Suppl. Material 3). Ka/Ks values greater than 1 suggest positive selection (Chen et al., 2022). All other genes had omega values substantially less than 1. Three genes, *pbf1*, *petN*, and *psbI*, had ENC values less than 35, which suggests strong codon usage bias (Wang et al., 2018). Additionally, RSCU values across plastid genes statistically differed among most codons for each amino acid. Only GCU (alanine) has a median value of 1.6 or greater, which is indicative of overrepresented codons (Fang et al., 2024), and CUG (leucine) has the lowest median value (0.35).

Of the models explored in *BioGeoBEARS*, the BayArea-like + J model is preferred based in AIC and AICc. Results provide evidence that *Lithospermum* originated in Eurasia and that *L. afromontanum* had a long-distance dispersal (LDD) event to southern Africa (Fig. 4). The New World species are the result of another LDD event, and the genus colonized eastern Mexico initially. Some clades diversified in that region, while others spread to adjacent regions across North America to subsequently radiate. The only exception is the ancestor of the two South American species, which dispersed to northern South America from eastern Mexico.

Discussion

Phylogenetics & Biogeography

Overall, relationships among outgroups are congruent with those from prior studies (Weigend et al., 2009; Cohen & Davis, 2012; Chacón et al., 2019; Noroozi et al., 2024), and

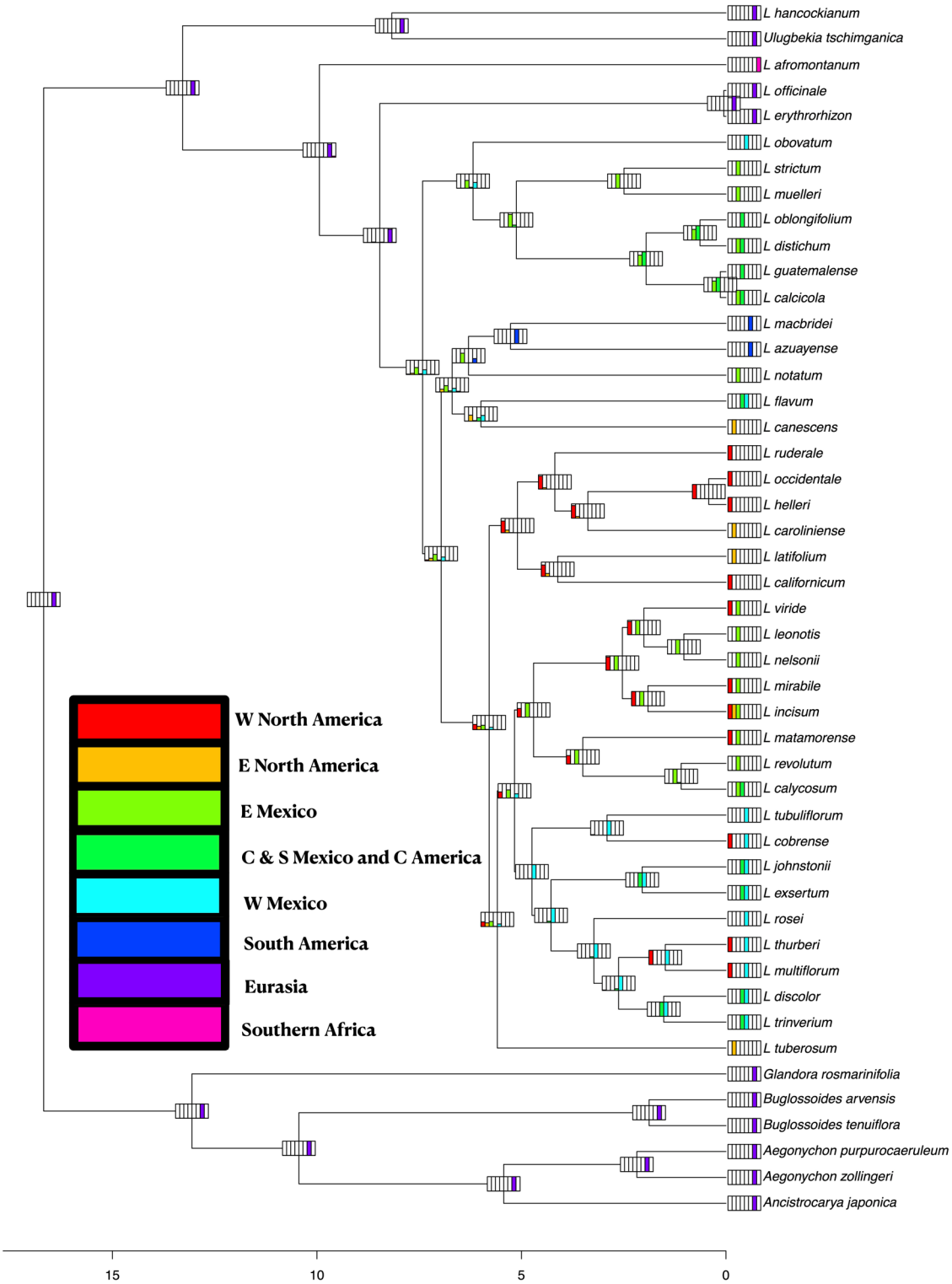


Fig. 4. Biogeographic reconstruction of *Lithospermum* and relatives based on *BayArea-like*+*J*. Colors correspond to biogeographic regions in the legend. C. is central, N. is north, S. is South, and W. is West.

these relationships receive strong support (100% BS). Differences are noted between the cpDNA and nrDNA regions, with only *Glandora* as sister to *Lithospermum*+*Ulugbekia* with the former, but with the latter, a clade of *Aegonychon*, *Ancistrocarya*, *Buglossoides*, and *Glandora* in this position (Fig. 3). These results suggest that among the outgroup

relationships, sampling with fewer DNA regions yields similar results to employing entire plastomes; however, the situation becomes more complicated when sampling from different genomes, which resolve incongruent relationships and provide evidence that the evolutionary history may be more complex than currently understood (Figs. 2, 3).

In the present phylogenies, *Lithospermum hancockianum* is sister to the monotypic *Ulugbekia* (or *Ulugbekia* and the rest of *Lithospermum*), and the clade of these two species is sister to the rest of *Lithospermum*. *Lithospermum hancockianum* has always been slightly unusual in the genus because it produces pink to blue to purple-red corollas, a feature unknown from any other species in the genus (Johnston, 1952), while the closely related *U. tschimganica* (B.Fedtsch.) Zakirov bears yellow corollas, a common color in *Lithospermum*. The two species differ in corolla color, but both are native to central and eastern Asia, have smooth white nutlets, and bear corollas with scattered glands and without faucal appendages, the presence of which are common in *Lithospermum* (Johnston, 1952). Johnston (1952) noted that both species have features similar to other Asian genera of Lithospermeae, such as *Arnebia*; however, phylogenetic evidence has demonstrated that these genera are not closely related (Chacón et al., 2019; Noroozi et al., 2024). While Johnston's thoughts and the present evolutionary relationships suggest that it could be better to treat both species in *Ulugbekia* and separate from *Lithospermum*, given that there may not be many features that unite *U. tschimganica* and *L. hancockianum*, this segregation would likely not benefit the systematics of *Lithospermum* or *Ulugbekia*. Other species that have traditionally been included in *Lithospermum*, such as *L. cobrense* Greene, produce corollas that lack faucal appendages and have glands (Johnston, 1952; Cohen, 2018), and this is the case for the New World genera that Cohen & Davis (2009) included in *Lithospermum*. The results of the present study agree with those of prior ones that have proposed to expand the circumscription of *Lithospermum* to include *Ulugbekia* (Weigend et al., 2009; Cecchi et al., 2014), in which case the genus could continue to be recognized by the smooth, white, lustrous nutlets. At the same time, it would be prudent to consider the position of *U. tschimganica* within *Lithospermum* as tentative because the plastome of *U. tschimganica* was incomplete, which could contribute to its long branch in the cpDNA + nrDNA phylogeny (Fig. 3) and only a small DNA region, ITS, was included for *L. hancockianum*. Additional data for these two species would provide further evidence for an expanded circumscription of *Lithospermum*.

In the present study, three other Old-World species of *Lithospermum*, *L. afromontanum*, *L. erythrorhizon*, and *L. officinale*, are in various configurations sister to the New-World species of the genus. In the phylogeny from the nrDNA + cpDNA data, the southern African species *L. afromontanum* is sister to *L. erythrorhizon*, *L. officinale*, and the New-World species (Fig. 3). Cohen (2012) resolved a similar relationship based on 10 KB of cpDNA data, but this differs from the results of Chacón et al. (2019), who reconstructed the southern African species arising from New World ones. While additional African species should be included in future analyses to gain a better understanding of the evolutionary relationships of this small radiation, more thorough character sampling and broader overall taxon sampling suggest that the Old-World species diverged prior to *Lithospermum* colonizing the New World.

Lithospermum started to diversify in the New World ca. 8.8 MYA in eastern Mexico, and this result is within one million years of the result Cohen (2022) estimated based on 298 nDNA regions. In the present study, the genus initially diversified in eastern Mexico prior to subsequent speciation in western Mexico and the western United States and Canada (USC) (Fig. 4). These results are broadly similar to those from Cohen (2022), even with different taxon and locus sampling; however, in the present study, the species native to the eastern USC originated multiple times from the western USC. In Cohen (2022), these eastern USC species are members of the same clade, with most diversifying in this region.

Lithospermum has also had a small radiation in the Andes in South America (Weigend et al., 2010). In the present study, the two South American species, *L. azuayense* and *L. macbridei* I.M.Johnst., are sisters, and the clade of these two species is sister to *L. notatum* (I.M.Johnst.) J.I.Cohen, a species native to eastern Mexico (Fig. 4). Amphitropical dispersals have occurred at least 18 times in Boraginaceae, particularly in Amsinckiinae (Simpson et al., 2017b), and given present geographic distributions, most of these would be hypothesized to have occurred between the west coasts of both continents. Dispersal from eastern Mexico to northwestern South America would likely be a less common dispersal route. The biogeographic results suggest that South American species originated via long-distance dispersal from eastern Mexico to northwestern South America approximately 6.8 MYA. This date is earlier than Weigend et al. (2009) resolved based on ITS and *trnL-trnF* sequence data, but it is prior to the final uplift of the Andes Mountains during the past five million years (Gregory-Wodzicki, 2000; Weigend et al., 2010). Given that *L. notatum* and some close relatives grow at higher elevations (up to 3,300 m, Cohen, 2018), new arrivals to the Andes could have already been adapted to the higher elevations of this region.

The study of Weigend et al. (2010) also placed a second South American species separate from the others from this continent. *Lithospermum mediale* I.M.Johnst. differs in its geographic distribution from other South American species, ranging from Mesoamerica to northern South America, farther north than all remaining species in the genus in South

America (Weigend et al., 2010; Cohen & Miller, 2012). It is the only species of the genus with this geographic distribution, but other species are found in Mesoamerica (Cohen & Miller, 2012; Cohen, 2018), such as *L. guatemalense* Donn. Sm., which is included in the present study. *Lithospermum guatemalense* is resolved as sister to *L. calcicola* B.L.Rob., which is native to eastern Mexico, and the biogeographic analyses suggest that *L. guatemalense* may have arisen via vicariance across Mexico. The present study has the broadest sampling of species of *Lithospermum* to date, but despite this fact, additional taxon sampling, especially of South African and South American species, would strengthen phylogenetic conclusions. It would be particularly helpful to employ more nDNA regions for this type of study as well, given that incongruence exists between genomes in *Lithospermum* (Fig. 2). The results from the present study provide a useful framework and hypothesis to better understand the evolutionary relationships and biogeographic history of the genus, especially in the New World.

Patterns of Incongruence

To date, researchers have employed various combinations of two to ten cpDNA regions (sometimes along with ITS) to explore the phylogenetics of *Lithospermum* and relatives, and taxon sampling has varied among these studies (Weigend et al., 2009; Cohen, 2011, 2012; Chacón et al., 2019). Across all of them, there are some consistent results, such as *Lithospermum* originating in the Old World and species with long-funnelform corollas (i.e., in the former *Macromeria*) being non-monophyletic (Cohen & Davis, 2009; Weigend et al., 2009), but apart from broad similarities, phylogenetic relationships, especially among New World members, have differed among studies (Weigend et al., 2009; Cohen, 2022).

Different results could be due to distinct taxon sampling in each study. This may especially be the case as Weigend et al. (2009) and Chacón et al. (2019) have included more species from across the geographic distribution of the genus (e.g., South America and South Africa), while Cohen & Davis (2012) and Cohen (2012) focused more on the North American species. Despite some similarities among smaller clades and sister species in these studies, differences abound not only from each other but also from results of the present study, which includes much more DNA sequence data (ca. 13 × as much). Noroozi et al. (2024) demonstrated that increased sequencing data across the plastome yields trees that are more similar to those from the entire plastome, and sampling that is only 10% of the plastome (which is most similar to sampling from prior studies of *Lithospermum*) only infrequently recovers a tree similar to that from the whole plastome. Such results suggest that while taxon sampling may play a role in the phylogenic trees that are recovered, it is likely that differences in nucleotide sampling have a greater impact. This is not to suggest that taxon sampling (both ingroup and outgroup) cannot influence, sometimes strongly, the resulting phylogenetic hypotheses, but the amount of sequence data sampled from across the plastid genome does matter. This impact is even more likely for groups that have experienced increased rates of diversification, as appears to be the case for *Lithospermum* in the New World (Chacón et al., 2019; Cohen, 2022).

Cohen (2022) used hundreds of nDNA regions to resolve the phylogeny of *Lithospermum*, and the results from the present study, using cpDNA and nrDNA, differ from those. He suggested that the different topologies recovered, within the nuclear genome and between nuclear and plastid genomes, were due to incomplete lineage sorting (ILS). The entire plastome-based phylogeny is distinct from that derived from the analysis of nDNA loci, but some small clades are generally similar. For example, both sets of loci recover a close relationship between *L. distichum*, *L. oblongifolium*, and *L. strictum* Lehm., which are all present in central Mexico; however, in the cpDNA analyses, *L. calcicola* and L. *obovatum* J.F.Macbr. are resolved in the same clade as these three species, and this is not the case in the nDNA tree. As a group, these five species do not share similar floral morphologies (e.g., medium-sized salverform white corollas with a yellow eye in *L. distichum*, long-funnelform green-yellow corollas in *L. oblongifolium*, and medium-sized, slender salverform yellow corollas in *L. strictum* Cohen, 2018), and *L. calcicola* and *L. obovatum* do not overlap geographically, although they do each overlap with some of the other three species. These results provide further evidence that it is likely ILS, not hybridization or sampling issues, that influenced the evolutionary history of *Lithospermum* as it radiated in the New World; however, further studies can continue to more thoroughly explore the underlying reasons for incongruence in the relationships among the New World members of *Lithospermum*.

The phylogeny derived from the nrDNA cistron has greater congruence with the nDNA tree of Cohen (2022), but differences still occur. For example, in the nrDNA tree, the aforementioned clade of five species also includes *Lithospermum calycosum* I.M.Johnst., a species with a similar geographic range to the others in the clade, but in the nDNA tree, this species is allied with two distylous species — *L. cobrense* and *L. tubuliflorum* Greene — each with a different corolla shape and native to Western Mexico and the United States. As with the plastome phylogeny, given distinct floral morphologies and current geographic distributions, hybridization does not seem a likely explanation for the topological differences from the nuclear genome. One possibility concerns the aligned length of the nrDNA cistron. This locus only includes ca. 10 KB of aligned nucleotides, and the nDNA tree is based on ca. 50 × that amount (Cohen, 2022); therefore, it may be possible that this situation is analogous

to the one with less cpDNA sequence data resulting in a different topology, and taxon sampling may also influence the recovery of a distinct topology. Another possibility is that concerted evolution could play a role as this occurs among the copies of the nrDNA cistron (Álvarez & Wendel, 2003), and the possibility of hybridization or ILS during the early diversification of the genus in the New World could have provided the alleles for the subsequent homogenizing process, which can start to take effect within less than a century (Tate et al., 2009).

Plastid Genome Structure & Patterns of Selection

Within *Lithospermum*, the plastid genome structure was quite consistent among species. This is congruent with the results of other studies of plastomes across Lithospermeae (Noroozi et al., 2024). One of the more variable features throughout *Lithospermum* concerns *ycf1* spanning the SSC and IRa boundary (Fig. 1), which is present in only some species (e.g., *L. cobrense*, *L. distichum*, *L. thurberi*, and others). While this feature was not examined using optimization, given its phylogenetic distribution, it appears to be evolutionarily labile throughout the genus.

Among *Lithospermum* and relatives, relative synonymous codon-usage (RSCU) values were statistically significant for most codons across most amino acids. While only some were above or below the 1.6 and 0.6 threshold that would suggest over- or under-representation, respectively (Jamil et al., 2022), the differences provide evidence for distinct ways in which species utilize codons in the plastome. For example, the codon GCU (alanine) was found to have the greatest median RSCU value, and this was also the case in other genera of Lithospermeae (Li et al., 2024; Noroozi et al., 2024; Okada & Watanabe, 2024; Wang et al., 2025). Additionally, the codon for leucine, CUG, had a low RSCU value across the genus and tribe. Based on these results, it appears that codon bias in *Lithospermum* is similar to that across Lithospermeae.

RNA editing is known to occur in some plastid genes (Freyer et al., 1997), including *petL* (Fiebig et al., 2004). This process could result in stabilizing selection, particularly for first and second codon positions, in order to ensure proper RNA editing. Some have suggested that natural selection acts on sites with RNA editing, although the influence of this process remains understudied in many lineages (Jobson & Qiu, 2008), including Boraginaceae. Of the genes recognized as under selection by *PAML* analyses, three (*psbF*, *psbJ*, and *psb1*) produce proteins involved in photosystem II (PS II), and two others (*petL* and *petN*) make proteins for the cytochrome b6f complex. The *psbI* gene is also involved in PS II. Robbins & Kelly (2025) investigated genes in photosystems I and II (PS I and II) and found that similar patterns of positive selection occurred for genes in each photosystem. This is the case even though only five genes for PS I are in the plastome, while 15 are for PS II. Their study did not include species of Boraginaceae, and Noroozi et al. (2024) and the present study, which focused on Lithospermeae, recognized genes only in PS II, not PS I, under positive selection. In contrast, Li & Wei (2022) identified *psaI*, a gene in PS I, under positive selection in species of the borage tribes Cynoglossoideae and Boragineae. This suggests that species of Lithospermeae may be undergoing different selection pressures for plastome genes compared to other clades of Boraginaceae. Given that many members of Boraginaceae have radiated in temperate and montane environments, different plastid genes could be under selection for photoprotection in distinct clades (e.g., Piot et al., 2018). These results suggest the potential for an intriguing area of study on plastid function and gene selection in across the family.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12228-025-09854-x.

Acknowledgements

The author would like to thank J. Pashnick, G. Plunkett, and two anonymous reviewers for helpful comments on the manuscript. DNA sequencing was conducted by SNPsaurus. Phylogenetic analyses were conducted using the Kettering University High Performance Cluster and the Center for High Performance Computing at the University of Utah.

Funding

This research was supported by the Kettering University and Weber State University, James Cohen.

Data Availability

DNA sequence data used in the present study is included in Supplementary File 1. One of the phylogenies from the present study is included in Supplementary File 2.

Declarations

Competing Interests

The author is an Associate Editor for *Brittonia* but was not involved in the peer review of this article or the decision to publish. The author declares that he had no known competing interests that could have appeared to undermine the objectivity or integrity of the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Literature Cited

Abascal, F., R. Zardoya, & M.J. Telford. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* 38: W7–W13.

Alexander, P.J., G. Rajanikanth, C.D. Bacon, & C.D. Bailey. 2007. Recovery of plant DNA using a reciprocating saw and silica-based columns. *Molecular Ecology Notes* 7: 5–9.

Álvarez, I. & J.F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.

Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

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

Bouckaert, R., T.G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, and N. De Maio. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Computational Biology 15: e1006650.

Cecchi, L., A. Coppi, H.H. Hilger, & F. Selvi. 2014. Non-monophyly of *Buglossoides* (Boraginaceae: Lithospermeae): Phylogenetic and morphological evidence for the expansion of *Glandora* and reappraisal of *Aegonychon*. *Taxon* 63: 1065–1078.

Chacón, J., F. Luebert, H.H. Hilger, S. Ovchinnikova, F. Selvi, L. Cecchi, C.M. Guilliams, K. Hasenstab-Lehman, K. Sutorý, M.G. Simpson, & M. Weigend. 2016. The borage family (Boraginaceae s. str.): A revised infrafamilial classification based on new phylogenetic evidence, with emphasis on the placement of some enigmatic genera. *Taxon* 65: 523–546.

Chacón, J., F. Luebert, F. Selvi, L. Cecchi, & M. Weigend. 2019. Phylogeny and historical biogeography of Lithospermeae (Boraginaceae): Disentangling the possible causes of Miocene diversifications. *Molecular Phylogenetics and Evolution* 141: 106626.

Chen, Q., H. Yang, X. Feng, Q. Chen, S. Shi, C.-I. Wu, & Z. He. 2022. Two decades of suspect evidence for adaptive molecular evolution—negative selection confounding positive-selection signals. *National Science Review* 9: nwab217.

Cohen, J.I. 2011. A phylogenetic analysis of morphological and molecular characters of *Lithospermum* L. (Boraginaceae) and related taxa: evolutionary relationships and character evolution. *Cladistics* 27: 559-580.

Cohen, J.I. 2012. Continuous characters in phylogenetic analyses: patterns of corolla tube length evolution in *Lithospermum* L. (Boraginaceae). *Biological Journal of the Linnean Society* 107: 442–457.

Cohen, J.I. 2018. A revision of the Mexican species of *Lithospermum* (Boraginaceae). *Annals of the Missouri Botanical Garden* 103: 200–257.

Cohen, J.I. 2022. Phylogenomics, floral evolution, and biogeography of *Lithospermum* L. (Boraginaceae). *Molecular Phylogenetics and Evolution* 166: 107317.

Cohen, J.I. 2025. Phylogenomics of Boraginaceae and Boraginales using lineage-specific and Angiosperms353 loci. *Systematic Botany* 49: 749–767.

Cohen, J.I. & J.I. Davis. 2009. Nomenclatural changes in *Lithospermum* (Boraginaceae) and related taxa following a reassessment of phylogenetic relationships. *Brittonia* 61: 101–111.

Cohen, J.I. & J.I. Davis. 2012. Molecular phylogenetics, molecular evolution, and patterns of clade support in *Lithospermum* (Boraginaceae) and related taxa. *Systematic Botany* 37: 490–506.

Cohen, J.I. & J.S. Miller. 2012. *Lithospermum*. in *Flora Mesoamericana, Rubiaceae a Verbenaceae.*, vol. 4. G. Davidse, M. Sousa M. and S. Knapp (eds.). Universidad Nacional Autonoma de Mexico, D. F., Mexico, Missouri Botanical Garden Press, St. Louis, USA and The Natural History Museum, London, London, England.

Danecek, P., J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M.O. Pollard, A. Whitwham, T. Keane, S.A. McCarthy, & R.M. Davies. 2021. Twelve years of SAMtools and BCFtools. *Gigascience 10*: giab008.

R Developement Core Team. 2009. A language and environment for statistical computing. http://www.R-project.org.

Díez Menéndez, C., P. Poczai, B. Williams, L. Myllys, & A. Amiryousefi. 2023. IRplus: An augmented tool to detect inverted repeats in plastid genomes. *Genome Biology and Evolution 15*: evad177.

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

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

Fang, J., Y. Hu, & Z. Hu. 2024. Comparative analysis of codon usage patterns in 16 chloroplast genomes of suborder Halimedineae. *BMC Genomics* 25: 945.

Fiebig, A., S. Stegemann, & R. Bock. 2004. Rapid evolution of RNA editing sites in a small non-essential plastid gene. *Nucleic Acids Research* 32: 3615–3622.

Freyer, R., M.-C. Kiefer-Meyer, & H. Kössel. 1997. Occurrence of plastid RNA editing in all major lineages of land plants. *Proceedings of the National Academy of Sciences* 94: 6285–6290.

Gabel, M.L. 1987. A fossil *Lithospermum* (Boraginaceae) from the Tertiary of South Dakota. *American Journal of Botany* 74: 1690–1693.

Goloboff, P.A. & S.A. Catalano. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* 32: 221–238.

Gregory-Wodzicki, K.M. 2000. Uplift history of the Central and Northern Andes: a review. *Geological Society of America Bulletin* 112: 1091–1105.

Jamil, Z., A. Uddin, S.S.M. Alam, A. Samanta, N. Altwaijry, M.A. Rauf, S. Ali, M.S. Khan, M.N. Asghar, & M. Hoque. 2022. Analysis of the compositional features and codon usage pattern of genes involved in human autophagy. *Cells* 11: 3203.

Jin, J.-J., W.-B. Yu, J.-B. Yang, Y. Song, C.W. DePamphilis, T.-S. Yi, & D.-Z. Li. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21: 1–31.

Jobson, R.W. & Y.-L. Qiu. 2008. Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? *Biology Direct* 3: 1–22.

Johnston, I.M. 1952. Studies in the Boraginaceae, XXIII A survey of the genus *Lithospermum*. *Journal of the Arnold Arboretum* 33: 299–366.

Jost, M. & S. Wanke. 2024. A comparative analysis of plastome evolution in autotrophic Piperales. *American Journal of Botany* 111: e16300.

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

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

Krueger, F. 2015. Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. Available online at: https://github.com/FelixKrueger/TrimGalore.

Li, Q. & R. Wei. 2022. Comparison of boraginales plastomes: insights into codon usage bias, adaptive evolution, and phylogenetic relationships. *Diversity* 14: 1104.

Li, T., S. Zhang, Y. Deng, & Y. Li. 2024. Comparative analysis of chloroplast genomes for the genus *Manglietia* Blume (Magnoliaceae): molecular structure and phylogenetic evolution. *Genes* 15: 406.

Matzke, N.J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology* 63: 951–970.

Minh, B.Q., H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. Von Haeseler, & R. Lanfear. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.

Noroozi, M., F. Ghahremaninejad, D. Bogler, J.M. Witherspoon, G.L. Ryan, J.S. Miller, M. Riahi, & J.I. Cohen. 2022. Parsing a plethora of pollen: the role of pollen size and shape in the evolution of Boraginaceae. *Cladistics* 38: 204–226.

Noroozi, M., F. Ghahremaninejad, M. Riahi, & J.I. Cohen. 2024. Phylogenomics and plastome evolution of Lithospermeae (Boraginaceae). *BMC Plant Biology* 24: 957.

Okada, T. & K. Watanabe. 2024. The complete chloroplast genome sequence of *Lithospermum erythrorhizon*: Insights into the phylogenetic relationship among Boraginaceae species and the maternal lineages of purple gromwells. *Plant Gene* 37: 100447.

Park, I., S. Yang, J.-H. Song, & B.C. Moon. 2020. Dissection for floral micromorphology and plastid genome of valuable medicinal borages *Arnebia* and *Lithospermum* (Boraginaceae). *Frontiers in Plant Science* 11: 606463.

Pease, J.B., J.W. Brown, J.F. Walker, C.E. Hinchliff, & 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.

Peden, J.F. 2000. Analysis of codon usage. Doctoral dissertation. University of Nottingham, Nottingham, England.

Piot, A., J. Hackel, P.-A. Christin, & G. Besnard. 2018. One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta* 247: 255–266.

POWO. 2025. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. https://powo.science.kew.org/. (Accessed: 3 April 2025).

Ramírez, F., D.P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F. Dündar, & T. Manke. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* 44: W160.

Ripma, L.A., M.G. Simpson, & K. Hasenstab-Lehman. 2014. Geneious! Simplified genome skimming methods for phylogenetic systematic studies: A case study in *Oreocarya* (Boraginaceae). *Applications in Plant Sciences* 2: 1400062.

Robbins, E.H. & S. Kelly. 2025. Widespread adaptive evolution in angiosperm photosystems provides insight into the evolution of photosystem II repair. *The Plant Cell* 37: koae281.

Saarela, J.M., S.V. Burke, W.P. Wysocki, M.D. Barrett, L.G. Clark, J.M. Craine, P.M. Peterson, R.J. Soreng, M.S. Vorontsova, & M.R. Duvall. 2018. A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. *PeerJ* 6: e4299.

Salinas, N.R. & D.P. Little. 2014. 2matrix: A utility for indel coding and phylogenetic matrix concatenation1. *Applications in Plant Sciences* 2: 1300083.

Simpson, M.G., C.M. Guilliams, K.E. Hasenstab-Lehman, M.E. Mabry, & L. Ripma. 2017a. Phylogeny of the popcorn flowers: Use of genome skimming to evaluate monophyly and interrelationships in subtribe Amsinckiinae (Boraginaceae). *Taxon* 66: 1406–v1420.

Simpson, M.G., L.A. Johnson, T. Villaverde, & C.M. Guilliams. 2017b. American amphitropical disjuncts: Perspectives from vascular plant analyses and prospects for future research. *American Journal of Botany* 104: 1600–1650.

Straub, S.C., M. Parks, K. Weitemier, M. Fishbein, R.C. Cronn, & A. Liston. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

Straub, S.C., J. Boutte, M. Fishbein, & T. Livshultz. 2020. Enabling evolutionary studies at multiple scales in Apocynaceae through Hyb-Seq. *Applications in Plant Sciences* 8: e11400.

Sun, J., S. Wang, Y. Wang, R. Wang, K. Liu, E. Li, P. Qiao, L. Shi, W. Dong, & L. Huang. 2022. Phylogenomics and genetic diversity of Arnebiae Radix and its allies (*Arnebia*, Boraginaceae) in China. *Frontiers in Plant Science* 13: 920826.

Tate, J.A., V.V. Symonds, A.N. Doust, R.J. Buggs, E. Mavrodiev, L.C. Majure, P.S. Soltis, & D.E. Soltis. 2009. Synthetic polyploids of *Tragopogon miscellus* and *T. mirus* (Asteraceae): 60 Years after Ownbey's discovery. *American Journal of Botany* 96: 979–988.

Thiers, B.M. updated continuously. Index Herbariorum. https://sweetgum.nybg.org/science/ih/ (Accessed: 3 April 2025).

Tillich, M., P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, & S. Greiner. 2017. GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45: W6–W11.

Vasimuddin, M., S. Misra, H. Li, & S. Aluru. 2019. *Efficient architecture-aware acceleration of BWA-MEM for multicore systems*. Pp 314–324 *in*: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), Rio de Janiero, Brazil.

Wang, L., H. Xing, Y. Yuan, X. Wang, M. Saeed, J. Tao, W. Feng, G. Zhang, X. Song, & X. Sun. 2018. Genome-wide analysis of codon usage bias in four sequenced cotton species. *Plos One* 13: e0194372.

Wang, S., Y. Wang, J. Sun, X. Cui, E. Li, R. Wang, Q. Li, P. Zhang, W. Dong, & L. Guo. 2025. Comparative chloroplast genome analyses provide new insights into molecular markers for distinguishing Arnebiae Radix and its substitutes (tribe Lithospermeae, Boraginaceae). *Phytomedicine* 136: 156338.

Weigend, M., M. Gottschling, F. Selvi, & H.H. Hilger. 2009. Marbleseeds are gromwells–Systematics and evolution of *Lithospermum* and allies (Boraginaceae tribe Lithospermeae) based on molecular and morphological data. *Molecular Phylogenetics and Evolution* 52: 755–768.

Weigend, M., M. Gottschling, H.H. Hilger, & N.M. Nürk. 2010. Five new species of *Lithospermum* L. (Boraginaceae tribe Lithospermeae) in Andean South America: Another radiation in the Amotape-Huancabamba zone. *Taxon* 59: 1161-1179.

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

Wicke, S., G.M. Schneeweiss, C.W. Depamphilis, K.F. Müller, & D. Quandt. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76: 273–297.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* 24: 1586–1591.

Ye, W.-Q., Z.-Y. Yap, P. Li, H.P. Comes, & Y.-X. Qiu. 2018. Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). *Molecular Phylogenetics and Evolution* 127: 978–987.

Zhang, D., F. Gao, I. Jakovlić, H. Zou, J. Zhang, W.X. Li, & G.T. Wang. 2020. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular ecology resources* 20: 348–355.

Zhang, H., M. Liu, & B. Zi. 2024. cubar: Codon Usage Bias Analysis. Downloadable from: https://mt1022.github.io/cubar/.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.