



Phylogenomics, floral evolution, and biogeography of *Lithospermum* L. (Boraginaceae)

James Cohen¹

Kettering University, Applied Biology, 1700 University Avenue, Flint, MI 48503, USA

ARTICLE INFO

Keywords:

Boraginaceae

Distily

Diversification

Heterostyly

Hybridization-based enrichment

Lithospermum

ABSTRACT

Lithospermum (Boraginaceae), a geographically cosmopolitan medium-sized genus, includes diverse floral morphology, with variation in corolla size and shape and in breeding system. Over the past decade, multiple studies have examined the evolutionary history of *Lithospermum*, with most utilizing DNA regions from the plastid genome and/or the nuclear ribosomal internal transcribed spacer. These studies have, in general, not resulted in well-resolved and well-supported phylogenies. In the present study, 298 nuclear DNA regions, amplified via target sequence capture, were utilized for phylogenetic reconstruction for *Lithospermum* and relatives in Boraginaceae, and patterns of floral evolution, species diversification, and biogeography were examined. Based on multiple phylogenetic methods, *Lithospermum* is resolved as monophyletic, and the New World species of the genus are also monophyletic. While minimal phylogenetic incongruence is resolved within the nuclear genome, incongruence between the nuclear and plastid genomes is recovered. This is likely due to incomplete lineage sorting during early diversification of the genus in the Americas approximately 7.8 million years ago. At least four shifts to longer corollas are identified throughout *Lithospermum*, and this may be due to selection for hummingbird-pollinated flowers, particularly for species in Mexico and the southwestern United States. In the New World, one clade of species of the genus diversified primarily across the United States and Canada, and another radiated throughout the mountains of Mexico.

1. Introduction

Lithospermum (Boraginaceae) is a medium-sized genus characterized by smooth, white, lustrous, erect nutlets (fruits) that resemble little pieces of porcelain. The genus is the second largest in Boraginoideae (Chacón et al., 2016), with approximately 80 species, and is present on all continents, except Australia and Antarctica. *Lithospermum* is most diverse in the New World, with a center of diversity in Mexico and the southwestern United States, and small radiations have occurred in South Africa (Cohen et al., 2019) and the Amotape-Huancabamba Zone in the northern Andes (Weigend et al., 2010).

In *Lithospermum* flower morphology is quite variable. The genus includes the longest flowers in Boraginaceae (Cohen, 2018; Johnston, 1952, 1954) as well as some of the shortest, with flowers ranging from one to 120 mm in length. Additionally, multiple species exhibit heterostyly, a breeding system characterized by two or three floral morphs in a population with reciprocal positions of the anthers and stigmas (Cohen, 2018; Ganders, 1979; Johnston, 1952; Weller, 1980). These floral features have been examined in a phylogenetic context using

chloroplast DNA (cpDNA) regions. In *Lithospermum*, Cohen (2011, 2012) resolved multiple independent origins of increasing flower size and of heterostyly, with numerous origins of the latter resolved throughout the family, making the breeding system a derived trait (Chacón et al., 2019; Cohen, 2014). de Vos et al. (2014) recovered an increased rate of diversification for heterostylous species of *Primula* due to decreased rates of extinction, not faster speciation; although heterostyly is ancestral in *Primula*, not derived, as is the case in *Lithospermum* and other genera in Boraginaceae. The influence of diversification and shifts in rates of evolution for flower size and of heterostyly have yet to be investigated using comprehensive and well-resolved phylogenies (especially based on nuclear DNA regions) across multiple taxa that bear these features. Elucidating these shifts will allow for a greater understanding of the role floral morphology has played in diversification of *Lithospermum* and other taxa that exhibit variation in floral length and breeding system.

Multiple studies have examined patterns of biogeography throughout the cosmopolitan geographic range of *Lithospermum*, and different results have been recovered depending on the taxon-gene

E-mail addresses: jamescohen@weber.edu, jcohen@kettering.edu.

¹ Current address: Weber State University, Department of Botany and Plant Ecology, 1415 Edvalson St., Ogden, UT 84404, USA.

<https://doi.org/10.1016/j.ympev.2021.107317>

Received 23 June 2020; Received in revised form 29 August 2021; Accepted 15 September 2021

Available online 20 September 2021

1055-7903/© 2021 Elsevier Inc. All rights reserved.

combinations employed (Chacón et al., 2019; Chacón et al., 2017; Cohen, 2011, 2012; Weigend et al., 2009). The genus is resolved to have originated in Eurasia, but the number and order of colonization events to North America, South America, and South Africa (and possibly Eurasia again) has varied depending on the phylogeny. Regardless of the order, the genus has radiated throughout the Americas, especially in Mexico and the southwestern United States, and in these areas it has diversified at higher elevations; however, studies of montane diversification of *Lithospermum* have, so far, been restricted to South America (Weigend et al., 2010). Consequently, as with patterns of morphological character evolution, a well-resolved phylogeny can help provide a greater comprehension of the biogeographical history of the genus as well as of speciation across mountain ranges of Mexico, which includes a notable biogeographic region, the Mexican Transition Zone, where Neotropical and Nearctic floras intersect (Corona et al., 2007; Morrone, 2017).

Recent updates to the taxonomy of *Lithospermum* date to 2009 when Cohen and Davis employed 10 cpDNA regions, nine of which were non-coding, to examine the phylogenetic relationships of the genus and its relatives. These authors resolved that five New World genera, *Lasiarrhenum*, *Macromeria*, *Onosmodium*, *Perittostema*, and *Psilolaemus*, were nested within *Lithospermum* making this genus paraphyletic. These five genera had been segregated from *Lithospermum* due to autapomorphic characters, primarily related to floral morphologies (Johnston, 1952, 1954); therefore, Cohen and Davis (2009) broadened the generic circumscription of *Lithospermum* to include these genera. Other studies (Chacón et al., 2019; Weigend et al., 2009) that have used different taxon and DNA region sampling have recovered phylogenies with similar results, but the evolutionary relationships among species of *Lithospermum* have differed depending on the taxon-gene combinations. This may be expected (e.g., Rosenberg and Kumar, 2003), but the different relationships make it challenging to resolve fundamental aspects of the phylogeny of *Lithospermum*, such as which species and clades are sisters. Consequently, these discordant topologies influence our ability to confidently reconstruct patterns of floral evolution and biogeography in the genus. One reason this is the case is that the backbone relationships of *Lithospermum* have generally not been well-supported and/or well-resolved, even with the use of approximately 10 kilobases (kb) of cpDNA sequence data (Cohen and Davis, 2012). Additionally, the use of nuclear ribosomal DNA (nrDNA) has not allowed for greater clarity and congruence among studies, methods, and genomes (Weigend et al., 2010; Weigend et al., 2009).

Most phylogenetic studies of Boraginaceae have utilized cpDNA sequence data, usually two or three regions, and the nrDNA internal transcribed spacer (ITS) (Chacón et al., 2017; Selvi et al., 2017; Weigend et al., 2013). For genera, such as *Selkirkia* and *Megacaryon*, this strategy has proved useful for elucidating evolutionary relationships (e.g., Holstein et al., 2016; Selvi et al., 2017), but in many others, including *Cynoglossum*, *Nonea*, and *Lithospermum*, it has not been adequate to resolve relationships among species (e.g., Chacón et al., 2019; Pourghorban et al., 2020; Selvi et al., 2006). Ripma et al. (2014), Simpson et al. (2017), and Mabry and Simpson (2018) have conducted phylogenomic studies that have involved entire chloroplast genomes in order to reconstruct a phylogeny of Amsinckieae and genera within the subtribe. This has resulted in a more well-supported and well-resolved phylogeny of the group, and entire plastid genomes have been employed in multiple other taxa to resolve phylogenetic relationships (e.g., Heckenhauer et al., 2019; Zhang et al., 2017). While the use of entire chloroplast genomes is a large forward leap for phylogenetics (i.e., 120 kb is much larger than 10 kb), the relationships resolved only reflect the evolutionary history from one, maternally inherited genomic region (Birky, 2008; Doyle, 1992). Another approach for phylogenetic analyses involves the use of target sequence capture, a method involving hybridization of RNA baits to regions of the genome for subsequent amplification and sequencing. This strategy can yield DNA sequences for hundreds of independent markers from across the three plant genomes, including the biparentally inherited nuclear genome (Cronn et al., 2012;

Lemmon et al., 2012; Lévillé-Bourret et al., 2018). Target sequence capture is a useful tool that can take advantage of transcriptomic and genomic information from across borages to identify appropriate DNA regions to design probes for capture and amplification, and the resulting sequence data can be used for reconstructing phylogenies and examining patterns of character evolution and of historical biogeography.

The present study is an examination of the phylogenomics of *Lithospermum* and relatives through the use of hundreds of DNA regions from across the nuclear genome. The resulting phylogeny is utilized to examine the evolutionary history of the genus and for investigations of patterns of character evolution, macroevolutionary shifts, and biogeography. Based on previous studies of *Lithospermum* and relatives, the present study includes four hypotheses: 1) the phylogeny of *Lithospermum* resolved from the nuclear DNA regions will differ from that of prior phylogenies reconstructed primarily from cpDNA data; 2) shifts in rates of evolution for corolla length will involve transitions from shorter to longer corollas; 3) a phylogeny based on nuclear DNA regions will resolve multiple origins of heterostyly, just as with the cpDNA data; and 4) *Lithospermum* originated in the Old World and colonized the New World once, which coincided with an increased rate of diversification.

2. Materials and methods

2.1. Taxon sampling and DNA isolation

A total of 54 species from across Boraginaceae were included in the present study, and eight, two, and one of these species included two, three, and four individuals, respectively (Table 1). Thirty-four species are from *Lithospermum*, the ingroup, and 20 are from other genera of Boraginaceae. The sampling represents members of both tribes of Boraginoideae and three tribes of Cynoglossoideae: Asperugeae, Rocheleae, and Cynoglosseae (sensu Chacón et al., 2016). Sampled species were either collected from wild populations or obtained from horticultural collections (Table 1), and sampling represents morphological diversity throughout the genus and family. Voucher specimens for collections specific to the present study were deposited at the University of Michigan Herbarium (MICH) or the Mary Carver Hall Herbarium at Weber State University (WSCO). Leaf material was preserved in silica gel prior to DNA isolation, which was conducted using either a modified CTAB extraction following the methodology of Cohen and Davis (2009) or the Norgen Plant/Fungi DNA Isolation Kit following the manufacturer's instructions (Norgen Biotek Corp., Thorold, ON, Canada).

2.2. Target sequence capture

Three hundred and forty-nine genes covering ca. 500,000 base pairs (bp) of genomic DNA were targeted for the present project. 80-mer RNA baits, tiled at 2X density, were designed by MYcroarray (now Arbor Biosciences, Ann Arbor, MI, USA) to capture these genes. Sequences for bait design were derived from two sources: the majority were identified as putative low-copy nuclear genes via BLAST searches, using default parameters, (Johnson et al., 2008) of arbitrary transcripts from the *Lithospermum multiflorum* transcriptome (Cohen, 2016a), and a small number were obtained from nuclear DNA regions of *Lithospermum erythrorhizon* deposited in GenBank. At MYcroarray, target-enrichment libraries for each sample were prepared following the recommended protocol of the company and then sequenced using an Ion Proton (ThermoFisher, Waltham, MA, USA). FASTQ files were generated using the default post-processing settings of the Torrent_Suite v4.4 (<https://github.com/iontorrent/TS>), and adapter and low-quality bases were trimmed from the 3' end of reads.

The FASTQ files for each sample were mapped to the original reads for bait design, using BWA-MEM (Li, 2013), as implemented in Galaxy (usegalaxy.org), following the simple Illumina mode. Using Geneious v8 (Kearse et al., 2012), a consensus sequence was produced, based on a minimum depth of 20X coverage per nucleotide, for each gene region for

Table 1
Sequencing and mapping statistics for species in analysis.

Species	Collection	# of reads	# of reads mapped	% mapped	# of bases 20X coverage	% of alignment with 20X coverage
<i>Aegonychon purpureocaeruleum</i>	Botanic Garden Meise 19,792,084	6,194,671	2,944,079	48%	293,420	54.79%
<i>Anchusa officinalis</i>	Cohen 179	3,083,797	1,559,380	51%	13,012	2.43%
<i>Borago officinalis</i>	Cohen 172	1,461,867	1,027,613	70%	13,117	2.45%
<i>Brunnera macrophylla</i>	Cohen 169	5,001,959	3,131,531	63%	20,724	3.87%
<i>Buglossoides arvensis</i>	Cohen 355	7,678,612	4,185,852	55%	234,672	43.82%
<i>Cerinth</i>	Plume 65	2,841,569	1,672,376	59%	20,089	3.75%
<i>Dasynotus daubenmirei</i>	Cohen 403	1,720,774	1,260,227	73%	12,325	2.30%
<i>Glandora rosmarinifolia</i>	Plume 73	2,560,603	960,370	38%	38,593	7.21%
<i>Hackelia floribunda</i>	Cohen 256	2,100,604	738,414	35%	8882	1.66%
<i>Hyunhia pulchra</i>	Cohen 260	3,701,663	2,478,508	67%	111,327	20.79%
<i>Hyunhia pulchra</i>	Kew 980,864	2,339,019	1,287,692	55%	9598	1.79%
<i>Lithospermum bejariense</i>	Cohen 375	3,753,959	2,226,944	59%	291,045	54.35%
<i>Lithospermum calcicola</i>	Cohen 191	4,005,534	2,761,458	69%	290,109	54.17%
<i>Lithospermum calcicola</i>	Cohen 187	1,386,828	733,178	53%	388,930	72.63%
<i>Lithospermum calycosum</i>	Cohen 197	1,696,482	1,163,009	69%	30,583	5.71%
<i>Lithospermum canescens</i>	Cohen and Straub 12	8,307,060	6,245,259	75%	385,350	71.96%
<i>Lithospermum canescens</i>	Cohen 1	1,583,690	1,192,004	75%	79,519	14.85%
<i>Lithospermum canescens</i>	Cohen 109	8,382,617	3,720,107	44%	363,860	67.94%
<i>Lithospermum carolinense</i>	Cohen 6	3,474,279	2,554,960	74%	175,311	32.74%
<i>Lithospermum cobrense</i>	Cohen 78	4,733,345	2,848,510	60%	331,530	61.91%
<i>Lithospermum cobrense</i>	Cohen 203	2,587,459	2,158,478	83%	383,562	71.62%
<i>Lithospermum discolor</i>	Cohen 216	618,098	416,610	67%	1867	0.35%
<i>Lithospermum distichum</i>	Cohen 202	4,432,787	3,065,661	69%	367,781	68.68%
<i>Lithospermum distichum</i>	Cohen 224	2,355,421	1,737,811	74%	383,859	71.68%
<i>Lithospermum erythrorhizon</i>	Cohen 173	2,144,860	1,828,317	85%	394,718	73.71%
<i>Lithospermum exsertum</i>	Cohen 217	9,134,145	7,011,828	77%	399,028	74.51%
<i>Lithospermum flavum</i>	Cohen 226	4,056,661	2,871,251	71%	391,945	73.19%
<i>Lithospermum helleri</i>	Cohen 132	5,011,170	3,679,067	73%	386,498	72.17%
<i>Lithospermum incisum</i>	Cohen 371	6,879,018	5,805,677	84%	390,788	72.97%
<i>Lithospermum incisum</i>	Cohen and Straub 35	1,300,124	905,896	70%	377,987	70.58%
<i>Lithospermum johnstonii</i>	Cohen 218	6,835,001	4,647,172	68%	385,885	72.06%
<i>Lithospermum latifolium</i>	Cohen 220	10,710,568	2,220,337	21%	376,921	70.38%
<i>Lithospermum latifolium</i>	Cohen 87	3,155,685	1,526,305	48%	388,862	72.61%
<i>Lithospermum leontotis</i>	Cohen 195	4,910,951	1,750,240	36%	387,548	72.37%
<i>Lithospermum macromeria</i>	Cohen 141	6,827,358	4,453,738	65%	386,047	72.09%
<i>Lithospermum mirabile</i>	Cohen 180	6,164,281	4,367,131	71%	396,972	74.13%
<i>Lithospermum mirabile</i>	Cohen and Straub 64	6,075,879	5,214,853	86%	397,569	74.24%
<i>Lithospermum mirabile</i>	Cohen 83	6,818,196	5,309,716	78%	386,240	72.12%
<i>Lithospermum molle</i>	Missouri Botanical Garden 2003–0671	9,423,321	7,418,219	79%	352,328	65.79%
<i>Lithospermum multiflorum</i>	Cohen 57	5,163,302	4,037,702	78%	256,621	47.92%
<i>Lithospermum multiflorum</i>	Cohen 81	7,819,234	6,135,288	78%	418,714	78.19%
<i>Lithospermum nelsonii</i>	Cohen 184	5,792,260	4,354,733	75%	392,480	73.29%
<i>Lithospermum notatum</i>	Cohen 188	2,113,891	1,169,908	55%	345,230	64.47%
<i>Lithospermum oblongifolium</i>	Cohen 201	6,282,654	3,839,135	61%	390,923	73.00%
<i>Lithospermum obovatum</i>	Cohen 208	3,719,047	2,918,678	78%	398,643	74.44%
<i>Lithospermum occidentale</i>	Cohen 122	2,779,697	2,064,877	74%	386,727	72.21%
<i>Lithospermum officinale</i>	Cohen 188	2,388,229	2,026,239	85%	398,246	74.36%
<i>Lithospermum revolutum</i>	Cohen 199	5,414,477	4,282,705	79%	399,026	74.51%
<i>Lithospermum rosei</i>	Cohen 207	6,972,074	4,641,506	67%	408,790	76.33%
<i>Lithospermum ruderales</i>	Cohen 31	3,415,917	2,709,194	79%	393,981	73.57%
<i>Lithospermum ruderales</i>	Cohen 32	7,811,729	6,385,897	82%	397,273	74.18%
<i>Lithospermum ruderales</i>	Cohen 156	2,653,223	1,323,405	50%	232,771	43.47%
<i>Lithospermum ruderales</i>	Cohen 157	3,341,702	2,512,883	75%	397,296	74.19%
<i>Lithospermum strictum</i>	Cohen 222	3,358,980	2,771,604	53%	391,875	73.18%
<i>Lithospermum trinervium</i>	Cohen 228	9,902,496	6,815,652	69%	404,139	75.47%
<i>Lithospermum tuberosum</i>	Cohen 108	4,611,953	2,622,672	57%	386,711	72.21%
<i>Lithospermum tubuliflorum</i>	M. Gonzalez 4001	5,948,947	1,072,458	18%	357,917	66.83%
<i>Lithospermum viride</i>	Cohen 82	7,186,465	4,999,277	70%	406,257	75.86%
<i>Lithospermum viride</i>	Cohen 86	10,803,774	7,971,846	74%	407,022	76.00%
<i>Mertensia bella</i>	Cohen 404	1,408,890	1,031,797	73%	7844	1.46%
<i>Moltkia petraea</i>	Botanic Garden Meise 2001260–54	8,392,929	1,499,292	18%	54,956	10.26%
<i>Oncaglossum pringlei</i>	Cohen 219	2,382,782	690,057	29%	14,172	2.65%
<i>Onosma rigida</i>	Brooklyn Botanical Garden 2006–1311	4,759,850	1,753,071	37%	64,391	12.02%
<i>Oreocarya crassipes</i>	Cohen 370	2,376,490	1,072,150	45%	13,278	2.48%
<i>Oreocarya fulvocanescens</i>	Cohen 394	2,905,690	1,871,266	64%	10,770	2.01%
<i>Paracaryum racemosum</i>	Cohen 259	3,076,684	444,878	14%	7648	1.43%
<i>Pontechium ruscicum</i>	DBG 990,415	8,263,856	1,780,087	22%	51,314	9.58%
<i>Pulmonaria angustifolia</i>	DBG 981,057	6,809,756	2,515,928	37%	22,179	4.14%
<i>Symphytum orientale</i>	Oxford Botanical Garden 0,001,865	3,001,370	2,079,763	69%	16,673	3.11%

each species, with 298 having sufficient sequence-taxon coverage for inclusion in phylogenetic analyses. Any nucleotide with less than 20X coverage was coded as N, and the appropriate IUPAC ambiguity code was used for polymorphic nucleotide positions. Sequences for each of the 298 DNA regions were aligned using MAFFT (Kato and Standley, 2013), as implemented in Geneious, with default settings. FASTQ files were submitted to GenBank (BioProject PRJNA641211), and alignments are available at TreeBASE (ID 28630).

2.3. Phylogenetic analyses

For each of the 298 DNA regions, Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI) phylogenetic analyses were undertaken with TNT (Goloboff and Catalano, 2016), RAXML v8.2.12 (Stamatakis, 2014), and MrBayes v3.2.5 (Ronquist et al., 2012), respectively. Analyses for each gene region in TNT, which were spawned from Winclada (Nixon, 2002), used the following settings: a random seed of 0 (random starting point), 1,000,000 trees held in memory, 1,000 parsimony ratchets with a 10% probability of upweighting and a 10% probability of downweighting, 1,000 iterations of tree drifting, and 100 rounds of tree fusing (Goloboff, 1999; Nixon, 1999). All most-parsimonious (MP) trees were retained, and a Nelson consensus tree was produced from all MP trees for each gene region. ModelTest-NG (Darriba et al., 2020) was used to identify the most appropriate model for sequence evolution for each individual DNA region, based on Bayesian Information Criterion (BIC), for ML analyses. Maximum Likelihood analyses for each gene region were undertaken with RAXML using the appropriate model of sequence evolution to search for the best-scoring ML tree, and 1,000 bootstrap replicates were performed. With MrBayes, each gene region was analyzed twice using the following parameters: 10 chains run for 5,000,000 generations, with chains sampled every 1,000 generations, a heating parameter of 0.2, a GTR + I + G model with four gamma categories, branch lengths with an exponential prior with a parameter of 10, an exponential shape of the gamma distribution with a parameter of 10, and burnin of 25%.

The results of the ML and BI analyses of the individual DNA regions were used for Coalescent Model (CM) and Bayesian Concordance (BC) analyses, respectively. The best-scoring ML trees and bootstrap files, from RAXML, were employed as input for Astral v5 (Zhang et al., 2018) to estimate species trees using a multi-species CM. For analyses in Astral, default settings were used, with the exception of 500 bootstrap replicates undertaken. Additionally, analyses were run with species that include multiple individuals both forced and not forced to be monophyletic (-a option). The hemiplasy risk factor (HRF) was calculated with P(e):P(o) based on the CM trees, with five different mutation rates in coalescent units (0.01, 0.001, 0.0000001, 0.00000001, and 0.000000001) to examine the impact of mutation rate on HRF (Guerrero and Hahn, 2018). For BC, the results of the two BI runs for each gene were analyzed together with mbsum with a 20% burnin, and the results were subsequently analyzed with BUCKy (Larget et al., 2010). Given the amount of missing data, five different sets of genes based on taxon sampling were used for analyses with BUCKy – 50, 100, 150, 200, and 225 genes (with increasing missing taxa with a greater number of DNA regions) – in order to examine the impact of taxon and gene sampling (Appendix A). For each of these sets of genes, the following parameters were employed in BUCKy, ten replicates with four chains run for 1,000,000 generations, and each of these analyses was conducted with four levels of gene-tree independence (alpha of 0.1, 1, 10, and 100) to investigate the influence of the level of discordance, on the resulting phylogeny. The resulting phylogenies were visualized with BUCKy's visualization tool (<http://ane-www.cs.wisc.edu/buckytools/buckytools.php>).

Along with analyses involving individual gene regions, analyses of matrices of concatenated gene regions also were conducted to explore the impact of missing data on the resulting phylogeny. For these analyses, four matrices were created: 1) all 298 aligned genes, 2) at least 60

individuals per aligned gene (nine genes), 3) at least 50 individuals per aligned gene (40 genes), and 4) at least 40 individuals per aligned gene (232 genes), and these are referred to as the all/298-gene, 9-gene, 40-gene, and 232-gene dataset, respectively (Appendix A). For each matrix, simple indel coding (Simmons and Ochoterena, 2000) was undertaken with 2matrix (Salinas and Little, 2014), resulting in an additional four matrices that also included presence/absence of gaps. Concatenated matrices were analyzed with MP (TNT), ML (RAXML), and BI (BEAST) (Drummond and Rambaut, 2007; Drummond et al., 2012a). TNT analyses of the eight matrices (all/298-gene, 9-gene, 40-gene, and 232-gene dataset with and without gaps) were the same as described above, but with the addition of random sectorial searches to the search strategy, with the ratchet, drifting, sectorial searches, and tree fusing finding the best tree 10 times before TBR max, swapping among the MP trees. Analyses of the four matrices of only DNA sequence data were conducted using RAXML v8 on CIPRES (www.phylo.org) or the Kettering University High-Performance Computing Cluster (KUHPC), with four partitioning approaches: 1) the entire matrix as one partition using GTR, GTR + I, CAT, and CAT + I, 2) each DNA region as its own partition using GTR, GTR + I, CAT, and CAT + I per partition, and 3 and 4) partitions identified with PartitionFinder2 (Lanfear et al., 2017) on CIPRES, based on models available in RAXML, using both the rcluster (Lanfear et al., 2014) and greedy algorithms to identify optimal partitioning schemes. The best-scoring ML tree was resolved via a thorough search, and 1,000 rapid bootstrap replicates were also run. Analyses with BEAST were conducted, using CIPRES, for the four DNA-only datasets. For these analyses, a Yule model of speciation and a GTR + I + G model of sequence evolution was employed. The analyses were run for 160 million generations, with samples taken every 1,000 generations. The output was visualized with Tracer v1.6 (Rambaut et al., 2018) to determine the number of trees to use as burnin for the consensus tree. The maximum clade credibility consensus tree was resolved with TreeAnnotator v1.8.4 (Helfrich et al., 2018), as implemented with CIPRES, using 64,000 trees as burnin and a posterior probability limit of 0.9. Results were compared among the various types of analytical methods as well as with results of a 10-cpDNA-region phylogeny of *Lithospermum* and relatives from Cohen (2012).

To explore measures of support for various clades, all of the individual DNA regions were concatenated and analyzed using the partition bootstrap (blockboot) and jackknife (blockjack) (Siddall, 2010), which treats each gene region as a unit for resampling. Additionally, resampling of partitions and of partitions and sites within partitions was conducted with IQ-TREE (Nguyen et al., 2015) on the KUHPC using 1,000 bootstrap and jackknife replicates with a GTR + I + G model of sequence evolution. Partitioned coalescent support (PCS) (Gatesy et al., 2017; Gatesy et al., 2019) also was examined with Astral v5 to examine the number of genes supporting different topologies, and these analyses were conducted with various combinations of RAXML consensus trees and individual gene trees using all 298 regions and the 9-gene, 40-gene, and 232-gene matrices. BOOSTER (Lemoine et al., 2018) was employed to calculate transfer bootstrap support.

To examine the potential role of hybridization in the genus, JML (Joly, 2012) was employed. JML uses trees generated via analyses in *beast (Drummond et al., 2012b) to test for patterns of introgression between species pairs. For the present study, one individual per species was arbitrarily selected and two matrices were constructed – concatenated sequences of 1) the 298 nuclear DNA regions and 2) the 10 cpDNA regions of Cohen and Davis (2012). Both matrices were culled to include the same taxon sampling, 34 species (32 of *Lithospermum*, one of *Aegonychon*, and one of *Buglossoides*). For *beast analyses, a birth–death model, a relaxed log-normal clock, and a GTR + I + G model of sequence evolution were used, and the analyses were run for 500 million generations, with samples taken every 5,000 generations. For JML, analyses were run with various combinations of DNA regions and trees (i.e., the nuclear DNA tree analyzed with nuclear DNA sequences and separately with cpDNA sequences, and the same with the cpDNA tree). For each

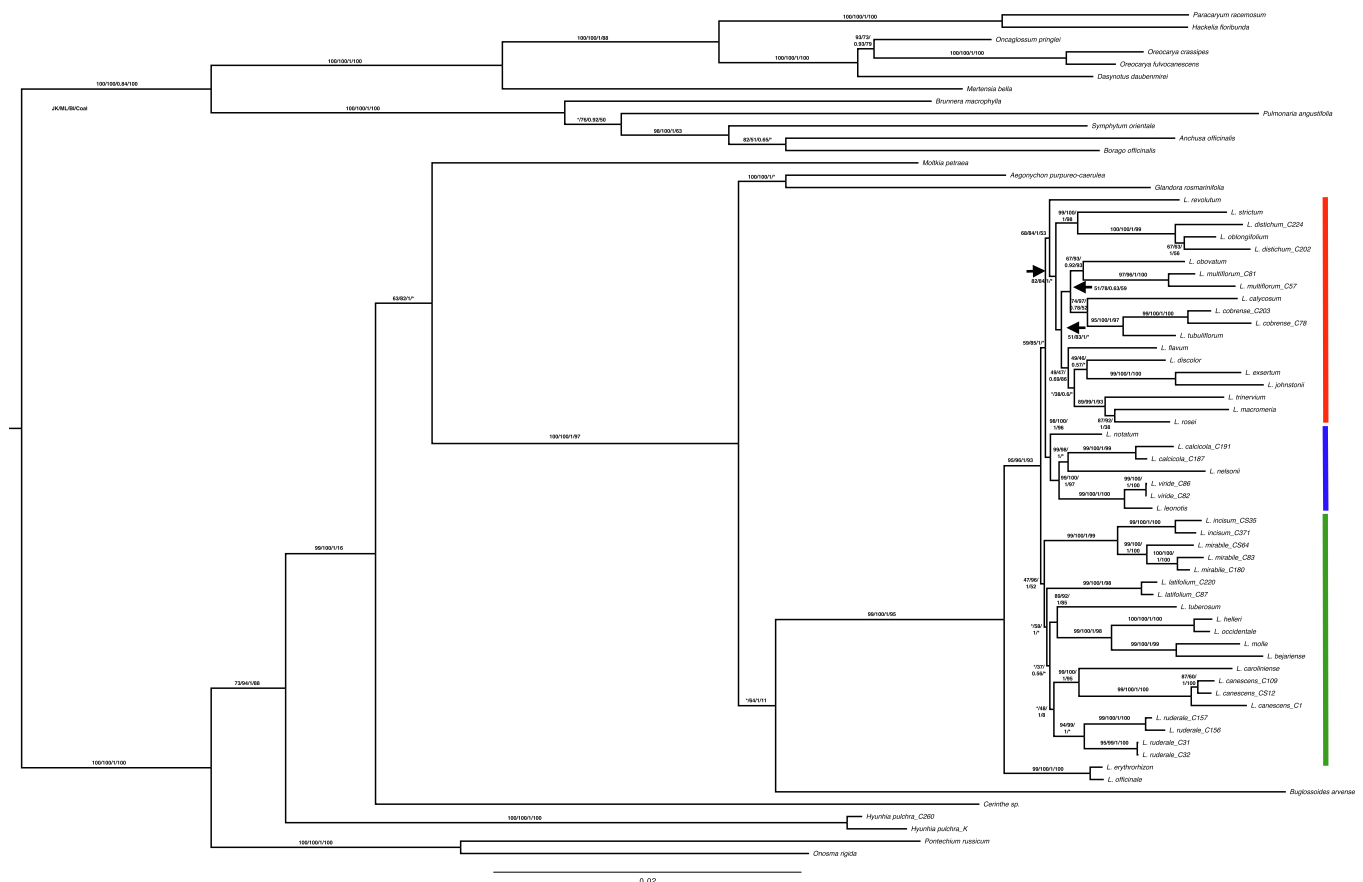
JML analysis, 90% of the *beast trees were treated as burnin, and the remaining 10,000 trees were used for simulations, with a significance level of 0.1 and a thinning of 4. The relative mutation rate was 1, and heredity scalar was 1 for nuclear DNA and 0.5 for cpDNA data. Results were compared within and among analyses. Species delimitations were explored, with a matrix of 295 individual DNA regions for *Lithospermum* and close relatives in *Aegonychon*, *Buglossoides*, and *Glandora*, using BPP (Yang, 2015). A01, A10, and A11 analyses were run on the KUHPC for 2.5 million generations, with trees sampled every 5th generation, and a 50,000 generation burnin.

2.4. Diversification, trait evolution, and biogeographic analyses

For analyses of diversification, evolution, and biogeography, a dated phylogeny was reconstructed with BEAST2 (Bouckaert et al., 2014), as implemented on the KUHPC, using only one arbitrarily selected individual per species of *Lithospermum* and close relatives in *Aegonychon*, *Buglossoides*, and *Glandora*. For this analysis, the dataset with all genes and no gaps was run for one billion generations, with trees stored every 1,000 generations, using a GTR + I + G model of sequence evolution and a fossilized birth–death model. The fossil outgroup of *Lithospermum dakotense* was used for dating (Gabel, 1987), and this fossil was placed at the crown of the New World species of *Lithospermum* and dated at a minimum of 6.6 million years ago (MYA) with a log normal distribution that includes a mean of 1.25 and standard deviation of 1 (Cohen, 2012). This analysis was run twice, and a maximum clade credibility consensus tree was resolved with TreeAnnotator v1.8.4, using 750,000 trees as burnin and a posterior probability limit of 0.9, for the result with the greatest posterior probability.

Patterns of speciation and extinction in *Lithospermum* and relatives were analyzed in Diversitree (FitzJohn, 2012), BAMM 2.5 (Mitchell and Rabosky, 2017), and MiSSE (Beaulieu and O'Meara, 2016). Analyses in Diversitree compared a full model and Yule model for speciation and extinction, and these Markov chain Monte Carlo (MCMC) analyses involved 1,000,000 steps with sampling every 10,000 steps. Following each analysis, a 400,000 step burnin was applied, and the probability density of the estimates of speciation and extinction was graphed. For analyses in BAMM, parameters include a global sampling fraction of 0.5, four MCMC chains run for 200 million generations, with sampling every 10,000 generations, and priors determined via setBAMMpriors in BAMMtools (Rabosky et al., 2014). To compare the results of patterns of diversification based on the nuclear DNA phylogeny in the present study with that of a phylogeny based on cpDNA with a different topology, but similar taxon sampling, the tree of Cohen (2012), which is based on 10 cpDNA regions reconstructed with BI in BEAST, was also analyzed with BAMM using the same parameters. All analyses were run for 200 million generations. The results of the BAMM runs were analyzed in BAMMtools, and an 80% burnin was used for the event data to understand shifts in rates of diversification. In BAMMtools, Rate Through Time plots were created for the two clades of Mexican species individually and collectively, for the US species, and for *Lithospermum* and close relatives (Fig. 1). MiSSE was employed to examine the number of distinct rates across the phylogeny and up to six potential rate categories were examined, with results also averaged across the rate categories to account for uncertainty in the models and a sampling fraction of 0.5. In MiSSE, the Akaike Information Criterion (AIC) was used to identify the preferred model.

Along with overall patterns of diversification in *Lithospermum*, the



evolutionary patterns of floral morphology were explored in multiple manners. All character data were taken from previous studies on *Lithospermum* (Cohen, 2011, 2012). Corolla type (i.e., long-funnelform flowers vs other types of flowers [Cohen, 2016b]) and heterostyly/homostyly were each treated as a binary character using BiSSE and HiSSE (Beaulieu and O'Meara, 2016; FitzJohn, 2012). BiSSE used the same parameters as described above for Diversitree, and given the greater number of parameters, different models for diversification (i.e., various combinations of equal vs unequal λ , μ , and q values) were analyzed in a ML framework and compared using a chi-squared analysis of AIC. Additionally, sampling fractions for each morphological character were included in the analyses. For analyses of each character in HiSSE, multiple net turnover, extinction fraction, and transition matrices were employed, and the various combinations included four, two, and one rate(s) for each net turnover and extinction fractions and transition matrices of 12 rates, eight rates (no dual transitions), three rates (one rate among hidden states and two rates for transitions of observed states), CID4, CID2, and BiSSE. This resulted in a total of 54 models examined, and the AIC was used to identify the preferred model. Corolla type was also examined, with MuSSE, using three different types (i.e., long-funnelform vs. campanulate/urceolate flowers vs. salverform/funnelform flowers), and following the same approach as for BiSSE. Additionally, patterns of floral morphology for flower type (both binary and multi-state) and heterostyly/homostyly were reconstructed with corHMM using the equal rates (ER), symmetrical (SYM), and all rates different (ARD) models (Beaulieu, 2020).

Along with corolla type treated as a qualitative character, corolla length was examined as a quantitative character using BAMM (Mitchell and Rabosky, 2017), I1ou (Khabbazian et al., 2016), SURFACE (Ingram and Mahler, 2013), and ParSplit (Didier et al., 2019). Quantitative data for the minimum, midpoint, and maximum corolla lengths were taken primarily from Cohen (2016b), and additional data, particularly for outgroups, were gathered from the literature (Brach and Song, 2006; Ferrero et al., 2011). Analyses in BAMM were the same as for diversification, including the creation of Rate Through Time plots, although analyses were run for only 100 million generations for the nDNA phylogeny. In I1ou, an OU model was fit to the three corolla-length datasets and BEAST2 phylogeny, and the number of rate shifts (both total and convergent) was estimated and visualized. SURFACE was employed to identify areas of the tree resolved to have undergone different and convergent selective regimes, for the three datasets of corolla lengths. Additionally, to investigate the data in an MP framework, evolutionary shifts were investigated with ParSplit. The three corolla lengths were examined using both inverse and identity functions, and these analyses were undertaken on the KUHPG.

Along with patterns of floral evolution, chromosome number evolution was examined. Chromosome numbers for thirteen species of *Lithospermum* and relatives were retrieved from the Chromosome Count Database (<http://ccdb.tau.ac.il>), and examined with ten models for patterns of chromosome evolution as implemented in ChromEvol (Glick and Mayrose, 2014), with models differing based on the rate of chromosome number increase, decrease, and duplication and demiploidy, whether change in chromosome number follows a constant or a linear function, and if there is a common base number for the clade. The most appropriate model was determined based on AIC.

Patterns of historical biogeography were examined for *Lithospermum* and close relatives. Geographic areas were broadly defined to include Eurasia, Eastern United States and Canada (east of 105° W), Western United States and Canada (west of 105° W), Eastern and northeastern Mexico (Sierra Madre Oriental and adjacent areas), Western and Northwestern Mexico (Sierra Madre Occidental and adjacent areas), and Central and Southern Mexico (Trans-Mexican Volcanic Belt, Sierra Madre del Sur, and adjacent areas). While some widespread species are present in more than one geographic area, these areas were delimited based on geographic ranges of species, as determined by the author and from floras and revisions (Boyd, 2002; Boyd, 2003; Cohen, 2018;

Johnston, 1952, 1954; Turner, 1995a,b), in order to capture appropriate biogeographic units across the genus. Using the dated phylogeny from BEAST2, the biogeographic history of the genus and close relatives was reconstructed using BioGeoBEARS, with the Dispersal-Extinction-Cladogenesis (DEC), Dispersal-Vicariance (DIVA)-like, and BayArea-like models, with and without the use of founder events (jump of BioGeoBEARS) (Matzke, 2012). The model that best fits the data was determined using a likelihood-ratio test, AIC, and corrected AIC (AICc). Analyses were run twice, with a maximum range size of three and of four.

3. Results

3.1. Taxon and sequence sampling

A total of 326,308,263 reads was produced among the 69 individuals, and statistics on the species and the DNA regions are in Table 1 and Appendix A, respectively. The greatest number of reads per individual was 10,803,774 for *L. viride*, and the smallest number was 618,098 for *L. discolor*. In general, over 50% of the reads mapped to the 349 genes; although, this number was less for 15 of the samples. Most of the species with poorer mapping were members of the outgroup. This was also the case for the percentage of aligned bases of the final alignment (535,530 bp in length) with greater than 20X coverage. Most members of the outgroup had fairly poor 20X coverage, ranging from 10,000 to 23,000 bp, which is less than 10% of the aligned sequence data; however, close relatives, such as *Aegonychon* and *Buglossoides* had 20X coverage closer to 50%. Most species of *Lithospermum* had at least 75% of the aligned sequence data with 20X coverage, but a few species (e.g., *L. calycosum* and *L. discolor*) had lower coverage. The mean coverage for bases covered by reads is 41X.

Of the 349 DNA regions included in the present study, 298 had sufficient sequence-taxon coverage for the final analyses, and this data set included all 69 individuals and a final aligned length of 535,530 bp (all-gene dataset). The 9-, 40-, and 232-gene matrices are 18,259 bp, 83,691 bp, and 444,500 bp in length, respectively. The longest aligned DNA region is 5,159 bp; the shortest is 463 bp; and the mean length of all DNA regions is 1,797 bp. Of the 298 included gene regions, the mean number of individuals included in an alignment is 44. The 9-, 40-, 232-, and all 298-gene matrices had nine, 34, 173, and 233 partitions, respectively, based on the rcluster algorithm, and eight, 31, 110, and 147 partitions based on the greedy algorithm.

3.2. Phylogenetic analyses

Outgroup relationships are similar, if not identical, among MP, ML, BI, CM, and BC analyses. Cynoglossoideae and Boraginoideae are each resolved as monophyletic, as are Boraginoideae and Lithospermeae. *Aegonychon purpureocaeruleum* and *Glandora rosmarinifolia* are sisters (81–100% bootstrap, jackknife, blockboot, and blockjack, support [BS, JK, BB, and BJ, respectively], 1.00 posterior probability [PP]). In ML, BI, and CM analyses, this clade is sister to one composed of *Buglossoides arvensis* and *Lithospermum* (61–100% BS, 1.00 PP); however, in MP analyses, this clade is sister to *Lithospermum*, and *B. arvensis* is sister to one comprising *A. purpureocaeruleum*, *G. rosmarinifolia*, and *Lithospermum* (99–100% BS and JK, 73–84% BB and BJ).

In all analyses, *Lithospermum* is monophyletic (100% BS and JK, 93% BB and BJ, 1.00 PP). The Eurasian species of *Lithospermum*, *L. erythrorhizon* and *L. officinale*, are sisters (100% BS, JK, BB, BJ, 1.00 PP), and the clade of these two species is sister to the other species of the genus (65–100% BS and JK, 64–66% BB and BJ, 1.00 PP), which in the present study are all native to the New World. The New World species are also monophyletic (95–96% BI and JK, 64–66% BB and BJ, 1.00 PP). Among the presently studied New World members of *Lithospermum*, three main clades are resolved, and relationships within these clades, based on the all-gene dataset, differ slightly among MP, ML, BI, CM, and

BC analyses. Multiple small differences are noted between ML, BI, and CM analyses, with most involving species or clades being successive sister to a clade rather than sister, or vice-versa, among analyses (Fig. 1, Appendix Fig. B.7). Two notable differences are the placement of *L. discolor* and the relationship of the clade that includes *L. nelsonii*, *L. notatum*, and three other species. *L. discolor* is sister to a clade composed of *L. exsertum* and *L. johnstonii* in ML and BI analyses (38–46% BS) but sister to a clade comprising individuals of *L. ruderales* in CM (10% BS) (*L. discolor* was not included in BC analyses due to low gene-region sampling). In ML and BI analyses, the clade of *L. nelsonii*, *L. notatum*, and three other species (Fig. 1, blue bar) is sister to a large clade of Mexican species comprising *L. revolutum*, *L. distichum*, and 14 other species (Fig. 1, red bar), and this relationship has low to moderate support (59–85% BS, 21–25% BB and BJ, 1.00 PP); however, in CM analyses, this five-species clade is sister to one composed of species in the US (Fig. 1, green bar), and this has low support (20%). The backbone of the clade of New World species of *Lithospermum* is the region of the phylogeny with HRF consistently at or near one among all mutation rates. Of the three aforementioned clades, two include species that are entirely or predominantly in Mexico, with one exception – *L. multiflorum*, a species that reaches its southern boundary in northern Mexico but is primarily in the United States – and the other clade is composed of species entirely or predominantly in the United States.

Among the clade comprising primarily US species (Fig. 1, green bar), five small clades are recovered that are congruent among the various methodologies and that receive high support (81–100% BS, JK, BB, and BJ), although relationships among these clades differ depending on the type of analysis. The clades include: 1) four individuals of *L. ruderales*, 2) *L. incisum* and *L. mirabile*, 3) *L. tuberosum* and the former members of *Onosmodium*, 4) *L. canescens* and *L. caroliniense*, and 5) two individuals of *L. latifolium*. While evolutionary relationships among these clades are well supported in BI analyses (1.00 PP), support values are less than 50% (BS, JK, BB, and BJ) in MP and ML analyses, with some different relationships recovered in CM analyses (Appendix Fig. B.7). In ML analyses, branch lengths among these clades are shorter than those for the Mexican species. All clades were supported by at least 63% transfer bootstrap support, with all but four having greater than 80%. Transfer bootstrap support was always equal to or greater than bootstrap support calculated by RAxML.

Among the various methodologies, evolutionary relationships reconstructed for each dataset are similar and congruent, but larger differences are observed among datasets. Most of the differences among the resulting consensus trees from the various datasets involve the placement of clades of two to five species of *Lithospermum* (Appendix B).

The use of partitions in ML analyses did not strongly influence resulting phylogenies, especially among the various partition schemes and models for a data matrix. For the all 298-gene, 9-gene, and 232-gene datasets, the three partition schemes involving multiple partitions resolved phylogenies similar to those with only one partition, and BS support values were similar as well. For the 40-gene dataset, a few differences are apparent between the partitioned datasets and that with only one partition, particularly for clades with low BS support (less than 50%). The most evident difference was that the clade that includes four individuals of *L. ruderales* was sister to the rest of the New World species, although the positions of *L. discolor* and *L. revolutum* also differ. In general, even with these slightly different phylogenetic relationships, BS support was similar to the dataset with only one partition.

With BC, the reconstructed phylogenies were fairly congruent among the taxon-gene sampling combinations, and this was also with greater gene tree independence (i.e., across alpha levels). For the 50- and 100-gene datasets, an alpha of 0.1 resulted in a consensus tree distinct from those resolved with larger alpha levels, but with the 150- and 200-gene datasets, the reconstructed phylogeny was the same across the four alpha levels. For the 50- and 100- gene datasets, the use of minimal gene-tree independence (alpha of 0.1) resulted in a phylogeny congruent with the other methods in the present study; however, with greater

gene-tree independence, the resulting phylogeny differed to a larger extent, with multiple clades resolved in positions not observed in other methods. Among the datasets, approximately 15 splits (nine in the 200-gene dataset) had a concordance factor greater than 0.5. The clades with these concordance factors were located towards the tips of the trees and only included two species or multiple individuals for the same species, with these results being similar to those for aforementioned clades that had high BS, JK, BB, BJ, and PP support values. The results from the various partitioned bootstrap and jackknife analyses were largely congruent with support values from ML bootstrapping of the entire, unpartitioned dataset in RAxML. Additionally, PCS recovered that the RAxML phylogeny based on the 298 DNA regions was supported by 220, 227, and 227 individual DNA regions compared to the phylogenies resolved using the 9-, 40-, and 232-gene matrices, respectively.

Among the four JML analyses, most species did not have a hypothesized history of hybridization, based on a posterior probability of less than 0.1. However, one species with a large amount of missing data, *L. discolor*, was identified most frequently as the species that may have had a history of hybridization. This is particularly the case in the nDNA and cpDNA trees analyzed with the nuclear DNA sequence data, although this was also recognized in other analyses as well. *Lithospermum helleri*, *L. calcicola*, and *L. ruderales* were most frequently recovered to have had histories of hybridization based on the cpDNA tree and the cpDNA sequence data. Analyses of the nDNA tree and the cpDNA sequence data resulted in approximately half of the species of the genus found to have over 10 hybridization events, with *L. helleri* and *L. calcicola* having the greatest number.

Analyses of BPP resolved a phylogeny that differed from those of other methods; however, most of the smaller clades are congruent, and multiple nodes of the backbone are unresolved and poorly supported (0.6–0.85 PP). The most apparent difference is that the clade that includes *L. incisum* and *L. mirabile* is sister to the other New World species of the genus, a placement not resolved in any other analysis. Analyses of species delimitation identified 35 or 36 species with the greatest posterior probabilities. Two species pairs – *L. erythrorhizon* and *L. officinale* and *L. helleri* and *L. occidentale* – are not resolved as distinct species based on A11 analyses, and the former is not supported by A10 analyses. All other species are supported by a PP of 1.00.

3.3. Diversification, trait evolution, and biogeography

The phylogeny resolved from analyses of BEAST2 is similar to those from analyses with more complete taxon sampling, and all clades are supported by greater than 0.98 PP (except *B. arvensis* as sister to the other outgroup taxa [0.66 PP]). *Lithospermum* is resolved to have diverged from sampled outgroups ca. 20.1 MYA, and the crown of the genus is dated to ca. 9.6 MYA. *Lithospermum* colonized the New World ca. 7.8 MYA, with most species originating over the next three million years. The parameters for this phylogeny all have an effective sample size greater than 200.

Patterns of diversification among *Lithospermum* and close relatives provide evidence that the genus had an approximately three-fold increase in the rate of diversification either at the crown of the genus ($f = 0.91$ for credible shifts in BAMM) or among the New World members ($f = 0.075$ for credible shifts in BAMM), and this increase in rate was followed by a subsequent decrease during latter diversification, among the studied species and each of the examined clades, due to both a slight increase in the rate of extinction and a decrease in the rate of speciation (Fig. 2). Similar patterns are observed in analyses of both BAMM and MiSSE. Analyses of the cpDNA phylogeny resolve no rate shifts with the greatest PP ($f = 0.52$); however, a rate shift in the same location as the nuclear DNA phylogeny has the second highest PP ($f = 0.48$). For analyses in BAMM, the effective sample sizes for the number of shifts and for the log likelihood are greater than 200.

Analyses of corolla shape as a binary character provide evidence that species with long-funnelform corollas have a rate of diversification two

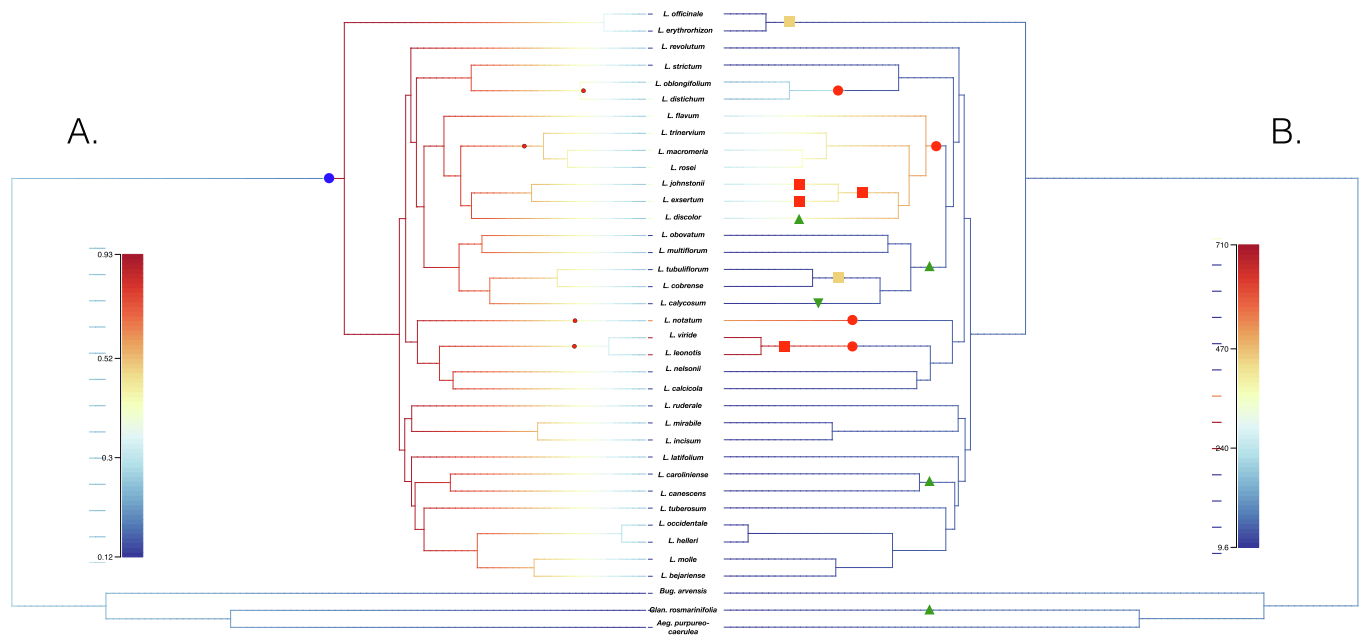


Fig. 2. Bayesian Inference phylogeny of 298 concatenated DNA regions for *Lithospermum* and relatives (*Aeg.* is *Aegonychon*, *Bug.* is *Buglossoides*, and *Glan.* is *Glandora*, *L.* is *Lithospermum*). A) Phylogeny resolves patterns of diversification of genus, blue circle denotes shift in diversification rate in *Lithospermum* as resolved by BAMM. Colors in tree correspond to associated legend. B) Phylogeny resolves patterns of corolla length evolution of genus, red circles denote significant shifts from BAMM analyses, boxes represent significant splits from ParSplits, with red boxes denoting more extreme shifts than orange boxes, green triangles indicate origins of heterostyly, and upside-down green triangles denote loss of heterostyly. Colors in tree correspond to associated legend. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to six times as large as for species with other types of corollas, with most of these rates occurring towards the tips of the phylogeny. Results differ with three, rather than two, corolla shapes included in the analyses (i.e., between BiSSE and MuSSE). In this situation, species with campanulate/urceolate corollas have a greater rate of speciation than the other two types of corollas, which have similar rates. As with corolla treated as a binary character, most transitions in corolla shape are to those that are not long-funnelform or campanulate/urceolate. For analyses in Diversitree, a full model was determined via AIC to be most appropriate, and for those with HiSSE, a model that includes two rates for turnover (one for each state), equal rates of extinction fraction, and no dual transitions was preferred.

Analyses based on I1ou and SURFACE, regardless of minimum, midpoint, or maximum values used, resolve two to four different rate shifts or regimes, with these shifts occurring along branches to clades with longer corollas (Fig. 2). One regime encompasses *L. macromeria* and *L. leonotis*, and another regime is present for *L. flavum* and *L. notatum*. The clade that includes *L. exsertum* and *L. johnstonii* is either included in the former or has its own regime based on different SURFACE analyses. Results from I1ou are similar, but *L. leonotis* has its own regime in analyses of the three datasets. All other species in the phylogeny have a separate regime for corolla-length evolution. Analyses in BAMM produce similar results, regardless of dataset. The clade that includes *L. flavum*, *L. macromeria*, and five other species has an increased rate of evolution for corolla length. The largest shift is resolved for the clade that includes *L. viride* and *L. leonotis*, and shifts in the rate of trait evolution are not resolved for *L. viride* using other methods. Analyses of the data based on the cpDNA phylogeny recovered similar results concerning rates of evolution, with four shifts in corolla length; however, the locations on the phylogeny differ due to distinct topologies. With the cpDNA tree, the greatest shift in rate is for the clade that includes *L. exsertum* and *L. johnstonii*.

The results from ParSplit analyses differ modestly from those of I1ou, SURFACE, and BAMM (Fig. 2). Five to eight shifts in corolla length are resolved, with many occurring among branches for species with long-

funnelform corollas, such as *L. johnstonii* and *L. exsertum*. However, other shifts occur among branches for shorter corollas, including the Eurasian species *L. officinale* and *L. erythrorhizon* and Texas species *L. helleri* and *L. occidentale*. It is notable that ParSplit was the only method employed that identified evolutionary shifts in rate toward both longer and shorter corollas. Most splits are identified with either the inverse or identity function used, but a split for *G. rosmarinifolia* and *A. purpureocaeruleum* is only recognized with identity.

The patterns of evolution of heterostyly and homostyly were also examined in a phylogenetic context, with at least three origins of the breeding system in *Lithospermum*. In BiSSE analyses, the full model for heterostyly resolved a bimodal distribution for speciation (λ), with homostyly being intermediate between the two peaks; however, the model with equal rates for λ was not identified as being statistically less preferable than the full model. While net diversification was similar between the two, only the full model results in higher net diversification for homostylous species compared to heterostylous ones. The results of HiSSE provide evidence that diversification increases for heterostylous species, which is reconstructed toward the tips of the phylogeny, and the same bimodal distribution from the full model in BiSSE is not recovered.

The preferred model for chromosome number evolution included a constant rate with no duplications, which involved two constant parameters (one each for chromosome gain and loss). In general, chromosome number decreased in *Lithospermum*. Thirty-one extant species and 15 nodes had an expectation of at least 0.5 for loss of a chromosome, and no nodes had this type of expectation for an increase in chromosome number. The ancestral haploid (n) chromosome number for *Lithospermum* is 15 or 16.

Analyses of biogeography, with both a maximum range of three or four areas, resolved that *Lithospermum* likely originated in Eurasia prior to colonizing the Americas (Fig. 3) given the plurality of Eurasia reconstructed and all outgroups and *L. officinale* and *L. erythrorhizon* resolved entirely in Eurasia. This was more pronounced with a maximum of three, rather than four, areas. The DEC + J model was

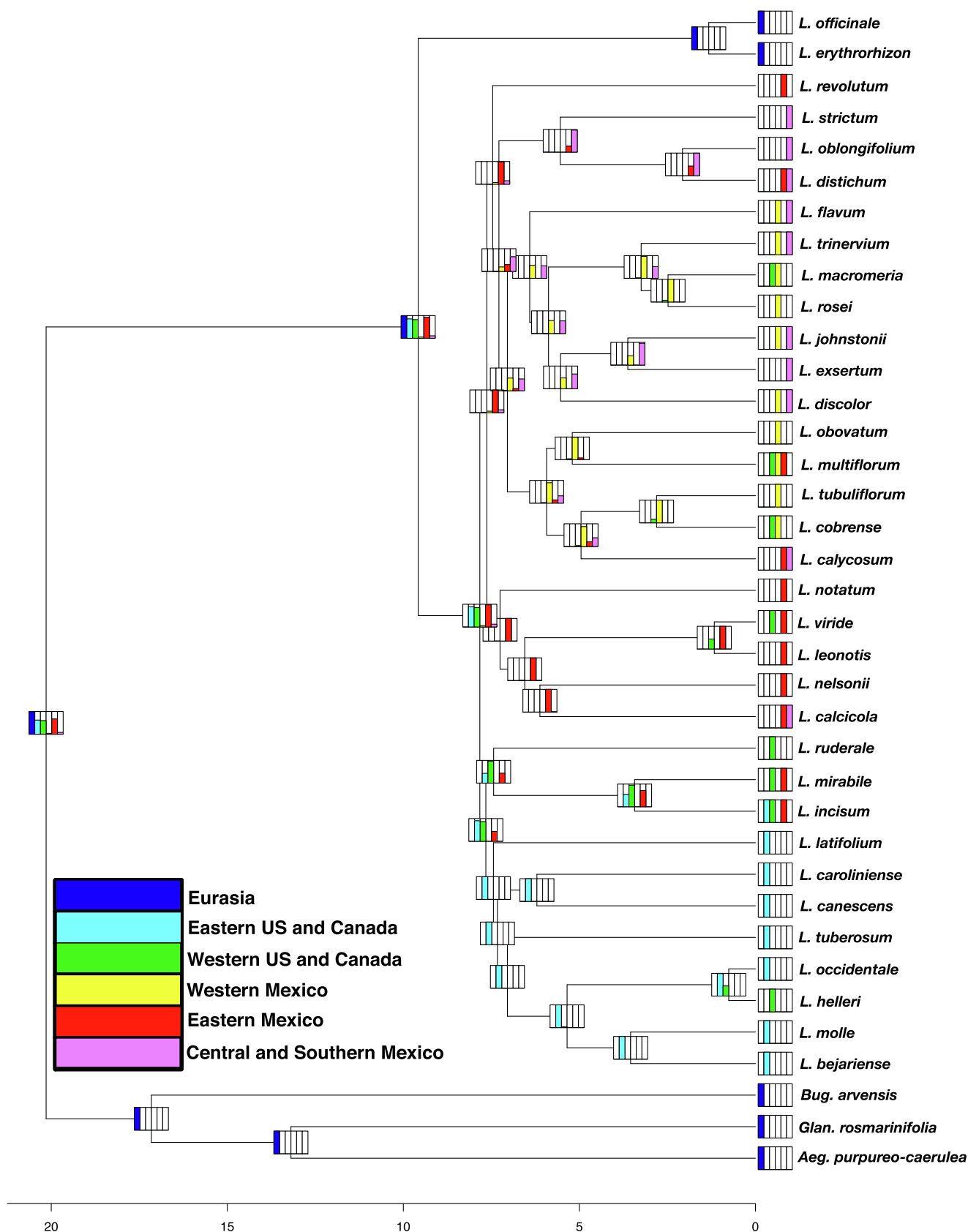


Fig. 3. Bayesian Inference phylogeny of 298 concatenated DNA regions for *Lithospermum* and relatives (*Aeg.* is *Aegonychon*, *Bug.* is *Buglossoides*, and *Glan.* is *Glandora*, *L.* is *Lithospermum*) with DEC + J biogeographic reconstruction using BioGeoBEARS. Ancestral areas at nodes are Eurasia (blue), eastern United States (US) and Canada (turquoise), western US and Canada (green), western Mexico (yellow), eastern Mexico (red), central and southern Mexico (pink), and probability of occupancy per area in bar graphs at nodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

identified as optimal for the data, based on AIC and AICc, and this model (along with others) provides evidence that upon reaching the New World, *Lithospermum* expanded throughout much of North America, initially occupying the eastern US and Canada and eastern Mexico. One clade of US and Canadian species colonized the western part of the countries prior to migrating back towards the east, and another one diversified across the eastern US. In Mexico, the species initially colonized the eastern region of the country, then central Mexico, and finally western Mexico, with extant species subsequently expanding their ranges after speciation to other areas of Mexico and, less frequently, the US. Figures from analyses of diversification and ancestral character reconstruction are included in Appendix C, and output concerning models from HiSSE and BioGeoBEARS is in Appendix D.

4. Discussion

4.1. Phylogenetics of *Lithospermum* and relatives

The relationships between the subfamilies and the tribes are congruent with those of previous analyses (e.g., Chacón et al., 2016), and relationships among members of Lithospermeae are similar to those from other studies (Chacón et al., 2019; Cohen and Davis, 2012); however, some differences are observed among genera. Given different taxon and gene-region sampling among analyses, these slight variations are unsurprising. One of the more notable differences is that in the present study, all phylogenetic analyses except MP resolved *Buglossoides arvensis* as sister to *Lithospermum*, and this is not the case in most other studies in which *Glandora* or *Ulugbekia* is resolved in this position (Chacón et al., 2019; Cohen and Davis, 2012). The current result could be due to multiple factors, such as the inclusion of only one species of *Glandora*, the lack of other close relatives of *Lithospermum* included, only 30–50% of the total DNA regions sampled among the outgroups, and the use of multiple nuclear DNA regions for phylogenetic inference, rather than the phylogenetic signal coming primarily from the plastid genome or the internal transcribed spacer (ITS), as is the case in previous studies. Additional sampling of species of *Lithospermum* and close relatives will aid in understanding if the nuclear genome is providing a distinct phylogenetic signal or if the results of the present study are due solely to taxon-sampling artifacts.

As hypothesized, the ingroup relationships differ from those resolved in previous studies, which are based primarily on cpDNA or cpDNA + ITS sampling (Chacón et al., 2019; Cohen and Davis, 2012; Weigend et al., 2010; Weigend et al., 2009). The use of the 298 nDNA regions resulted in a well-resolved and generally well-supported phylogeny of the genus (Fig. 1), which tended not to be the case in prior studies, especially those that only employed a few DNA regions (Chacón et al., 2019; Weigend et al., 2009). The nDNA phylogeny departs in various manners from those of prior analyses, although some smaller clades are resolved in multiple studies as well as in the present one. These include the species formerly in *Onosmodium*, with *L. tuberosum* as sister, and various sister species pairs, such as *L. trinervium* and *L. rosei*, *L. distichum* and *L. oblongifolium*, *L. exsertum* and *L. johnstonii*, *L. cobrense* and *L. tubuliflorum*, and *L. officinale* and *L. erythrorhizon*.

Overall, relationships in *Lithospermum* based on nDNA data are quite congruent among various methods of analyses in the present study. The largest differences among analyses involve the placement of a clade of Mexican species and the relationship of *L. discolor*, with the distinct positions for this particular species likely due to the large quantity of missing data for *L. discolor* (Table 1). Along the backbone of the phylogeny of the New World species of *Lithospermum*, lower support and the slight incongruence are observed between the MP, ML, BI, and CM phylogenies, which could be the result of biological issues that occurred over the three million years of the early diversification of the genus in the Americas, a time period during which most of the sampled lineages of the genus arose (Figs. 1–3). Possible biological explanations include incomplete lineage sorting, hybridization, rate heterogeneity, and rapid

radiation, with incomplete lineage sorting during a rapid radiation being the most likely situation. This possibility is the result of other hypotheses being unlikely. Hybridization does not appear to play a large role in the evolution of the genus, based on results of the analyses with JML. Indeed, the species identified as being most likely to have been involved in hybridization are *L. calycosum* and *L. discolor*, the two species with the most missing data (Table 1), suggesting that it is missing data, not past hybridization, that is influencing this result. The short branch lengths could support rapid radiation, and Chacón et al. (2019) identified *Lithospermum* as being one of four clades in Lithospermeae with increased rates of diversification, a result also recovered in the present study (Fig. 2A). Given that the results are similar among methods and models, rate heterogeneity likely did not play a substantial role in the minimal phylogenetic incongruence from the nuclear genome. Incomplete lineage sorting could have played a role, and both the slight differences of the backbone of the phylogeny of *Lithospermum* between MP, ML, BI, and CM (e.g., species being successive sisters rather than solely sister to a particular clade) and the high HRF in this region of the tree may support this. Furthermore, approximately 75% of DNA regions in the present study support the 298-gene tree (Fig. 1) as opposed to an alternative phylogeny, which provides evidence for a fairly strong evolutionary signal and minimal incongruence.

Due to the presence of minimal differences among the various methods for phylogenetic analysis, it also seems possible that methodological issues, such as sampling artifacts, suboptimal partitioning scheme(s), and missing data, would play a role in producing the recovered incongruence among methods. While taxon sampling was similar among the multiple data sets, the amount of sequence data was quite different, with the amount of sequence data increasing with a lower threshold for the number of species per DNA region. Results from the four datasets provide evidence that increasing sequence data helps to provide greater stability in the topology of the tree. Additionally, given the lack of variation recovered based on the different partitioning schemes, it seems unlikely that partitioning may have had much of an impact on the phylogeny and support values across the tree. Taken together, results suggest that increased sampling should help to solve the issue and aid in the identification in whether the minor incongruence is due to biological or analytical reasons, or both.

A larger issue than the incongruence among the nuclear genome and the methods in the present study concerns incongruence between phylogenies from nDNA and those primarily based on cpDNA regions, particularly the dataset of Cohen and Davis (2012) that includes 10 cpDNA regions and has resolved a comprehensive molecular phylogeny of the genus. Multiple biological possibilities exist for the underlying mechanisms of the patterns of incongruence between the nuclear and plastid genomes, such as issues with orthology/paralogy, horizontal gene transfer, hybridization (including chloroplast capture), and incomplete lineage sorting. For the present study, care was taken to minimize potential issues of orthology and paralogy arising, and therefore this likely did not influence incongruence. Not only were the loci in the present study putative single-copy regions but also only one consensus sequence per individual was used, with any nucleotide differences treated as polymorphisms in the consensus sequence. Given the results of Kates et al. (2018), it is expected that the use of consensus sequences, rather than allele phasing, would have negligible impact on phylogeny reconstruction. Additionally, while the consensus sequences could represent the consensus of recently diverged paralogs, it would be challenging to distinguish these from alleles, at least without further examination and the use of longer reads to appropriately identify variation across loci. For similar reasons, horizontal gene transfer seems an unlikely cause for the incongruence. While it could be possible that recent hybridization played a role to resolve the observed patterns, this also appears doubtful. Along with the results from JML analyses that do not recognize strong support for hybridization (including chloroplast capture), the phylogeny from the nuclear DNA regions resolves multiple groups of close relatives that have overlapping geographic ranges, but

that do not have much ecological overlap. Although, there are exceptions for species that have similar evolutionary relationships between the nuclear and plastid phylogenies (e.g., *L. distichum* and *L. oblongifolium*). In these situations, the floral morphology frequently differs between close relatives suggesting that pollinators would not visit the flowers of both species, particularly for groups of Mexican species in which floral morphology of these close relatives is quite divergent (Cohen, 2018). For example, *L. distichum* bears salverform corollas that are white with a yellow center and corolla tube that is 3–7 mm in length, while *L. oblongifolium* produces long-funnelform corollas that are green to green-yellow and a corolla tube that is 16–37 mm in length. Additionally, species with similar floral morphologies and overlapping geographic ranges, such as *L. multiflorum* and *L. obovatum*, have established incompatibility systems (cf. Ganders, 1979; Ralston, 1994), reducing the possibility of interspecific gene flow. These geographic and floral barriers would reduce the possibility of recent hybridization among species.

Consequently, incomplete lineage sorting, particularly during the early diversification of the genus in the New World, would be the most likely biological cause for incongruence between the two genomes. Many of the lineages of *Lithospermum* arose during the first three million years of the genus in the Americas. Twenty-three species of *Lithospermum* originated during this period, with much speciation occurring in the same geographic area for groups of species (Figs. 1–3). This may have led to alleles and/or plastid genomes shared during speciation of closely related, but not sister, species resolving the disparate evolutionary patterns between genomes.

Because the present study includes multiple individuals for 10 species of *Lithospermum*, it is possible to begin to evaluate the monophyly of some species of the genus. All species are resolved as monophyletic, with the exception of *L. distichum*. This suggests that in the genus most species are monophyletic (although, this should still be considered preliminary given limited sampling of individuals), and this result sheds further light on the incongruence between genomes. Given that the plastid genome has one quarter the effective population size of the nuclear genome (Hamilton, 2011), the presence of putative monophyly for the nuclear genome, among species of *Lithospermum*, provides evidence that there has been sufficient time for coalescence to occur among species.

As with the incongruence in the nuclear genome, the explanation of lineage sorting during early diversification as the cause for incongruence between genomes is due to the lack of a compelling explanation, based in biology, for other causes of incongruence among the molecular data – this hypothesis is the result of other causes being ruled out, rather than strong support for this phenomenon. It is also possible that the nDNA and cpDNA will resolve similar relationships as more data from plastid genomes become available. The amount of cpDNA sequence data Cohen and Davis (2012) included was less than 3% of that in the present study and ca. 7% of an entire plastid genome, and this is the most researchers have used to date for studies on the phylogenetics of *Lithospermum*. With a greater quantity of data from the plastid genome (e.g., whole plastomes), similar relationships may be resolved, and the observed incongruence could be more of a methodological issue than a biological one. Alternatively, the greater quantity of data could provide another explanation for incongruence between genomes.

The present study is the first in Boraginaceae to include hundreds of nuclear gene regions and to employ target sequence capture. Many previous studies of *Lithospermum* did not produce resolved phylogenies of the genus with high support among branches or even with resolution across the genus (Chacón et al., 2019; Weigend et al., 2009). The present study not only resulted in a well-resolved phylogeny of the genus but also, overall, a well-supported phylogeny. This provides further evidence that within challenging taxa, increasing gene sampling can result in greater support and resolution (Massoni et al., 2014; Nicholls et al., 2015; Parks et al., 2009). In *Lithospermum*, the increase in support is primarily toward the tips of the phylogeny and among smaller clades that include ca. three to six species, and this is observed among all

methods. At the same time, the sampling included in the present study is not a panacea as there are regions of the tree with lower support, that are unresolved in the MP phylogeny, and that differ depending on the methodology employed (Fig. 1).

4.2. Diversification and species

Upon arriving in the Americas, *Lithospermum* rapidly diversified. Three lines of evidence support this: 1) incongruence among methods and between genomes (particularly along the backbone of the phylogeny of the New World species of *Lithospermum*), 2) short branch lengths and high HRF of the backbone of phylogeny of the New World species, and 3) analyses of patterns of macroevolution of *Lithospermum* and relatives. Analyses of diversification from the present study resolved that a significant shift in the rate of diversification occurred at the crown of the genus (Fig. 2A). This pattern had the highest posterior probability based on the tree from the nDNA data, but the patterns were not recovered as having the highest posterior probability using the cpDNA phylogeny, which could result from different rates of evolution of these DNA regions or divergent taxon sampling between the studies. Based on the nDNA phylogeny, the rate of diversification decreased as the genus radiated throughout the New World. The results of these types of diversification analyses may need to be treated with caution for multiple reasons (e.g., Louca and Pennell, 2020), including the limited sampling of *Lithospermum*, which was accounted for in analyses; however, results are congruent with those of Chacon et al. (2019). This suggests that it may be possible to resolve patterns of diversification with smaller datasets and using species arbitrarily sampled from across a phylogeny, possibly with the use of fossil data included in phylogenetic analyses. The slower rate of diversification after colonization of the Americas, which is due to both less speciation and greater extinction (more the former than the latter), is congruent with multiple other taxa that also exhibit a decrease in diversification over time (Moen and Morlon, 2014). During this slowdown across the genus, species of *Lithospermum* were able to exploit new niches and develop novel morphological features, particularly related to pollination (Cohen, 2016b), and subsequently, this could have resulted in the decreased rate of speciation as niches were filled. Additionally, chromosome number is resolved to have decreased throughout diversification of the genus, and perhaps the loss of genomic material may have also have contributed to the slowdown in diversification as larger chromosome numbers are associated with increasing invasiveness (Pandit et al., 2014), which would have been taking place during early diversification of the genus in the Americas.

Thirty-four species of *Lithospermum* were included in the present study, and analyses in BPP (Yang, 2015) recognized 32 species as having the greatest posterior probability. The two pairs of species that were not identified as distinct were the Eurasian species – *L. erythrorhizon* and *L. officinale* – and two Texas species – *L. helleri* and *L. occidentale*. While this may suggest that these species pairs should each be combined into one species, it seems that this analysis is rather an indication that these species should be examined in greater detail. It is understandable to recover each of these two species pairs as indistinct because the species in each pair are morphologically similar and have overlapping geographic ranges (Brach and Song, 2006; Turner, 1995a). Indeed, Johnston (1952) recognized *L. erythrorhizon* as a “very close relative of *L. officinale*,” and even stated that some western individuals of the former are more similar to the latter and may even belong to it. Species formerly in the genus *Onosmodium*, such as *L. helleri* and *L. occidentale*, have been included in various taxonomic schemes (Das, 1965; Turner, 1995a; Weakley et al., 2011) that frequently include overlapping or adjacent quantitative features and geographic ranges. Both groups would benefit from further study involving a greater number of individuals across their geographic ranges as well as examinations of morphometric characters to carefully characterize species boundaries (or lack thereof).

4.3. Floral evolution

Lithospermum includes three suites of floral features (Cohen, 2016b, 2018), and species with salverform or funnellform corollas with included sexual organs are resolved as ancestral. Multiple origins of long-funnelform corollas with exserted sexual organs and of campanulate or urceolate corollas with included anthers and exserted stigmas are resolved, a result also recovered by Cohen (2016b) with the use of only cpDNA sequence data, and these results support the stated hypothesis. The increase in corolla length among various clades is likely the result of adaptation to particular pollinators, such as hawkmoths and hummingbirds, which are known to visit species of *Lithospermum* (Abrahamczyk and Renner, 2015; Armstrong, 1987; Boyd, 2004; Cohen, 2018; Grant and Grant, 1970). Gómez et al. (2016) found that across Brassicaceae, corolla shape integration increased as corolla depth increased, and a similar pattern may be observed in *Lithospermum*. Species in the genus with longer corollas also tend to have other longer floral features, such as filaments and styles, suggesting a high degree of floral integration. Additionally, an increase in floral size may be more important than other features, such as corolla shape, in encouraging floral visitors who are attracted to the larger floral displays. This may be especially important for the species of Mexico and the southwestern United States, the geographic range of most of the species that have undergone floral rate shifts in the present study, as larger flowers may have allowed these species to take advantage of a greater diversity of pollinators, such as hummingbirds, that were migrating to and diversifying in the same geographic area at approximately the same time as species of *Lithospermum* (Licóna-Vera and Ornelas, 2017).

Heterostyly also evolved multiple times in *Lithospermum*, with at least three independent origins of the breeding system (Fig. 2B), which supports the hypothesis on the patterns of evolution of the breeding system, although this number is fewer than Cohen (2011) resolved based on only cpDNA sequence data. The results of the full-model BiSSE analyses of heterostyly resolve a bimodal distribution for speciation (λ), with one peak greater than that for homostylous species and the other less. This suggests that in some clades of *Lithospermum*, heterostyly is allowing for a higher rate of speciation, while in others, it is less. This is seen among the at least three origins of the breeding system throughout the genus. One origin is at the base of a clade that includes five species (*L. calycosum*, *L. cobrense*, *L. multiflorum*, *L. obovatum*, and *L. tubuliflorum*), four of them are heterostylous, while the other two origins involve one and two species, respectively. The clade that includes the four heterostylous species is present primarily in northwestern Mexico and the southwestern United States, and most of these species have relatively small geographic ranges within the area as well as overlapping floral features – salverform to funnellform yellow corollas with included sexual organs (Cohen, 2018). Indeed, the geographic ranges of most of these species are smaller than those from the other origins of heterostyly (Cohen, 2018; Johnston, 1952). Barriers to effective pollination throughout the mountains of northwestern Mexico and the southwestern United States as well as the presence of self- and intramorph incompatibility in heterostyly in *Lithospermum* (Ganders, 1979) may have facilitated greater opportunities for the establishment of new species than occurred in other geographic areas where the breeding system is present, which includes southern Mexico and the eastern United States and Canada.

4.4. Biogeography

Biogeographical analyses of *Lithospermum* and close relatives provide evidence that the crown group of the genus likely originated in Eurasia ca. 9.6 MYA and began to diversify in the New World at ca. 7.8 MYA (Fig. 3). In order to migrate from the Old to the New World, the ancestors dispersed from Eurasia across the Atlantic Ocean to eastern North America, including the eastern US, Canada, and Mexico. Given the timing of this dispersal event, it may have occurred via long-distance

dispersal. While this could have happened during one event, there is evidence to suggest that areas of the North Atlantic Land Bridge (NALB) were still emergent until the late Miocene (Denk et al., 2010), which could have allowed for a stepping-stone pattern of dispersal. Unlike during the Eocene, when the NALB may have served as a means of dispersal for tropical plant species (Davis et al., 2002), during the Miocene, the NALB could have provided a pathway for plants already adapted to cooler and drier environments to travel between North America and Eurasia across the Atlantic Ocean. Other taxa of Boraginaceae were also migrating from Eurasia to North America at approximately the same time. Chacón et al. (2017) recovered *Omphalodes* and *Hackelia* as dispersing to North America across the Atlantic Ocean, while other taxa, like *Mertensia*, migrated to western North America via Beringia (Nazaire et al., 2014).

In North America, *Lithospermum* first colonized the eastern US, Canada, and Mexico, and subsequently one clade continued to speciate in Mexico while the other diversified in the US and Canada (Fig. 3), which is congruent with the hypothesized order of events. The US and Canadian species primarily diversified in eastern areas, with some species expanding their ranges to more western regions. The Mexican species initially diversified in the Sierra Madre Oriental. One clade migrated and diversified throughout the mountain ranges of the country: southward to the Trans-Mexican Volcanic Belt (TMVB) and Sierra Madre del Sur and then northward through the Sierra Madre Occidental and subsequently into the Trans-Pecos, Rocky Mountains, and Intermountain West. This demonstrates that many of the western US species present in the flora are the result of migration northwards from Mexico. This may have occurred more recently as North America warmed and glaciers retreated during the Pleistocene (Mastretta-Yanes et al., 2015).

Corona et al. (2007) found that the TMVB represents a transitional region between Nearctic and Neotropical biogeographic regions, and the TMVB is associated with the former for *Lithospermum* in the present study. This region has served as a setting for diversification for other plant (Ruiz-Sanchez and Specht, 2013) and animal taxa (e.g., Bryson and Riddle, 2012), and it seems to have played a similar role in *Lithospermum* as the genus diversified throughout the mountains of Mexico. Indeed, the formation of the TMVB, which occurred at approximately the same time as *Lithospermum* colonized and diversified across eastern North America (Ferrari et al., 2012; Mastretta-Yanes et al., 2015), may have more easily allowed the genus to migrate to appropriate habitats in western Mexico. Mesoamerican or South American species were not included in the present study, and future analyses can examine the biogeographic history of these Neotropical species in relation to those resolved for the currently investigated Nearctic species of *Lithospermum*. Given the transitional nature of the TMVB, the Mesoamerican and South American species may have a biogeographical pattern distinct from the species presently studied.

Much of the speciation in the New World species of *Lithospermum* occurred between 7.8 and 4.9 MYA, during a time when North America was becoming cooler and drier (Chacón et al., 2017; Herbert et al., 2016). These types of conditions could have facilitated speciation of *Lithospermum* throughout North America via expanded montane forests at lower elevations (Nazaire et al., 2014). Additionally, multiple Mexican species of *Lithospermum* are pollinated by hummingbirds, which were also radiating concurrently (Licóna-Vera and Ornelas, 2017). These newly diverse pollinators could have provided appropriate conditions for selection of the longer corolla tubes in species of the genus (Abrahamczyk and Renner, 2015). Furthermore, during cooler periods of glaciation, species of *Lithospermum* likely retreated down mountain ranges and colonized lower elevation geographic regions, and upon warming and recolonizing higher elevations, new species may have formed via parapatry (Mastretta-Yanes et al., 2015). This pattern may have resulted in more recent diversification for the genus, particularly in central and western Mexican mountain ranges (Fig. 3).

The present study recovered slightly different results from others that have examined the biogeography of *Lithospermum* and relatives, and

there are three primary reasons that this may be the case. First, the present study includes more (nuclear) DNA sequence data for reconstructing the phylogeny of the genus and relatives, and this resulted in the New World species forming a monophyletic group, which was not the case in some other studies on the genus (Chacón et al., 2019; Chacón et al., 2017). In the present study there was only one dispersal event from the Old to the New World, while in prior analyses, researchers resolved two dispersal events from the New to the Old World – one to Europe and one to Africa. Second, while Ree and Sanmartín (2018) provide evidence that DEC + J, the preferred model in the present study, is not statistically appropriate, Klaus and Matzke (2020) suggest the opposite. Should DEC + J be inappropriate for the present study, this could influence the results and interpretation; however, recent studies on the biogeography of Boraginaceae have also used this model (Chacón et al., 2019; Chacón et al., 2017), and similar results are recovered with DEC, the next-preferred model. Third, in the present study, the fossil nutlet *L. dakotense* was placed at the base of the New World species of *Lithospermum* because the nutlet was found in North America. Chacón et al. (2017) placed the fossil at the crown node of *Lithospermum*, which would be appropriate given that the Old World species are nested among the New World species of the genus. Additionally, these authors used a slightly older age (12–9 MYA) than in the present study (minimum age of 6.6 MYA, which is congruent with prior analyses of the genus [e.g., Cohen, 2012; Weigend et al., 2009]). Overall, the dates of origin of the genus and of the New World members are quite similar, despite the differences in topology, fossil placement, and dating, suggesting that *Lithospermum* arose and diversified in the middle to late Miocene and radiated in North America throughout the late Miocene and early Pliocene amid the cooler and drier conditions of the region at that time. Chacón et al. (2017) included South American and South African species in their analyses, and future studies can help establish if similar biogeographical patterns to those they recovered are resolved with more nuclear DNA regions.

CRedit authorship contribution statement

James Cohen: Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Visualization, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author would like to thank J. Enk and J. M. Rouillard for help with the target sequence capture and sequencing. Galaxy and CIPRES provided computational resources for the phylogenetic analyses. S. Turgman-Cohen was incredibly helpful with assistance setting up and running multiple phylogenetic analyses. S. Hunter provided assistance with BUCKy visualization. J. B. Landis, S. C. K. Straub, and one anonymous reviewer gave thoughtful and constructive feedback on the manuscript.

Funding

This work was supported by funds from Texas A&M International University and Kettering University, and a National Science Foundation Major Instrumentation Project Award (Award Number 1725938) allowed for the acquisition of the Kettering University High-Performance Computing Cluster (KUHPC) that was used for phylogenetic analyses.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ympev.2021.107317>.

References

- Abrahamczyk, S., Renner, S.S., 2015. The temporal build-up of hummingbird/plant mutualisms in North America and temperate South America. *BMC Evol. Biol.* 15, 104.
- Armstrong, D.P., 1987. Economics of breeding territoriality in male Calliope Hummingbirds. *Auk* 104, 242–253.
- Beaulieu, J., 2020. corHMM. <https://github.com/thej022214/corHMM>.
- Beaulieu, J.M., O'Meara, B.C., 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst. Biol.* 65, 583–601.
- Birky, C.W., 2008. Uniparental inheritance of organelle genes. *Curr. Biol.* 18, R692–R695.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, e1003537.
- Boyd, A., 2002. Morphological analysis of Sky Island populations of *Macromeria viridiflora* (Boraginaceae). *Syst. Bot.* 116–126.
- Boyd, A.E., 2003. Phylogenetic relationships and corolla size evolution among *Macromeria* (Boraginaceae). *Syst. Bot.* 28, 118–129.
- Boyd, A.E., 2004. Breeding system of *Macromeria viridiflora* (Boraginaceae) and geographic variation in pollinator assemblages. *Am. J. Bot.* 91, 1809–1813.
- Brach, A.R., Song, H., 2006. eFloras: New directions for online floras exemplified by the Flora of China Project. *Taxon* 55, 188–192.
- Bryson Jr, R.W., Riddle, B.R., 2012. Tracing the origins of widespread highland species: a case of Neogene diversification across the Mexican sierras in an endemic lizard. *Biol. J. Linn. Soc.* 105, 382–394.
- Chacón, J., Luebert, F., Hilger, H.H., Ovchinnikova, S., Selvi, F., Cecchi, L., Williams, C. M., Hasenstab-Lehman, K., Sutorý, K., Simpson, M.G., 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., Luebert, F., Selvi, F., Cecchi, L., Weigend, M., 2019. Phylogeny and historical biogeography of *Lithospermeae* (Boraginaceae): Disentangling the possible causes of Miocene diversifications. *Mol. Phylog. Evol.* 141, 106626.
- Chacón, J., Luebert, F., Weigend, M., 2017. Biogeographic events are not correlated with diaspore dispersal modes in Boraginaceae. *Front. Ecol. Environ.* 5, 26.
- 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). *Biol. J. Linn. Soc.* 107, 442–457.
- Cohen, J.I., 2014. A phylogenetic analysis of morphological and molecular characters of Boraginaceae: evolutionary relationships, taxonomy, and patterns of character evolution. *Cladistics* 30, 139–169.
- Cohen, J.I., 2016a. De novo sequencing and comparative transcriptomics of floral development of the distylous species *Lithospermum multiflorum*. *Front. Plant Sci.* 7, 1934.
- Cohen, J.I., 2016b. Floral evolution in *Lithospermum* (Boraginaceae): independent origins of similar flower types. *Bot. J. Linn. Soc.* 180, 213–228.
- Cohen, J.I., 2018. A revision of the Mexican species of *Lithospermum* (Boraginaceae). *Ann. Missouri Bot. Gard.* 103, 200–257.
- Cohen, J.I., Davis, J.I., 2009. Nomenclatural changes in *Lithospermum* (Boraginaceae) and related taxa following a reassessment of phylogenetic relationships. *Brittonia* 61, 101–111.
- Cohen, J.I., Davis, J.I., 2012. Molecular phylogenetics, molecular evolution, and patterns of clade support in *Lithospermum* (Boraginaceae) and related taxa. *Syst. Bot.* 37, 490–506.
- Cohen, J.I., Manning, J.C., Euston-Brown, D., 2019. *Lithospermum sylvestre* (Boraginaceae): A new species from the Baviaanskloof, Eastern Cape, South Africa. *Bothalia-African Biodiversity & Conservation* 49, 1–5.
- Corona, A.M., Toledo, V.H., Morrone, J.J., 2007. Does the Trans-mexican Volcanic Belt represent a natural biogeographical unit? An analysis of the distributional patterns of Coleoptera. *J. Biogeogr.* 34, 1008–1015.
- Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J.V., Udall, J., 2012. Targeted enrichment strategies for next-generation plant biology. *Am. J. Bot.* 99, 291–311.
- Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294.
- Das, T.L., 1965. A taxonomic revision of the genus *Onosmodium*. Kansas State University, Manhattan, Kansas, p. 80.
- Davis, C.C., Bell, C.D., Mathews, S., Donoghue, M.J., 2002. Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. *PNAS* 99, 6833–6837.
- de Vos, J.M., Hughes, C.E., Schneeweiss, G.M., Moore, B.R., Conti, E., 2014. Heterostyly accelerates diversification via reduced extinction in primroses. *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* 281, 20140075.
- Denk, T., Grímsson, F., Zetter, R., 2010. Episodic migration of oaks to Iceland: Evidence for a North Atlantic “land bridge” in the latest Miocene. *Am. J. Bot.* 97, 276–287.

- Didier, G., Chabrol, O., Laurin, M., 2019. Parsimony-based test for identifying changes in evolutionary trends for quantitative characters: implications for the origin of the amniotic egg. *Cladistics* 35, 576–599.
- Doyle, J.J., 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot.* 144–163.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012a. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973.
- Drummond, A.J., Xie, W., Heled, J., 2012b. Bayesian inference of species trees from multilocus data using* BEAST. *Mol. Biol. Evol.* 29, 1969–1973.
- Ferrari, L., Orozco-Esquivel, T., Manea, V., Manea, M., 2012. The dynamic history of the Trans-Mexican Volcanic Belt and the Mexico subduction zone. *Tectonophysics* 522, 122–149.
- Ferrero, V., Chapela, I., Arroyo, J., Navarro, L., 2011. Reciprocal style polymorphisms are not easily categorised: the case of heterostyly in *Lithodora* and *Glandora* (Boraginaceae). *Plant Biol.* 13, 7–18.
- FitzJohn, R.G., 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* 3, 1084–1092.
- Gabel, M.L., 1987. A fossil *Lithospermum* (Boraginaceae) from the tertiary of South Dakota. *Am. J. Bot.* 74, 1690–1693.
- Ganders, F.R., 1979. Heterostyly in *Lithospermum cobrense* (Boraginaceae). *Am. J. Bot.* 66, 746–748.
- Gatesy, J., Meredith, R.W., Janecka, J.E., Simmons, M.P., Murphy, W.J., Springer, M.S., 2017. Resolution of a concatenation/coalence kerfuffle: partitioned coalescence support and a robust family-level tree for Mammalia. *Cladistics* 33, 295–332.
- Gatesy, J., Sloan, D.B., Warren, J.M., Baker, R.H., Simmons, M.P., Springer, M.S., 2019. Partitioned coalescence support reveals biases in species-tree methods and detects gene trees that determine phylogenomic conflicts. *Mol. Phylog. Evol.* 139, 106539.
- Glick, L., Mayrose, I., 2014. ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* 31, 1914–1922.
- Goloboff, P.A., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* 15, 415–428.
- Goloboff, P.A., Catalano, S.A., 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* 32, 221–238.
- Gómez, J.M., Torices, R., Lorite, J., Klingenberg, C.P., Perfectti, F., 2016. The role of pollinators in the evolution of corolla shape variation, disparity and integration in a highly diversified plant family with a conserved floral bauplan. *Ann. Bot.* 117, 889–904.
- Grant, V., Grant, K.A., 1970. A hummingbird-pollinated species of Boraginaceae in the Arizona flora. *PNAS* 66, 917–919.
- Guerrero, R.F., Hahn, M.W., 2018. Quantifying the risk of hemiplasy in phylogenetic inference. *PNAS* 115, 12787–12792.
- Hamilton, M., 2011. Population Genetics. John Wiley & Sons, Oxford, UK.
- Heckenhauer, J., Paun, O., Chase, M.W., Ashton, P.S., Kamariah, A., Samuel, R., 2019. Molecular phylogenomics of the tribe Shoreaeae (Dipterocarpaceae) using whole plastid genomes. *Ann. Bot.* 123, 857–865.
- Helfrich, P., Riebel, E., Abrami, G., Lücking, A., Mehler, A., 2018. TreeAnnotator: versatile visual annotation of hierarchical text relations. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Herbert, T.D., Lawrence, K.T., Tzanova, A., Peterson, L.C., Caballero-Gill, R., Kelly, C.S., 2016. Late Miocene global cooling and the rise of modern ecosystems. *Nat. Geosci.* 9, 843–847.
- Holstein, N., Chacon, J., Hilger, H.H., Weigend, M., 2016. No longer shipwrecked—*Selkirkia* (Boraginaceae) back on the mainland with generic rearrangements in South American “*Omphalodes*” based on molecular data. *Phytotaxa* 270, 231–251.
- Ingram, T., Mahler, D.L., 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods Ecol. Evol.* 4, 416–425.
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., Madden, T.L., 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9.
- Johnston, I.M., 1952. Studies in the Boraginaceae, XXIII A survey of the genus *Lithospermum*. *J. Arnold Arbor.* 33, 299–366.
- Johnston, I.M., 1954. Studies in the Boraginaceae, XXVI Reevaluations of the genera of the Lithospermeae. *J. Arnold Arbor.* 35, 1–81.
- Joly, S., 2012. JML: testing hybridization from species trees. *Mol. Ecol. Resour.* 12, 179–184.
- Kates, H.R., Johnson, M.G., Gardner, E.M., Zerega, N.J., Wickett, N.J., 2018. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*. *Am. J. Bot.* 105, 404–416.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649.
- Khabbazian, M., Kriebel, R., Rohe, K., Ané, C., 2016. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods Ecol. Evol.* 7, 811–824.
- Klaus, K.V., Matzke, N.J., 2020. Statistical comparison of trait-dependent biogeographical models indicates that Podocarpaceae dispersal is influenced by both seed cone traits and geographical distance. *Syst. Biol.* 69, 61–75.
- Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol. Biol.* 14, 82.
- Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773.
- Larget, B.R., Kotha, S.K., Dewey, C.N., Ané, C., 2010. BUCKY: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26, 2910–2911.
- Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* 61, 727–744.
- Lemoine, F., Entfellner, J.-B.-D., Wilkinson, E., Correia, D., Felipe, M.D., De Oliveira, T., Gascuel, O., 2018. Renewing Felsenstein’s phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456.
- Léveillé-Bourret, É., Starr, J.R., Ford, B.A., Moriarty Lemmon, E., Lemmon, A.R., 2018. Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Syst. Biol.* 67, 94–112.
- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
- Licona-Vera, Y., Ornelas, J.F., 2017. The conquering of North America: dated phylogenetic and biogeographic inference of migratory behavior in bee hummingbirds. *BMC Evol. Biol.* 17, 126.
- Louca, S., Pennell, M.W., 2020. Extant timetrees are consistent with a myriad of diversification histories. *Nature* 580, 502–505.
- Mabry, M.E., Simpson, M.G., 2018. Evaluating the monophyly and biogeography of *Cryptantha* (Boraginaceae). *Syst. Bot.* 43, 53–76.
- Massoni, J., Forest, F., Sauquet, H., 2014. Increased sampling of both genes and taxa improves resolution of phylogenetic relationships within Magnoliidae, a large and early-diverging clade of angiosperms. *Mol. Phylog. Evol.* 70, 84–93.
- Mastretta-Yanes, A., Moreno-Letelier, A., Piñero, D., Jorgensen, T.H., Emerson, B.C., 2015. Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *J. Biogeogr.* 42, 1586–1600.
- Matzke, N.J., 2012. Founder-event speciation in BioGeoBEARS package dramatically improves likelihoods and alters parameter inference in Dispersal-Extinction-Cladogenesis (DEC) analyses. *Front. Biogeogr.* 4, 210.
- Mitchell, J.S., Rabosky, D.L., 2017. Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts. *Methods Ecol. Evol.* 8, 37–46.
- Moen, D., Morlon, H., 2014. Why does diversification slow down? *Trends Ecol. Evol.* 29, 190–197.
- Morrone, J.J., 2017. Neotropical Biogeography: Regionalization and Evolution. CRC Press, Boca Raton, Florida.
- Nazaire, M., Wang, X.Q., Hufford, L., 2014. Geographic origins and patterns of radiation of *Mertensia* (Boraginaceae). *Am. J. Bot.* 101, 104–118.
- Nguyen, L.-T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.
- Nicholls, J.A., Pennington, R.T., Koenen, E.J., Hughes, C.E., Hearn, J., Bunnefeld, L., Dexter, K.G., Stone, G.N., Kidner, C.A., 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Front. Plant Sci.* 6, 710.
- Nixon, K., 2002. WinClada ver. 1.00. 08. Published by the author, Ithaca, NY.
- Nixon, K.C., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15, 407–414.
- Pandit, M.K., White, S.M., Pocock, M.J., 2014. The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis. *New Phytol.* 203, 697–703.
- Parks, M., Cronn, R., Liston, A., 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7, 1–17.
- Pourghorban, Z., Salmaki, Y., Weigend, M., 2020. Phylogenetic relationships within the subtribe Cynoglossinae (Cynoglossodeae: Boraginaceae): new insights from nuclear and plastid DNA sequence data. *Plant Syst. Evol.* 306, 1–15.
- Rabosky, D.L., Grudler, M., Anderson, C., Title, P., Shi, J.J., Brown, J.W., Huang, H., Larson, J.G., 2014. BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* 5, 701–707.
- Ralston, B.E., 1994. Phylogenetic systematics and the evolution of mating systems in *Lithospermum* (Boraginaceae). Northern Arizona University, Flagstaff, Arizona.
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904.
- Ree, R.H., Sanmartín, I., 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *J. Biogeogr.* 45, 741–749.
- Ripma, L.A., Simpson, M.G., Hasenstab-Lehman, K., 2014. Geneious! Simplified genome skimming methods for phylogenetic systematic studies: A case study in *Oreocarya* (Boraginaceae). *Appl. Plant Sci.* 2, 1400062.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542.
- Rosenberg, M.S., Kumar, S., 2003. Taxon sampling, bioinformatics, and phylogenomics. *Syst. Biol.* 52, 119.
- Ruiz-Sanchez, E., Specht, C.D., 2013. Influence of the geological history of the Trans-Mexican Volcanic Belt on the diversification of *Nolina parviflora* (Asparagaceae: Nolinoideae). *J. Biogeogr.* 40, 1336–1347.
- Salinas, N.R., Little, D.P., 2014. 2matrix: A utility for indel coding and phylogenetic matrix concatenation. *Appl. Plant Sci.* 2, 1300083.

- Selvi, F., Bigazzi, M., Hilger, H.H., Papini, A., 2006. Molecular phylogeny, morphology and taxonomic re-circumscription of the generic complex *Nonea/Elizaldia/Pulmonaria/Paraskevia* (Boraginaceae-Boragineae). *Taxon* 55, 907–918.
- Selvi, F., Cecchi, L., Hilger, H.H., Coppi, A., 2017. A reappraisal of the genus *Megacaryon* (Boraginaceae, Lithospermeae) based on molecular, morphological, and karyological evidence. *Syst. Biodivers.* 15, 552–563.
- Siddall, M.E., 2010. Unringing a bell: metazoan phylogenomics and the partition bootstrap. *Cladistics* 26, 444–452.
- Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381.
- Simpson, M.G., Williams, C.M., Hasenstab-Lehman, K.E., Mabry, M.E., Ripma, L., 2017. Phylogeny of the popcorn flowers: Use of genome skimming to evaluate monophyly and interrelationships in subtribe Amsinckieae (Boraginaceae). *Taxon* 66, 1406–1420.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Turner, B.L., 1995a. Synopsis of the genus *Onosmodium* (Boraginaceae). *Phytologia* 78, 39–60.
- Turner, B.L., 1995b. Synoptical study of the genus *Macromeria* (Boraginaceae). *Phytologia* 77, 393–407.
- Weakley, A.S., LeBlond, R.J., Sorrie, B.A., Witsell, C.T., Estes, L.D., Gandhi, K., Mathews, K.G., Ebihara, A., 2011. New combinations, rank changes, and nomenclatural and taxonomic comments in the vascular flora of the southeastern United States. *J. Bot. Res. Inst. Tex.* 5, 437–455.
- Weigend, M., Gottschling, M., Hilger, H.H., Nürk, N.M., 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.
- Weigend, M., Gottschling, M., Selvi, F., Hilger, H.H., 2009. Marbleseeds are gromwells—Systematics and evolution of *Lithospermum* and allies (Boraginaceae tribe Lithospermeae) based on molecular and morphological data. *Mol. Phylogen. Evol.* 52, 755–768.
- Weigend, M., Luebert, F., Selvi, F., Brokamp, G., Hilger, H.H., 2013. Multiple origins for Hound's tongues (*Cynoglossum* L.) and Navel seeds (*Omphalodes* Mill.)—The phylogeny of the borage family (Boraginaceae s. str.). *Mol. Phylogen. Evol.* 68, 604–618.
- Weller, S.G., 1980. Pollen flow and fecundity in populations of *Lithospermum carolinense*. *Am. J. Bot.* 67, 1334–1341.
- Yang, Z., 2015. The BPP program for species tree estimation and species delimitation. *Curr. Zool.* 61, 854–865.
- Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* 19, 153.
- Zhang, M.-Y., Fritsch, P.W., Ma, P.-F., Wang, H., Lu, L., Li, D.-Z., 2017. Plastid phylogenomics and adaptive evolution of *Gaultheria* series *Trichophyllae* (Ericaceae), a clade from sky islands of the Himalaya-Hengduan Mountains. *Mol. Phylogen. Evol.* 110, 7–18.