









RESEARCH ARTICLE

Anatomy of a mega-radiation: Biogeography and niche evolution in *Astragalus*

Ryan A. Folk¹  | Joseph L. M. Charboneau²  | Michael Belitz³  | Tajinder Singh¹ | Heather R. Kates³  | Douglas E. Soltis^{3,4,5,6}  | Pamela S. Soltis^{3,4,5}  | Robert P. Guralnick^{3,5}  | Carolina M. Siniscalchi^{1,7} 

¹Department of Biological Sciences, Mississippi State University, Mississippi State, MS, USA

²Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

³Florida Museum, University of Florida, Gainesville, FL, USA

⁴Genetics Institute, University of Florida, Gainesville, FL, USA

⁵Biodiversity Institute, University of Florida, Gainesville, FL, USA

⁶Department of Biology, University of Florida, Gainesville, FL, USA

⁷General Libraries, Mississippi State University, Mississippi State, MS, USA

Correspondence

Ryan A. Folk, Department of Biological Sciences, Mississippi State University, 295 E. Lee Blvd., P.O. Box GY, Mississippi State, MS 39762 USA.
Email: rfolk@biology.msstate.edu

Abstract

Premise: *Astragalus* (Fabaceae), with more than 3000 species, represents a globally successful radiation of morphologically highly similar species predominant across the northern hemisphere. It has attracted attention from systematists and biogeographers, who have asked what factors might be behind the extraordinary diversity of this important arid-adapted clade and what sets it apart from close relatives with far less species richness.

Methods: Here, for the first time using extensive phylogenetic sampling, we asked whether (1) *Astragalus* is uniquely characterized by bursts of radiation or whether diversification instead is uniform and no different from closely related taxa. Then we tested whether the species diversity of *Astragalus* is attributable specifically to its predilection for (2) cold and arid habitats, (3) particular soils, or to (4) chromosome evolution. Finally, we tested (5) whether *Astragalus* originated in central Asia as proposed and (6) whether niche evolutionary shifts were subsequently associated with the colonization of other continents.

Results: Our results point to the importance of heterogeneity in the diversification of *Astragalus*, with upshifts associated with the earliest divergences but not strongly tied to any abiotic factor or biogeographic regionalization tested here. The only potential correlate with diversification we identified was chromosome number. Biogeographic shifts have a strong association with the abiotic environment and highlight the importance of central Asia as a biogeographic gateway.

Conclusions: Our investigation shows the importance of phylogenetic and evolutionary studies of logistically challenging “mega-radiations.” Our findings reject any simple key innovation behind high diversity and underline the often nuanced, multifactorial processes leading to species-rich clades.

KEYWORDS

abiotic, *Astragalus*, diversification, edaphic, Fabaceae, legume, mega-genus, species richness

Astragalus (milk-vetch or locoweed; Fabaceae: Papilionoideae; Azani et al., 2017a), with ~3100 species (Maassoumi, 2020; POWO, 2023), is currently the largest clade in land plants treated at the generic rank. While mega-genera arise primarily as historical artifacts from taxonomic revision (Frodin, 2004) and are not directly comparable to less species-rich genera (Hennig, 1966; Sanderson and Wojciechowski, 1996), they are clearly related to morphological evolution and often arise from

or persist due to a paucity of easily recognizable morphological variation that would delimit more manageable units, ultimately adding to the logistical challenge of their study. Aside from species number, the *Astragalus* clade is remarkable as a major nodulating-legume radiation in temperate arid areas, particularly central Asia and western North America, where it is one of the most ecologically important legumes. *Astragalus* species, relatively uniform in terms of morphology

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *American Journal of Botany* published by Wiley Periodicals LLC on behalf of Botanical Society of America.

(primarily herbaceous perennials) and niche biology (primarily occurring in arid and semi-arid areas), are morphologically highly similar to other members of the Astragalean clade (e.g., *Oxytropis*, *Colutea*), offering no clear adaptive explanation as to their extraordinary diversity and ecological importance.

The ecological dominance, species richness, and narrow endemism of *Astragalus* in many semi-arid areas of the northern hemisphere and a smaller disjunct portion of the southern hemisphere have attracted the attention of biogeographers and ecologists (Scherson et al., 2008; Samad et al., 2014; Hardion et al., 2016; Amini et al., 2019; Azani et al., 2019; Závěská et al., 2019; Plenk et al., 2020; Li et al., 2021; Wang et al., 2021; Maassoumi and Ashouri, 2022). Perhaps the major biogeographic finding of molecular phylogenetic studies in *Astragalus*, suspected previously based on chromosome counts (Ledingham, 1960; Barneby, 1964), is the recovery of a clade comprising nearly all North American *Astragalus*, informally termed the Neo-Astragalus clade (Wojciechowski et al., 1993, 1999), which renders species in the eastern hemisphere a paraphyletic group. The Neo-Astragalus clade is distinguished by chromosomal processes. Polyploidy, probably often allopolyploidy (Bartha et al., 2013; Závěská et al., 2019; Plenk et al., 2020; Bagheri et al., 2022), is thought to be the dominant chromosomal process in eastern hemisphere species (Lewis et al., 2005), where counts of $n = 8, 16$, and 24 (interpreted as euploid) predominate. By contrast, aneuploidy prevails in the Neo-Astragalus clade (Wojciechowski et al., 1993, 1999), with the aneuploid series $n = 11–15$ (Spellenberg, 1976) forming a “nearly perfect correlation” with biogeography according to Wojciechowski et al. (1999, p. 426).

Rapid chromosomal evolution, in the form of long aneuploid series seen in some plant groups, has been associated with responses to environmental stress (De Storme and Mason, 2014; Ackerfield et al., 2020) and has been interpreted as promoting recombination in the face of ecological change more generally (Levin, 1975; Grant, 1981). While the Neo-Astragalus clade is not clearly morphologically delimited (Wojciechowski et al., 1993), the strong biogeographic and chromosomal pattern is in line with niche biology. Barneby (1964) observed that, by contrast with the dominance of Eurasian *Astragalus* in the highlands, most North American species, believed by Barneby to have developed in situ, seem to specialize in dry, lowland environments. The genus is also associated with poor soils, for which its nodulating ability (Afkhani et al., 2018; Kates et al., 2022 [preprint]) would serve as a key adaptation. Edaphic specialization is thought to be a major force in plant biogeography (Raven, 1964; Axelrod, 1972; Lichter-Marck and Baldwin, 2023), and while few large-scale investigations of edaphic ecology have been done for nodulating legumes (Tamme et al., 2021; Doby et al., 2022; Siniscalchi et al., 2022), soil environments control the distribution of root nodule symbiosis by limiting the availability of both plant and especially bacterial partners and affecting partner choice (Sprent et al., 2017; Dladlu et al., 2018; Rathi et al., 2018; Ardley and Sprent, 2021). Differing availability of symbiotic partners at the regional

scale is thought to be a major factor limiting the dispersal process in nodulating plants (Parker et al., 2006; Simonsen et al., 2017). Therefore, it appears that a continental shift to the Americas, likely before the radiation of today's major clades, was associated with ecological specialization to challenging low-nutrient soils and harsh climatic conditions, along with more rapid chromosomal evolution.

Finally, related to questions of biogeography and chromosomal change are diversification mechanisms in *Astragalus*. Given its high species richness, the traditional explanation was that *Astragalus* experienced particularly high diversification rates (Barneby, 1964; Polhill, 1981; Lewis et al., 2005), and the genus was among the earliest in plants where explicit rates of diversification were calculated from molecular phylogenetic data (Sanderson and Wojciechowski, 1996). Remarkably, the primary finding of Sanderson and Wojciechowski was that *Astragalus* is not characterized by extraordinary rates of diversification, but instead shares uniform, relatively high rates of diversification with other members of the broader “Astragalean” clade (tribe Galageae pro parte; Lewis et al., 2005; see also Azani et al., 2017a; Choi et al., 2022), a group distinctive in its ecology but lacking any clear morphological innovations that might relate to diversification. Use of more recent probabilistic methods to estimate diversification in *Astragalus* have focused on recently diverged subclades (Azani et al., 2019, using BMM [Rabosky, 2014]) or on crown dates of diversification events using time calibration (Hardion et al., 2016; Bagheri et al., 2017). A reevaluation of the uniform diversification hypothesis of Sanderson and Wojciechowski (1996) is warranted with more complete sampling and robust phylogenetic trees enabling the investigation of diversification rates both within and between genera, as well as innovative diversification methods that more fully use the branch length information from molecular phylogenetics beyond species count data.

Both intrinsic and extrinsic factors that shape diversity in *Astragalus* have been posited, but empirical work has been piecemeal. Areas of species richness and endemism have recently been reported for Eurasian *Astragalus* (Maassoumi and Ashouri, 2022), and *Astragalus* has also been the subject of biogeographic and diversification analyses, particularly in specific subclades (Sanderson and Wojciechowski, 1996; Scherson et al., 2008; Azani et al., 2019; Su et al., 2021). A needed next step is to test, across the entire clade, the associations of biogeographic processes such as lineage dispersal with chromosome evolution and shifts in niche occupancy.

Here, we used a combination of comparative methods and spatial phylogenetics to dissect the drivers of diversity in *Astragalus*. We tested six main hypotheses: We first tested whether (1) the high species diversity of *Astragalus* is due to uniform, relatively high rates of diversification (cf. Sanderson and Wojciechowski, 1996) and not to bursts of diversification within the genus. We then tested whether (2) *Astragalus* diversification is primarily associated with cold and arid habitats or with (3) shifts in soil type and

chromosome number (4). Finally, we tested whether (5) shifts in habitat are associated with shifts between biogeographic regions and (6) whether *Astragalus* originated in West Asia with direct connections to other realms from this ancestral region.

MATERIALS AND METHODS

Taxon sampling

The taxa sampled are part of the NitFix sequencing initiative (<https://nitfix.org/>; Kates et al., 2022 [preprint]), with DNA sequence data reanalyzed for the purpose of this study. DNA was sampled from herbarium specimens using a sample management protocol reported previously (Folk et al., 2021a). In total, 939 taxa were included, with 847 in the ingroup. Outgroup sampling was extensive across the Astragalean clade sensu Lewis et al. (2005) to yield a strong biogeographic foundation. Representatives of *Clianthus*, *Colutea*, *Eremosparton*, *Erophaca*, *Lessertia*, *Smirnowa*, *Sphaerophysa*, *Sutherlandia*, *Swainsona*, and the *Astragalus* segregates *Biserrula*, *Podlechiella*, and *Phyllolobium* were included. *Oxytropis*, biogeographically similar to *Astragalus*, was extensively sampled (71 taxa). Samples of the segregate genera *Ophiocarpus* and *Barnebyella* (Lewis et al., 2005) were not available, but these are almost certainly better treated as *Astragalus* (Kazempour Osaloo et al., 2003). As the most distant outgroup, *Wisteria* (sensu Lewis et al. 2005, a member of the inverted repeat-lacking clade [IRLC] but outside the Astragalean clade; Wojciechowski et al., 1999, 2004) was used to root the trees. Appendix S1 provides accession information; assignments to clade and sections given the phylogenetic results (see below) are available in Appendix S2. Sampling statistics for the phylogeny are reported in the Results.

Molecular methods

DNA was extracted from herbarium specimens according to a modified CTAB protocol (Doyle and Doyle, 1987) reported previously (Folk et al., 2021a) and performed at the University of Florida (Gainesville, FL, USA). Extracts were then cleaned of remaining secondary compounds with SPRI beads and submitted for preparation of Illumina TruSeq libraries, omitting a size selection step to account for DNA degradation. We used a Hyb-Seq approach (Weitemier et al., 2014) targeting 100 low-copy nuclear loci, with data generated as part of the NitFix sequencing initiative (Kates et al., 2022 [preprint]). Library construction and target capture were performed by Rapid Genomics (Gainesville, FL, USA). Details concerning the bait set have been reported previously (Folk et al., 2021a; Kates et al., 2022 [preprint]).

Phylogenomic assembly

Sequences were assembled using the HybPiper v1.3 pipeline (Johnson et al., 2016). Given the wide phylogenetic scope of the probe set used for hybridization, spanning the four taxonomic orders that comprise the nitrogen-fixing clade of angiosperms, target recovery was achieved with the use of an amino acid reference for each locus, instead of DNA, with the BLAST option for read mapping. HybPiper flags loci for which multiple contigs were assembled as potential paralogs on a per-species basis, deciding the putative “ortholog” by assessing sequencing depth and contig length; default coverage thresholds were used. The `hybpiper_stats.py` script was used to extract assembly information; given the low percentage of flagged paralogs (96% of the samples had 5% or fewer of loci flagged as potentially paralogous; paralogy flags are summarized in Appendix S3), we did not employ further paralog resolution methods, instead using the potential “ortholog” selected by the pipeline. Assembly success in terms of number of assembled loci is summarized in Appendix S4. General assembly statistics are available on the project GitHub (https://github.com/ryanafolk/astragalus_niche_biogeo/blob/main/tree_quality_control/qc_astragalus_rename.csv).

Phylogenomic analysis

Individual gene trees were obtained using maximum likelihood in RAXML v.8.2.11 (Stamatakis, 2014), with 100 bootstraps and using the GTR+G nucleotide substitution model. Gene trees were then used either as inferred (no branches collapsed) or applying a branch-collapsing criterion, requiring bootstrap support <10% per a previous benchmarking study (Zhang et al., 2017). Using gene trees either with or without the branch-collapsing criterion, a species tree was obtained with the multispecies coalescent method in ASTRAL-III (Zhang et al., 2018), with branches annotated with support values calculated as local posterior probabilities (LPPs). Because coalescent species trees do not have terminal branch lengths without population sampling and internal branches measure branch length partly in terms of population size, branch lengths were re-optimized in RAXML using the option “-f e” to yield branch lengths in expected per-site substitutions. Because this tree was used for divergence time estimation, which could be biased by gene-tree discord (Burbrink and Pyron, 2011), quartet scores were also calculated both globally and per branch to identify areas of the tree with high discord. These quartet scores measure the percentage of gene trees that agree with the species tree and the branch, respectively.

Two methods were used to date the tree: penalized likelihood as implemented in treePL (Smith and O'Meara, 2012) and a relative rate framework (RRF) as implemented in RelTime (Tamura et al., 2012) on the pre-calculated branch lengths. Secondary calibration points were extracted from the chronogram of Azani et al. (2019),

consisting of four calibrations: a root constraint (minimum: 32.33 million years ago [Mya], maximum: 44.82 Mya), *Oxytropis* crown group (min: 1.3 Mya, max: 6.3 Mya), an *Astragalus* crown group (min: 11.47 Mya, max = 16.95 Mya), and an *Oxytropis/Astragalus* split (min: 11.93 Mya, max: 20.33 Mya). These secondary calibration points were implemented as minimum and maximum constraints. The “prime” option was used to calculate the initial parameters, and the analysis was conducted with cross-validation, which favored a smoothing rate of 100. In RelTime, the constraint scheme was the same except that (1) the root constraint was omitted due to the differing treatment of outgroups in RelTime, and (2) constraints were implemented under a uniform distribution and default parameters.

Diversification rates

Estimations of diversification rates have been controversial (Moore et al., 2016; Louca and Pennell, 2020), so we used a combination of approaches here including both parametric and semiparametric methods and estimates across timescales including the present day (cf. Louca and Pennell, 2020). To obtain semiparametric estimates of speciation rate (cf. Title and Rabosky, 2019) across *Astragalus* species, we used tip speciation rates according to the DR statistic of Jetz et al. (2012), calculated using the function of Sun et al. (2020b). To estimate shifts in diversification rate regimes, we used BAMM v. 2.5.0 (Rabosky, 2014). Priors followed recommendations from the R package BAMMtools, with additional settings as follows: expected shifts = 1 per documentation, segLength (likelihood grain) = 0.02, minimum clade size for shifts = 2 (i.e., shifts constrained to internal branches). Incomplete sampling was accounted for using the taxonomic data from the website *Astragalus* of World (Maassoumi, 2020). Because clade membership of unsequenced taxa is not certain (see Results), we assigned different sampling probabilities for Neo-*Astragalus* (0.4472) and the Eurasian clades (0.2041), as this was the lowest phylogenetic level at which applying taxonomic species count data for previously unsequenced species was feasible. Because outgroup sampling practices violate sampling assumptions of many birth–death models (Stadler, 2009; Spasojevic et al., 2021) and models that can incorporate imbalanced sampling still have poor performance under sparse sampling of higher-level taxa (Sun et al., 2020a), diversification estimation was limited to the ingroup, the *Astragalus* s.s. Markov chain Monte Carlo (MCMC) was run in four chains for 64.4 million generations (stopped based on ESS > 400 and strong convergence), $\delta T = 0.01$, swap period = 1000 generations, and sampling every 100,000th generation; further parameters followed BAMM documentation (<http://bamm-project.org/>).

Assembly of locality data

We assembled a data set of georeferenced occurrence records for species of *Astragalus* by first downloading all

occurrence records in Fabales from the online biodiversity discovery platforms iDigBio (iDigBio.org, 2020) and the Global Biodiversity Information Facility (GBIF.org, 2020). Using a list of all recognized Fabales species names and their associated synonyms, we aggregated all records of synonyms to accepted species names. This Fabales data set was then filtered to include only species in *Astragalus* and *Oxytropis* and further sampled outgroups. Next, occurrence records were subjected to a data cleaning process that relied heavily on the R package CoordinateCleaner (Zizka et al., 2019). In this process, records were removed if they were missing coordinates, or if latitude values were not between -90 and 90 or if longitude values were not between -180 and 180 . We also removed records if coordinates had equal longitude and latitude values, were within 500 m of the geographic centroid of political countries or provinces, were within 500 m of the GBIF headquarters, had either zero longitude or latitude, or were within 100 m of locations in a global database of biodiversity institutions that includes zoos, botanical gardens, herbaria, universities, and museums (Zizka et al., 2019). Records were also removed if the minimum absolute distance between a record and all other records of the same species was greater than 1000 km. Duplicate records with the same longitude and latitude values were filtered to retain only one record.

Assembly of niche attributes

We assembled a set of niche attributes selected to best capture properties of interest for *Astragalus*, such as climate and soil, while selecting parameters that show minimal correlation at global scales (Pearson's $R^2 < 0.7$). Using previously published code (Folk et al., 2019), we extracted niche attributes for 18 predictors representing climate, soil, elevation, and land cover. Bioclimatic data (Hijmans et al., 2005) were BIO1, BIO2, BIO3, BIO4, BIO7, BIO12, BIO15, BIO17. From the larger set of 19 Bioclim variables, we chose a limited set of predictors that cover both absolute values and variance in temperature in precipitation. The temperature variables represent mean annual temperature (BIO1), diurnal variability in temperature (BIO2, BIO3), and seasonal variability in temperature (BIO4, BIO7). The precipitation variables represent annual precipitation (BIO12) and seasonal variation in precipitation (BIO15 and BIO17). Elevational data were derived from GTOPO30 (EROS Center, 2018). Soil data derived from SoilGrids (Hengl et al., 2017) were nitrogen and carbon content, sand, and coarse fragment percentage, and the most likely of the 29 soil classifications of WRB 2006 (World Reference Base; FAO, 2006) for each grid cell. Finally, land-cover data (Tuanmu and Jetz, 2014) comprised herbaceous percentage of land cover (i.e., grassland) and coniferous (“needle-leaf”) and broadleaf forest percentage of land cover. For the quantitative variables (all attributes except for soil type classification), the niche occupancy of a species was summarized by the mean of all extracted occurrence values. For soil type, a categorical variable, the mode was used.

Ancestral reconstructions of environmental data were implemented in the R package *phytools*, with model choice via corrected Akaike information criterion (AIC_c); the model set included Brownian motion, Ornstein-Uhlenbeck (OU), and early-burst models for quantitative variables; OU models were selected without exception. For categorical soil type, only the equal-rates model could be implemented due to the high number of states. Estimates of phylogenetic signal were also implemented using *phytools*, using the lambda test (Pagel, 1999). This method measures phylogenetic signal as the parameter λ , with $\lambda = 0$ indicating no phylogenetic signal and $\lambda = 1$ indicating phylogenetic signal of the strength expected under Brownian motion; the significance of phylogenetic signal ($\lambda \neq 0$) is implemented as a likelihood ratio test.

Biogeography

Coding of geographic regions followed a previously developed shape file and Python code for regionalization (Folk et al., 2021b). Briefly, considering the regions most important to *Astragalus*, we recognized North and South America as distinct regions, and Africa was inclusive of the southern Mediterranean Basin. Western Europe and the northern Mediterranean were considered a region combined with boreal Eurasia, distinct from West Asia (the most important area for *Astragalus* in terms of species richness), the latter delimited from Anatolia east to the Altai, north to the Pontic area and south to the Arabian Peninsula following suggestions by Folk et al. (2021b) (Figure 3). East Asia and South Asia were recognized together as a further distinct region. In total, seven regions were recognized as summarized in Figure 3. Biogeographic scoring was manually checked with information from Plants of the World Online (POWO, 2023) and *Astragalus* of the World website (Maassoumi, 2020). We used a model comparison framework in BioGeoBEARS (Matzke, 2013) to select among DIVA-like, BAYAREA-like, and DEC biogeographic models; inclusion of parameterizations of jump dispersal (j) have proven controversial (Ree and Sanmartín, 2018; Matzke, 2022), so two favored models are presented in the Results, one considering a model set with and the other without j . The maximum range size was set to five, following the largest range size observed in extant taxa.

Chromosome number reconstruction

We used reported chromosome counts to investigate chromosome evolution in *Astragalus*, particularly whether Neo-*Astragalus* and other clades differ in the relative prevalence of aneuploidy vs. polyploidy. We gathered chromosome data from the Chromosome Counts Database (CCDB; Rice et al., 2015; <http://ccdb.tau.ac.il/>) and the *Astragalus* taxonomic database (Maassoumi, 2020), yielding 418 taxa with data. Chromosome ancestral states were

calculated using the ChromEvol model (Glick and Mayrose, 2014) as implemented in RevBayes (Höhna et al., 2016), considering rates of chromosome gains, losses, and polyploidization. The calibrated tree was trimmed to match the species for which counts were available, using the *drop.tip* function in the R package *ape* (Paradis et al., 2004). The analysis was run for 50,000 generations, with a 25% burn-in. Analysis results were summarized and plotted using the R package *RevGadgets* (Tribble et al., 2022).

Trait-associated diversification

Tests of trait-associated diversification were implemented both within the BAMM method and external to BAMM using tip speciation rates under a model-free paradigm. BAMM rate-trait association tests were implemented in STRAPP (Rabosky and Huang, 2016), which is a semiparametric permutation test using assigned BAMM diversification regimes to test for trait-associated regime shifts. Semiparametric tip rate were tested using ESSim (Harvey and Rabosky, 2018) for continuous data and FiSSE (Rabosky and Goldberg, 2017) for binary data. All three tests are designed to control for evolutionary pseudoreplication, which can result in spurious significance in state-dependent speciation extinction (SSE)-type models (see Rabosky and Goldberg, 2015). Multistate data (soil type) were recoded to binary characters to satisfy requirements for FiSSE as described in Results.

RESULTS

Sampling

We included 939 accessions, including 92 outgroups in the final phylogenetic tree. Of these, 373 taxa were assignable to Neo-*Astragalus* (44.7% of the total described diversity based on accepted taxa used by Maassoumi, 2020) and 474 taxa assignable to the Eurasian ("Old World") clades (20.4% of the total). While the Eurasian species are therefore under-sampled, species included here represent all recognized major clades (Groups A to F; Kazempour Osaloo et al., 2003, 2005; Azani et al., 2017b; Su et al., 2021). Of the names matched to recognized groups in the *Astragalus* of World taxonomic database (Maassoumi, 2020), this sampling effort comprises 16.3% of species in Group A, 15.1% of Group B, 50% of Group C, 16.5% of Group D, 12.6% of Group H, 7.6% of Group G, and 14.2% of Group F. Based on DNA sequence quality and manual comparison of phylogenetic placements with those expected from previous literature, 21 taxa in the phylogeny were excluded from downstream analysis. These exclusions mostly pertained to the Neo-*Astragalus* clade, and these are enumerated in the GitHub project repository (https://github.com/ryanafolk/astragalus_niche_biogeo). For downstream analyses using the phylogeny (e.g., biogeographical and environmental

analyses), we also randomly subsampled the tree terminals to one accession per species (thus eliminating multiple subspecific taxa and other cases of multiple samples per species), resulting in a tree with 818 tips.

Phylogenetic relationships

Relationships recovered here within the Astragalean clade differed from summary relationships reported by Wojciechowski et al. (2000), Lewis et al. (2005), and Azani et al. (2019). However, our results agree with those of Su et al. (2021), with a strongly supported monophyletic *Astragalus* s.s. (local posterior probability [LPP] of 0.99; hereafter, these analyses refer to the gene tree analysis without collapsed branches) sister to the remainder of the Astragalean clade, comprising *Oxytropis*+*Coluteoid* clade sensu Su et al. (2021). These relationships were primarily well resolved, although the clade supporting *Biserrula epiglottis* as sister to the remainder of the Astragalean clade exclusive of *Astragalus* s.s. was only moderately supported (LPP 0.79).

Within *Astragalus*, we recovered an overall set of clades similar to those of Kazempour Osaloo et al. (2005). The clade names in Figures 1–3 follow the system of Groups A–H for Eurasian species in Kazempour Osaloo et al. (2003, 2005). The backbone was well-supported with all LPP > 0.96. As seen earlier, Group A was sister to the rest of the genus. Groups C and B, whose branching order was unresolved in earlier studies, are successively sister to the remainder of *Astragalus*. Group G, recovered as a paraphyletic grade of two clades named G1 and G2 in Kazempour Osaloo et al. (2005) and as a polytomy with Group G3 in Kazempour Osaloo et al. (2003), was recovered as monophyletic here, although with limited shared taxon sampling among these studies. We did not recover Group E; two species of this small group sampled here (*A. sinicus*, *A. alpinus*), also included in Kazempour Osaloo et al. (2005) with two other taxa, were recovered here in (respectively) a clade of Asian species sister to the remainder of Group C and embedded in Group D. Group H was recovered as sister to Group G with low support (LPP < 0.5). Clades corresponding to Groups D and F were recovered as sister (LPP 0.88), and this clade was sister to Neo-Astragalus (LPP 0.96). Groups A–D were not supported as monophyletic (LPP < 0.5; although confidently supported subclades were recovered), but the remainder were well-supported (LPP > 0.95).

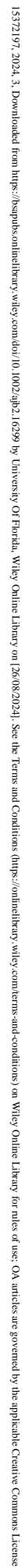
Additional species that had not been previously sequenced have been tentatively assigned to Groups A–H in the *Astragalus* taxonomic database (Maassoumi, 2020). Our results partly agree with these taxonomic assignments, with many exceptions; complete taxonomic subsectional and clade information can be found in Appendix S2. Group F contained the most species previously assigned to other groups, especially of species assigned to Group B (Appendix S2).

Our results were similar to recent studies of Asian *Astragalus* (Azani et al., 2017b; Su et al., 2021), with our Group F corresponding to the Hypoglottis and Diholcos clades of these studies. The Contortuplicata and Hamosa clades correspond to the Asian clade recovered within our Group C, and the Phaca clade corresponds to our Group A. Our study finds Neo-Astragalus as an independent clade not placed within Diholcos (Group F), but Azani et al. (2017b) and Su et al. (2021) had limited sampling of Neo-Astragalus. The Glottis clade, sister to all other *Astragalus* in Azani et al. (2017b) and Su et al. (2021), was recovered here as closer to *Oxytropis* among the outgroups (represented by *Astragalus epiglottis* = *Biserrula epiglottis*). We also did not recover a Pseudosesbanella clade as sister to the remainder of *Astragalus* exclusive of the Glottis clade (Su et al., 2021); instead, the species we sampled of the Pseudosesbanella clade as previously recognized are early diverging within Group A.

Gene tree support and discord

To explore the effect of different data trimming decisions, we ran coalescent species tree inference with and without gene-branch-collapsing criteria. Interestingly, the branch-collapsed analysis had decreased backbone support, with most backbone nodes having LPP > 0.9 but lower than in the analysis without branch-collapsing criteria. The average LPP for the total tree was 0.605 without branch-collapsing criteria and 0.567 with collapsing criteria. The branch-collapsed tree overall was similar in backbone relationships to the main result (previous section), but relationships within the Neo-Astragalus clade contained numerous differing clade resolutions with low support (LPP < 0.5) and included several species that in the main result and taxonomic assignments belong to group F (Appendix S5). Thus, gene-branch-collapsing criteria resulted in a violation of the monophyly of Neo-Astragalus by including several Eurasian species, a result not seen before (Wojciechowski et al., 1999). Observing more problems with an analysis using branch-collapsing criteria contrasts with the recommendations of Zhang et al. (2017); it is possible that this problem is due to the inference of gene trees on large numbers of similar sequences, where bootstrap measures can display downward bias and underestimate accuracy (Lemoine et al., 2018). Thus, the main phylogenetic analysis was favored for the remainder of the analyses.

The overall quartet score for the tree with branch-collapsing criteria was 0.6730 (~67% of gene tree quartets are concordant with the species tree), indicating a moderate level of discord. Plotting quartet scores across the phylogeny (Appendix S6) demonstrated similarly moderate levels of discord throughout the tree. However, the Neo-Astragalus clade was marked by extremely high discord in the backbone, with quartet scores near zero (i.e., few to no gene trees agree with the species tree at branches representing the Neo-Astragalus backbone). High gene tree heterogeneity in the Neo-Astragalus backbone was consistent with the sensitivity of these branches to branch-collapsing criteria (above). A tanglegram comparing analyses



with and without branch-collapsing criteria further demonstrates an impact on topology, which is almost entirely restricted to the Neo-Astragalus clade (Appendix S7).

The hypothesis of uniform diversification-rate regimes was definitively rejected (posterior probability ~ 0 for zero shifts). Rates varied through time, with a shallow peak in

the late Miocene (~10 Mya) and a strong peak at the Plio-Pleistocene boundary (~2.5 Mya; Figure 3). The first burst was approximately contemporaneous with the divergence of Groups A–H and Neo-Astragalus, while the second represents nested radiations early in the history of each major clade (Figure 3). While eight shifts were recovered in the best configuration for the diversification-rate regime (Figure 3), these were not clearly associated with any attributes studied here. BIO2 (BAMM STRAPP $P = 0.094$; ESSim $P = 0.7373$) and BIO3 (STRAPP $P = 0.056$; ESSim

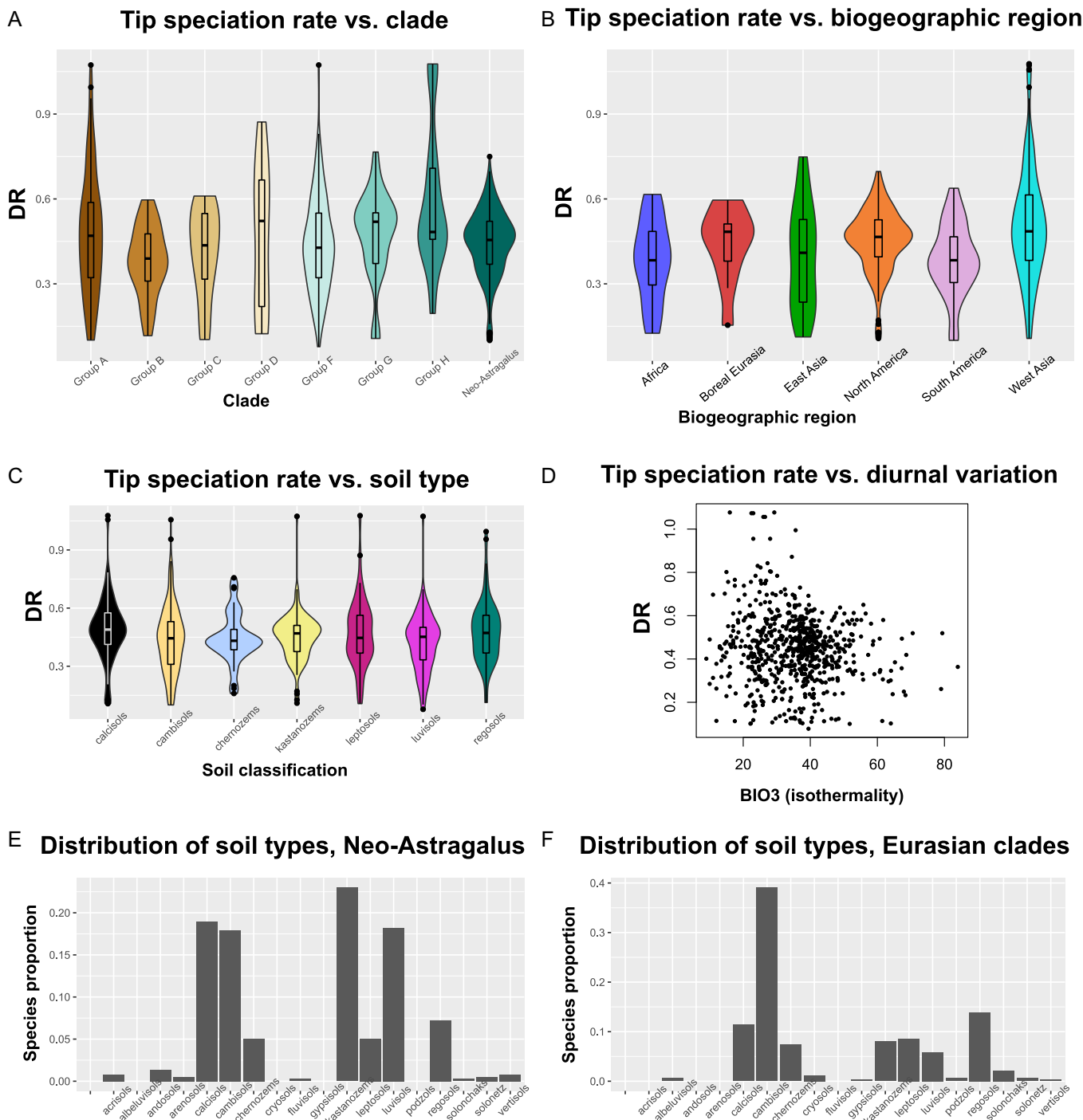


FIGURE 2 (A–D) Distribution of tip speciation rates, (A) vs. named clade, (B) vs. biogeographic region (colors are the same as in Figure 3; note Oceania with one outgroup species is omitted), (C) vs. soil type (colors are the same as in Figure 1C), and (D) vs. isothermality. (E, F) Distribution of species occupancy of soil types (E) for Neo-Astragalus and (F) for the Eurasian clades.

$P = 0.5435$; Figure 3D), the environmental characteristics closely associated with Neo-Astragalus, were not significantly associated with diversification. Soil characteristics also were not associated with diversification (STRAPP $P = 0.07$; FiSSE $P = 0.9391$; soil types were recoded as calcisols vs. other soil types for FiSSE to yield a binary contrast; Figure 3C). Ploidal levels did not differ in

diversification rate based on chromosome counts (STRAPP $P = 0.799$; here counts $[n] > 13$ were considered polyploid following the ChromEvol result). Finally, tip speciation rates showed no obvious relationship with biogeographic region (Figure 3B) and did not differ among the major recognized clades of *Astragalus* (Figure 3A). Overall, despite BAMM detecting bursts of diversification, they are seen in most of

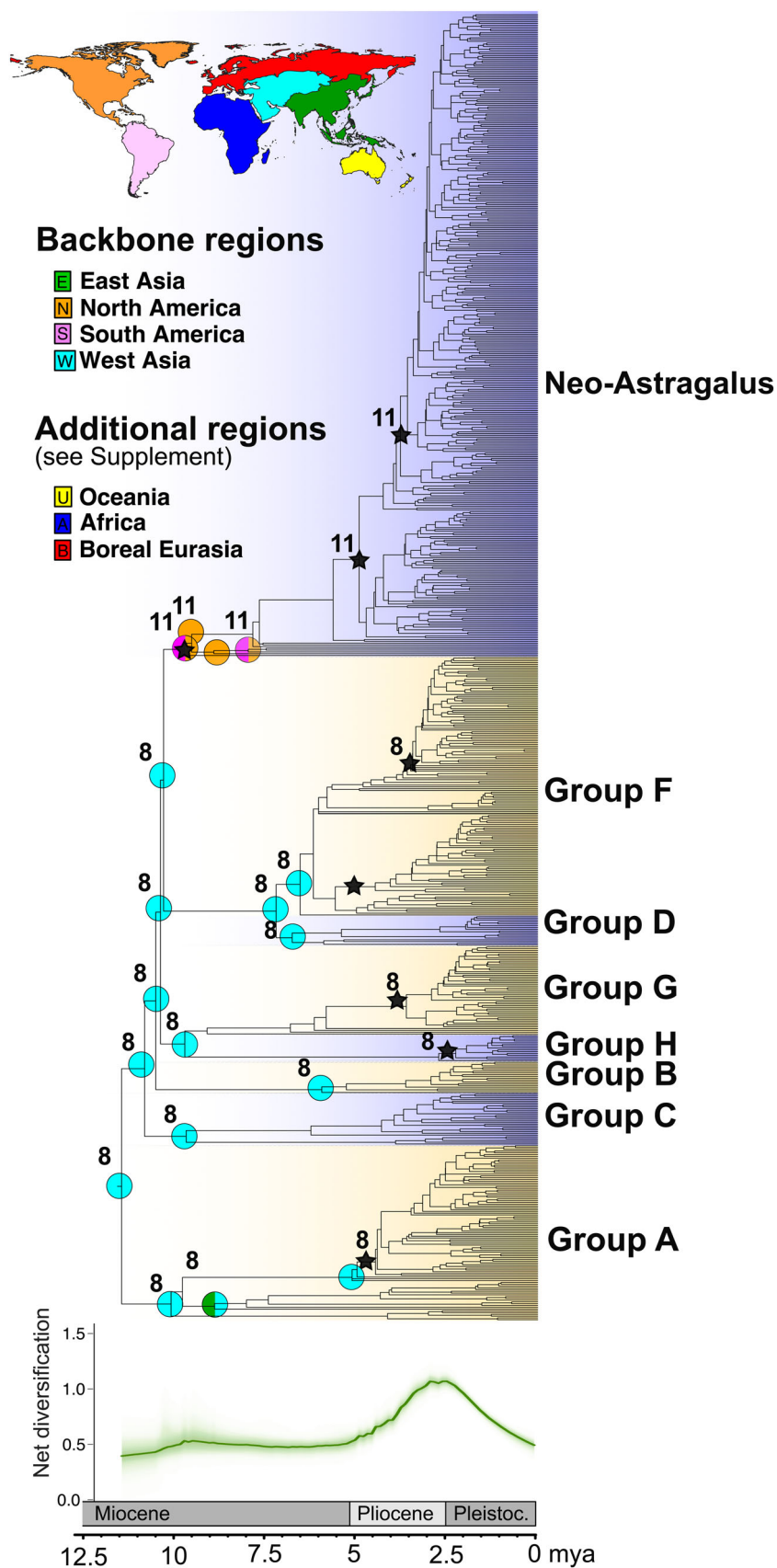


FIGURE 3 (See caption on next page).

the species-rich clades (Figure 3) and resulted in similarly high rates of diversification across *Astragalus* (Appendix S8). The only data set significant with respect to diversification was chromosome number (STRAPP $P = 0.027$; however, no chromosome shifts occurred directly on branches experiencing diversification bursts except for the stem lineage of Neo-Astragalus: Figure 3). These results suggest variable but overall high diversification rates in *Astragalus* that are not obviously linked to particular ecological strategies or other features investigated here, consistent with the idea of a non-adaptive radiation (Givnish, 2015).

Phylogenetic signal in quantitative niche attributes

All measured niche attributes showed strong evidence of phylogenetic niche conservatism (all highly significant, $\lambda > 0.5$ in all cases; see Appendix S9). While strong phylogenetic structuring may be the rule for niche attributes, in most cases this involved small, recently diverged clades (Figure 1; Appendix S10; cf. Wiens et al. [2010] for discussion of niche conservatism and phylogenetic scale). However, ancestral reconstruction (Figure 1A, B; performed under the favored OU model) revealed that diurnal temperature range and particularly isothermality (BIO2 and BIO3) were remarkable in clearly delimiting the Neo-Astragalus clade of the Americas from Eurasian clades (Figure 1B); no other niche attributes corresponded so strongly with geography. While strong diurnal temperature swings are associated with arid environments, this phylogenetic distribution was highly distinct from that seen for aridity, temperature, and precipitation (Figure 1A; Appendix S10). Likewise, elevation (Appendix S10), although an important factor, does not delimit the major clades of *Astragalus*. Biogeographic region alone explains a remarkable 51.8% and 51.7% of the variance in diurnal range and isothermality, respectively (ANOVA adjusted R^2). Sampling random cells from the predictor layers in the range of Neo-Astragalus and the Eurasian clades as a null demonstrates that this difference cannot be attributed to overall continental climatic differences in these variables (t -test, BIO2: $P = 0.6394$; BIO3: $P = 0.4816$).

Because environmental trends were also associated with clade membership (Figure 1A) and biogeographic domain and could have simply been driven by phylogenetic structuring of distributions, we performed a two-way MANOVA (Appendix S11; here, including only the quantitative niche attributes), showing that taxonomic group (using the taxonomy of the database *Astragalus* of

world; $P < 2e-16$) and biogeographic domain ($P < 2e-16$) have independent explanatory power for niche occupancy, as does their interaction ($P = 0.0007413$). This finding suggests that there is both clade- and area-based niche specialization.

Phylogenetic signal in soil type

The only other environmental characteristic that delimited Neo-Astragalus from Eurasian species and showed similar deep-level phylogenetic structuring to BIO2 and BIO3 was the FAO soil classification, which is based on physical and chemical characteristics that are rooted primarily in soil formation processes such as erosion and alluvial deposition. Direct measures of soil fertility such as carbon and nitrogen content did not show similar phylogenetic structuring, suggesting this result was independent of direct nutrient challenge. Overall, calcisols, cambisols, chernozems, kastanozems, leptosols, luvisols, and regosols, most of which are poor soils of northern and arid regions, account for nearly all species diversity (Figure 2E, F). Neo-Astragalus species occur in calcisol-enriched soils (defined partly by high calcium content and characteristic of the southwestern arid United States), kastanozems (a humus-rich soil associated with grasslands), and luvisols (a typical clay-rich temperate soil type; Figure 2E). Eurasian species primarily occur in cambisols and regosols (Figure 2F), soils associated with high erosion that are also found to a lesser extent to support Neo-Astragalus species. Cambisols, the most common soil association for the genus and a common soil type in cold and arid regions, are reconstructed with high probability as ancestral in *Astragalus* and *Oxytropis*, their most recent common ancestors, and the entire backbone of *Astragalus* including Neo-Astragalus and all other major named clades (Figure 1C).

Chromosome reconstruction

Chromosome reconstructions in ChromEvol (Figure 3; Appendix S12) definitively supported a base number of $n = 8$ for the Astragalean clade. This base number was largely retained throughout the eastern hemisphere grade and outgroup genera. The exceptions, considering only internal phylogenetic nodes, are a shift (with low posterior probability) to $n = 6$ in a small clade of *Oxytropis* and a shift to $n = 7$ in a small clade within *Astragalus* Group C. A shift to $n = 11$ occurred in the ancestor of Neo-Astragalus, which was followed by a shift to $n = 12$ in a clade of South

FIGURE 3 Biogeography and diversification of *Astragalus* based on the main analysis without branch-collapsing criteria and treePL divergence-time estimation. Ancestral biogeographic regions are indicated by colored circles according to the map legend (colors are the same as in Figure 2B); only *Astragalus* s.s. is shown here, and only backbone nodes and crown nodes of named clades are plotted. See Appendix S13 for full result. Black stars denote all recovered BAMM diversification shifts; see Appendix S8 for branches colored by diversification rate. Boldfaced numbers indicate reconstructed ancestral chromosome numbers, plotted for all nodes with diversification shift and/or biogeographic region plotted; see Appendix S12 for full result.

American species and three clades of North American species, and to $n = 13$ in a clade of South and North American species (Appendix S12). ChromEvol also detected numerous ancestral polyploidization events. These were distributed in small clades in *Oxytropis* and in *Astragalus* Groups B, G, and D (Appendix S12).

Biogeographic results

The DEC model was favored against other models without a jump parameter (ΔAIC_c 97 compared to the second-best BAYAREA-LIKE model); DEC+J was the favored model overall, with ΔAIC_c 48). DEC (Appendix S13) and DEC+J models (Appendix S14) were largely similar; both are discussed below when they disagree concerning major clades. We recovered a strongly supported origin of *Astragalus* in West Asia, dated by treePL to ~11.47 ya (14.17 Mya in RelTime) (hereafter main text refers to treePL dates with RelTime dates in parentheses; crown ages in both cases), with these ages consistent with the variation in estimates from previous studies focusing on late Miocene origins. For example, Azani et al. (2019) recovered a crown age of 14.16 My (11.47–16.95 95% highest posterior density [HPD] in their study) and also reconstructed West Asian origins; Su et al. (2021) recovered a crown age of 12.51 My (9.45–16.00 95% HPD). Among named clades traditionally recognized in *Astragalus*, all named eastern hemisphere clades (Groups A–H) retained the ancestral West Asia distribution at the crown divergence.

Dispersals to the other two important regions for the eastern hemisphere grade of *Astragalus*, boreal Eurasia and Africa, occur almost solely through western Asia, and in both dating analyses dispersal occurred primarily in the last 2–3 my (see the detailed result in Appendix S13). Likewise, *Astragalus* shows no examples of the eastern Asia–eastern North America floristic disjunction despite a strong presence in eastern Asia (primarily in the Himalaya–Qinghai–Tibet area, less often associated with disjunct distributions than eastern China). Almost without exception, eastern Asian clades descend by vicariance from ancestors with eastern and western Asian distributions; *Astragalus* clade A has a single example of a boreal-eastern Asia vicariant event.

Almost all *Astragalus* species of the Americas have long been known to form the Neo-Astragalus clade (Wojciechowski et al., 1993), and our more densely sampled results are consistent with the recovery of this clade. In the DEC model, a single dispersal from western Asia led to a broad ancestral distribution across the Americas, reconstructed at 9.84 Mya (11.99 My in RelTime; both estimates older than the much younger dates of 4.4 Mya recovered by Scherson et al. [2008] and 4.36 Mya by Azani et al. [2019]). This dispersal to the Americas was immediately followed by a vicariance across North and South America, with the North American lineage leading to most of the diversity of Neo-Astragalus. Consistent with the purpose of the model,

DEC+J instead proposed a jump dispersal to South America in the ancestor, immediately followed by a dispersal to North America. Thus, the inference of vicariance at the time of the most recent common ancestor of the Neo-Astragalus clade was sensitive to model choice, although the overall pattern was similar between the two analyses. Most South American species represent at least 13 subsequent dispersals to South America from North America with dates between 2.5–4 Mya in both dating analyses, four of which represent clades of 2–17 species. While this scenario conflicts with that of Scherson et al. (2008), who found evidence for two dispersals from North to South America, the difference is attributable partly to differing resolution of clades (neither of Scherson's South American clade was recovered here) but especially to improved sampling in our study, which included several South American taxa never studied before. Overall, this timeline and the predominant north-to-south direction are consistent with current understanding of the history of amphitropical disjunctions in plants. As reviewed by Simpson et al. (2017), ~81% of amphitropical disjuncts involve north-to-south dispersal, although this bias varies by habitat and was highest (~100%) in high-latitude plants and not evident (~46%) in desert plants. Long-distance dispersal is currently favored over vicariance as the mechanism to explain amphitropical disjunction, and it is thought that dispersal in this disjunction reflects the routes of migratory birds as the most likely dispersal vector (Simpson et al., 2017). Based on our current sampling, only a single species of Neo-Astragalus left the Americas, with *Astragalus polaris* extending from Alaska to the Russian Far East.

Approximately 15 North American *Astragalus* species, generally of more mesic environments than other species of the genus (Barneby, 1964), fall outside the Neo-Astragalus clade. These examples of secondary connections with North America, which include both endemics and species with large Northern Hemisphere distributions, are primarily descended from a broad range of ancestors distributed in boreal regions, and unlike Neo-Astragalus, date to no earlier than the early Pleistocene, as predicted respectively by Barneby (1964) and Wojciechowski et al. (1999).

DISCUSSION

Diversification of *Astragalus*

Sanderson and Wojciechowski (1996) were the first to quantify diversification rates in *Astragalus* and related genera; their primary hypothesis was that the Astragalean clade was characterized by relatively high but uniform rates of diversification. Considering that the methodologies available for investigating diversification were limited in the 1990s, the analyses they performed were more restricted than those considered here, primarily testing differences between genera rather than at the species level within genera. Nevertheless, the hypothesis of uniform

diversification rates is rejected here both considering rates within *Astragalus* and between *Astragalus* and other genera. Early-diverging clades within *Astragalus* share diversification regimes with outgroup taxa, while the Coluteoid clade and *Oxytropis* each have distinct diversification rate regimes, as do numerous core clades of *Astragalus*.

The rates we found, with a mean of between 0.5 and 1 speciation event per lineage per million years depending on the period (Figure 3) and rates as high as 7 speciation events/my (Appendix S8), are in line with those identified previously for the genus (cf. Wojciechowski et al., 1999; Scherson et al., 2008) and on the high end of what is known for flowering plants in general (reviewed by Lagomarsino et al. [2016]), well befitting the high diversity of the genus. Although the rates are high, the timeline of diversification is conventional for a temperate radiation (reviewed by Folk et al. [2020]). Two peaks of diversification in the late Miocene and Plio-Pleistocene accord with earlier work in *Astragalus* (Azani et al., 2019) and are typical of other major temperate clades representing diverse lineages of angiosperms (Arakaki et al., 2011; Folk et al., 2019; Sun et al., 2020b), a timing thought to reflect two major global historical cooling and aridification events that nevertheless favored temperate vegetation (Folk et al., 2019, 2020).

We could not identify any clear abiotic factors among temperature, precipitation, topography, or soil variables that are associated with diversification. While diversification appears to be relatively heterogeneous, our overall scenario nevertheless agrees with Sanderson and Wojciechowski (1996) that *Astragalus* and the entire Astragalean clade have relatively high diversification rates. Sanderson and Wojciechowski (1996) appeal to parallel origins of growth form to withstand abiotic stress and similar origins of chemical adaptations against herbivory to explain the high diversity of the group, yet abiotic ecological factors at least are rejected by our analyses. As well as assessing biotic interactions and chemical or physiological aspects, future work should focus on broadening the phylogenetic scope outside of the Astragalean clade, which may identify alternative drivers from what we have seen here (Graham et al., 2018), as Sanderson and Wojciechowski (1996) focused on the Astragalean clade as a whole.

Climate evolution in *Astragalus*

The most clear and surprising result of our investigation of niche attributes was the importance of diurnal temperature variation and its close tracking of biogeographic provinces. Given that we did not find a direct association of diurnal temperature range with ploidy or diversification, biogeographic area remains as the only significant associate with this niche attribute. While infrequently studied in plants, the importance of diurnal temperature range for determining plant distributions has been emphasized in previous studies. Ramsay (2001) and Zhao et al. (2018) found diurnal temperature variation to be an important

predictor of plant growth form, a result they attributed to physiological stress adaptation; likewise, there are interactions between diurnal temperature and plant germination (Liu et al., 2013) and with plant size (Myster and Moe, 1995). In some investigations, latitudinal diversity patterns have been partly attributed to diurnal temperature variation in plants and microbes, attributed to filtering out species with narrow physiological tolerances within highly variable biomes (Hu et al., 2019; Lancaster and Humphreys, 2020). At the level of species rather than clades, important variables identified in the *Astragalus* of Eurasia have focused on seasonal rather than diurnal variation (Yang et al., 2020) or soil and topographical variables (Safaei et al., 2018; Aghajanolou et al., 2021; Baumberger et al., 2021), so phylogenetic scale is an important consideration (Graham et al., 2018). However, studies in Neo-Astragalus have previously identified temperature variation as important for the clade; for instance, Jones et al. (2021) identified BIO4 as the most important PCA loading in the *A. sabulosus* group of the Colorado Plateau; similarly, BIO4 was the second-most important variable in a distribution modeling study of *A. utahensis*, and BIO3 was also included in the favored model (Baer and Maron, 2020).

Our *Astragalus* results indicate the importance of diurnal variation but are surprising for two reasons. First, the strict biogeographic divide is remarkable and does not represent an equivalent divide in the background environmental variation of the eastern and western hemispheres. This finding suggests that dispersal to the Americas was associated with niche specialization. Second, the primary correlate identified in the literature with diurnal variation is growth form, with high diurnal variation classically associated with low, cushion-like growth forms (Ramsay, 2001). *Astragalus*, by contrast, is remarkable for a morphological conservatism that in many cases fails to demark major clades (Sanderson and Wojciechowski, 1996). Although cushion forms are indeed prevalent in Neo-Astragalus, we did not have sufficient trait data to pursue this hypothesis further. The exact physiological adaptations relating to diurnal temperature variation occurring in *Astragalus* are therefore unclear; nevertheless, the status of this clade as including some of the most important nodulating plants in its range suggests a possible photosynthetic explanation rooted in water stress. As argued in several previous works (Adams et al., 2016; Pellegrini et al., 2016; Doby et al., 2022), plants engaging in root nodule symbiosis exist on a distinct area of the leaf economic spectrum as embodied by their differing nitrogen investments (McKey, 1994; Crews, 1999). High nitrogen investment in photosynthesis in legumes appears to relate primarily to confronting water stress (Adams et al., 2016), suggesting that nodulation may serve as part of a spectrum of adaptations in *Astragalus* that explains their success in distinctive freezing, arid areas. However, nitrogen-fixing symbiosis has seen limited study in *Astragalus*, and the implication of this important biogeographic trait for ecological specialization merits further study.

Soil biology

Edaphic ecology, while still understudied (Anacker, 2014; Rajakaruna, 2004, 2018; Hulshof and Spasojevic, 2020), is often singled out as one of the most important biogeographic factors in plants (Raven, 1964; Axelrod, 1972). Investigation of the distribution of soil niche across the *Astragalus* phylogeny and biogeographic regions confirms the existence of clade-level and region-level specialization. Aside from diurnal temperature, soil classification (which captures primarily how soil formed rather than nutrient availability) turned out to be one of the most phylogenetically conserved habitat attributes, much more so than individual chemical attributes like carbon and nitrogen (which have been identified as important in *Astragalus* at the species rather than clade level; cf. Safaei et al., 2018; Yang et al., 2020).

While legume biogeography is most often understood in terms of biomes and continents (Lewis et al., 2005), the distributions of nodulating legumes are clearly related to edaphic ecology (Sprenst et al., 2017). Soil rhizosphere environments and the physiological limitations that soil environments impress upon rhizosphere bacteria and fungi represent potential constraints on plant distributions (Parker et al., 2006; Simonsen et al., 2017). Spatially robust microbiome studies on nodulating plants remain few; one study of *Astragalus* considered soil properties but found geographic distance the most important predictor of biogeographic structure for rhizospheres (Li et al., 2021). Nevertheless, evidence from other legume species suggests that edaphic ecology may be a key factor dictating symbiotic choice (Rathi et al., 2018). Besides symbiont choice, the edaphic setting interacts with water stress because soils differ in their water retention characteristics, potentially leading to an increased importance of root nodule symbioses.

Given the few large-scale studies of edaphic characteristics in nodulating plants (Siniscalchi et al., 2022), progress documenting basic edaphic specializations are needed first if we are to make progress in understanding major contrasts such as that seen in *Astragalus* and test whether the importance of edaphic specialization can be attributed to differences in root nodule symbiosis or other factors. This study represents the first in-depth investigation of soil biology across *Astragalus* (see Wang and Wang, 2013; Wang et al., 2021 for studies on individual species), and indeed, to our knowledge, the first broad-scale phylogenetic investigation of recognized FAO (Food and Agricultural Organization) soil classifications in a large number of plant species. Further studies using gridded soil classification data are warranted to investigate the patterns we have recovered in other plant groups; it will be interesting to discover whether other clades display similar evidence for radiations in specialized soil environments.

Chromosome variation in *Astragalus*

Similarly to previous investigations, this study showed base chromosome numbers are closely associated with biogeographic patterns in *Astragalus*, with a major shift in base number associated with radiation in the Americas and more limited base shifts associated with many South American species. Higher base numbers as seen in Neo-*Astragalus* were often traditionally interpreted as old polyploids with subsequent aneuploid reduction (Grant, 1981: p. 367), but the hypothesis of ancestral polyploidy was not consistent with our reconstruction, which revealed only aneuploid increase. Instead, consistent with Wojciechowski et al. (1999), the eastern and western hemispheres are distinguished by distinct chromosome evolutionary mechanisms. Eurasian and African species are characterized by frequent polyploidy, whereas all chromosome evolution for Neo-*Astragalus* in the Americas was reconstructed as aneuploidy in our model.

There was marginal evidence for an association between chromosome number and diversification rates, but diversification shifts are not directly associated with base number shifts. Investigations into the relationships between chromosome evolution and either biogeography (e.g., Rice et al., 2019) or diversification (e.g., Landis et al., 2018) have yielded mixed results. Polyploidy has often been seen as a major source of evolutionary innovation in the face of stress, and aneuploidy, while less well studied, has similarly been associated with abiotic stress (Folk et al., 2020; Van de Peer et al., 2021). The major association between chromosomes and biogeographic province in *Astragalus* may likewise relate to survival of abiotic stress through mechanisms promoting genetic variation.

Biogeography of *Astragalus*

As suggested by Maassoumi and Ashouri (2022), western Asia and especially the area comprising modern-day Iran, is the center of Eurasian *Astragalus* diversity, both geographically and in terms of species richness and diversity of major lineages. Our results support this conclusion by suggesting western Asia as the ancestral area of *Astragalus* and as an essential crossroads for achieving its modern-day distribution. Dispersal to the Americas and other regions of Eurasia and Africa mostly depend on ancestors that were distributed in western Asia. Only transitions between North and South America are substantially independent of western Asia; otherwise, movement into eastern Asia, Africa, and boreal Eurasia occurred from western Asia without exception in the biogeographic reconstruction.

We found more complex connections between North and South America than recovered in previous studies, likely reflecting the denser taxon sampling used here. It has been long known that South American species are quite similar to North American species, seemingly more so than to each other, and certainly much more so than to the more

numerous central Asian relatives (Johnston, 1947). This problem was previously studied by Scherson et al. (2008), who suggested a North American origin, followed by two dispersals to South America. However, the study by Scherson et al. (2008) was limited by sampling (only 44 species of Neo-Astragalus) and did not use statistical biogeographic methods; likewise, studies sampling standard ITS and chloroplast markers have generally suffered from limited resolution caused by very low nucleotide divergence (Wojciechowski et al., 1999). The denser taxon sampling approach here points to a more complex story, favoring vicariance rather than a solely North American scenario for the earliest history of the clade and supporting a larger number of subsequent dispersals from North to South America.

Statistical support was lower for subclades in Neo-Astragalus compared to the other major *Astragalus* clades, which were relatively well resolved. The greater evidence for paralogs in this clade (Appendix S3), the sensitivity of detailed relationships in this clade to gene-branch-collapsing criteria (see Results), and the very high levels of discord that localize specifically within the Neo-Astragalus backbone, suggest that more work is needed to resolve Neo-Astragalus relationships and decisively place shallower geographic transitions in this clade in the context of complex evolutionary processes. Moreover, localization of high copy numbers in the Neo-Astragalus clade might suggest ancestral mass duplication, such as a tetraploid common ancestor (perhaps $n = 16$), followed by ancestral aneuploid decrease to arrive at the reconstructed $n = 11$ for the most recent common ancestor, accounting for the greater complexity of this clade in terms of copy number. However, this hypothesis would require other methods, particularly high-quality genome assemblies, to test more directly. Otherwise, our results are similar to those reported previously for *Astragalus*, with a decisive western Asian origin and a confident *Astragalus* backbone allowing us to place the early transition to the Americas.

***Astragalus*, *Oxytropis*, and Astragalean clade relationships**

Most previous phylogenetic work in *Astragalus* has used few genetic loci and focused either on representative sampling (Wojciechowski et al., 1999; Scherson et al., 2008) or on more extensive sampling of focal clades (Scherson et al., 2008; Bartha et al., 2013; Bagheri et al., 2017; Amini et al., 2019; Azani et al., 2017b, 2019) and areas (Su et al., 2021). Our analysis primarily corroborates previous phylogenetic findings across *Astragalus* while greatly improving sampling of Eurasian species. The phylogenetic placements we recovered validate and better resolve the “Old-World groups” of Kazempour Osaloo et al. (2003, 2005). As mentioned in the Results, while the delimitation of named clades is similar to previously published work, the well-supported backbone we recovered

is discordant with two recent studies that also resolved backbone relationships with high support (Azani et al., 2017b; Su et al., 2021); interestingly, our findings are more similar to earlier work (Kazempour Osaloo et al., 2005). Two explanations may account for discordant results compared to Azani et al. (2017b) and Su et al. (2021): first, these two studies included very few Neo-Astragalus species, thus lacking the sampling to place Neo-Astragalus confidently relative to eastern hemisphere clades, therefore possibly explaining the anomalous placement of Neo-Astragalus. Second, these studies share a focus on plant genomic loci with special evolutionary properties not shared with low-copy nuclear genes, with Su et al. (2021) using whole plastid genomes and Azani et al. (2017b) using ITS (the nuclear internal transcribed spacer) and plastid *trnK/matK*; Kazempour Osaloo et al. (2005) only included ITS sequences. Plastid genes are usually uniparentally inherited, not recombinationally independent from one another, and are often treated as single genes for the purpose of phylogenetic analysis (Doyle, 1992; but see Gonçalves et al., 2020); likewise, ribosomal genes are subject to gene conversion with potential phylogenetic effects (Baldwin et al., 1995). Hence, the previous primary use of genomic compartments with special inheritance properties underlines the value of alternative evidence from the nuclear genome. Future studies should investigate the potential for cytonuclear phylogenetic conflict, which may involve deep relationships (Sun et al., 2015; Stull et al., 2020, 2023) and could be expected given that hybridization is thought to be present in *Astragalus* (Bartha et al., 2013).

Finer-scale clades approximately correspond to sectional taxonomy (Appendix S2), but there are many exceptions at the species level, and results will be of service in revising subgeneric taxonomy with additional focused sampling. However, Barneby's (1964) sectional classification of Neo-Astragalus is not supported, with many well-supported clades that are discordant with sectional delimitation (see Appendix S2). However, limited backbone support in Neo-Astragalus in this analysis means that much of the Neo-Astragalus tree remains unresolved despite the increase in marker and taxon sampling. These groups should be tested further with intensive phylogenomic analyses and additional marker sets to improve resolution.

Among our outgroups, we sampled 71 species of *Oxytropis*, a monophyletic segregate of *Astragalus* more prevalent in the Arctic (Meyers, 2012; Sandanov et al., 2022) that has been studied previously using small numbers of chloroplast and nuclear ribosomal markers (Archambault and Strömvik, 2012; Meyers, 2012; Kholina et al., 2016, 2020; Tekpinar et al., 2016a, b) and to a limited extent, phylogenomic data (Shahi Shavvon et al., 2017). Biogeography of the *Oxytropis* species of the Americas was investigated by Archambault and Strömvik (2012) and Tekpinar et al. (2016b). Both studies suggested multiple dispersals to the Americas from Eurasia but were limited by very little nucleotide variation and were unable to specifically identify the origin of this dispersal. We found

four dispersals to North America, two each in the two major *Oxytropis* clades recovered. Interestingly, two of them in one of the clades (most recent common ancestor of [*Oxytropis pilosa*, *Oxytropis yunnanensis*]) accord with *Astragalus* in being dispersals from western Asia (Appendix S13), but the remaining two dispersal events in the other clade are classic eastern Asia–eastern North America disjunctions (Wen, 1999), a biogeographic pattern never observed in *Astragalus*. It is possible that, rather than strict long-distance dispersal, transitions of *Astragalus* to the Americas from western Asia involved eastern Asian distributional areas that were transient and quickly extirpated. While such a pattern is similar to distributional patterns seen in *Oxytropis*, the pattern would likewise be inconsistent with a standard eastern Asia–eastern North America disjunction pattern. Overall, *Oxytropis* would be an excellent subject of further work because the relationships we recovered do not accord well with subgeneric taxa (e.g., Malyshev, 2008; Archambault and Strömviik, 2012), but assessing detailed relationships in this genus are outside the scope of this study, and we lack the sampling to test the monophyly of existing taxa further.

In terms of subtribal relationships, our analysis divides the Astragalean clade (currently tribe Galegeae pro parte) into two sister clades: *Astragalus* and the Coluteoid clade, a series of genera around and including *Oxytropis*; relationships in the Coluteoid clade are similar to those reported in Zhang et al. (2009). Our results support the segregation of the sister genera *Biserrula* and *Podlechiella* (Kazempour Osaloo et al., 2003), both primarily distributed in the Mediterranean basin, and the primarily northern Asian *Phyllobium*.

CONCLUSIONS

Our results confidently identify western Asia as the area of origin of *Astragalus* and the key area for practically all major intercontinental biogeographic events. Temperature niche, edaphic ecology, and chromosome processes are remarkably distinct in the Neo-Astragalus clade of the Americas, yet neither these nor biogeographic regions could be related clearly to variation in diversification rates. Two diversification peaks, both corresponding to Miocene and Pliocene–Pleistocene global cooling events, could reflect temperature-dependent diversification, although such an association would be inconsistent with trait-associated diversification tests. Our major niche result, the biogeographic importance of diurnal temperature variation, has never been suspected in *Astragalus* and is not often studied in plants. Yet, like other factors, diurnal temperature variation does not have a clear relationship to diversification processes despite being closely associated with a species-rich clade. These results highlight the need to broadly study abiotic parameters at large scales in major radiations and to explicitly test their potential importance in

the diversification process. Further study would clarify whether *Astragalus* corresponds to a non-adaptive radiation (Rundell and Price, 2009; Givnish, 2015), which seems to match well on the basis of abiotic factors.

AUTHOR CONTRIBUTIONS

R.A.F. and C.M.S. conceived the study with input from all authors; C.M.S., R.A.F., M.B., and T.S. collected and/or analyzed data. J.L.M.C. contributed to data curation, analysis, and interpretation. D.E.S., P.S.S., R.P.G., and H.R.K. contributed to data analysis and interpretation. R.A.F. drafted the first version of this manuscript; all authors contributed to and approved the final version.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, award DEB-1916632. Sample collection and generation of sequence data were additionally supported by the Department of Energy, award DE-SC0018247. The authors thank herbarium curators for permission to perform DNA sampling from specimens (a full list of participating herbaria may be found in Appendix S1), particularly primary contributors Jim Solomon (MO) and Barbara Thiers (NY). Members of LPWG are thanked for consultation on taxonomic updates. The authors also thank the reviewers for their input, which substantively contributed to the final manuscript.

DATA AVAILABILITY STATEMENT

Scripts and data to produce the analyses, including phylogenetic trees, are available on GitHub at https://github.com/ryanafolk/astragalus_niche_biogeo; a permanent stable version of this repository is available on Zenodo at <https://doi.org/10.5281/zenodo.10382448>. Sequence data are deposited at SRA under BioProjects PRJNA1064968, PRJNA1021556, PRJNA1021608, PRJNA1021620, PRJNA1022015, PRJNA1022023, PRJNA1022025, PRJNA1022027, PRJNA1022029, PRJNA1022030, PRJNA1022032, PRJNA1022138, PRJNA1022141, PRJNA1022147, and PRJNA1022323; individual SRA accession numbers are listed in Appendix S1.

ORCID

Ryan A. Folk  <http://orcid.org/0000-0002-5333-9273>

Joseph L. M. Charboneau  <http://orcid.org/0000-0003-0136-735X>

Michael Belitz  <http://orcid.org/0000-0002-8162-5998>

Heather R. Kates  <http://orcid.org/0000-0001-8490-2201>

Douglas E. Soltis  <http://orcid.org/0000-0001-8638-4137>

Pamela S. Soltis  <http://orcid.org/0000-0001-9310-8659>

Robert P. Guralnick  <http://orcid.org/0000-0001-6682-1504>

Carolina M. Siniscalchi  <http://orcid.org/0000-0003-3349-5081>

REFERENCES

- Ackerfield, J., A. Susanna, V. Funk, D. Kelch, D. S. Park, A. H. Thornhill, B. Yildiz, et al. 2020. A prickly puzzle: generic delimitations in the *Carduus-Cirsium* group (Compositae: Cardueae: Carduinae). *Taxon* 69: 715–738.
- Adams, M. A., T. L. Turnbull, J. I. Sprent, and N. Buchmann. 2016. Legumes are different: leaf nitrogen, photosynthesis, and water use efficiency. *Proceedings of the National Academy of Sciences, USA* 113: 4098–4103.
- Afkhami, M. E., D. Luke Mahler, J. H. Burns, M. G. Weber, M. F. Wojciechowski, J. Sprent, and S. Y. Strauss. 2018. Symbioses with nitrogen-fixing bacteria: nodulation and phylogenetic data across legume genera. *Ecology* 99: 502.
- Aghajani, F., H. Mirdavoudi, M. Shojaei, E. Mac Sweeney, A. Mastinu, and P. Moradi. 2021. Rangeland management and ecological adaptation analysis model for *Astragalus curvirostris* Boiss. *Horticulturae* 7: 67.
- Amini, E., S. Kazempour-Osaloo, A. A. Maassoumi, and H. Zare-Maivan. 2019. Phylogeny, biogeography and divergence times of *Astragalus* section *Incarni* DC. (Fabaceae) inferred from nrDNA ITS and plastid *rpl32-trnL* (UAG) sequences. *Nordic Journal of Botany* 37: e02059.
- Anacker, B. L. 2014. The nature of serpentine endemism. *American Journal of Botany* 101: 219–224.
- Arakaki, M., P.-A. Christin, R. Nyffeler, A. Lendel, U. Eggli, R. M. Ogburn, E. Spriggs, et al. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences, USA* 108: 8379–8384.
- Archambault, A., and M. V. Strömvik. 2012. Evolutionary relationships in *Oxytropis* species, as estimated from the nuclear ribosomal internal transcribed spacer (ITS) sequences point to multiple expansions into the Arctic. *Botany* 90: 770–779.
- Ardley, J., and J. Sprent. 2021. Evolution and biogeography of actinorhizal plants and legumes: A comparison. *Journal of Ecology* 109: 1098–1121.
- Axelrod, D. I. 1972. Edaphic aridity as a factor in angiosperm evolution. *American Naturalist* 106: 311–320.
- Azani, N., M. Babineau, C. D. Bailey, H. Banks, A. R. Barbosa, R. B. Pinto, J. S. Boatwright, et al. 2017a. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: The Legume Phylogeny Working Group (LPWG). *Taxon* 66: 44–77.
- Azani, N., A. Bruneau, M. F. Wojciechowski, and S. Zarre. 2017b. Molecular phylogenetics of annual *Astragalus* (Fabaceae) and its systematic implications. *Botanical Journal of the Linnean Society. Linnean Society of London* 184: 347–365.
- Azani, N., A. Bruneau, M. F. Wojciechowski, and S. Zarre. 2019. Miocene climate change as a driving force for multiple origins of annual species in *Astragalus* (Fabaceae, Papilionoideae). *Molecular Phylogenetics and Evolution* 137: 210–221.
- Baer, K. C., and J. L. Maron. 2020. Ecological niche models display nonlinear relationships with abundance and demographic performance across the latitudinal distribution of *Astragalus utahensis* (Fabaceae). *Ecology and Evolution* 10: 8251–8264.
- Bagheri, A., A. Akhavan Roofgar, Z. Nemati, and F. R. Blattner. 2022. Genome size and chromosome number evaluation of *Astragalus* L. sect. *Hymenostegis* Bunge (Fabaceae). *Plants* 11: 435.
- Bagheri, A., A. A. Maassoumi, M. R. Rahiminejad, J. Brassac, and F. R. Blattner. 2017. Molecular phylogeny and divergence times of *Astragalus* section *Hymenostegis*: an analysis of a rapidly diversifying species group in Fabaceae. *Scientific Reports* 7: 14033.
- Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell, and M. J. Donoghue. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277.
- Barneby, R. C. 1964. Atlas of North American *Astragalus*. *Memoirs of the New York Botanical Garden* 13: 1–1188.
- Bartha, L., N. Dragoş, A. Molnár-V., and G. Sramkó. 2013. Molecular evidence for reticulate speciation in *Astragalus* (Fabaceae) as revealed by a case study from sect. *Dissitiflori*. *Botany* 91: 702–714.
- Baumberger, T., A. Baumel, P.-J. Dumas, J. Ugo, L. Keller, E. Dumas, T. Tatoni, et al. 2021. Is a restricted niche the explanation for species vulnerability? Insights from a large field survey of *Astragalus tragacantha* L. (Fabaceae). *Flora* 283: 151902.
- Burbrink, F. T., and R. A. Pyron. 2011. The impact of gene-tree/species-tree discordance on diversification rate estimation. *Evolution* 65: 1851–1861.
- Choi, I. S., D. Cardoso, L. P. de Queiroz, H. C. de Lima, C. Lee, T. A. Ruhlman, R.K. Jansen, M. F. Wojciechowski. 2022. Highly resolved papilionoid legume phylogeny based on plastid phylogenomics. *Frontiers in Plant Science* 13: 823190.
- Crews, T. E. 1999. The presence of nitrogen fixing legumes in terrestrial communities: evolutionary vs ecological considerations. *Biogeochemistry* 46: 233–246.
- De Storme, N., and A. Mason. 2014. Plant speciation through chromosome instability and ploidy change: cellular mechanisms, molecular factors and evolutionary relevance. *Current Plant Biology* 1: 10–33.
- Dludlu, M. N., S. B. M. Chimphango, C. H. Stirton, and A. M. Muasya. 2018. Distinct edaphic habitats are occupied by discrete legume assemblages with unique indicator species in the Cape Peninsula of South Africa. *Journal of Plant Ecology* 11: 632–644.
- Doby, J. R., D. Li, R. A. Folk, C. M. Siniscalchi, and R. P. Guralnick. 2022. Aridity drives phylogenetic diversity and species richness patterns of nitrogen-fixing plants in North America. *Global Ecology and Biogeography* 31: 1630–1642.
- Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Doyle, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144–163.
- EROS Center. 2018. USGS EROS Archive - Digital elevation - Global 30 arc-second elevation (GTOPO30). Earth Resource Observation and Science Center, U. S. Geological Survey, Sioux City, SD, USA. Website: <https://doi.org/10.5066/F7DF6PQS> [accessed 20 June 2018].
- FAO. 2006. FAO soils portal. Food and Agriculture Organization, Rome, Italy. Website: <https://www.fao.org/soils-portal/data-hub/soil-classification/world-reference-base/en/> [accessed 07 August 2023].
- Folk, R. A., H. R. Kates, R. LaFrance, D. E. Soltis, P. S. Soltis, and R. P. Guralnick. 2021a. High-throughput methods for efficiently building massive phylogenies from natural history collections. *Applications in Plant Sciences* 9: e11410.
- Folk, R. A., C. M. Siniscalchi, and D. E. Soltis. 2020. Angiosperms at the edge: extremity, diversity, and phylogeny. *Plant, Cell & Environment* 43: 2871–2893.
- Folk, R. A., R. L. Stubbs, N. J. Engle-Wrye, D. E. Soltis, and Y. Okuyama. 2021b. Biogeography and habitat evolution of Saxifragaceae, with a revision of generic limits and a new tribal system. *Taxon* 70: 263–285.
- Folk, R. A., R. L. Stubbs, M. E. Mort, N. Cellinese, J. M. Allen, P. S. Soltis, D. E. Soltis, and R. P. Guralnick. 2019. Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. *Proceedings of the National Academy of Sciences, USA* 116: 10874–10882.
- Frodin, D. G. 2004. History and concepts of big plant genera. *Taxon* 53: 753–776.
- GBIF.org. 2020. GBIF occurrence download. Website: <https://doi.org/10.15468/dl.yfddmm> [accessed 22 May 2020].
- Givnish, T. J. 2015. Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolution. *New Phytologist* 207: 297–303.
- Glick, L., and I. Mayrose. 2014. ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Molecular Biology and Evolution* 31: 1914–1922.
- Gonçalves, D. J. P., R. K. Jansen, T. A. Ruhlman, and J. R. Mandel. 2020. Under the rug: abandoning persistent misconceptions that obfuscate organelle evolution. *Molecular Phylogenetics and Evolution* 151: 106903.
- Graham, C. H., D. Storch, and A. Machac. 2018. Phylogenetic scale in ecology and evolution. *Global Ecology and Biogeography* 27: 175–187.
- Grant, V. 1981. Plant speciation. Columbia University Press, NY, NY, USA.

- Hardion, L., P.-J. Dumas, F. Abdel-Samad, M. Bou Dagher Kharat, B. Surina, L. Affre, F. Médail, et al. 2016. Geographical isolation caused the diversification of the Mediterranean thorny cushion-like *Astragalus* L. sect. *Tragacantha* DC. (Fabaceae). *Molecular Phylogenetics and Evolution* 97: 187–195.
- Harvey, M. G., and D. L. Rabosky. 2018. Continuous traits and speciation rates: alternatives to state-dependent diversification models. *Methods in Ecology and Evolution* 9: 984–993.
- Hengl, T., J. Mendes de Jesus, G. B. M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, et al. 2017. SoilGrids250m: global gridded soil information based on machine learning. *PloS one* 12: e0169748.
- Hennig, W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana, IL, USA.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.
- Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65: 726–736.
- Hu, A., Y. Nie, G. Yu, C. Han, J. He, N. He, S. Liu, et al. 2019. Diurnal temperature variation and plants drive latitudinal patterns in seasonal dynamics of soil microbial community. *Frontiers in Microbiology* 10: 674.
- Hulshof, C. M., and M. J. Spasojevic. 2020. The edaphic control of plant diversity. *Global Ecology and Biogeography* 19: 320.
- iDigBio.org. 2020. iDigBio occurrence download. Website: <http://www.idigbio.org/portal>. Query: {"core_type": "records", "rq": {"geopoint": {"type": "exists"}, "order": "Fabales"}, "form": "dwca-csv", "core_source": "indexterms", "mediarecord_fields": null, "record_fields": null, "mq": null}. [accessed 22 May 2020].
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. *Nature* 491: 444–448.
- Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. J. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4: apps.1600016.
- Johnston, I. M. 1947. *Astragalus* in Argentina, Bolivia and Chile. *Journal of the Arnold Arboretum* 28: 336–374.
- Jones, M. R., D. E. Winkler, and R. Massatti. 2021. The demographic and ecological factors shaping diversification among rare *Astragalus* species. *Diversity & Distributions* 27: 1407–1421.
- Kates, H. R., B. C. O'Meara, R. LaFrance, G. W. Stull, E. K. James, D. Conde, S. Liu, et al. 2022. Two shifts in evolutionary lability underlie independent gains and losses of root-nodule symbiosis in a single clade of plants. *Nature Communications bioRxiv* <https://doi.org/10.1101/2022.07.31.502231> [preprint].
- Kazempour Osaloo, S., A. A. Maassoumi, and N. Murakami. 2003. Molecular systematics of the genus *Astragalus* L. (Fabaceae): phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacers and chloroplast gene *ndhF* sequences. *Plant Systematics and Evolution* 242: 1–32.
- Kazempour Osaloo, S., A. A. Maassoumi, and N. Murakami. 2005. Molecular systematics of the Old World *Astragalus* (Fabaceae) as inferred from nrDNA ITS sequence data. *Brittonia* 57: 367–381.
- Kholina, A. B., M. M. Kozyrenko, E. V. Artyukova, D. V. Sandanov, and E. A. Andrianova. 2016. Phylogenetic relationships of the species of *Oxytropis* DC. subg. *Oxytropis* and *Phacoxytropis* (Fabaceae) from Asian Russia inferred from the nucleotide sequence analysis of the intergenic spacers of the chloroplast genome. *Russian Journal of Genetics* 52: 780–793.
- Kholina, A. B., M. M. Kozyrenko, E. V. Artyukova, V. V. Yakubov, M. G. Khoreva, E. A. Andrianova, and O. A. Mochalova. 2020. Phylogenetic relationships of *Oxytropis* section *Arctobia* of northeast Asia according to sequencing of the intergenic spacers of chloroplast and ITS of nuclear genomes. *Russian Journal of Genetics* 56: 1424–1434.
- Lagomarsino, L. P., F. L. Condamine, A. Antonelli, A. Mulch, and C. C. Davis. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist* 210: 1430–1442.
- Lancaster, L. T., and A. M. Humphreys. 2020. Global variation in the thermal tolerances of plants. *Proceedings of the National Academy of Sciences, USA* 117: 13580–13587.
- Landis, J. B., D. E. Soltis, Z. Li, H. E. Marx, M. S. Barker, D. C. Tank, and P. S. Soltis. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany* 105: 348–363.
- Ledingham, G. F. 1960. Chromosome numbers in *Astragalus* and *Oxytropis*. *Canadian Journal of Genetics and Cytology* 2: 119–128.
- Lemoine, F., J.-B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila Felipe, T. De Oliveira, and O. Gascuel. 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556: 452–456.
- Levin, D. A. 1975. Pest pressure and recombination systems in plants. *American Naturalist* 109: 437–451.
- Lewis, G. P., B. Schrire, B. Mackinder, and M. Lock. 2005. Legumes of the World. Royal Botanic Gardens, Kew, UK.
- Lichter-Marck, I. H., and B. G. Baldwin. 2023. Edaphic specialization onto bare, rocky outcrops as a factor in the evolution of desert angiosperms. *Proceedings of the National Academy of Sciences, USA* 120: e2214729120.
- Li, Y., Y. Yang, T. e Wu, H. Zhang, G. Wei, and Z. Li. 2021. Rhizosphere bacterial and fungal spatial distribution and network pattern of *Astragalus mongholicus* in representative planting sites differ the bulk soil. *Applied Soil Ecology* 168: 104114.
- Liu, K., J. M. Baskin, C. C. Baskin, H. Bu, G. Du, and M. Ma. 2013. Effect of diurnal fluctuating versus constant temperatures on germination of 445 species from the eastern Tibet Plateau. *PloS one* 8: e69364.
- Louca, S., and M. W. Pennell. 2020. Extant timetrees are consistent with a myriad of diversification histories. *Nature* 580: 502–505.
- Maassoumi, A. A. 2020. *Astragalus* of the world, version 1 [online]. Website: <http://astragalusofworld.com> [accessed 28 June 2023]
- Maassoumi, A. A., and P. Ashouri. 2022. The hotspots and conservation gaps of the mega genus *Astragalus* (Fabaceae) in the Old-World. *Biodiversity and Conservation* 31: 2119–2139.
- Malyshev, L. I. 2008. Phenetics of the subgenera and sections in the genus *Oxytropis* DC. (Fabaceae) bearing on ecology and phylogeny. *Contemporary Problems of Ecology* 1: 440–444.
- Matzke, N. J. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* 5: 242–248.
- Matzke, N. J. 2022. Statistical comparison of DEC and DEC+J is identical to comparison of two ClaSSE submodels, and is therefore valid. *Journal of Biogeography* 49: 1805–1824.
- McKey, D. 1994. Legumes and nitrogen: the evolutionary ecology of a nitrogen-demanding lifestyle. In J. L. Sprent and D. McKey [eds.], *Advances in legume systematics*, Part 5, 211–228. Chicago University Press, Chicago, IL, USA.
- Meyers, Z. J. 2012. A contribution to the taxonomy and phylogeny of *Oxytropis* section *Arctobia* (Fabaceae) in North America. M.S. thesis, University of Alaska, Fairbanks, AK, USA.
- Moore, B. R., S. Höhna, M. R. May, B. Rannala, and J. P. Huelsenbeck. 2016. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proceedings of the National Academy of Sciences, USA* 113: 9569–9574.
- Myster, J., and R. Moe. 1995. Effect of diurnal temperature alternations on plant morphology in some greenhouse crops—a mini review. *Scientia Horticulturae* 62: 205–215.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401: 877–884.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.
- Parker, M. A., W. Malek, and I. M. Parker. 2006. Growth of an invasive legume is symbiont limited in newly occupied habitats. *Diversity & Distributions* 12: 563–571.

- Pellegrini, A. F. A., A. C. Staver, L. O. Hedin, T. Charles-Dominique, and A. Tourgee. 2016. Aridity, not fire, favors nitrogen-fixing plants across tropical savanna and forest biomes. *Ecology* 97: 2177–2183.
- Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Website: <https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:330028-2> [accessed 07 August 2023].
- Plenk, K., W. Willner, O. N. Demina, M. Höhn, A. Kuzemko, K. Vassilev, and M. Kropf. 2020. Phylogeographic evidence for long-term persistence of the Eurasian steppe plant *Astragalus onobrychis* in the Pannonian region (eastern Central Europe). *Flora* 264: 151555.
- Polhill, R. M. 1981. Galegeae. In R. M. Polhill and P. H. Raven [eds.], *Advances in legume systematics*, part 1, 367–370. Chicago University Press, Chicago, IL, USA.
- Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS one* 9: e89543.
- Rabosky, D. L., and E. E. Goldberg. 2017. FiSSE: a simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution* 71: 1432–1442.
- Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology* 64: 340–355.
- Rabosky, D. L., and H. Huang. 2016. A robust semi-parametric test for detecting trait-dependent diversification. *Systematic Biology* 65: 181–193.
- Rajakaruna, N. 2018. Lessons on evolution from the study of edaphic specialization. *Botanical Review* 84: 39–78.
- Rajakaruna, N. 2004. The edaphic factor in the origin of plant species. *International Geology Review* 46: 471–478.
- Ramsay, P. M. 2001. Diurnal temperature variation in the major growth forms of an Ecuadorian páramo plant community. In P. M. Ramsay [ed.], *The ecology of Volcán Chiles: high-altitude ecosystems on the Ecuador–Colombia border*, 101–112. Pebble & Shell Publications, Plymouth, UK.
- Rathi, S., N. Tak, G. Bissa, B. Chouhan, A. Ojha, D. Adhikari, S. K. Barik, et al. 2018. Selection of *Bradyrhizobium* or *Ensifer* symbionts by the native Indian caesalpinoid legume *Chamaecrista pumila* depends on soil pH and other edaphic and climatic factors. *FEMS Microbiology Ecology* 94: fty180.
- Raven, P. H. 1964. Catastrophic selection and edaphic endemism. *Evolution* 18: 336–338.
- Ree, R. H., and I. Sanmartín. 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography* 45: 741–749.
- Rice, A., L. Glick, S. Abadi, M. Einhorn, N. M. Kopelman, A. Salman-Minkov, J. Mayzel, et al. 2015. The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers. *New Phytologist* 206: 19–26.
- Rice, A., P. Šmarda, M. Novosolov, M. Drori, L. Glick, N. Sabath, S. Meiri, et al. 2019. The global biogeography of polyploid plants. *Nature Ecology & Evolution* 3: 265–273.
- Rundell, R. J., and T. D. Price. 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution* 24: 394–399.
- Safaei, M., M. Tarkesh, H. Bashari, and M. Bassiri. 2018. Modeling potential habitat of *Astragalus verus* Olivier for conservation decisions: a comparison of three correlative models. *Flora* 242: 61–69.
- Samad, F. A., A. Baumel, M. Juin, D. Pavon, S. Siljak-Yakovlev, F. Médail, and M. Bou Dagher Kharrat. 2014. Phylogenetic diversity and genome sizes of *Astragalus* (Fabaceae) in the Lebanon biogeographical crossroad. *Plant Systematics and Evolution* 300: 819–830.
- Sandanov, D. V., A. S. Dugarova, E. P. Brianskaia, I. Y. Selyutina, N. I. Makunina, S. V. Dudov, V. V. Chepinoga, and Z. Wang. 2022. Diversity and distribution of *Oxytropis* DC. (Fabaceae) species in Asian Russia. *Biodiversity Data Journal* 10: e78666.
- Sanderson, M. J., and M. F. Wojciechowski. 1996. Diversification rates in a temperate legume clade: Are there 'so many species' of *Astragalus* (Fabaceae)? *American Journal of Botany* 83: 1488–1502.
- Scherson, R. A., R. Vidal, and M. J. Sanderson. 2008. Phylogeny, biogeography, and rates of diversification of New World *Astragalus* (Leguminosae) with an emphasis on South American radiations. *American Journal of Botany* 95: 1030–1039.
- Shahi Shavvon, R., S. Kazempour Osaloo, A. A. Maassoumii, F. Moharrek, S. Karaman Erkul, A. R. Lemmon, E. M. Lemmon, et al. 2017. Increasing phylogenetic support for explosively radiating taxa: the promise of high-throughput sequencing for *Oxytropis* (Fabaceae). *Journal of Systematics and Evolution* 55: 385–404.
- Simonsen, A. K., R. Dinnage, L. G. Barrett, S. M. Prober, and P. H. Thrall. 2017. Symbiosis limits establishment of legumes outside their native range at a global scale. *Nature Communications* 8: 14790.
- Simpson, M. G., L. A. Johnson, T. Villaverde, and C. M. Williams. 2017. American amphitropical disjuncts: perspectives from vascular plant analyses and prospects for future research. *American Journal of Botany* 104: 1600–1650.
- Siniscalchi, C. M., H. R. Kates, P. S. Soltis, D. E. Soltis, R. P. Guralnick, and R. A. Folk. 2022. Testing the evolutionary drivers of nitrogen-fixing symbioses in challenging soil environments. *bioRxiv* <https://doi.org/10.1101/2022.09.27.509719> [preprint].
- Smith, S. A., and B. C. O'Meara. 2012. treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28: 2689–2690.
- Spasojevic, T., G. R. Broad, I. E. Sääksjärvi, M. Schwarz, M. Ito, S. Korenko, and S. Klopstein. 2021. Mind the outgroup and bare branches in total-evidence dating: a case study of pimpliform Darwin wasps (Hymenoptera, Ichneumonidae). *Systematic Biology* 70: 322–339.
- Spellenberg, R. 1976. Chromosome numbers and their cytotaxonomic significance for North American *Astragalus* (Fabaceae). *Taxon* 25: 463–476.
- Sprent, J. I., J. Ardley, and E. K. James. 2017. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytologist* 215: 40–56.
- Stadler, T. 2009. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* 261: 58–66.
- Stamatakis, A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Stull, G. W., K. K. Pham, P. S. Soltis, and D. E. Soltis. 2023. Deep reticulation: the long legacy of hybridization in vascular plant evolution. *Plant Journal* 114: 743–766.
- Stull, G. W., P. S. Soltis, D. E. Soltis, M. A. Gitzendanner, and S. A. Smith. 2020. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *American Journal of Botany* 107: 790–805.
- Su, C., L. Duan, P. Liu, J. Liu, Z. Chang, and J. Wen. 2021. Chloroplast phylogenomics and character evolution of eastern Asian *Astragalus* (Leguminosae): tackling the phylogenetic structure of the largest genus of flowering plants in Asia. *Molecular Phylogenetics and Evolution* 156: 107025.
- Sun, M., R. A. Folk, M. A. Gitzendanner, P. S. Soltis, Z. Chen, D. E. Soltis, and R. P. Guralnick. 2020a. Estimating rates and patterns of diversification with incomplete sampling: a case study in the rosids. *American Journal of Botany* 107: 895–909.
- Sun, M., R. A. Folk, M. A. Gitzendanner, P. S. Soltis, Z. Chen, D. E. Soltis, and R. P. Guralnick. 2020b. Recent accelerated diversification in rosids occurred outside the tropics. *Nature Communications* 11: 3333.
- Sun, M., D. E. Soltis, P. S. Soltis, X. Zhu, J. G. Burleigh, and Z. Chen. 2015. Deep phylogenetic incongruence in the angiosperm clade Rosidae. *Molecular Phylogenetics and Evolution* 83: 156–166.
- Tamme, R., M. Pärtel, U. Kõljalg, L. Laanisto, J. Liira, Ü. Mander, M. Moora, et al. 2021. Global macroecology of nitrogen-fixing plants. *Global Ecology and Biogeography* 30: 514–526.
- Tamura, K., F. U. Battistuzzi, P. Billings-Ross, O. Murillo, A. Filipiński, and S. Kumar. 2012. Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences* 109: 19333–19338.

- Tekpinar, A., S. K. Erkul, Z. Aytaç, and Z. Kaya. 2016a. Phylogenetic relationships among native *Oxytropis* species in Turkey using the *trnL* intron, *trnL-F* IGS, and *trnV* intron cpDNA regions. *Turkish Journal of Botany* 40: 472–479.
- Tekpinar, A., S. K. Erkul, Z. Aytaç, and Z. Kaya. 2016b. Phylogenetic relationships between *Oxytropis* DC. and *Astragalus* L. species native to an Old World diversity center inferred from nuclear ribosomal ITS and plastid *matK* gene sequences. *Turkish Journal of Biology* 40: 250–263.
- Title, P. O., and D. L. Rabosky. 2019. Tip rates, phylogenies, and diversification: What are we estimating, and how good are the estimates? *Methods in Ecology and Evolution* 10: 821–834.
- Tribble, C. M., W. A. Freyman, M. J. Landis, J. Y. Lim, J. Barido-Sottani, B. T. Kopperud, S. Höhna, and M. R. May. 2022. RevGadgets: An R package for visualizing Bayesian phylogenetic analyses from RevBayes. *Methods in Ecology and Evolution* 13: 314–323.
- Tuanmu, M.-N., and W. Jetz. 2014. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling: consensus land cover. *Global Ecology and Biogeography* 23: 1031–1045.
- Van de Peer, Y., T.-L. Ashman, P. S. Soltis, and D. E. Soltis. 2021. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* 33: 11–26.
- Wang, Q., C. Chen, Z. Pang, C. Li, D. Wang, Q. Ma, and J. Wu. 2021. The role of the locoweed (*Astragalus variabilis* Bunge) in improving the soil properties of desert grasslands. *Rangeland Journal* 43: 47–54.
- Wang, Z.-B., and Q.-Y. Wang. 2013. Cultivating erect milkvetch (*Astragalus adsurgens* Pall.) (Leguminosae) improved soil properties in Loess Hilly and Gullies in China. *Journal of Integrative Agriculture* 12: 1652–1658.
- Weitemier, K., S. C. K. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* 2: apps.1400042.
- Wen, J. 1999. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annual Review of Ecology and Systematics* 30: 421–455.
- Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I. Damschen, et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* 13: 1310–1324.
- Wojciechowski, M. F., M. Lavin, and M. J. Sanderson. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* 91: 1846–1862.
- Wojciechowski, M. F., M. J. Sanderson, B. G. Baldwin, and M. J. Donoghue. 1993. Monophyly of aneuploid *Astragalus* (Fabaceae): evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* 80: 711–722.
- Wojciechowski, M. F., M. J. Sanderson, and J.-M. Hu. 1999. Evidence on the monophyly of *Astragalus* (Fabaceae) and its major subgroups based on nuclear ribosomal DNA ITS and chloroplast DNA *trnL* intron data. *Systematic Botany* 24: 409–437.
- Wojciechowski, M. F., M. J. Sanderson, K. P. Steele, and A. Liston. 2000. Molecular phylogeny of the 'temperate herbaceous tribes' of papilionoid legumes. *Advances in Legume Systematics* 9: 277–298.
- Yang, M., Z. Li, L. Liu, A. Bo, C. Zhang, and M. Li. 2020. Ecological niche modeling of *Astragalus membranaceus* var. *mongolicus* medicinal plants in Inner Mongolia, China. *Scientific Reports* 10: 12482.
- Záveská, E., C. Maylandt, O. Paun, C. Bertel, B. Frajman, and P. Schönschetter. The Steppe Consortium. 2019. Multiple auto- and allopolyploidisations marked the Pleistocene history of the widespread Eurasian steppe plant *Astragalus onobrychis* (Fabaceae). *Molecular Phylogenetics and Evolution* 139: 106572.
- Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.
- Zhang, C., E. Sayyari, and S. Mirarab. 2017. ASTRAL-III: Increased scalability and impacts of contracting low support branches. *Comparative Genomics* (Vol. 10562, pp. 53–75). Springer International, Cham, Switzerland.
- Zhang, M., Y. Kang, L. Zhou, and D. Podlech. 2009. Phylogenetic origin of *Phyllolobium* with a further implication for diversification of *Astragalus* in China. *Journal of Integrative Plant Biology* 51: 889–899.
- Zhao, Y., H. Cao, W. Xu, G. Chen, J. Lian, Y. Du, and K. Ma. 2018. Contributions of precipitation and temperature to the large scale geographic distribution of fleshy-fruited plant species: Growth form matters. *Scientific Reports* 8: 17017.

SUPPORTING INFORMATION

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

Appendix S1. Accession table.

Appendix S2. Phylogeny, broken in three parts for display as indicated.

Appendix S3. Number of HybPiper paralog flags per sample, mapped across all samples on the pre-curation phylogeny.

Appendix S4. Number of loci assembled to at least 50% of the reference, mapped across all samples on the pre-curation phylogeny.

Appendix S5. Phylogeny, broken in three parts for display.

Appendix S6. Quartet scores plotted across the *Astragalus* phylogeny.

Appendix S7. Tanglegram comparing the analysis with no gene tree filtering to that with branches with bootstrap support <10 collapsed.

Appendix S8. Plot of BAMM net diversification rate shifts.

Appendix S9. Results for phylogenetic signal tests.

Appendix S10. Ancestral reconstructions for the nine predictor variables not shown in the main text.

Appendix S11. Detailed results for two-way MANOVA, with variance partitioned by taxonomic group, biogeographic region, and their interaction.

Appendix S12. Phylogeny, broken in two parts for display.

Appendix S13. Complete biogeographic result under the DEC model (cf. Figure 1 in main text).

Appendix S14. Complete biogeographic result under the DEC+J model.

How to cite this article: Folk, R. A., J. L. M.

Charboneau, M. Belitz, T. Singh, H. R. Kates, D. E. Soltis, P. S. Soltis, R. P. Guralnick, and C. M. Siniscalchi. 2024. Anatomy of a mega-radiation: biogeography and niche evolution in *Astragalus*. *American Journal of Botany* 111(3): e16299. <https://doi.org/10.1002/ajb2.16299>