New Phytologist Supporting Information

Article Title: Global analysis of Poales diversification – parallel evolution in space and time into open and closed habitats

Authors: Tammy L Elliott, Daniel Spalink, Isabel Larridon, Alexandre Rizzo Zuntini, Marcial Escudero, Jan Hackel, Russell L Barrett, Santiago Martín-Bravo, José Ignacio Márquez-Corro, Carolina Granados Mendoza, Aluoneswi C. Mashau, Katya J Romero-Soler, Daniel A Zhigila, Berit Gehrke, Caroline Oliveira Andrino, Darren M Crayn, Maria S Vorontsova[4], Félix Forest, William J Baker, Karen L Wilson, David A Simpson, A Muthama Muasya

Article acceptance date: 3 November 2023

.

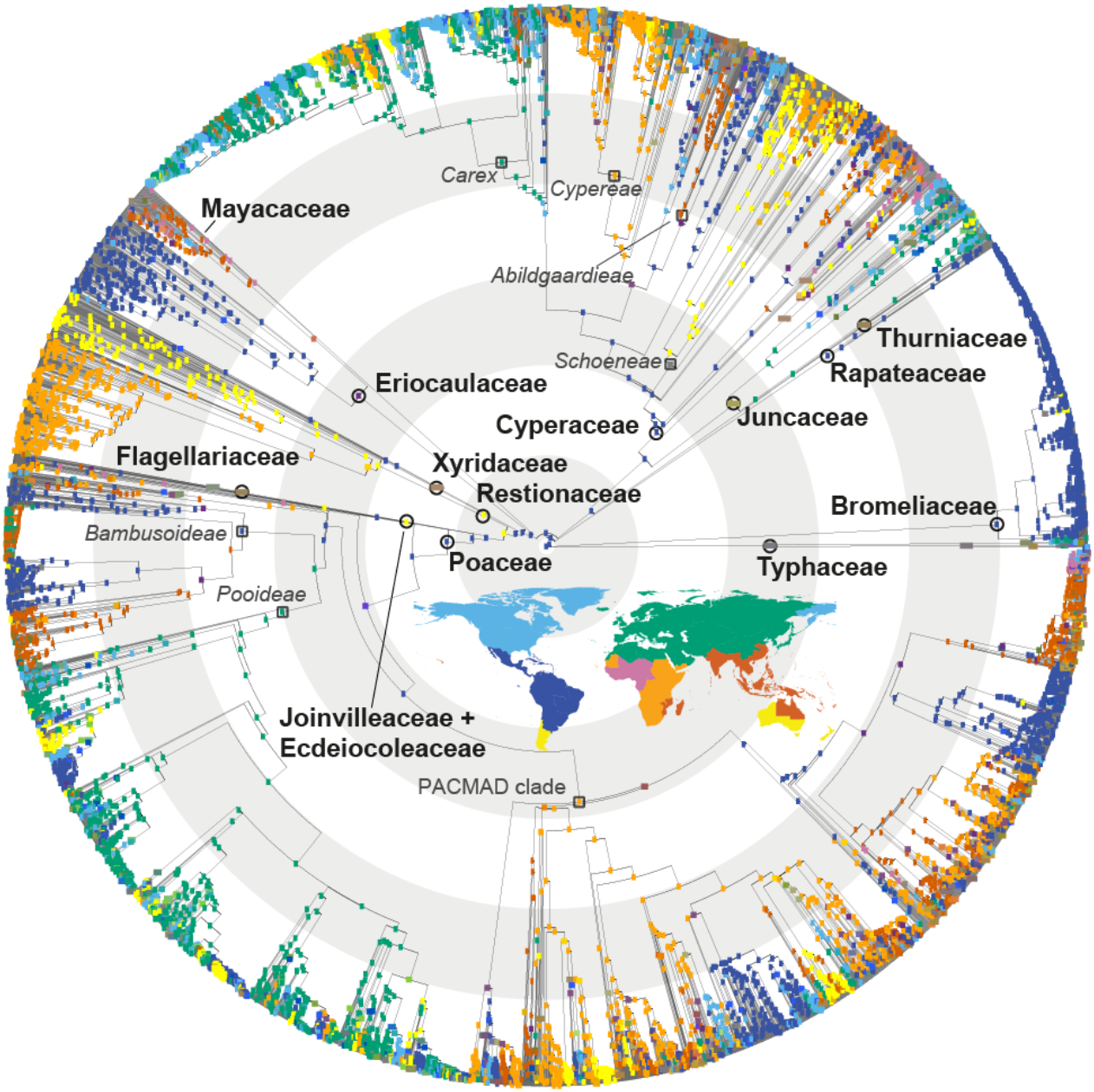

Figure S1. Ancestral area reconstruction within Poales based on seven regions, obtained using the DIVA model in BioGeoBEARS.

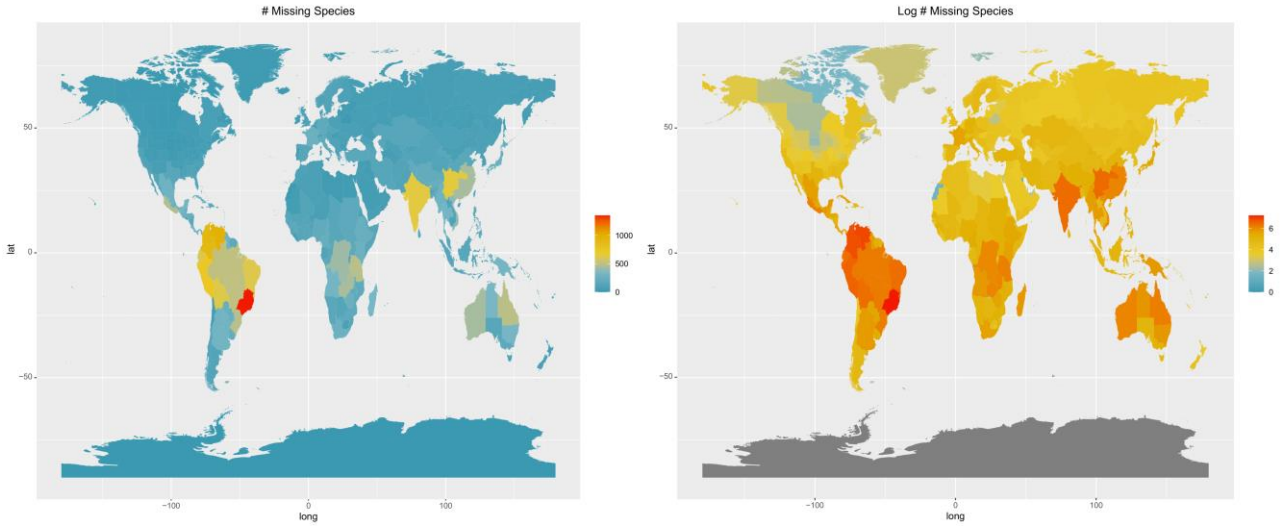

Figure S2. The number of species of Poales missing from the phylogenetic dataset compared to the number listed in the World Checklist of Vascular Plants (WCVP) as of 28 February 2022, mapped per botanical region.

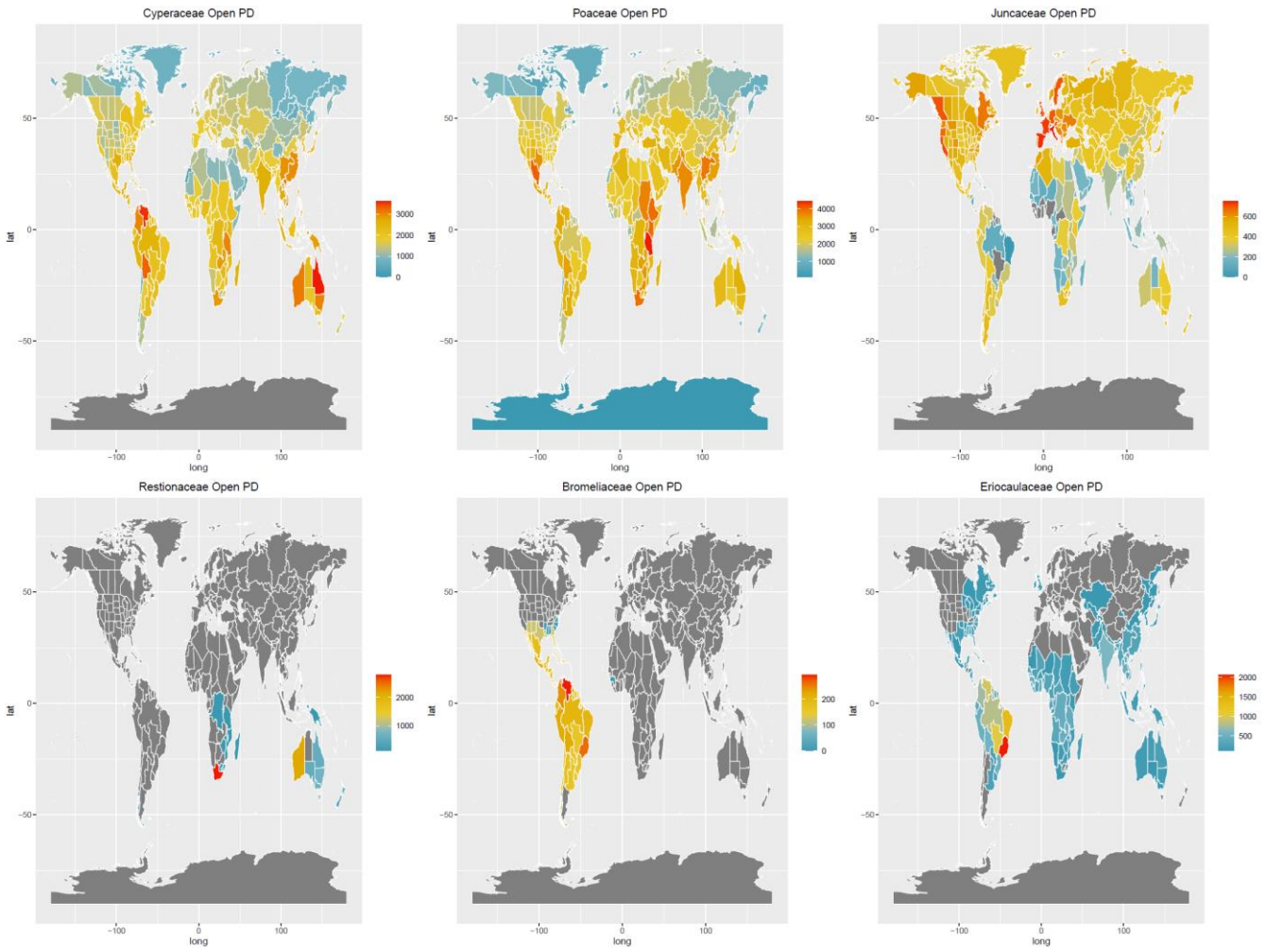

Figure S3. Phylogenetic diversity (PD) of the six largest Poales families categorised into open and closed habitats.

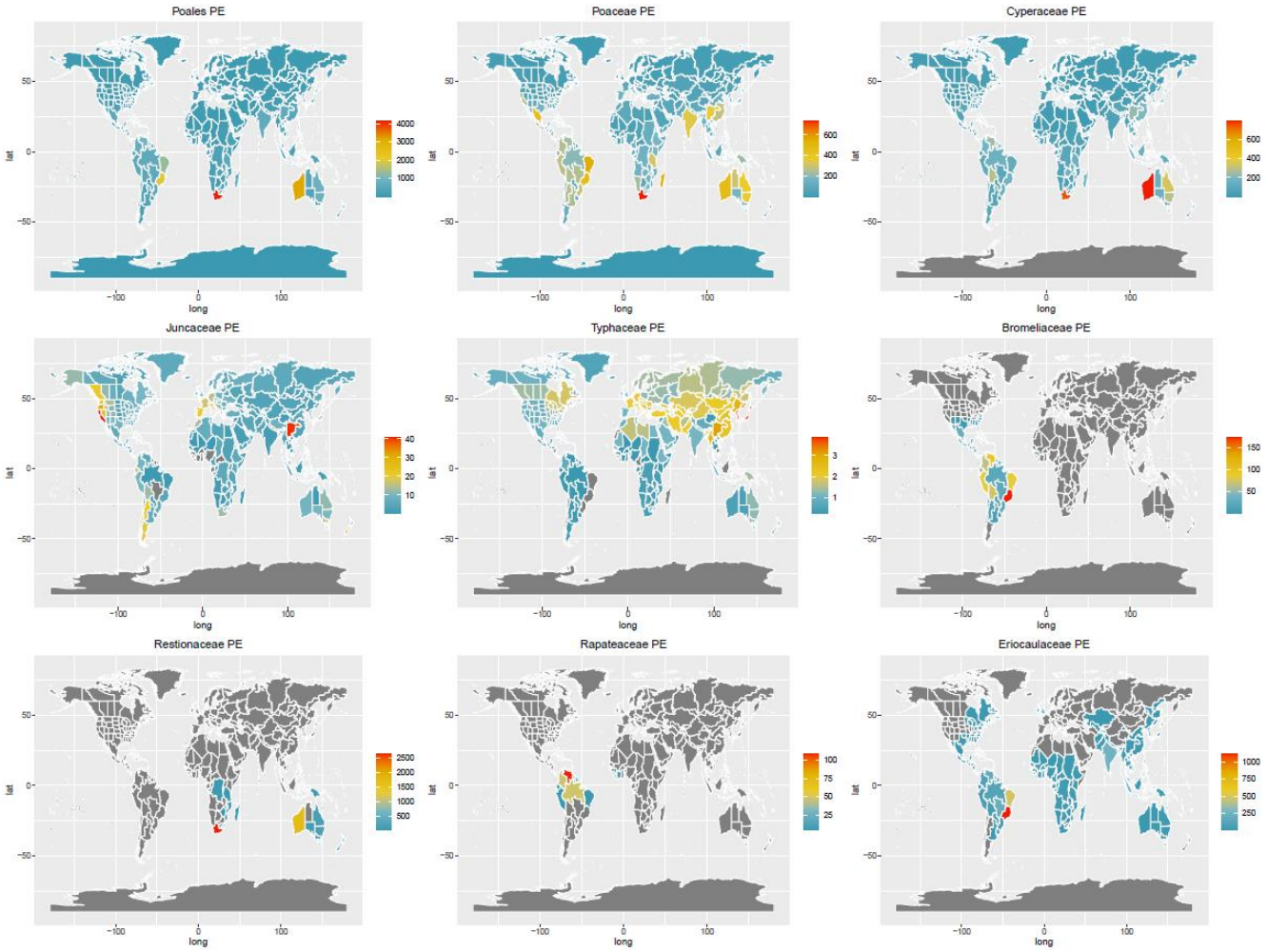

Figure S4. Phylogenetic endemicity (PE) mapped per botanical region for Poales and eight families with the highest number of species in the dataset.

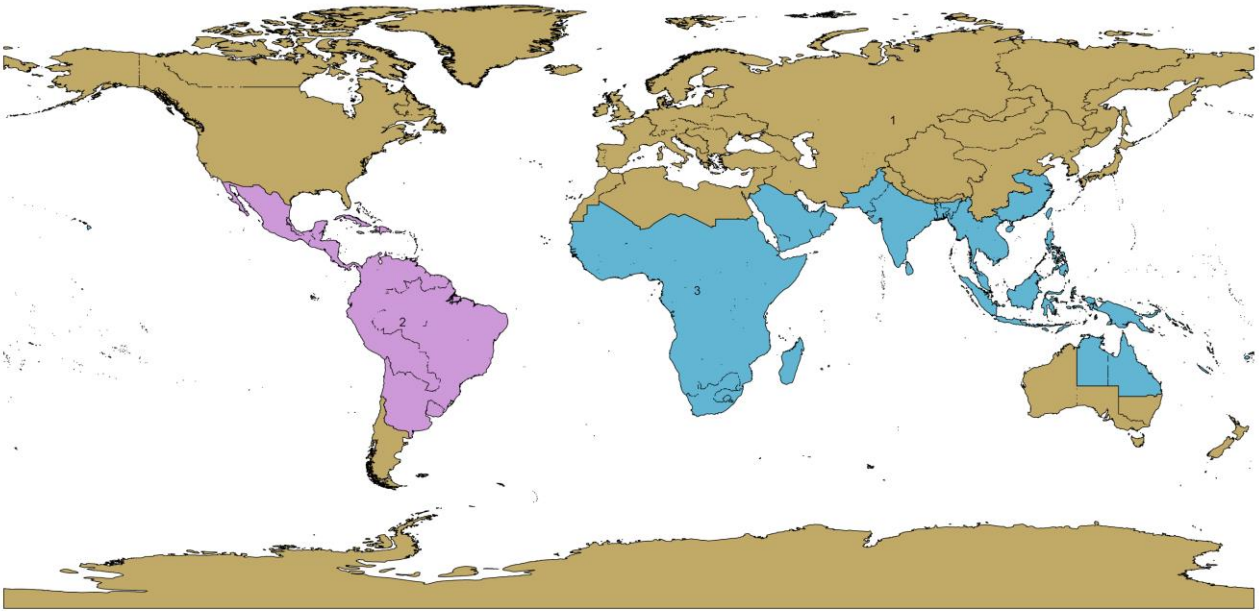

Figure S5. Poales botanical regions grouped into three ‘floristic kingdoms’ based on phylogenetic beta diversity, indicated by different colours and numbers

Table S2. Calibrations used in the treePL (Smith & O'Meara, 2012) configuration file.

Clade	min	max	Backbone	tip_subclade1	tip_subclade2	Reference	Note
Poales	120,1	120,1	yes	(all)	(all)	Givnish et al. 2018	secondary; crown age used
Bromeliaceae	19,9	19,9	yes	Brocchinia_prismatica	Portea_fosteriana	Givnish et al. 2018	secondary; crown age used
Typhaceae	70,4	70,4	yes	Typha	Sparganium	Givnish et al. 2018	secondary; crown age used
Rapateaceae	44,6	44,6	yes	Stegolepis_hitchcockii	Rapatea_paludosa	Givnish et al. 2018	secondary; crown age used; Potarophytum-Rapatea-Stegolepis
Thurniaceae	33,9	33,9	no	Thurnia_sphaerocephala	Prionium_serratum	Givnish et al. 2018	secondary; crown age used; not constrained in backbone - only 1
Juncaceae	67,8	67,8	yes	Juncus_pauciflorus	Luzula_elegans	Givnish et al. 2018	secondary; crown age used
Juncaceae_Cyperaceae_Thurniaceae	107,4	107,4	yes	Thurnia_sphaerocephala	Cyperus_diffusus	Givnish et al. 2018	secondary; crown age used
Juncaceae_Cyperaceae	90,8	90,8	yes	Juncus_pauciflorus	Cyperus_diffusus	Givnish et al. 2018	secondary; crown age used
Xyridaceae	92,5	92,5	yes	Xyris_jupicai	Abolboda_grandis	Givnish et al. 2018	secondary; crown age used
Eriocaulaceae	66,6	66,6	yes	Comanthera_kegeliana	Eriocaulon_australe	Givnish et al. 2018	secondary; crown age used
Restionaceae	104,6	104,6	yes	Hopkinsia_anoectocolea	Restio_wittebergensis	Givnish et al. 2018	secondary; crown age used
graminids	106,6	106,6	yes	Flagellaria_neocaledonica	Andropogon_tracyi	Givnish et al. 2018	secondary; crown age used
Ecdeiocoleaceae	70,1	70,1	yes	Ecdeiocolea_monostachya	Georgeantha_hexandra	Givnish et al. 2018	secondary; crown age used
Cyperaceae	75	88	yes	Hypolytrum_longifolium	Schoenus_exilis	inclusive of Spalink et al., 2016, Givnish et al., 2018	Secondary
Carex	34	38	no	Carex_moupinensis	Carex_longii	Jiménez-Mejías et al. (2016)	Fossil: Carex colwellensis
Carex_Vignea_clade	16	23	no	Carex_gibba	Carex_tribuloides	Jiménez-Mejías et al. (2016)	Fossil: Carex marchica
Cyperus	24	32	no	Cyperus_prolifer	Cyperus_nipponicus	Spalink et al., 2016	Secondary
Eleocharis	31	41	no	Eleocharis_robbinsii	Eleocharis_spiralis	Spalink et al., 2016	Secondary
Fimbristylis	30	40	no	Fimbristylis_compacta	Fimbristylis_densa	Spalink et al., 2016	Secondary
Rhynchospora	40	50	no	Rhynchospora_corniculata	Rhynchospora_megalocarpa	Spalink et al., 2016	Secondary
Scleria	38	48	no	Scleria_brownii	Scleria_virgata	Smith et al., 2010	Fossil
Poaceae	75	95	yes	Anomochloa_marantoidea	Andropogon_tracyi	Gallaher et al. 2021	secondary
BOP	65	85	no	Agrostis_lenis	Streptogyna_americana	Gallaher et al. 2021	secondary
PACMAD	45	70	no	Sartidia_jucunda	Andropogon_tracyi	Gallaher et al. 2021	secondary

References

Gallaher TJ, Peterson PM, Soreng RJ, Zuloaga FO, Li DZ, Clark LG, Tyrrell CD, Welker AD, Kellogg EA, Teisher JK. 2022. Grasses through space and time: An overview of the biogeographical and macroevolutionary history of Poaceae. Journal of Systematics and Evolution **60:** 522–569.

Givnish TJ, Zuluaga A, Spalink D, Soto Gomez M, Lam VK, Saarela JM, Sass C, Iles WJ, De Sousa DJ, Leebens-Mack J, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. American Journal of Botany **105:** 1888-1910.

Jim´enez-Mej´ıas P, Martinetto E, Momohara A, Popova S, Smith SY, Roalson EH. 2016. A commented synopsis of the pre-Pleistocene fossil record of Carex (Cyperaceae). The Botanical Review **82:** 258–345.

Smith SY, Collinson ME, Rudall PJ, Simpson DA. 2010. Cretaceous and Paleogene fossil record of Poales: review and current research. Diversity, phylogeny, and evolution in the monocotyledons (ed. by Seberg O, Peterson G, Barford AS, Davis JI), pp. 333–356. Aarhus University Press, Lancaster.

Spalink D, Drew BT, Pace MC, Zaborsky JG, Starr JR, Cameron KM, Givnish TJ, Sytsma KJ. 2016. Biogeography of the cosmopolitan sedges (Cyperaceae) and the area-richness correlation in plants. Journal of Biogeography **43:** 1893–1904.

Table S4. Comparison of six ancestral area reconstruction models based on BioGeoBEARS analyses for Poales.

Model	Log-likelihood	Number of para	Rate of dispers	Rate of extinct	Jump dispersa	AICc
DEC	-21114	2	0,026	0,15	0	42233
DEC + J	-21879	3	0,013	1,00E-12	0,0058	43765
DIVALIKE + J	-23148	3	0,014	3,30E-09	0,007	46303
DIVALIKE	-23831	2	0,017	0,0037	0	47666

AICc: corrected Akaike's information criterion; dispersal-extinction-cladogenesis (DEC); DIVALIKE (Dispersal-Vicariance Analysis).

Table S5. Results from corHMM ancestral state reconstructions.

Table S5a. Models (2 rates vs. 1 rate and ARD vs. SYM), AIC of the best model, deltaAIC in comparison with the best fit model are indicated.

Trait	Model	AIC	deltaAIC
Open / Closed	2 Rates ARD	8814,2	NA
	1 Rate ARD	9740,6	926,4
	2 Rates SYM	9707,8	893,6
	1 Rate SYM	10906,1	2091,9

Table S5b. Model parameters associated with the 2 Rates ARD model.

Trait	Model	1R1 to 2R	1R1 to 1R	2R1 to 1R	2R1 to 2R	1R2 to 1R	1R2 to 2R	2R2 to 2R	2R2 to 1R2
Open / Closed	2 Rates	0,31	0,07	0,51	0,07	0,005	1E-09	0,005	0,01

corHMM: Hidden Markov Models of Character Evolution; AIC: Akaike information criterion; ARD: all-rates-different; SYM: symmetrical model; 1: state 1; 2: state 2; R1: rate regime 1; R2; rate regime 2

corHMM: Hidden Markov Models of Character Evolution; AIC: Akaike information criterion; ARD: all-rates-different; SYM: symmetrical model; 1: state 1; 2: state 2; R1: rate regime 1; R2; rate regime 2

The evolutionary history of important traits for Poales was reconstructed using Generalized Hidden Markov models, as implemented in the function corHMM of R package corHMM v.2.8 (Boyko & Beaulieu, 2021), to estimate the transition rates and ancestral state of several binary characters across the Poales tree phylogeny.

For each open / closed habitat binary trait, we ran the following Markov models:

- symmetric rate (SYM: one transition rate category; one parameter): transition rate (1 parameter); no hidden states
- all rates differ (ARD: one transition rate category; two parameters): transition rate for each regime (2 parameters); no hidden states
- symmetric rate (SYM: two transition rate categories; four parameters): transition rate for each regime (2 parameters); backward rate connecting two transition rate categories (1 parameter); forward rate connecting two transition rate categories (1 parameter); 2 hidden states; and
- all rates differ ARD: two transition rate categories; six parameters): two transition rates for each transition rate regime (4 parameters); two rates connecting the two transition rate categories (2 parameters); 2 hidden states.

Notes S1. Additional details on the phylogenomic backbone reconstruction.

We produced a family-level phylogenomic backbone using nuclear data from 353 loci (Angiosperms353; Johnson *et al.,* 2019). The sampling for the backbone aimed towards 50% of the currently accepted genera and involved new data produced and samples mined from public repositories. The genomic data production was conducted following Baker *et al.* (2022), with DNA extractions (mostly from herbarium materials) using CTAB (Doyle & Doyle, 1987). We used the NEBNext Ultra II DNA Library Prep kit (New England Biolabs) for standard pair-ended library preparation and libraries were hybridised with myBaits Angiosperms353 v1 probe kit (Arbor Biosciences).

The sequence recovery from raw data (target enrichment and mined reads) started with reads being trimmed for short and/or low-quality sequences using Trimmomatic (Bolger *et al.*, 2014) and then assembled with a *de novo* approach implemented in HybPiper v.1.3.1 (Johnson e*t al.*, 2016). In HybPiper, trimmed reads were initially binned into genes using BLASTN, which were assembled into scaffolds using SPADES (Bankevich *et al.*, 2012), and the coding regions later extracted with Exonerate (Slater & Birney, 2005). For assembled datasets (i.e., whole genomes and transcriptomes) sequence recovery followed Baker *et al.* (2022).

We inferred the phylogenomic backbone using a multi-species coalescent framework (MSC) based on individual gene trees. Sequences were aligned in MAFFT (Katoh & Standley, 2013) in einsi-mode, with gappy sites (> 90% missing data) removed using Phyutility (Smith & Dunn, 2008). Gene trees were inferred using IQ-TREE 2 (Minh *et al.*, 2020), with support assessed via UltraFast bootstrap (UFBS; Hoang *et al.*, 2018). TreeShrink (Mai & Mirarab, 2018) was used to identify outliers that significantly increased tree space. Alignment and tree building was repeated for those genes with outlier trees. All gene trees were subsequently trimmed for poorly supported branches (UFBS < 30%) and used as input for the MSC analysis in ASTRAL-III (Zhang *et al.,* 2018). To obtain a species tree with branch lengths proportional to the genetic distance, we first ranked the genes according to the congruence of their resulting trees to the species tree using SortaDate (Smith *et al.,* 2018) and then concatenated the alignments of the 25 most congruent genes. Using the MSC species tree as topological constraint and this concatenated alignment, a new phylogram was inferred in IQ-Tree 2. For more details on library preparation and data analyses, please refer to Baker *et al.* (2022).

References

Baker WJ, Bailey P, Barber V, Barker A, Bellot S, Bishop D, Botigué LR, Brewer G, Carruthers T, Clarkson JJ, *et al*. 2022. A comprehensive phylogenomic platform for exploring the Angiosperm tree of life. *Systematic Biology* **71:** 301–319.

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

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

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

Hoang DT, Chernomor O, Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35: 518–522.

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* **4**: 1600016.

Johnson MG, Pokorny L, Dodsworth S, Botigue LR, Cowan RS, Devault A, Eiserhardt WL, Epitawalage N, Forest F, Kim JT, *et al*. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* **68:** 594–606.

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

Mai U, Mirarab S. 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* **19**: 272.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Haeseler A, Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37:** 1530–1534.

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

Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24:** 715–716.

Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PLOS ONE* **13**: e0197433. https://doi.org/10.1371/journal.pone.0197433

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19:** 153.

Notes S2. Justification for selecting dispersal-extinction-cladogeneis (DEC) model of ancestral estimation.

Because phyloregions are determined by spatial patterns of lineage turnover, they reflect – but do not necessarily conform to – discrete geologic boundaries typically used in ancestral area estimations (e.g., Martín-Bravo *et al*., 2019). However, these phyloregions present a data-driven hypothesis for the spatial relationship of areas as they relate to the biogeographical processes of dispersal and vicariance in Poales, and are thus well-suited for ancestral estimations. We *a priori* selected the dispersal-extinction-cladogeneis (DEC) model of ancestral estimation (Ree *et al.,* 2005; Ree & Smith, 2008) instead of other available models (e.g., DIVA, Ronquist *et al.*, 1997; BayArea, Landis *et al.*, 2013), because our expectation is that the parameters of this model are best suited to the particular biology and distribution of Poales. For example, we expect both cladogenetic sympatry and vicariance to be important processes in Poales, particularly when descendent lineages diverge within only a portion of the ancestral range (i.e., subset sympatry) and when vicariant events unevenly split an ancestral range between two descendent ranges (i.e., narrow vicariance). The former scenario is not modelled by DIVA, while the latter is not modeled by BayArea. Given that the BayArea model does not parameterize vicariant speciation but instead allows widespread sympatric speciation, we do not consider it a reasonable model for the global analysis of a clade that spans 120 millions years of evolution. Indeed, the BayAreaLIKE model places the Poales in a nearly cosmopolitan range for the first 40 million years of its evolution, which is neither supported by fossil data nor biologically plausible. Many Poalean lineages are exceptionally good dispersers and able to migrate across typical migration barriers (e.g., oceans; Linder *et al.*, 2018; Martín-Bravo *et al.*, 2019, Spalink *et al.*, 2019; Larridon *et al*., 2021, Benítez-Benítez *et al.*, 2021), while lineages with species with poor dispersal ability tend to be restricted to single or physically adjacent phyloregions (e.g., Rapateaceae, Bromeliaceae). Highly parameterized models – with time-stratification or with geographic dispersal multipliers – are unlikely to be a good fit for all clades in the exceptionally diverse Poales. Thus subsequent analyses are based on the DEC model.

References

Benítez-Benítez C, Martín-Bravo S, Bjorå CS, Gebauer S, Hipp AL, Hoffmann MH, Luceño M, Pedersen TM, Reznicek A, Roalson E. 2021. Geographical vs. ecological diversification in *Carex*

section *Phacocystis* (Cyperaceae): Patterns hidden behind a twisted taxonomy. *Journal of Systematics and Evolution* **59**: 642–667.

Landis MJ, Matzke NJ, Huelsenback JP. 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology* **62**: 789–804.

Larridon, I., Spalink, D., Jiménez-Mejías, P., Márquez-Corro, J.I., Martín-Bravo, S., Muasya, M. and Escudero, M., 2021. The evolutionary history of sedges (Cyperaceae) in Madagascar. *Journal of Biogeography* **48**: 917–932.

Linder HP, Lehmann CER, Archibald S, Osborne CP, Richardson DM. 2018. Global grass (Poaceae) success underpinned by traits facilitating colonization, persistence and habitat transformation. *Biological Reviews* **93:** 1125–1144.

Martín-Bravo S, Jiménez-Mejías P, Villaverde T, Escudero M, Hahn M, Spalink D, Roalson EH, Hipp AL, the Global *Carex* Group. 2019. A tale of worldwide success: Behind the scenes of *Carex* (Cyperaceae) biogeography and diversification. *Journal of Systematics and Evolution* **57**: 695–718.

Ree RH, Moore BR, Webb CO, Donoghue MJ. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* **59**: 2299–2311.

Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* **57**: 4–14.

Ronquist F. 1997. Dispersal–vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* **46**: 195–203.

Spalink D, MacKay R, Sytsma KJ. 2019. Phylogeography, population genetics, and distribution modelling reveal vulnerability of *Scirpus longii* (Cyperaceae) and the Atlantic Coastal Plain Flora to climate change. *Molecular Ecology* **28**: 2046-2061.