SPECIAL ISSUE ARTICLE

Botany
American Journal of
BOTANICAL SOCIETY OF AMERICA

Investigating historical drivers of latitudinal gradients in polyploid plant biogeography: A multiclade perspective

Eric R. Hagen[1,2] | Thais Vasconcelos[1,3] | James D. Boyko[1,3,4] | Jeremy M. Beaulieu[1]

[1]Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

[2]Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON M5S 3B2, Canada

[3]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

[4]Michigan Institute for Data Science, University of Michigan, Ann Arbor, MI 48109, USA

Correspondence
Eric R. Hagen, Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA.
Email: eric.hagen@utoronto.ca

Abstract

Premise: The proportion of polyploid plants in a community increases with latitude, and different hypotheses have been proposed about which factors drive this pattern. Here, we aimed to understand the historical causes of the latitudinal polyploidy gradient using a combination of ancestral state reconstruction methods. Specifically, we assessed whether (1) polyploidization enables movement to higher latitudes (i.e., polyploidization precedes occurrences in higher latitudes) or (2) higher latitudes facilitate polyploidization (i.e., occurrence in higher latitudes precedes polyploidization).

Methods: We reconstructed the ploidy states and ancestral niches of 1032 angiosperm species at four paleoclimatic time slices ranging from 3.3 million years ago to the present, comprising taxa from four well-represented clades: Onagraceae, Primulaceae, *Solanum* (Solanaceae), and Pooideae (Poaceae). We used ancestral niche reconstruction models alongside a customized discrete character evolution model to allow reconstruction of states at specific time slices. Patterns of latitudinal movement were reconstructed and compared in relation to inferred ploidy shifts.

Results: No single hypothesis applied equally well across all analyzed clades. While significant differences in median latitudinal occurrence were detected in the largest clade, Poaceae, no significant differences were detected in latitudinal movement in any clade.

Conclusions: Our preliminary study is the first to attempt to connect ploidy changes to continuous latitudinal movement, but we cannot favor one hypothesis over another. Given that patterns seem to be clade-specific, more clades must be analyzed in future studies for generalities to be drawn.

KEYWORDS
biogeography, diploidization, latitudinal gradient, Onagraceae, phylogenetic comparative methods, Poaceae, polyploidy, Primulaceae, Solanaceae, whole genome duplication

Polyploidy—the state of having more than two complete sets of chromosomes—has continually shaped the evolutionary history of flowering plants. Indeed, whole-genome duplications have been identified along the stem leading to all flowering plants, and many of these events occurred throughout some of the most diverse and some of the most depauperate clades nested within (Jiao et al., 2011). Through comparisons of diploid and polyploid plants, polyploidy appears linked to a variety of evolutionary changes, including novel phenotypic traits (Levin, 1983), ecological relationships (Segraves, 2017), and macroevolutionary patterns (e.g., Mayrose et al., 2011; Soltis et al., 2014). In biogeography, polyploidy is largely studied in the context of latitudinal and elevational gradients, in which polyploids tend to compose larger proportions of the flora at higher latitudes and elevations than at lower ones (Stebbins, 1950; Brochmann et al., 2004; Rice et al., 2019). The latitudinal polyploidy gradient (LPG) has long been observed in individual clades (e.g., Löve and Löve, 1943, 1949), and recent studies incorporating large amounts of distribution data across clades have largely confirmed the generality of this pattern (Rice et al., 2019).

Proposed mechanisms responsible for the creation of the LPG can be divided into two categories. First, conditions of poleward environments may lead to higher rates of polyploid formation at higher latitudes. Harsh environmental conditions such as cold stress are known to induce polyploidy (De Storme and Geelen, 2014; Lohaus and Van de Peer, 2016), and the fragmented ranges of poleward environments could lead to allopolyploid formation via repeated contacts after range expansion (Stebbins, 1985). Second, various adaptations of polyploids lead them to preferentially move into poleward environments at rates higher than those of diploids. Polyploids are believed to have generally greater colonizing ability than diploids due to higher rates of self-compatibility (Bierzychudek, 1985; Barringer, 2007) and phenotypic plasticity (Price et al., 2003; Leitch and Leitch, 2008). Thus, in the time since freezing conditions began at northern latitudes during the Pliocene (Mudelsee and Raymo, 2005), the LPG could have been generated by plant lineages generally moving to higher latitudes after polyploidization events.

These two scenarios, which we call the centers of polyploidization and centers of arrival hypotheses, respectively, are not mutually exclusive: Some clades could theoretically show the former pattern, and other clades the latter. One potential first step toward understanding which mechanism best explains observed patterns in angiosperms is by analyzing the biogeographical history in the context of ploidy transitions in clades that present variation in ploidy level and widespread distribution across latitudinal zones. Four good candidates for this initial investigation are Onagraceae (evening primrose family), Primulaceae (primrose family), Solanaceae (tomato family), and Poaceae (grass family). These specific clades are well suited for such an analysis because they exhibit cosmopolitan distributions, a large proportion of their species have available ploidy data, and they possess well-sampled molecular phylogenies (Särkinen et al., 2013; De Vos et al., 2014; Spriggs et al., 2014; Freyman and Höhna, 2019). They also represent different branches of the angiosperm tree of life: Onagraceae are rosids, whereas Solanaceae and Primulaceae are asterids, and Poaceae are monocots (APG IV, 2016). Thus, if we find that the same patterns are observed in all or most of the four clades, it is more likely that these mechanisms may explain the LPG in angiosperms more generally.

We also note that there remains the possibility that the LPG emerges passively. For example, Rice et al. (2019) found that the global distribution of polyploid plants is strongly correlated with climate, though they suggest that this relationship is mainly indirect, because polyploids tend to be perennial (Van Drunen and Husband, 2019), herbaceous plants (i.e., chamaephytes) that are low to the ground and able to survive the harsh conditions of poleward environments (Raunkiaer, 1934). However, while present-day climatic variables do correlate with biogeographic patterns, the modern distributions of plants largely result from past climate changes (Normand et al., 2011). Additionally, correlations between specific traits and environmental variables may be shaped more by shared evolutionary history among species sharing those traits rather than functional relationships

(Svenning and Skov, 2007; Ma et al., 2016; Sundaram and Leslie, 2021), so phylogenetic information must be considered as well. In any event, teasing apart the evidence for each scenario across flowering plants would provide invaluable clues about the historical causes for the LPG.

Here, we conducted what we believe is the first attempt to discern the historical causes that create the LPG pattern in widely distributed clades. Specifically, we analyzed the distributions of plants in historical and phylogenetic context to determine how plants in specific clades move across latitudes after ploidy transitions. By examining the timing of reconstructed ploidy changes and biogeographic movements in four clades, we tested support for the centers of arrival hypothesis, in which range movement toward higher latitudes happens most often after polyploidization events (Figure 1A), and the centers of polyploidization hypothesis, in which polyploids form mostly at poleward environments and subsequently stay or move toward the equator (Figure 1B).

MATERIALS AND METHODS

Phylogenetic and ploidy data sets

We opted to use a multiclade approach for this work with the aim to discern biological generalities and clade-specific patterns (e.g., Boyko et al., 2023; Vasconcelos, 2023). The main reason for this approach is to reduce the impact of sampling bias in subsequent analyses of ancestral state and ancestral range reconstructions by focusing on clades that are particularly well sampled, as opposed to using super-matrix trees (e.g., Smith and Brown, 2018) that are unevenly sampled. These biases are also caused by available ploidy data being skewed toward certain taxonomic groups, particularly those studied in the Global North (Marks et al., 2021), and the fact that available GBIF data are incomplete and spatially clustered (Beck et al., 2014).

The Onagraceae tree contains 292 species (c. 45% sampling; 186 with ploidy data), Primulaceae contains 263 species (c. 9.4% sampling [Xu and Chang, 2017]; 141 with ploidy data), Pooidae contains 1312 species (c. 40.6% sampling [Soreng et al., 2017]; 748 with ploidy data), and *Solanum* contains 441 species (c. 33.3% sampling; 256 with ploidy data) (Appendices S1–S4). The Pooidae and *Solanum* trees were pruned from larger phylogenies of Poaceae and Solanaceae, respectively, because the larger Poaceae and Solanaceae trees had data coverage of less than 50% for ploidy data, and pruning to include only these lower taxonomic rankings allowed us to focus on clades that are particularly data-rich. These four clades were selected for this study because they are comparatively large, are well represented in our ploidy data set, are geographically widespread, come from different parts of the phylogeny of angiosperms (different major clades: rosids, asterids, and monocots), and because preliminary analyses recovered a relatively large number of recent polyploidization and diploidization events in the Quaternary, our focal time period.
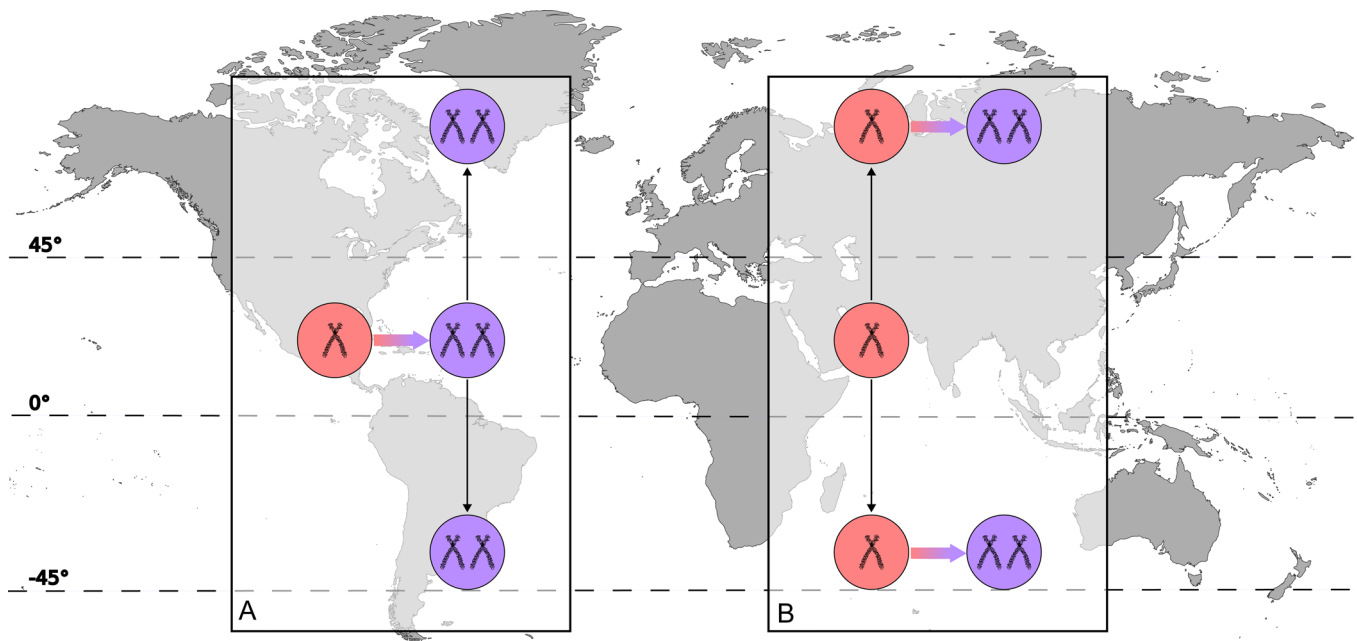
FIGURE 1 Conceptual diagram of historical biogeographic patterns expected to be observed under the centers of arrival hypothesis (A) and the centers of polyploidization hypothesis (B). Under the centers of arrival scenario, polyploidization occurs across the globe but is followed by higher rates of antiequatorial movement relative to diploids, thus creating the latitudinal polyploidy gradient (LPG). Under the centers of polyploidization scenario, the LPG is created by higher rates of polyploidization in poleward environments.

Ploidy data were extracted from the supplementary data of Rice et al. (2019), which is contained in individual ChromEvol output files separated by genus. While the accuracy of ploidy inferences based on ChromEvol analyses can vary (Glick and Mayrose, 2014), Rice et al. (2019) described a conservative procedure that allowed the most robust inferences possible. We combined these individual files into a master table and filtered it for species represented in our four phylogenies. In our analysis, we defined a "polyploid" narrowly to specifically refer to a neopolyploid (i.e., newly formed polyploids; Ramsey and Schemske, 2002), following the methodology of Rice et al. (2019). Neopolyploids are cytologically distinct from their diploid progenitors, and they have undergone whole-genome multiplication sufficiently recently that they retain additive genome sizes of their parents as well as distinguishable subgenomes (Mandáková and Lysak, 2018). In contrast, mesopolyploids and paleopolyploids are species that underwent polyploidization further in the past and have undergone diploidization, resulting in decreased genome size and in genome restructuring. We used this definition because (1) the LPG is a gradient of plants that *are* polyploid (i.e., neopolyploids) rather than of plants that *behave* like polyploids (in the sense of gaining advantageous traits rather than chromosomal behavior) and (2) we examined latitudinal changes after inferred events of both polyploidization and diploidization, so it did not make sense to consider re-diploidized plants in our analysis as polyploids, that is, paleopolyploids (see section Integrating trait evolution models with reconstructions of past climatic niches below).

Distribution data

We downloaded all occurrence points available on GBIF that were based on preserved specimens (i.e., excluding human observations) for the four focal clades in our study (GBIF.org 2022; see section Data Availability Statement). We then removed inaccuracies following protocols similar to those of Boyko et al. (2023). Our final occurrence point database had 331,434 points, including 11,200 for Primulaceae, 43,408 for *Solanum*, 66,365 for Onagraceae, and 210,461 for Pooideae. After filtering to include only those phylogenetically represented species with ploidy data and sufficient occurrence points (3 or more), we analyzed 107 species in Primulaceae, 218 in *Solanum*, 164 in Onagraceae, and 543 in Pooideae, for a grand total of 1032 species.

Integrating trait evolution models with reconstructions of past climatic niches

Investigating the history of ploidy transitions in a biogeographical context comes with challenges. The first is that most models that connect biogeographic shifts with discrete trait evolution require modeling areas discretely rather than continuously (e.g., Ree and Smith, 2008; Goldberg et al., 2011; Caetano et al., 2018), but the reconstructions they create are usually coarse and contain few areas. A second challenge is the need to incorporate information about historical plant distributions, which is particularly difficult due to the large number of biotic and abiotic factors that can potentially

influence a species' geographic range. For instance, Rice et al. (2019) included paleoclimatic data from the Last Glacial Maximum (LGM; 21 kya) in their analysis, but these data were only used in the context of correlating deglaciation extent with ploidy distributions, and they implicitly assumed that ranges remained unchanged to the present. Because it is unclear whether the LPG may be caused by climatic factors or other biogeographic causes (e.g., Stebbins, 1985), we opted to instead model range evolution and ploidy evolution over time separately and test for connections between the two post hoc.

We began by modeling ploidy shifts along the phylogeny of each clade during the past c. 3.3 million years (Myr). While ploidy can be reconstructed from fossils with preserved cuticle (McElwain and Steinthorsdottir, 2017), fossil data is too sparse for a large-scale study. For this reason, we opted to use corHMM (Beaulieu et al., 2013; Boyko and Beaulieu, 2021) with modified functions that allow for ancestral state reconstruction at specific time slices rather than at nodes. The modified function allows ancestral states to be inferred at particular time slices, which is apt for our data set because we are interested in ploidy states at particular times that correspond to the PaleoClim database (Brown et al., 2018), rather than at asynchronous branching points (i.e., the nodes of a phylogeny) as is the default of the software (see Figure 2). Since our model of range changes over time (machuruku) from which we estimate latitudinal movement depends on climatic information from PaleoClim, which only goes back 3.3 Myr, we only examined reconstructed latitudinal movement as it relates to whole-genome multiplications from 3.3 Myr ago (Ma) to the present. For each phylogeny, we tested three model structures: ER (equal transition rates between diploid state and polyploid state), ARD (transition rates between diploidy and polyploidy

are allowed to vary), and a unidirectional structure where reversal to diploidy was disallowed after polyploidization. We evaluated support for each model using the Akaike information criterion (AIC; Akaike, 1974; Burnham and Anderson, 2002).

It should be noted that some models of ploidy evolution (e.g., Robertson et al., 2011) disallow reversals to diploidy (i.e., diploidization, or the reorganization of the genome that returns a plant to a diploid, or "diploid-like", state after whole-genome multiplication) based on arguments that ploidy evolution is significantly asymmetrical (e.g., Stebbins, 1971; Meyers and Levin, 2006). However, much research suggests that reversals to diploidy are prevalent in flowering plants (Mandáková and Lysak, 2018), and other models of ploidy evolution reflect this (Zenil-Ferguson et al., 2019). In our view, one advantage of using models like corHMM to model ploidy evolution is the ability to reconstruct not only polyploidization but also diploidization. Although we have no expectation of how species will move latitudinally following diploidization, it may be illuminating to compare movement between species that polyploidize as opposed to diploidize, as well as stay polyploid or diploid, as a "control" group.

Once corHMM models were run, we used a novel ancestral state reconstruction function to calculate the marginal probabilities of anagenetic taxa occurring at the time slices for which we had paleoclimatic data. Once ploidy shifts had been modeled, we reconstructed the range evolution of lineages in each tree using the R package machuruku (Guillory and Brown, 2021), a tool for phylogenetic niche modeling that allows for continuous reconstruction of ranges at time slices with paleoclimatic data and visualization of inferred spatial distributions. We reconstructed ranges at four time slices based on data from PaleoClim (Brown et al., 2018): the Last Interglacial (LIG, c. 130 ka),
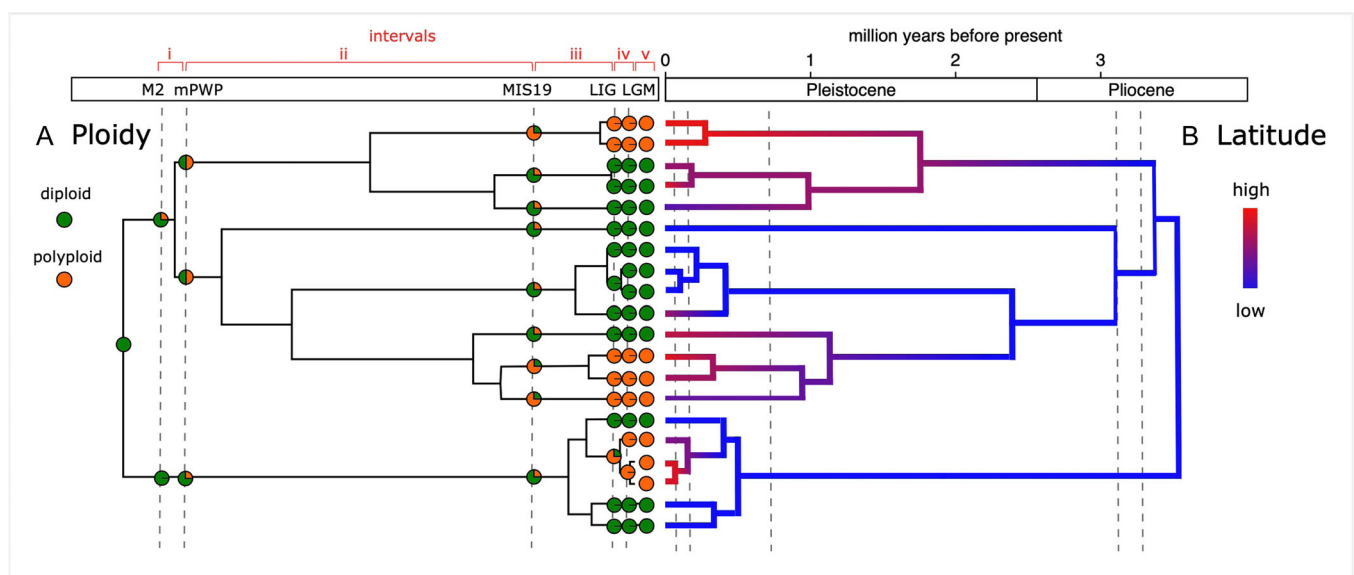
FIGURE 2 Conceptual figure showing our method of correlating inferred ploidy shifts at paleoclimatic time slices with estimated latitudinal changes, allowing for the connection of ploidy shifts to biogeographic movement. This scenario depicts the expectation under the centers of arrival hypothesis in which (A) shifts in ploidy are followed by (B) antiequatorial latitudinal movement. The time intervals spanned in (A) correspond to the time slices for which we collected paleoclimatic data: Marine Isotope Stage M2 (M2), the mid-Pliocene Warm Period (mPWP), Marine Isotope Stage 19 (MIS19), and the Last Interglacial (LIG), and Last Glacial Maximum (LGM).

Marine Isotope Stage 19 (MIS19, c. 787 ka), the mid-Pliocene Warm Period (mPWP, c. 3.205 Ma), and Marine Isotope Stage M2 (M2, c. 3.3 Ma), all using the spatial resolution of 10 arc-minutes (~20 km). For each time slice, we first estimated tip response curves to each climatic variable using the function machu.1.tip.resp, then estimated the ancestral niches of each taxon extant at each time slice with machu.2.ace, and finally projected the ancestral climatic niche models for each slice onto maps containing paleoclimatic variables with machu.3.anc.niche. We ran the machu.3.anc.niche function with the clip.Q option set to False, which produces models including less suitable areas but which prevented the function from returning NA results for some lineages.

Biogeographic analyses

To assess biogeographic movements through time, we first examined the extent to which the LPG distribution pattern is present among the four separate clades using both ANOVAs and phylogenetic ANOVAs (Revell, 2012). We then parsed latitudinal changes between time slices concurrent with different ploidy transitions, characterizing species ranges by their median latitudes. We divided possible ploidy transitions into four possible groups, which we refer to as ploidy status categories, depending on whether ploidy changes happened or not within a given time slice: (1) staying diploid, (2) staying polyploid, (3) diploidization, and (4) polyploidization. Lineages that did not change ploidy (i.e. "staying diploid" and "staying polyploid") are used as null hypotheses against which we can compare species that changed ploidy. The centers of arrival hypothesis would be supported when movement toward higher latitudes occurs more frequently after polyploidization than any other category of ploidy change. On the other hand, the centers of polyploidization hypothesis will be supported when the starting latitudes at the time slice when polyploidization occurs is significantly higher than for the other ploidy change categories. For each category, we tested for significant trends in movement (absolute latitudinal change) using a simple sign test (Conover, 1971), employing the binom.test function in R (R Core Team, 2022) to compare median latitudes at the beginning and end of each time slice. To account for the magnitude of change in addition to whether movement was generally equatorial or antiequatorial, we also conducted Wilcoxon signed-rank tests (Wilcoxon, 1945) on the same data, both with and without phylogenetic weights incorporated.

To quantitatively compare whether latitudinal movements across all time slices significantly differed between species that polyploidized and those that diploidized, we conducted two-sided Kolmogorov–Smirnov tests (Smirnov, 1939). Finally, to determine whether species that polyploidize possess ranges at significantly different latitudes relative to species in the other three ploidy status categories, we used phylogenetic ANOVA (Revell, 2012) to compare reconstructed median starting latitudes and latitudinal change across species. Comparisons were restricted to within clades and within the same time slices.

RESULTS

Model selection and transition rates

Goodness of fit of corHMM models with different assumptions of transitions rates (i.e. equal rates, ER; all rates different, ARD; and unidirectional, uni) between ploidy states within the four clades were mixed (Onagraceae, $AIC_{ER} = 80.63$, $AIC_{ARD} = 82.58$, $AIC_{uni} = 105.48$; Primulaceae, $AIC_{ARD} = 82.96$, $AIC_{ER} = 87.89$, $AIC_{uni} = 92.74$; Solanum, $AIC_{uni} = 103.0$, $AIC_{ER} = 103.31$, $AIC_{ARD} = 103.31$; Pooideae, $AIC_{ARD} = 498.46$, $AIC_{ER} = 503.09$, $AIC_{uni} = 586.63$); however, the model with the best fit in all groups always allowed some transitions between the two states. In other words, the unidirectional model was never favored. Ploidy transitions were reconstructed using ARD in Primulaceae and Pooideae, as it was favored by >2 AIC units (Burnham and Anderson, 2002), and with ER in Onagraceae and Solanum, because we defaulted to the model with fewest parameters since no model was favored by AIC comparison. Inferred rates of polyploidization were 0.019 transitions Myr^{-1} in Onagraceae, 0.09 transitions Myr^{-1} in Primulaceae, 0.02 transitions Myr^{-1} in Solanum, and 0.21 transitions Myr^{-1} in Pooideae. Rates of diploidization were 0.019 transitions Myr^{-1} in Onagraceae, 0.23 transitions Myr^{-1} in Primulaceae, 0.02 transitions Myr^{-1} in Solanum, and 0.14 transitions Myr^{-1} in Pooideae. The four phylogenies, with marginal reconstructions of ploidy states at nodes rather than time slices, are depicted in Appendices S1–S4. Pooideae was the clade with the highest number of estimated ploidy transitions with 43 polyploidizations and 53 diploidizations across all times slices; the next largest number of events was in Primulaceae, with seven polyploidizations and five diploidizations. Onagraceae underwent three polyploidizations and one diploidization, while we reconstructed 12 polyploidizations and no diploidizations in Solanum. Very few events of polyploidization and diploidization were recovered during the M2 and LIG slices, likely due to the short durations of those slices (c. 100 kyr each). For that reason, our discussion on latitudinal movements in relation to changes in ploidy category is mainly based on results across all time slices.

## Relationships between ploidy shifts and latitudinal movements

ANOVA results indicate that every clade except Pooideae showed significantly higher present-day absolute latitudes in polyploids relative to diploids. Thus, we were able to recover the LPG pattern for most clades; namely, polyploids generally were located at higher latitudes than diploids (Figure 3). However, phylogenetic ANOVA revealed no

significant differences between absolute latitudes in diploids and polyploid species in any clade (phylogenetic ANOVA: Onagraceae $F = 27.77$, $P = 0.257$; Primulaceae $F = 7.168$, $P = 0.124$; Solanum $F = 8.574$, $P = 0.171$; Pooideae $F = 0.143$, $P = 0.856$). These results suggest that the observed differences between ploidy states among species were not more different than expected by chance alone, and they are likely due to ploidy shifts being less labile than changes in latitude.

When we correlated inferred ploidy shifts at particular time slices with estimated starting latitudes for that time slice, we found mixed support for the centers of polyploidization hypotheses across clades. Across all clades, phylogenetic paired ANOVA detected almost no significant differences between the starting latitudes of lineages that polyploidize in a given time slice and lineages in the other ploidy status categories (Figure 4A–C, E). The only significant comparison was between species that polyploidized vs. stayed diploid in Pooideae, where those that polyploidize had significative higher starting latitudes than those that stay diploid ($F = 14.109$, $P = 0.011$; Figure 4D). Since corHMM reconstructed only one diploidization event in Onagraceae, and none in Solanum, comparisons between the polyploidized and diploidized ploidy status categories could not be conducted in these clades.

Regarding support for centers of arrival, phylogenetic ANOVA did not indicate a significant difference in latitudinal movement between lineages that diploidized as opposed to polyploidized in any clade nor across all clades (Figure 5). Binomial sign tests detected no significant directional movement in any of the four clades. Wilcoxon signed-rank tests, which account for both direction and magnitude of movement, were not significant for any clade except Pooideae, in which movement was significantly different between species that diploidized and polyploidized both with ($P = 0.0358$) and without ($P = 0.019$) phylogenetic correction. In this case, lineages that polyploidized tended to move equatorially, rather than antiequatorially. Kolmogorov–Smirnov tests indicate that shifts in median latitudes after polyploidization are not significantly different from changes in median latitudes after diploidization. Species that polyploidized did show noticeable spikes in northward movement relative to other groups in some clades and time slices (Appendix S5), but these findings were countered by the mostly nonsignificant Wilcoxon and Kolmogorov–Smirnov test results.
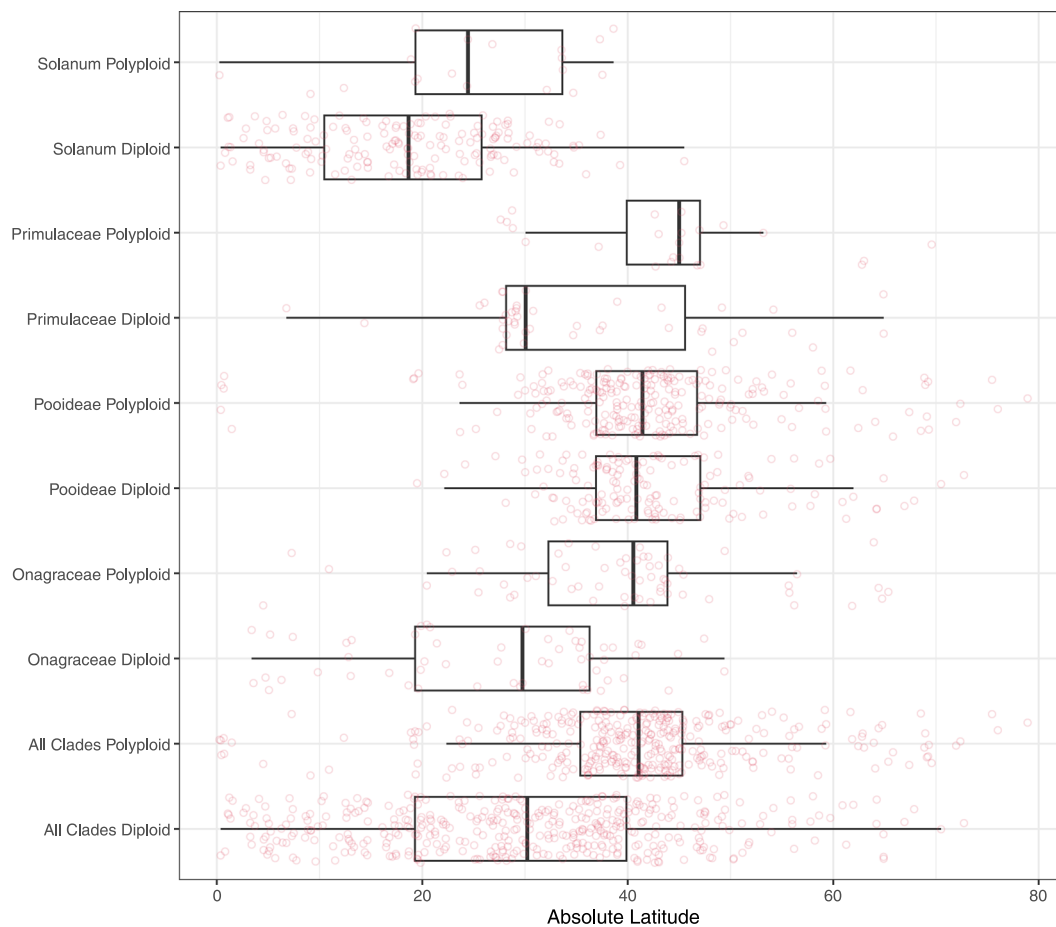
FIGURE 3 Boxplot showing present-day absolute latitudes of all plants in our data set by ploidy and by clade. Underlying distributions are shown in red.
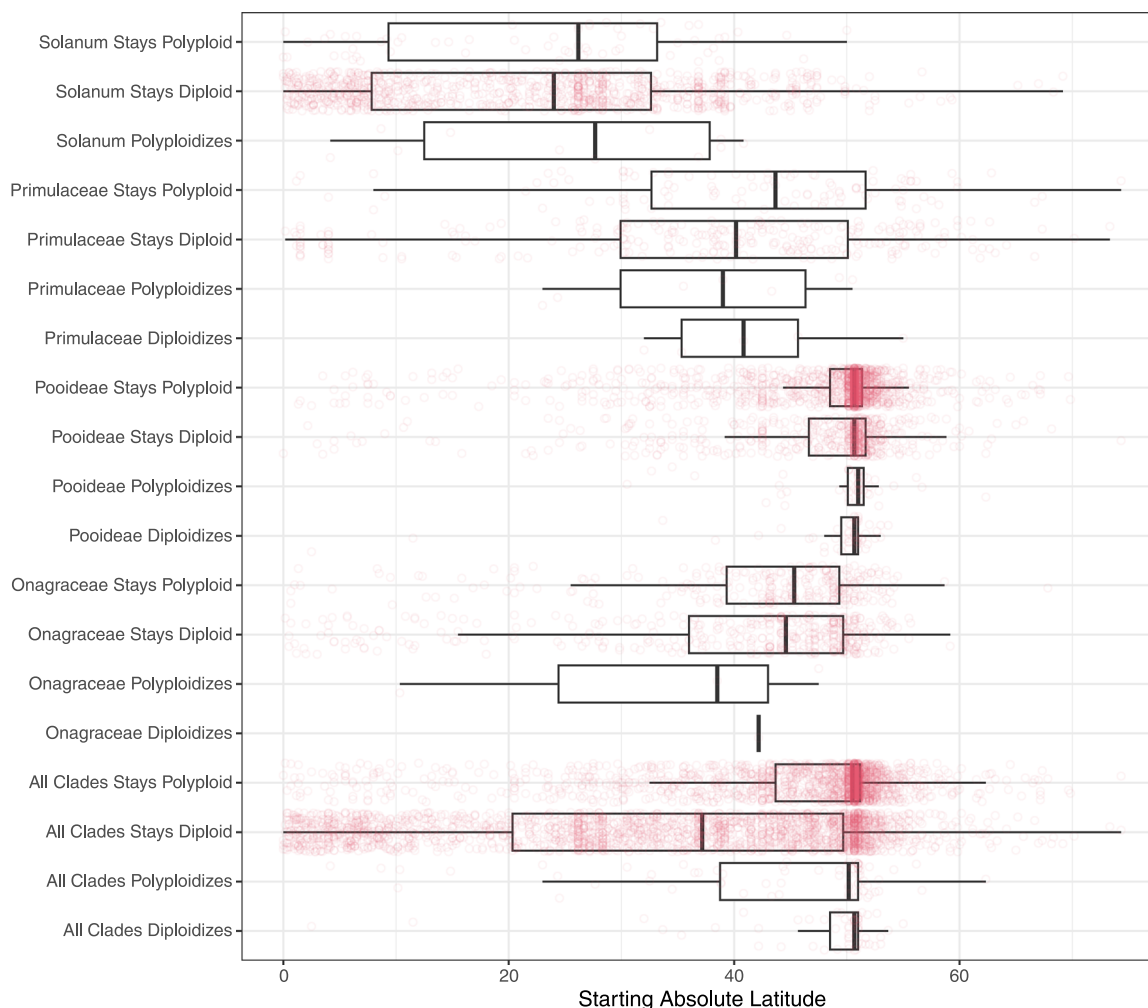
FIGURE 4 Boxplots of the starting absolute latitudes across each of the four ploidy groups, across all time slices, divided by clade. Underlying distributions are shown in red.

DISCUSSION

Insights into the causes of the LPG

Our study aimed to determine whether the LPG distribution pattern in four flowering plant clades was better explained by greater rates of origination in or movement into poleward environments by polyploid species relative to diploid ones. At this point, given that neither hypothesis was strongly supported in any of the four clades we examined, we do not favor one explanation over the other when it comes to general patterns in these groups. However, we note that a lack of support for the LPG once relationships between modern day latitude and ploidy take the phylogeny into account (Figure 3). That could mean that many of the polyploid species in higher latitudes are closely related, which could also be interpreted as support for the centers of polyploidization hypothesis.

The general lack of support for either greater rates of movement or origination at high latitudes relative to diploids

may accord with a third hypothesis to explain the LPG: that of higher latitude environments being centers of survival for polyploid plants. In this scenario, polyploids do not originate or move to poleward latitudes at higher rates relative to diploids. Rather, polyploids and diploids originate at the same rates in high latitude environments, but diploids go extinct more frequently than polyploids (see Stebbins, 1984). In this case, the harsh environmental conditions hypothesized under centers of polyploidization to create the conditions for higher rates of unreduced gamete formation, and thus polyploidization, instead filter out diploids in favor of polyploids, perhaps due to polyploidy conferring beneficial traits to tolerate abiotic stresses (Tossi et al., 2022). Our current analysis is unable to detect this possible pattern because we did not examine diversification rates; additionally, strong support for centers of survival would require finding higher rates of extinction in diploids at high latitudes relative to polyploids at high latitudes. We hope that future work will address this and other possible mechanisms for creating the LPG, both in our clades of interest and beyond.
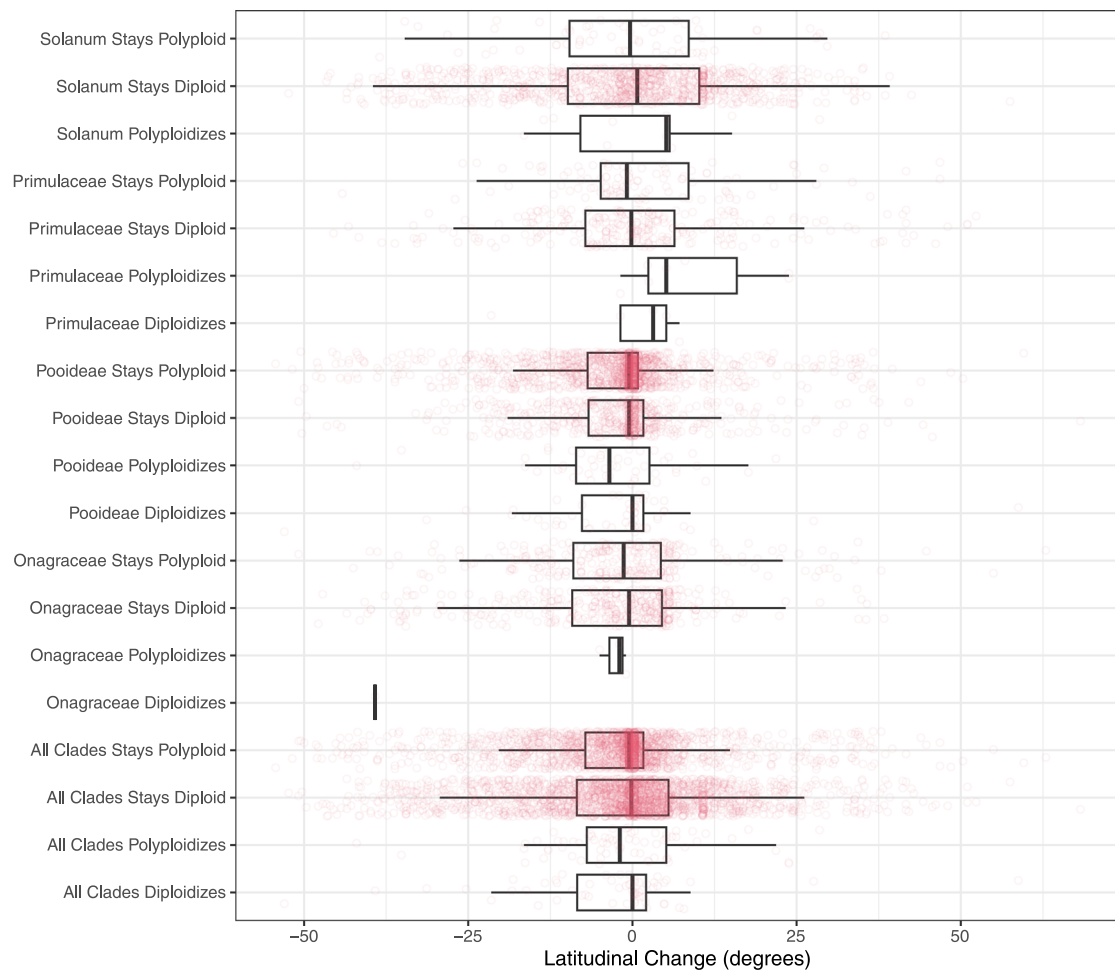
FIGURE 5 Boxplots of the change in median latitude across clades, averaged across all time slices and separated by ploidy status category. Underlying distributions are shown in red.

We hasten to acknowledge that the study conducted here is very much a preliminary one. Despite our unclear results, we hope that phylogenetic-informed ecological niche modeling will continue to be used to study both the LPG and other biogeographic patterns. Such methods would be improved by the introduction of more sophisticated ancestral state reconstruction. In machuruku, ancestral characters are estimated assuming a simple Brownian motion model of evolution, and parameters underlying the evolutionary model are not free for the user to adjust or conduct model selection procedures. Fortunately, new work is being conducted to study whether bioclimatic variables are correlated with diversification rate changes (Zhang et al., 2021) and allow for selective models of climatic evolution such as Ornstein–Uhlenbeck models with hidden states (Boyko et al., 2023).

Clade-specific patterns

We were surprised to find little difference between movements in polyploidized vs. diploidized species in

Solanum because this is the only one of our groups that is distributed primarily in the southern rather than northern hemisphere (Olmstead and Palmer, 1997). It is possible that their Andean center of richness causes species to move elevationally rather than latitudinally, though there is a noticeable spike in antiequatorial movement in lineages that polyploidized during the MIS19 (Appendix S5). The Andean distribution of *Solanum* may also explain the equatorial movement seen in the clade during the mPWP. In the temperate clades, latitudinal differences are difficult to decipher, possibly due to the narrow and biased GBIF ranges centered on Europe (see Beck et al., 2014). The largest group with the most reconstructed ploidy shifts, Pooideae, showed the most significant results by far, with the most tests showing significant differences in latitudinal movement among groups, though movement is most clearly observed in species that diploidized during the LIG (Appendix S5). Possibly the other, smaller clades with few reconstructed ploidy shifts leave us with little statistical power to detect associations between ploidy and latitudinal movement. It may also be the case that unique adaptations to cold stress (Schubert et al., 2019) and arid environments

(Das et al., 2021) found in several species within the clade enable movement to higher latitudes.

Alternatively, it is possible that the biogeographic patterns displayed by species in each ploidy status category, which compose the LPG, arose before the time scale we studied, such as during one of the Pliocene glaciation events in northern latitudes before the M2 (De Schepper et al., 2014). In this scenario, species may not exhibit significant movement in the present day or recent geologic past due to niches already being filled in polar environments. Additionally, there is the possibility that LPG distribution patterns are created via polyploid formation due to secondary contacts of previously isolated populations confined to glacial refugia (Stebbins, 1984, 1985). Though this hypothesis is very similar to the centers of polyploidization hypothesis, testing it would require comparing polyploid frequencies in deglaciated areas to nondeglaciated areas rather than a simple latitudinal comparison. If this hypothesis proves to be explanatory, it would explain the lack of movement mostly observed in temperate clades, which possess ranges that overlap with potential glacial refugia (Comes and Kadereit, 1998).

While we did find support for antiequatorial movement in Pooideae in species that polyploidized relative to those that diploidized, and the opposite pattern in Primulaceae, these findings may be better explained by methodological limitations rather than clade-specific traits. While rates of diploidization vary across species (Li et al., 2021), it is likely that full genome reorganization requires much more time than was included in in the 3.3 Myr for which we possessed paleoclimatic data (Lynch and Conery, 2003; Landis et al., 2018). Future studies may benefit from examining longer time scales than we considered here.

Caveats

Our study is not without important caveats. First, ploidy levels of tropically distributed plant species remain largely uncharacterized relative to those with temperate distributions (Husband et al., 2013; Vasconcelos, 2023). This pattern is reflected in a large European bias in the distributions of plants included in this study. Additionally, our interpretations of how ploidy changes relate to subsequent latitudinal movements in each clade are limited by the available resolution of paleoclimatic data through time. For example, species that exhibit small amounts of latitudinal change after ploidy change may have transitioned soon before the end of the time slice, and movement in the subsequent time slice that may be caused by the ploidy change would not be detected by our methods. In other words, it is possible that latitudinal movement may occur after a "lag" (Schranz et al., 2012). While the lag hypothesis focuses on gaps between polyploidy and diversification, if lags are often required for the "success" of polyploids, this necessity may also explain delayed ecological shifts or phenotypic shifts that enable range expansion and alteration.

We do not believe that incorporating diversification into our modeling of ploidy changes and latitudinal movement would significantly affect our results, because we restricted our analysis to a relatively small time period (3.3 Ma to the present) and to range shifts occurring in the immediate aftermath of neopolyploidization events. Our methods specifically tested for trends in latitudinal movement after ploidy changes within specific windows of time, and we did not examine the possibility of latitudinal changes after paleopolyploidization events, which would have required a much longer time frame and an incorporation of lags, which have so far been modeled only indirectly (Kellogg, 2016) and with great difficulty (see Smith et al., 2018). However, we do believe it is possible that distinguishing between auto- and allopolyploid events may shed light on patterns in latitudinal movement after ploidy shifts because allopolyploids in particular have been theorized to arise via secondary contacts in deglaciated environments (Stebbins, 1984, 1985; Brochmann et al., 2004). We did not distinguish and separately analyze allopolyploids and autopolyploids in this study because there is currently little information available on the hybridization histories of plants that possess ploidy data, and phylogenetic approaches used to infer ploidy are currently unable to distinguish between the two (see Halabi et al., 2023). A recent smaller-scale study which was able to include hybridization histories for all included species found that allopolyploids do not consistently tend to occupy higher-latitude or more extreme environments relative to their diploid progenitors (Mata et al., 2023).

Finally, the unevenness of our historical data makes it difficult to solidly connect ploidy shifts to subsequent latitudinal changes. Our time slices of interest range widely in size: the gap between the M2 and the mPWP is smaller than 100 kyr, while the largest gap between the mPWP and MIS19 is almost 2.5 Myr. The large number of ploidy shifts detected between the mPWP and the MIS19 could be attributed to climatic changes because mean annual temperature declined during this period (Lisiecki and Raymo, 2005) or to the relatively long period between these time slices. Additionally, while the inclusion of phylogeny in reconstructing ancestral ranges will, in theory, produce better predictions, estimates can be spurious in cases where closely related species on the phylogeny exhibit widely disjunctive ranges. In the case of Pooideae, species in the genus *Aciachne* were reconstructed to have a very high median latitude around 50° north in the MIS19 and prior. However, all three species of the genus included in our study have present-day median latitudes around −10° south of the equator, making such large shifts suspect. This discrepancy is likely driven by the biogeographic influence of closely related genera like *Oryzopsis*, in which all three of the species included in our data set were estimated to occupy median latitudes around the range of 40° to 50° north from the M2 to the present, and *Piptocheatium*, in which species ranges vary widely. As examples, *P. lasianthum* currently occurs in southern Brazil and northeast Argentina, while *P. avenaceum* occurs from Mexico to southeast Canada (POWO, 2023).

CONCLUSIONS

Our first examination of the historical causes of latitudinal gradient distribution patterns in the varying cytotypes of Onagraceae, Primulaceae, *Solanum*, and Pooideae found clade-specific differences in support of our two hypotheses of interest, specifically that LPG patterns are driven more by polyploid origination at higher latitudes or by polyploid movement to higher latitudes. When comparing the median latitudes and latitudinal movement across species that stayed polyploid, stayed diploid, polyploidized, and diploidized in individual time slices, we found significant differences in our largest clade, Pooideae. We also found significant differences in starting latitudes across clades, though the latitudinal relationship between species that polyploidized vs. diploidized varied. While we were able to detect LPG distribution patterns in differences between median latitudes occupied by species that stay polyploid as opposed to stay diploid, we likely lack sufficient data to detect differences between species that polyploidize as opposed to diploidize. We hope that this question will be further studied using similar methods with larger, more inclusive clades.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

Thais Vasconcelos is an Associate Editor of the *American Journal of Botany* but took no part in the peer review and decision processes for this paper.

DATA AVAILABILITY STATEMENT

All scripts and data generated by this study are available from the following Dryad Digital Repository: https://doi.org/10.5061/dryad.hx3ffbgnm. Occurrence data are available from GBIF [GBIF.org occurrence downloads: https://doi.org/10.15468/dl.pw2qns (23 July 2021); https://doi.org/10.15468/dl.yesy2v (23 July 2021); https://doi.org/10.15468/dl.vqm9q3 (23 July 2021); https://doi.org/10.15468/dl.gqu424 (23 July 2021); https://doi.org/10.15468/dl.3ucjgk (23 July 2021); https://doi.org/10.15468/dl.78shpr (23 July 2021); https://doi.org/10.15468/dl.f9pq57 (23 July 2021)].

ORCID

Eric R. Hagen http://orcid.org/0000-0002-9447-7664
Thais Vasconcelos http://orcid.org/0000-0001-9991-7924
James D. Boyko http://orcid.org/0000-0003-0952-169X

REFERENCES

Akaike, H. 1974. A new look at statistical model identification. *IEEE Transactions on Automatic Control AU*-19: 716–722.

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.

Barringer, B. C. 2007. Polyploidy and self-fertilization in flowering plants. *American Journal of Botany* 94: 1527–1533.

Beaulieu, J. M., B. C. O'Meara, and M. J. Donoghue. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62: 725–737.

Beck, J., M. Böller, A. Erhardt, and W. Schwanghart. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics* 19: 10–15.

Bierzychudek, P. 1985. Patterns in plant parthenogenesis. *Experientia* 41: 1255–1264.

Brochmann, C., A. K. Brysting, I. G. Alsos, L. Borgen, H. H. Grundt, A.-C. Sheen, and R. Elven. 2004. Polyploidy in arctic plants. *Biological Journal of the Linnean Society* 82: 521–536.

Boyko, J. D., and J. M. Beaulieu. 2021. Generalized hidden Markov models for phylogenetic comparative datasets. *Methods in Ecology and Evolution* 12: 468–478.

Boyko, J. D., E. R. Hagen, J. M. Beaulieu, and T. Vasconcelos. 2023. The evolutionary responses of life-history strategies to climatic variability in flowering plants. *New Phytologist* 240: 1587–1600.

Brown, J. L., D. J. Hill, A. M. Dolan, A. C. Carnaval, and A. M. Haywood. 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Scientific Data* 5: 180254.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and inference: a practical information–theoretic approach, 2nd ed. Springer-Verlag, NY, NY, USA.

Caetano, D. S., B. C. O'Meara, and J. M. Beaulieu. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* 72: 2308–2324.

Comes, H. P., and J. W. Kadereit. 1998. The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science* 3: 432–438.

Conover, W. J. 1971. Practical nonparametric statistics. John Wiley, NY, NY, USA.

Das, A., A. Prakash, N. Dedon, A. Doty, M. Siddiqui, and J. C. Preston. 2021. Variation in climatic tolerance, but not stomatal traits, partially explains Pooideae grass species distributions. *Annals of Botany* 128: 83–95.

De Schepper, S., P. L. Gibbard, U. Salzmann, and J. Ehlers. 2014. A global synthesis of the marine and terrestrial evidence for glaciation during the Pliocene Epoch. *Earth-Science Reviews* 135: 83–102.

De Storme, N., and D. Geelen. 2014. The impact of environmental stress on male reproductive development in plants: biological processes and molecular mechanisms. *Plant, Cell & Environment* 37: 1–18.

De Vos, J. M., R. O. Wüest, and E. Conti. 2014. Small and ugly? Phylogenetic analyses of the "selfing syndrome" reveal complex evolutionary fates of monomorphic primrose flowers. *Evolution* 68: 1042–1057.

Freyman, W. A., and S. Höhna. 2019. Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae lineages. *Systematic Biology* 68: 505–519.

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.

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology* 60: 451–465.

Guillory, W. X., and J. L. Brown. 2021. A new method for integrating ecological niche modeling with phylogenetics to estimate ancestral distributions. *Systematic Biology* 70: 1033–1045.

Halabi, K., A. Shafir, and I. Mayrose. 2023. PloiDB: the plant ploidy database. *New Phytologist* 240: 918–927.

Husband, B. C., S. J. Baldwin, and J. Suda. 2013. The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. *In* I. J. Leitch, J. Greilhuber, J. Dolezel, and J. Wendel [eds.], Plant genome diversity, vol. 2, 255–276. Springer, Vienna, Austria.

Jiao, Y, N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr, P. E. Ralph, L. P. Tomsho, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

Kellogg, E. A. 2016. Has the connection between polyploidy and diversification actually been tested? *Current Opinion in Plant Biology* 30: 25–32.

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.

Leitch, A. R., and I. J. Leitch. 2008. Genomic plasticity and the diversity of polyploid plants. *Science* 320: 481–483.

Levin, D.A. 1983. Polyploidy and novelty in flowering plants. *American Naturalist* 122: 1–25.

Li, Z., M. T. W. McKibben, G. S. Finch, P. D. Blischak, B. L. Sutherland, and M. S. Barker. 2021. Patterns and processes of diploidization in land plants. *Annual Review of Plant Biology* 72: 387–410.

Lisiecki, L. E., and M. E. Raymo. 2005. A Pliocene–Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography* 20: PA1003.

Lohaus, R., and Y. Van de Peer. 2016. Of dups and dinos: evolution at the K/Pg boundary. *Current Opinion in Plant Biology* 30: 62–69.

Löve, A., and D. Löve. 1943. The significance of differences in the distribution of diploids and polyploids. *Hereditas* 29: 145–163.

Löve, A., and D. Löve. 1949. The geobotanical significance of polyploidy. I. Polyploidy and latitude. *In* R. B. Meyer and Y. Goldschmidt [ed.], Portugaliae Acta Biologica, série A special volume, 273–352. Instituto Botânico de Lisboa, Lisbon, Portugal.

Lynch, M., and J. S. Conery. 2003. The evolutionary demography of duplicate genes. *In* A. Meyer and Y. Van de Peer [eds.], Genome evolution: gene and genome duplications and the origin of novel gene functions, 35–44. Springer, Dordrecht, Germany.

Ma, Z., B. Sandel, and J.-C. Svenning. 2016. Phylogenetic assemblage structure of North American trees is more strongly shaped by glacial–interglacial climate variability in gymnosperms than in angiosperms. *Ecology and Evolution* 6: 3092–3106.

Mandáková, T., and M. A. Lysak. 2018. Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology* 42: 55–65.

Marks, R. A., S. Hotaling, P. B. Frandsen, and R. VanBuren. 2021. Representation and participation across 20 years of plant genome sequencing. *Nature Plants* 7: 1571–1578.

Mata, J. K., S. L. Martin, and T. W. Smith. 2023. Global biodiversity data suggest allopolyploid plants do not occupy larger ranges or harsher conditions compared with their progenitors. *Ecology and Evolution* 13: e10231.

Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333: 1257.

McElwain, J. C., and M. Steinthorsdottir. 2017. Paleoecology, ploidy, paleoatmospheric composition, and developmental biology: a review of the multiple uses of fossil stomata. *Plant Physiology* 174: 650–664.

Meyers, L. A., and D. A. Levin. 2006. On the abundance of polyploids in flowering plants. *Evolution* 60: 1198–1206.

Mudelsee, M., and M. E. Raymo. 2005. Slow dynamics of the Northern Hemisphere glaciation. *Paleoceanography* 20: PA4022.

Normand, S., R. E. Ricklefs, F. Skov, J. Bladt, O. Tackenberg, and J.-C. Svenning. 2011. Postglacial migration supplements climate in determining plant species ranges in Europe. *Proceedings of the Royal Society, B, Biological Sciences* 278: 3644–3653.

Olmstead, R. G., and J. D. Palmer. 1997. Implications for the phylogeny, classification, and biogeography of *Solanum* from cpDNA restriction site variation. *Systematic Botany* 22: 19–29.

POWO. 2023. Plants of the world online. Royal Botanic Gardens, Kew, UK. Website: http://www.plantsoftheworldonline.org/

Price, T. D., A. Qvarnström, and D. E. Irwin. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society, B, Biological Sciences* 270: 1433–1440.

R Core Team. 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: https://www.R-project.org/

Ramsey, J., and D. W. Schemske. 2002. Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* 33: 589–639.

Raunkiaer, C. 1934. The life forms of plants and statistical plant geography; being the collected papers of C. Raunkiaer. Clarendon Press, Oxford, UK.

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

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.

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.

Robertson, K., E. E. Goldberg, and B. Igić. 2011. Comparative evidence for the correlated evolution of polyploidy and self-compatibility in Solanaceae. *Evolution* 65: 139–155.

Särkinen, T., L. Bohs, R. G. Olmstead, and S. Knapp. 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology* 13: 1–15.

Schranz, M. E., S. Mohammadin, and P. P. Edger. 2012. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Current Opinion in Plant Biology* 15: 147–153.

Schubert, M., L. Grønvold, S. R. Sandve, T. R. Hvidsten, and S. Fjellheim. 2019. Evolution of cold acclimation and Its role in niche transition in the temperate grass subfamily Pooideae. *Plant Physiology* 180: 404–419.

Segraves, K. A. 2017. The effects of genome duplications in a community context. *New Phytologist* 215: 57–69.

Smirnov, N. V. 1939. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow University Mathematics Bulletin* 2: 3–26.

Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 302–314.

Smith, S. A., J. W. Brown, Y. Yang, R. Bruenn, C. P. Drummond, S. F. Brockington, J. F. Walker, et al. 2018. Disparity, diversity, and duplications in the Caryophyllales. *New Phytologist* 217: 836–854.

Soltis, D. E., M. C. Segovia-Salcedo, I. Jordon-Thaden, L. Majure, N. M. Miles, E. V. Mavrodiev, W. Mei, et al. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytologist* 202: 1105–1117.

Soreng, R. J., P. M. Peterson, K. Romaschenko, G. Davidse, J. K. Teisher, L. G. Clark, P. Barberá, et al. 2017. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: an update and a comparison of two 2015 classifications. *Journal of Systematics and Evolution* 55: 259–290.

Spriggs, E. L., P.-A. Christin, and E. J. Edwards. 2014. C₄ photosynthesis promoted species diversification during the Miocene grassland expansion. *PloS One* 9: e97722.

Stebbins, G. L. 1950. Variation and evolution in plants. Columbia University Press, NY, NY, USA.

Stebbins, G. L. 1971. Chromosomal evolution in higher plants. Addison-Wesley, London, UK.

Stebbins, G. L. 1984. Polyploidy and the distribution of the arctic-alpine flora: new evidence and a new approach. *Botanica Helvetica* 94: 1–13.

Stebbins, G. L. 1985. Polyploidy, hybridization, and the invasion of new habitats. *Annals of the Missouri Botanical Garden* 72: 824–832.

Sundaram, M., and A. B. Leslie. 2021. The influence of climate and palaeoclimate on distributions of global conifer clades depends on geographical range size. *Journal of Biogeography* 48: 2286–2297.

Svenning, J.-C., and F. Skov. 2007. Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology Letters* 10: 453–460.

Tossi, V. E., L. J. Martinez Tosar, L. E. Laino, J. Iannicelli, J. J. Regalado, A. S. Escandón, I. Baroli, et al. 2022. Impact of polyploidy on plant tolerance to abiotic and biotic stresses. *Frontiers in Plant Science* 13: 869423.

Van Drunen, W. E., and B. C. Husband. 2019. Evolutionary associations between polyploidy, clonal reproduction, and perenniality in the angiosperms. *New Phytologist* 224: 1266–1277.

Vasconcelos, T. 2023. A trait-based approach to determining principles of plant biogeography. *American Journal of Botany* 110: e16127.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1: 80–83.

Xu, Z., and L. Chang. 2017. Primulaceae. *In* Z. Xu and L. Chang [eds.], Identification and control of common weeds, 3, 51–81. Springer, Singapore.

Zenil-Ferguson, R., J. G. Burleigh, W. A. Freyman, B. Igić, I. Mayrose, and E. E. Goldberg. 2019. Interaction among ploidy, breeding system and lineage diversification. *New Phytologist* 224: 1252–1265.

Zhang, X., J. B. Landis, Y. Sun, H. Zhang, N. Lin, T. Kuang, X. Huang, et al. 2021. Macroevolutionary pattern of Saussurea (Asteraceae) provides insights into the drivers of radiating diversification. *Proceedings of the Royal Society, B, Biological Sciences* 288: 20211575.

SUPPORTING INFORMATION

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

Appendix S1. Onagraceae phylogeny.

Appendix S2. Primulaceae phylogeny.

Appendix S3. Pooideae phylogeny.

Appendix S4. Solanum phylogeny.

Appendix S5. Latitudinal change by clade.