



# Origin and diversification of flax and their relationship with heterostyly across the range

Enrique Maguilla  | Marcial Escudero | José Ruíz-Martín | Juan Arroyo

Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Sevilla, Spain

## Correspondence

Enrique Maguilla, Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Apartado 1095, 41080 Sevilla, Spain.

Email: emagsal@gmail.com

## Funding information

Ministerio de Ciencia, Innovación y Universidades; Spanish Government, Grant/Award Number: FJCI-2017-32314 and BES-2008-003946; FEDER, Grant/Award Number: CGL2013-45037-P and PGC2018-099608-B-100

Handling Editor: Gerald Schneeweiss

## Abstract

**Aim:** Understanding plant diversity and how different traits have shaped the current biodiversity setting across the world is one of the major challenges for evolutionary biology. In this sense, genus *Linum* have been broadly studied as a model of species showing heterostyly, but no studies have attempted to investigate possible correlations between heterostyly and diversification. Thus, in this study we aim to explore the origin and diversification of the genus *Linum* and to determine its centre of diversity and potential source for the colonisation of other areas. Also, we aim to study how heterostyly may have shaped diversification rates and colonisation patterns in the genus.

**Location:** Worldwide (focused on the Western Palearctic).

**Taxon:** Genus *Linum* L. (Linaceae).

**Methods:** We analysed nuclear ITS and plastid *ndhF*, *matK* and *trnL-F* DNA regions in a total of 103 samples of 93 different species of the genus *Linum*, as well as seven samples from other Linaceae. We performed divergence time analysis in BEAST2 under a birth-death tree model, then used the resulting tree for an ancestral area reconstruction using the R package “BioGeoBEARS.” Finally, we estimated diversification rates using BAMM and examined the correlation between diversification rates and geographic ranges and heterostyly.

**Results:** *Linum* s.l. diverged during the late Eocene to mid Oligocene (27.2–38.29 Ma at 95% highest posterior density interval) in the Western Palearctic, where most species diversified. Within-area speciation is the main mechanism of diversification in the genus. Most dispersal events occurred from the Western Palearctic to other regions, probably through long-distance dispersal (LDD). No changes were found regarding diversification rates in specific clades or in relation with reproductive system (heterostyly) or geographic areas. Colonisation of new areas was achieved mostly by monomorphic lineages, whereas heterostylous species mostly remained in the ancestral Western Palearctic area.

**Main conclusions:** The Western Palearctic acted as a source of dispersal in seven out of nine total dispersal events but never as a sink. All species or lineages that colonised new areas after LDD were monomorphic for style length, as predicted by the theory on reproductive traits of colonising species. Thus, heterostyly is shown to affect

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Journal of Biogeography* published by John Wiley & Sons Ltd.



establishment success in a new area. Neither biogeographic changes nor analysed trait changes can explain speciation nor extinction rates in *Linum*. Our results confirm that the evolution of heterostyly is related to the paleogeographic history and are not consistent with the idea that a key innovation or “dispersification” has shaped the diversification patterns in *Linum*.

#### KEYWORDS

biogeography, BiSSE, breeding system, colonization, GeoSSE, life history, *Linum*, long-distance dispersal, Mediterranean, style polymorphism

## 1 | INTRODUCTION

Explaining correlates of plant diversity across the World and its historical setting and possible associated biological traits is a major challenge for plant evolutionary biology. The Mediterranean Basin is one of the richest biodiversity hotspots and has the third highest rate of endemism amongst plant species (about 13,000 out of 25,000 species), following the Tropical Andes in South America (20,000 species) and Sundaland in Southeast Asia (15,000; Myers et al., 2000). Characterised by a highly seasonal climate, species growth is limited during the warm, dry summer season in this region (Thompson, 2020). Additionally, due to its location between temperate and tropical climates, the Mediterranean Basin has served as a refugium for many species during the Tertiary and the Quaternary glaciations and source for later colonisation of adjacent areas during interglacial periods (Thompson, 2020). This fact affected the evolution of many plant lineages by facilitating recurrent hybridisation, isolation and adaptation in alternate climates (Avice, 2000; Maguilla et al., 2017; Schneeweiss et al., 2017). Moreover, adjacent areas such as the Irano-Turanian floristic region, ranging from western Asian subregions to central Asia, have been described as a source of biodiversity which later colonised the Mediterranean region (Manafzadeh et al., 2014; Peterson et al., 2019). Thus, in general, the flora of the Mediterranean Basin is often seen as the result of convergent lineages in a melting pot or crossroads (e.g. Feliner, 2014; Rodríguez-Sánchez et al., 2008), rather than a centre of lineage dispersal.

One of the best ways to explore the assembly of plant biodiversity hotspots is through deep analysis of key plant groups. In this respect, *Linum* L. is a good model system because it has a distribution pattern and some biological traits that make it suitable to test evolutionary hypotheses about its place of origin and the role of key traits in its diversification. *Linum* is the largest genus within the family Linaceae DC. ex Perleb (Dressler et al., 2014; McDill & Simpson, 2011). This genus includes about 180 species (of approximately 255 species comprising Linaceae) with a nearly cosmopolitan distribution (Dressler et al., 2014). The family Linaceae is found in temperate to tropical regions, mainly in the Northern Hemisphere, and the Mediterranean region and Southwestern Asia are the centre of biodiversity of *Linum* (Dressler et al., 2014). Although Linaceae sensu Dressler et al. (2014) has been shown to be monophyletic (McDill &

Simpson, 2011), *Linum* is paraphyletic, with genera *Cliococca* Bab., *Hesperolinon* (A. Gray) Small, *Sclerolinon* C.M. Rogers and *Radiola* Hill nested within the core *Linum* (hereafter *Linum* s.l.; McDill et al., 2009; Ruiz-Martín et al., 2018; Schneider et al., 2016).

The genus *Linum* has been studied as a model group for disentangling the evolution of floral polymorphisms since the seminal work by Darwin (1877) on within-population variation in reproduction in plants. Experimental and observational work by Darwin in several *Linum* species has made this genus a classic representative of plants with one of the most common flower polymorphisms: heterostyly (Barrett, 2019; Barrett & Shore, 2008). Heterostyly is defined as the occurrence of two or three different floral morphs within a given population and mainly occurs due to reciprocal differences in the placement of the stigmas and anthers in the flower (Barrett, 1992, 2019; Darwin, 1877; Ruiz-Martín et al., 2018). The underlying mechanism was described by Darwin to promote outcrossing (Darwin, 1877; Barrett, 1992; Barrett & Shore, 2008; but see alternative hypotheses for the maintenance and loss of heterostyly in Charlesworth & Charlesworth, 1979; Lloyd & Webb, 1992). For example, if heterostyly promotes outcrossing, which has been suggested to favour diversification more than selfing (Stebbins, 1957; but see Igic & Busch, 2013), heterostyly could thus be expected to promote diversification (de Vos, Hughes, et al., 2014) and could be considered a key innovation trait (Haller et al., 2014; Vamosi & Vamosi, 2011). Similar arguments were posed for suggesting a high diversification rate in dioecious clades, but testing this hypothesis has got only mixed results (Vamosi et al., 2018).

The genus *Linum* has both high biodiversity in different regions and wide occurrence of heterostyly (Ruiz-Martín et al., 2018), providing an opportunity to investigate how heterostyly could be correlated with macro-evolutionary processes shaping diversification. The only previous empirical test of the hypothesis that heterostyly promotes diversification was conducted in Primulaceae, where heterostyly is also common (de Vos, Hughes, et al., 2014). Thus, in exploring the importance of heterostyly in diversification, we contribute to better understanding the role of plant traits in explaining angiosperm diversification (Dodd et al., 1999), Darwin's “abominable mystery” (Davies et al., 2004). Here, we add a biogeographical component, given the wide range of the genus and the differential distribution of heterostyly and species diversity of this genus

(Ruiz-Martín et al., 2018). Ruiz-Martín et al. (2018) suggested that there may be a correlation between style polymorphism and distribution; the proportion of *Linum* species that are heterostylous is substantially higher in the Western Palearctic than in the Americas and southern Africa. This could be in part due to the lack of specific pollinators that promote outcrossing (by legitimate pollen transfer) between different morphs, which may differ amongst regions (Busch, 2011; Lloyd & Webb, 1992) or higher colonisation capacity of the non-heterostylous species because they might behave in part as selfers more frequently than heterostylous species (see Baker's law; Baker, 1974).

There has been a long-standing interest in the effect of reproductive systems on the construction of the species range (Moeller et al., 2017). Thus, Baker's law has been invoked to explain the higher frequency of selfers in isolated territories at several spatial scales (Cheptou, 2012). However, the high incidence of dioecy, an obligate outcrossing breeding system, in small, isolated populations and on oceanic islands has been considered as evidence against this rule. However, it should be distinguished if dioecy is an ancestral trait in colonising lineages or if it evolved after colonisation (Pannell et al., 2015). The case of heterostyly has some parallels to dioecy, as it divides the population into two (or three in tristily) mating groups, thus increasing the chances of a limitation of mates. Consequently, heterostyly is expected to be rare in isolated areas (Crawford et al., 2011; Watanabe & Sugawara, 2015). If heterostyly is associated with both heteromorphic incompatibility and reciprocal herkogamy, which requires specialist pollination, a heterostylous species will be a super-outcrosser; this is in strong contrast to monomorphic species, particularly if the latter are self-compatible. In other words, it can be argued that Baker's law can also be applied to heterostyly, hence predicting that nonheterostylous species should colonise new territories more easily and that heterostyly should be rarer in colonised areas. Although there are several studies that indirectly address biogeographical questions in heterostylous groups (e.g. Ferrero et al., 2012; Guggisberg et al., 2006; Watanabe & Sugawara, 2015), including *Linum* (McDill et al., 2009), we know of no previous attempts to explicitly test biogeographical hypotheses of the association between migration, transitions of heterostyly and diversification. In addition, we aimed to test the hypothesis of "dispersification" (Moore & Donoghue, 2007), which proposes that dispersal is followed by a shift in diversification rates of species in the newly colonised area, assuming that these species are pre-adapted to the new niche, such as a harsher pollination environment. Concerning dispersal, we adhere to its biogeographical meaning, that is, the ability to colonise a new area, which includes both diaspore dispersal and establishment. In *Linum*, there are no apparent differences across species, whether heterostylous or not, with respect to seed dispersal abilities, thus dispersal mostly implies the ability to establish once diaspores reach a new area. Based on the above considerations, we expect that (1) clades where heterostyly is common have high diversification rates in contrast to those clades where heterostyly is scarce and (2) that diversification rates are high in clades after dispersal (dispersification) and low in clades without dispersal. We have

no a priori expectations concerning the outcome of the combination of both factors.

In the present work, we aim to reconstruct the biogeographic history of the genus *Linum* s.l. The specific objectives of this study are (1) to elucidate the origin of the most recent common ancestor (MRCA) of the whole genus and of the main clades within *Linum*, (2) to determine the centre of diversification and source for later colonisation of other regions around the world, (3) to estimate diversification rates across the phylogeny and (4) to determine whether dispersed species were monomorphic or polymorphic and to evaluate the influence of heterostyly (monomorphic vs. polymorphic species) and of geographic area (within vs. outside the Palearctic) on colonisation and diversification. To do this, we used updated phylogenetic information. Although we included a significant number of species from around the world, we concentrated particularly on species from the Palearctic, given that this is the region with the highest species richness in the genus.

## 2 | MATERIALS AND METHODS

### 2.1 | Sampling

Sequence data for this study were downloaded from NCBI GenBank. We included all accessions used in a recent phylogenetic study published by Ruiz-Martín et al. (2018), comprising 103 *Linum* samples from 93 different species (approximately 52% of all *Linum* species sensu Dressler et al., 2014). In addition, we included seven samples from different genera within Linaceae (*Anisadenia pubescens* Griff., *Cliococca selaginoides* (Lam.) C.M. Rogers and Mildner, *Hesperolinon micranthum* (A. Gray) Small, *Radiola linooides* Roth., *Reinwardtia indica* Dumort, *Sclerolinon digynum* (A. Gray) C.M. Rogers and *Tirpitzia sinensis* (Hemsl.) Hallier f.), and three samples from genera of sister families as outgroup (*Humiria balsamifera* Aubl., *Hypericum perforatum* L. and *Viola pubescens* Aiton; Table S1 in Appendix S1 in Supporting Information). In total, whenever available, we used four DNA regions for each sample: the nuclear ITS (113 sequences), plastid *ndhF* (114) and *matK* (87) genes, plus *trnL-F* spacer (113; see Ruiz-Martín et al., 2018 for more details). We categorised species as polymorphic (including any state related with style length polymorphism) or monomorphic for style length, according to the information provided in Ruiz-Martín et al. (2018) or, when unavailable in that work, from taxonomic descriptions of species. Unlike in other style-polymorphic groups, heterostyly has been frequently considered a useful diagnostic character in *Linum*, so there is sound information on this trait in floras and taxonomic studies (e.g. Ockendon & Walters, 1968). We coded as polymorphic any species with at least two discrete morphs for style length. This included mostly heterostylous species: The most common state was distyly (46 species), and the three least common states were each expressed in a single species—3D distyly (*L. suffruticosum*; see Armbruster et al., 2006 for a precise description of this rare type), tristily (*L. hirsutum*) and style dimorphism (*L. grandiflorum*). Species with only one morph were coded as



monomorphic. This included the most common states of approach herkogamy (the style protruding the stamens; 19 species) and homomorphism (stamens and style of similar length, also called by some authors homostyly; 20 species) and the less frequent horizontal herkogamy (styles and stamens separated in the horizontal plane of the flower; four species) and reverse herkogamy (styles shorter than stamens; three species). Some examples of this floral variety are shown in Figure S1 in Appendix S1. Most of these states are related with proposed steps in models of the evolution of heterostyly (Charlesworth & Charlesworth, 1979; Lloyd & Webb, 1992).

## 2.2 | Divergence time analysis

We performed a divergence time estimation analysis in BEAST 2.4.0 (Bouckaert et al., 2014) using all four DNA regions combined into a single matrix on the one hand, and considering two partitions (ITS and plastid) on the other hand. Three independent Markov Chain Monte Carlo (MCMC) runs were carried out with 100 million generations each, and GTR+I+G for ITS and GTR+I+G for plastid regions in the partitioned analysis, based on results from jModel-Test 2.1.3 (Darriba et al., 2012). We used a log normal relaxed clock model and birth-death for the tree model (Drummond et al., 2006). Two calibration points were included in our analyses: a secondary calibration point for the stem node of the Linaceae family (a mean of 93.5 Ma and 95% confidence interval between 88 and 97 Ma; Bell et al., 2010) following a normal distribution, and a fossil record for the crown node of the genus *Linum*, including additional genera gathered within *Linum* with a log-normal distribution with an estimated mean age of 33.9 Ma and an interval of 33.9–37.2 Ma (Cavagnetto & Anadón, 1996; but see details in Ruiz-Martín et al., 2018). Results were visualised to assess convergence in Tracer 1.6 (Rambaut & Drummond, 2013), and a maximum clade credibility (MCC) tree was reconstructed using TreeAnnotator 2.4.0 (Bouckaert et al., 2014), removing the first 20% of the trees from each run as burn-in.

## 2.3 | Ancestral areas reconstruction and biogeographic events

We pruned the calibrated phylogeny (Figure S2 in Appendix S1) for our analyses, keeping only one sample per monophyletic species. For polyphyletic species (*L. densiflorum* P.H. Davis, *L. punctatum* C. Presl and *L. tenue* Desf.), we kept one sample per clade in which the species were present. A total of 96 species (100 samples, considering duplicates of all polyphyletic species, plus both subspecies of *L. austriacum* L.) constituted the core *Linum* in our analyses. In addition to *Linum* species, we included samples from *C. selaginoides*, *H. micranthum*, *S. digynum* and *R. linoides*. Finally, although we removed all outgroup species considered by Ruiz-Martín et al. (2018), we maintained the sister clade to the core *Linum* because its inclusion could shed light on the origin of the MRCA to the core *Linum*. This clade

was represented mainly by Asian species: *A. pubescens*, *R. indica* and *T. sinensis*.

Ancestral areas reconstruction was performed using the package “BioGeoBEARS 1.1” (Matzke, 2013) in R 3.2.2 (R Core Team, 2015) with the MCC tree from BEAST. Absence or presence of species was coded for six biogeographic realms: (1) Western Palearctic, (2) Eastern Palearctic (including the Indo-Malaysian realm), (3) Afrotropic, (4) Nearctic, (5) Neotropic and (6) Australasia (see realm codification for each species in Table S1 in Appendix S1). Limits for these regions were based on Udvardy (1975; with modifications by Olson et al., 2001). Nevertheless, due to the predominantly Mediterranean distribution of *Linum*, we divided the Palearctic into the Western Palearctic, where most species of the genus occur, and Eastern Palearctic, where few species are distributed, using the Ural Mountains as the boundary between the two regions. Moreover, we included the Indo-Malaysian realm in the Eastern Palearctic because all species in this region are also in the Eastern Palearctic. The realms considered mostly coincide with continental plates that were also clearly separated during the Oligocene, when *Linum* s.l. originated and began differentiating (see below), and the Indo-Malaysian and Eastern Palearctic realm were tightly connected (Scotese, 2014).

“BioGeoBEARS” can implement three main reticulate models (Ronquist & Sanmartín, 2011): Dispersal-Extinction-Cladogenesis (DEC; Ree & Smith, 2008), Dispersal-Vicariance Analysis (DIVA-like; Ronquist, 1997) and BayArea-like (Landis et al., 2013). We decided not to consider a BayArea-like model in our analyses because it would have assumed modes of speciation that we considered unrealistic in view of our study group and codification of biogeographic areas. The remaining DEC and DIVA-like models were implemented with two free parameters—“*d*” (range expansion) and “*e*” (range contraction). A third parameter “*j*” (founder event or jump dispersal), which allows the analysis to consider cladogenetic dispersal, was not included in our analyses, since it forces models to have greater tendency towards explaining the data entirely by cladogenetic events and inferring, in some cases, anagenetic rates of 0 (Ree & Sanmartín, 2018), which could lead to unrealistic results.

Additionally, testing for the importance of changes in dispersal rates through time, we repeated all approaches above constrained with a combination of connectivity matrices. The dispersal matrices were based on those by Dupin et al. (2017), which set different rates depending on the distance between continents (Table S2 in Appendix S1). Variation in plant dispersal between North (Nearctic realm) and South America (Neotropical realm), between 24 and 10 Ma, was codified in the matrices based on Bacon et al. (2015; Table S2 in Appendix S1). Matrices were scaled using an additional free parameter, “*w*”. This parameter reduces the effect of subjectivity in the establishment of specific values by the researcher in the matrices. In total, four models were tested: DEC and DIVA-like, each one unconstrained and constrained (with dispersal multiplier matrices and free parameter “*w*”). We selected the best fit model with the highest Akaike Information Criterion weight (AICw; Akaike, 1974).

Phylogenetic uncertainty was evaluated for the best fitting model by running independent “BioGeoBEARS” analyses in 100

post-burn-in trees from BEAST and estimating the average probabilities from all trees for the reconstruction of a consensus tree.

We used the biogeographical stochastic mapping (BSM), as used in Matzke (2014) and implemented in “BioGeoBEARS 1.1” package in R 3.2.2 (Dupin et al., 2017; R Core Team, 2015) to characterise biogeographical events between areas (number and type of events). Mean and standard deviation were estimated after 50 BSM runs.

## 2.4 | Diversification analyses

Bayesian analyses of macroevolutionary mixtures (BAMM; Rabosky et al., 2013, 2014; Shi & Rabosky, 2015) were used to quantify diversification rates over time across the *Linum* phylogeny. Analyses were done in BAMM 2.5 as implemented in the R package “BAMMtools” (Rabosky, Grudler, et al., 2014). Our analysis allowed shifts in diversification rates and changes over time within each regime. We ran BAMM for five million generations. We used the R package “coda” (Plummer et al., 2006) to check MCMC convergence. The “BAMMtools” R package was used to process the results and summarise the parameters of the models with the highest posterior probabilities. In a recent study, Louca and Pennell (2020) concluded that because extant time trees are consistent with multiple diversification scenarios, they are not appropriate by themselves to estimate diversification rates. However, Morlon et al. (2020) concluded that Louca and Pennell's (2020) results do not undermine hypothesis-driven model selection procedures and that diversification rates may be investigated from extant time trees using a data-driven procedure, without a priori hypothesis implemented in the model of the analysis.

To test whether heterostyly or geographic range is correlated with diversification rates in *Linum*, we used two different state-dependent speciation and extinction (SSE) approaches to estimate speciation and extinction rates in relation to (i) heterostyly as a binary morphological trait (monomorphic vs. polymorphic) using a binary-state speciation and extinction model (BiSSE; Maddison et al., 2007) and (ii) geographic range as a binary trait (Western Palearctic—the origin based on BioGeoBEARS results—vs. the rest other realms) using a geographic model (GeoSSE; Goldberg et al., 2011). Both analyses were done in the R package “diversitree” (Fitzjohn, 2012). For the BiSSE model, we compared the fits of the full models versus constrained models (equal rates of transition/dispersal, equal speciation rates or equal extinction rates). We also performed an ancestral state reconstruction for style monomorphism versus polymorphism using the BiSSE model. For the GeoSSE model, we compared the fits of the full models versus constrained models (equal speciation and extinction rates or null speciation rates in the widespread range). We estimated the parameters of the models (BiSSE and GeoSSE) using a Bayesian approach. To do this, we ran the full model for the BiSSE model and a constrained model for GeoSSE (widespread speciation was not allowed as only one species was coded as widespread). Then, we ran MCMC analyses with one million generations each.

## 2.5 | Evolution of reproductive system, geographic distribution and life history

Because we did not find statistically significant trait-dependent diversification after the BiSSE analyses (see Section 3), we used Pagel's model (Pagel, 1994) for the correlated evolution of two binary traits as implemented in the function `fitPagel` of the R package “phytools” (Revell, 2012). We ran the analyses twice, once assuming equal rates of transitions between character states and once with different rates (model “ER” and “ARD”). We also ran the analyses assuming that the variable *x* depended on variable *y* and vice versa. We tested whether evolution was dependent or independent amongst three traits: reproductive system (as coded in the BiSSE analyses), geographic distribution (as coded in the GeoSSE analyses) and annual versus perennial life history (as coded in Ruiz-Martín et al., 2018).

## 3 | RESULTS

### 3.1 | Divergence times

The crown node for *Linum* s.l. (which includes the genera *Cliococca*, *Hesperolinon*, *Radiola* and *Sclerolinon*) was estimated to have originated during the late Eocene to mid Oligocene (a mean of 32.55 Ma; 27.2–38.26 Ma at 95% HPD; Figure S2 in Appendix S1). Two main clades were differentiated within *Linum*. One originated during the early Oligocene to late Miocene (25.99 Ma; 19.88–31.71 at 95% HPD), containing species mainly from the Western Palearctic. The other clade originated during the Miocene (15.73 Ma; 9.42–22.15 at 95% HPD) and was composed of taxa from the Western Palearctic, Afrotropic, Nearctic and Neotropic (Figure S2 in Appendix S1).

### 3.2 | Ancestral areas

The unconstrained DIVA-like model was the most supported model, with an AIC<sub>c</sub> weight of 0.53 (Table 1, but see Figure S3 in Appendix S1 for results under unconstrained DEC model, which had similar results). This model had an extinction rate (*e*) of zero ( $1e-12$  events/Myr) and anagenetic dispersal rate (*d*) of 0.0039 events/Myr (Table 1). The summarised result of the ancestral area reconstruction under the DIVA-like model in the 100 post-burn-in MCC trees was almost identical to the obtained using the single consensus MCC tree (Figure S4 in Appendix S1).

The origin of the MRCA of *Linum* s.l. was predicted to have occurred in the Western Palearctic (Figure 1). Two main clades were derived from the MRCA, one with species distributed mainly in the Western Palearctic (including *Linum* sections *Dasylinum* (Planch.) Juz., and *Linum*) and another comprising species from the Western Palearctic, Afrotropic and the Nearctic and Neotropic realms (sections *Cathartolinum* (Rchb.) Griseb., *Linopsis* (Rchb.) Engelm. and *Syllinum* Griesb.). The MRCA of this second clade was predicted to be from the Western Palearctic and the



TABLE 1 Biogeographical models tested in “BioGeoBEARS”

Model	LnL	Free params	d	e	w	AIC <sub>c</sub>	AIC <sub>c</sub> wt
No constraints							
DEC	-64.64	2	0.0028	1.0e-12	1	133.4	0.16
<b>DIVA-like</b>	<b>-63.44</b>	<b>2</b>	<b>0.0039</b>	<b>1.0e-12</b>	<b>1</b>	<b>131</b>	<b>0.53</b>
Dispersal multipliers							
DECc	-64.08	3	0.0038	1.0e-12	0.39	134.4	0.097
DIVA-likec	-63.32	3	0.0041	1.0e-12	0.035	132.9	0.21

Note: Dispersal-extinction-cladogenesis (DEC) and dispersal-vicariance like model (DIVA-like) were tested. Additionally, all analyses were performed adding dispersal multiplier matrices with shifts over time as input, based on Dupin et al. (2017). In these cases, “w” was considered as a free parameter to reduce subjectivity induced by the assignment of a given value for dispersal between geographic realms. LnL indicates Log-likelihood for each model tested, numbers of free parameters included as “Free Params,” “d” for anagenetic dispersal rate, “e” for extinction, Akaike’s information criterion corrected for small sample size (AIC<sub>c</sub>), Akaike’s information criterion weights for all four models (AIC<sub>c</sub>wt). The best fitted model is in bold text.

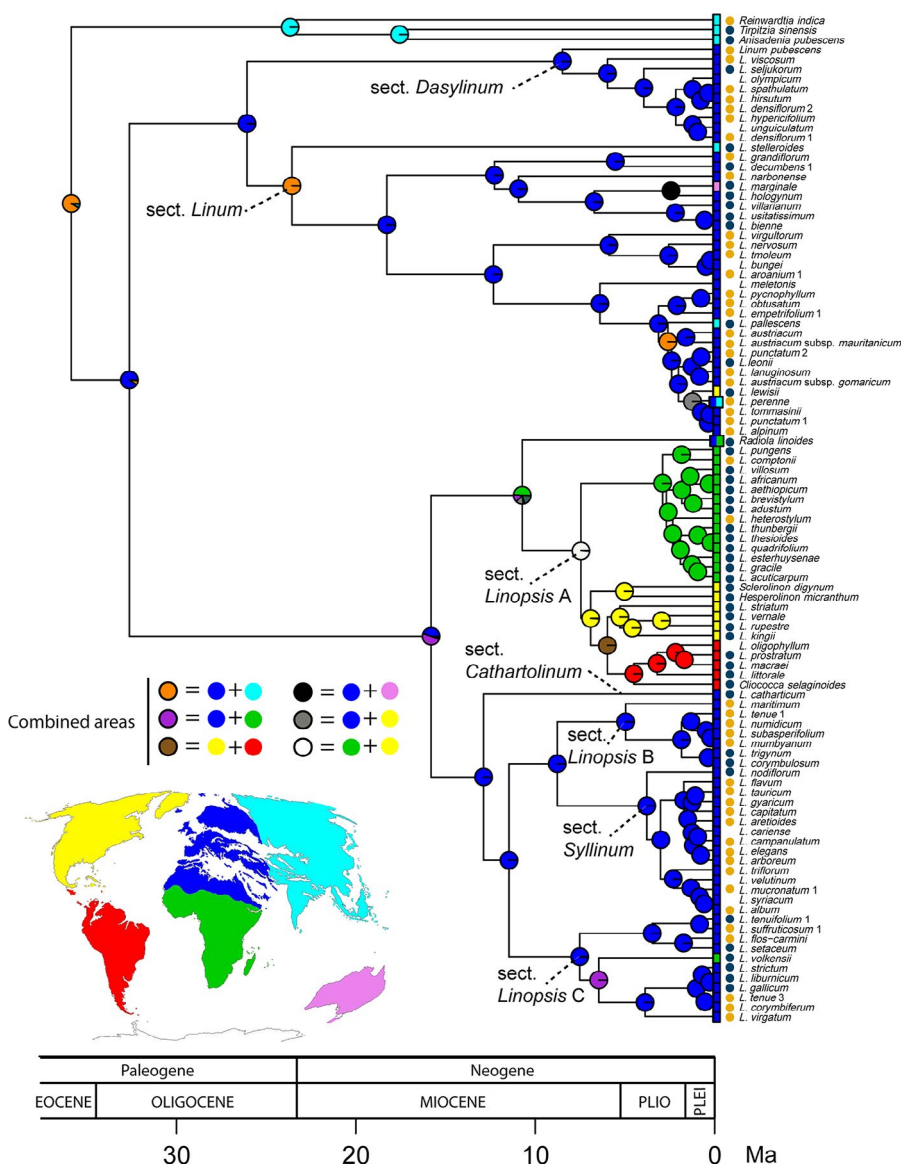
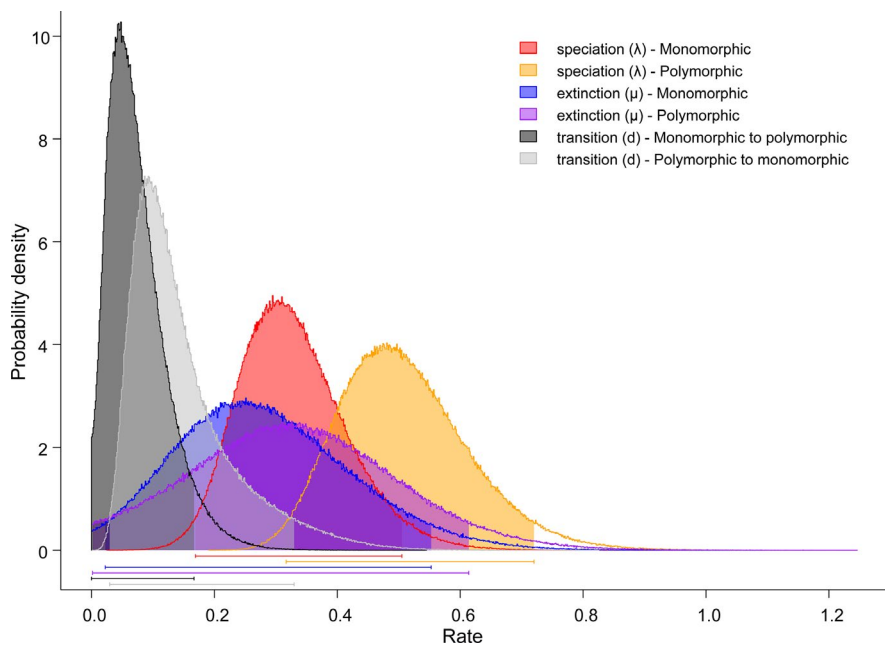


FIGURE 1 Divergence time estimation in *Linum* s.l., showing results of the ancestral area reconstruction in “BioGeoBEARS” (Matzke, 2013) under the DIVA-like model. Each node represents the probability of the most recent common ancestor (MRCA) to that clade occurring in each area or combination of areas as in the map and colour legend. Tip labels indicate the species names and in a coloured square, the area where the species occurs, with colours corresponding to the map from the Oligocene, and whether the species is monomorphic (dark blue) or polymorphic (yellow) in circles, when information was available. The scale along the bottom of the figure indicates the periods from the origin of the phylogeny to the present in millions of years ago (Ma)

	Western Palearctic	Eastern Palearctic	Afrotropic	Nearctic	Neotropic	Australasia	
Western Palearctic	-	3 (0)	2.16 (0.62)	1.22 (0.42)	0	1 (0)	7.38 (73.21%)
Eastern Palearctic	0	-	0	0	0	0	0
Afrotropic	0.66 (0.69)	0	-	0.78 (0.42)	0	0	1.44 (14.29%)
Nearctic	0	0	0.26 (0.49)	-	1 (0)	0	1.26 (12.5%)
Neotropic	0	0	0	0	-	0	0
Australasia	0	0	0	0	0	-	0
	0.66 (6.55%)	3 (29.76%)	2.42 (24.01%)	2 (19.84%)	1 (9.92%)	1 (9.92%)	10.08 (100%)

**FIGURE 2** Average number of dispersal events per source and sink after 50 biogeographic stochastic mappings (BSM), performed in “BioGeoBEARS” (Matzke, 2013) under the DIVA-like model, indicating standard deviation in parentheses. Warmer colours indicate higher event frequency. The marginal row (bottom) and column (right) indicate the sum and percentage of events in each area as the origin (row) or the destination (column) of dispersal events



**FIGURE 3** Posterior probability density plot of speciation ( $\lambda$ ), extinction ( $\mu$ ) and transition or dispersal rate to another area ( $d$ ), comparing monomorphic and polymorphic species of *Linum* s.l., using the MCC tree from BEAST analysis under the BiSSE model (see colour legend in the figure). X axis indicates rates for each analysed parameter in event/Myr, Y axis probability in % and lower bars in colours correspond to the 95% credibility intervals for each parameter

Afrotropic or, perhaps, only the Western Palearctic (Figure 1). Diversification of the genus *Linum* was inferred to be mostly due to within-area speciation events (a mean of  $92.8 \pm 0.45$  events predicted after 50 BSM runs; Figure 1). Vicariance events were less frequent ( $9.2 \pm 0.45$ ), which were the causes of diversification in the Afrotropic and clades from the Nearctic and Neotropic in *Linum* (Figure 1 and Figure S5 in Appendix S1). Finally, a mean of  $10.08 \pm 0.27$  anagenetic dispersal events was found. Most of them were from the Western Palearctic to other regions (73.21% of all dispersal events; Figures 1 and 2; but see Figures S5 and S6 in Appendix S1 for alternative models).

### 3.3 | Diversification analyses

Three possible scenarios of evolutionary regimes were found. The scenario with no changes in diversification rates was clearly preferred, with a probability ( $p$ ) of 0.80. The second-best model, with a  $p$  of 0.20,

considered one shift in diversification rates in two alternative locations: an instantaneous increase of diversification in the sister clade of *Linum* ( $p = 0.15$ ), or less probable, an instantaneous decrease of diversification rates in *Linum* with a  $p$  of 0.056. The model with the highest posterior probability had a speciation rate ( $\lambda_{m1}$ ) of 0.473 lineages per million years (lin./Myr), an extinction rate ( $\mu_1$ ) of 0.540 lin./Myr and a speciation growth parameter of 0.028 (Figure S7 in Appendix S1).

The BiSSE and GeoSSE models yielded similar results (Figure 3 and Figure S8 in Appendix S1). Comparing the AIC between constrained and unconstrained BiSSE and GeoSSE models did not show that any model was significantly more supported than another (Table S3 in Appendix S1). In the Bayesian approach for the BiSSE model, we found no differences in extinction or transition/dispersal rates, although there were slight differences in the speciation rate, which was higher in style polymorphic species (although the difference was not significant, Figure 3). Character reconstruction of heterostyly under the BiSSE model did not resolve the state for this character in older nodes but did for more recent nodes (Figure 4). Thus, we

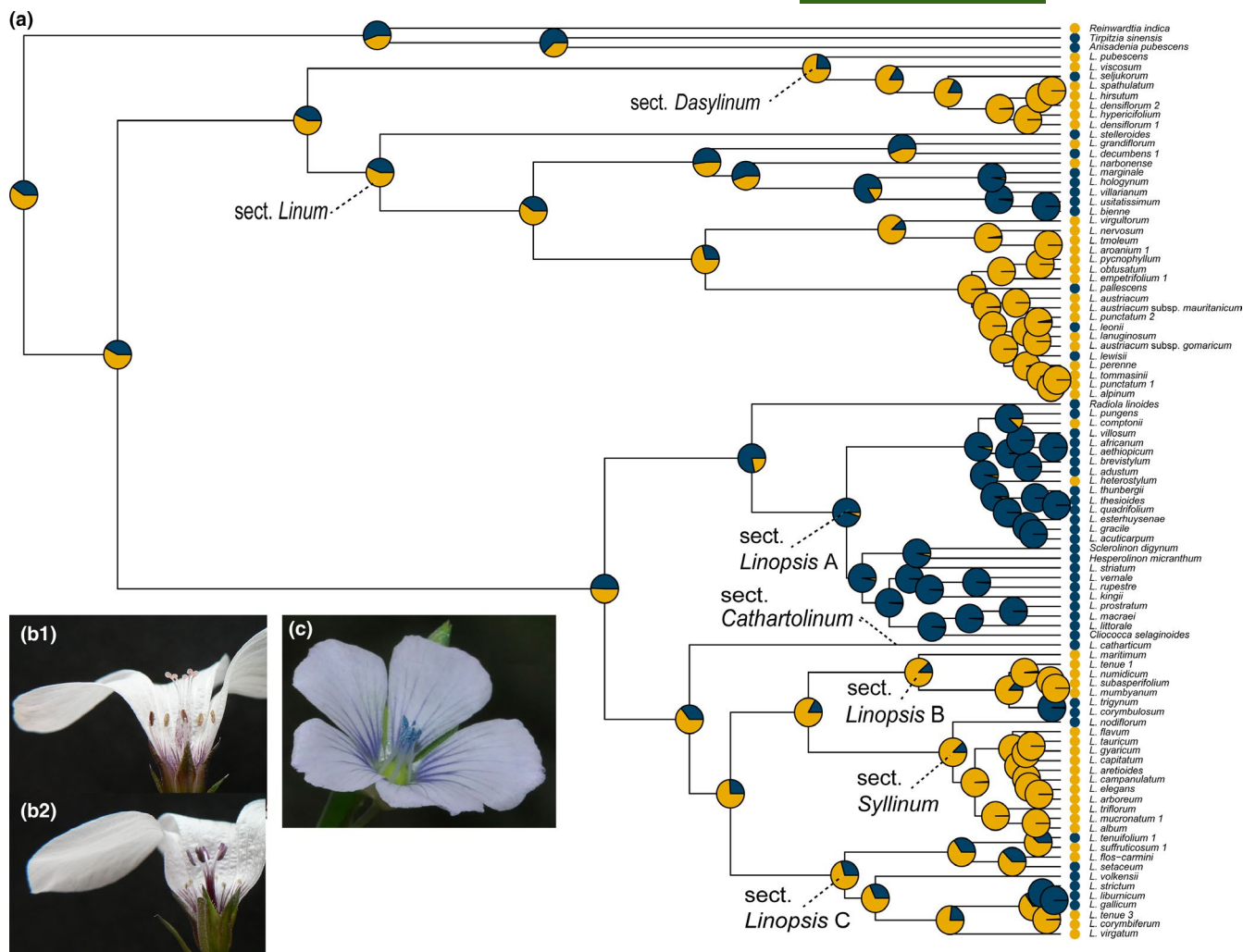


FIGURE 4 (a) Ancestral state reconstruction of heterostyly in *Linum* s.l. using the maximum clade credibility (MCC) tree from the BEAST analysis under the BISSE model. Colours indicate monomorphic (dark blue) and polymorphic (yellow) states. (b) Example of style polymorphic species *L. suffruticosum* with long-styled (b1) and short-styled (b2) flowers and (c) style monomorphic species *L. bienne*. Pictures by J. Arroyo (b) and M. Luceño (c). See Figure S1 in Appendix S1 for more examples of floral morphs

cannot confirm the state of the MRCA of the genus *Linum* s.l. nor the MRCA of the two main clades in the phylogeny. However, our results suggest that most likely state of the MRCA of most of the main or major nodes within *Linum* was polymorphic. The exception was the MRCA of the *Linum* section *Linopsis* "A" plus *R. linoides*, which contains species from the Afrotropic, the Nearctic and the Neotropic, which have been predicted to be monomorphic (Figure 4). The Bayesian approach for the GeoSSE model showed no differences in dispersal, extinction or speciation rates for source (Western Palearctic) versus sink (the rest) areas (Figure S8 in Appendix S1, but see also Table S3 in Appendix S1).

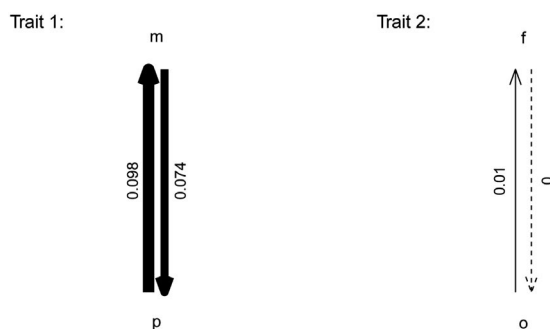
### 3.4 | Joint evolution of reproductive system, geographic distribution and life history

The best model explaining the evolution of the reproductive system (heterostyly vs. style monomorphism) and geographic distribution

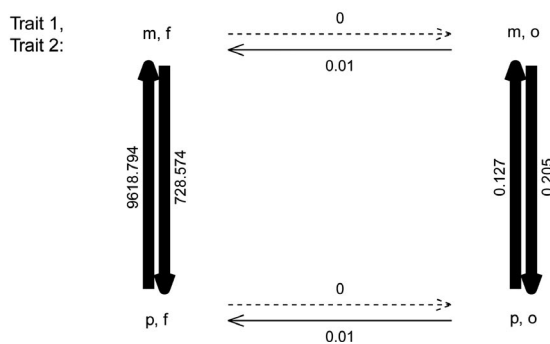
was the dependent model in which reproductive system depended on geographic distribution following a model for rates of transition where all rates were different (ARD). Specifically, the best model inferred only transitions from the Palearctic to the rest of the areas. The rate of transition from monomorphism to polymorphism was almost twice the rate from polymorphism to monomorphism in the Palearctic. In the rest of the areas, the rate of transition from polymorphism to monomorphism was more than 10 times higher than from monomorphism to polymorphism (Figure 5; see Figures S9 and S10 and Tables S4–S6 in Appendix S1 for more details). An extraordinarily high rate of monomorphism-polymorphism transition was inferred outside the Palearctic, especially from polymorphism to monomorphism. We associate this with the coincidence of transitions in style polymorphism between very closely related species (very short branches) outside the Palearctic. The best model to explain the evolution of annual versus perennial life history and geographic distribution was the independent model with equal rates of transition (ER; Tables S4–S6 in Appendix S1). The best model to



## (a) Independent model AIC = 142.4



## (b) Dependent (Heterostyly | Distribution) model AIC = 132.1



**FIGURE 5** Results from Pagel's binary correlation test of reproductive system (monomorphic versus polymorphic species) and area of origin (Western Palearctic vs. the rest of the areas) in *Linum* s.l. Rates are proportional to arrow thickness, with thicker arrows representing higher rates. m = monomorphic, p = polymorphic, o = Palearctic and f = rest of the areas

explain the evolution of reproductive system and life history was the dependent model in which reproductive system depended on life history with the ER model for rates of transition. Specifically, the best model inferred a single rate of transition (0.0926) for life history trait but two rates of transitions for reproductive system. The rate of transition in reproductive system was five times faster in annual (a rate of 0.3087) than in perennial (rate = 0.0599) species (AIC dependent model = 192.9526, AIC independent model = 193.8328), although this was only marginally supported ( $p = 0.0897$ ). Finally, we found that all lineages in the phylogeny that have dispersed out of the area of origin were monomorphic (Figure 1).

## 4 | DISCUSSION

### 4.1 | Origin and diversification of *Linum* s.l.

The best-fitting model from "BioGeoBEARS" (DIVA-like, AICw = 0.53; Table 1) placed the origin of the MRCA of the genus *Linum* s.l. in the Western Palearctic during the late Eocene to mid Oligocene (27.2–38.26 Ma at 95% HPD; Figure S2 in Appendix S1). When the MRCA of *Linum* diversified into two major lineages, one of these clades,

comprising sections *Dasylinum* and *Linum*, diversified in the early Oligocene to late Miocene (19.88–31.71 at 95% HPD; Figure S1 in Appendix S1) and remained in the Western Palearctic. The other major clade, constituted of sections *Cathartolinum*, *Linopsis* and *Syllinum*, originated during the Miocene (9.42–22.15 at 95% HPD; Figure S2 in Appendix S1), and we cannot discern whether the MRCA of this clade remained distributed in the Western Palearctic or spread into the Afrotropic (Figure 1). These dates are similar to those obtained by Ruiz-Martín et al. (2018), although dates in our results are generally somewhat more recent. They dated the crown node for the genus at 35.37 Ma (33.95–43.31 Ma at 95% HPD), and the crown nodes of each of the two major clades at 30.38 Ma (23.65–38.59 Ma) and 19.7 (11.48–29.49 Ma). We used almost the same sequence data as Ruiz-Martín et al. (2018), with the modifications of excluding a sample from genus *Hugonia* L. and keeping only one sample of each species except for nonmonophyletic species. However, two aspects of our methods that can explain the differences in our divergence time results, namely, (1) using two partitions of DNA matrices (nuclear and plastid) instead of the fully concatenated matrix with a single evolutionary model used in Ruiz-Martín et al. (2018) and (2) using a Birth-Death tree model instead of the Yule tree model used by Ruiz-Martín et al. (2018; see results). Moreover, dates for *Linum* s.l. and two major clades within the genus given by McDill et al. (2009) are even older than our results and those by Ruiz-Martín et al. (2018). For example, they dated the crown node for the genus between 41.63 and 46.24 Ma, falling in the mid-Eocene. Thus, after comparing different models of divergence time estimations and including data partitions for each DNA region, each with its own substitution model of evolution, our study gives more accurate divergence time estimations for *Linum* s.l. and major clades, which are younger in general than published in previous studies. This might be critical as the geomorphological and tectonic setting is different in each time slice, particularly in the Old World, which has more connected landmasses. In other words, erroneously dating the timing of divergence might have important implications in terms of biogeographical interpretations.

Currently, most species of the genus occur in the Western Palearctic. Some isolated taxa have dispersed and speciated in different areas, for example *L. stelleroides* Planch. and *L. pallescens* Bunge in the Eastern Palearctic, *L. volkensisii* Engl. in the Afrotropic, *L. lewisii* Pursh in the Nearctic and *L. marginale* A. Cunn. ex Planch. in Australasia (Figure 1). Nevertheless, these taxa did not diversify after colonisation, instead constituting long terminal branches in the phylogeny; there probably was a stasis in diversification after colonisation. The main exception to this stasis after colonisation is the species of the clade constituted by section *Linopsis* A, where the colonisation into new areas (Afrotropic, Nearctic and Neotropic) has also led to a diversification process, leading to monophyletic clades with various species in those areas (Figure 1). Most of the dispersal events (seven out of nine events) were from the Western Palearctic (approximately 75% of dispersal events; Figure 2): three to the Eastern Palearctic, two to the Afrotropic, one to the Nearctic and another to Australasia. Apart from these events, there was one dispersal event from the Afrotropic realm



and another from the Nearctic (Figure 2). This highlights the fact that the area of origin of the MRCA of *Linum* s.l. (Western Palearctic) was probably not the destination of any dispersal events (Figure 2).

Two dispersal events from the Western Palearctic to colonise the Afrotropic were predicted (Figure 2). The first, by the MRCA of the clade constituted by *Linopsis* A plus *R. linooides* (ca. 15.73 Ma; Figure 1 and Figure S2 in Appendix S1), coincides in time with a period when the Western Palearctic and the African platform (the Afrotropic) were connected, constituting a corridor for animal and plant lineages (Meulenkamp & Sissingh, 2003; Rosenbaum et al., 2002; Santos-Gally et al., 2012). The second event was by the MRCA of *L. volkensis* and the clade constituted by *L. strictum* L., *L. liburnicum* auct., *L. gallicum* L., *L. tenue* 3, *L. corymbiferum* Desf. and *L. virgatum* Schousboe (Figure 1). Divergence time analysis does not resolve the timing for this node (Figure S2 in Appendix S1), although the position within the phylogeny suggests that this dispersal to the Afrotropic took place after the establishment of the Sahara Desert (7.2–11.6 Ma; Zhang et al., 2014). This implies a long-distance dispersal (LDD) event crossing the Sahara. Similar LDD across the Sahara (once established) has occurred in other plant species, both from the north (e.g. *Arabis alpina*, Assefa et al., 2007; Koch et al., 2006) and from the south (e.g. *Erica arborea*, Désamoré et al., 2011; see Pokorný et al., 2015 for more biogeographic examples of this area). Other likely LDD include those to the Nearctic and Australasian realms (Figures 1 and 2), as those territories were already separated. In *Linum*, LDD has been described to be facilitated by the formation of mucilage around seeds after hydration, which allows them to adhere to animals for dispersal (Kreitschitz et al., 2015; Sorensen, 1986), as well as facilitate establishment (Yang et al., 2012). The crossing from the Nearctic to the Neotropic (North America to South America) could have occurred over the Isthmus of Panama, as this node was dated after the closure of the isthmus (Figure 1 and Figure S2 in Appendix S1; Bacon et al., 2015), a well-known migration process (Bacon et al., 2015; Marshall, 1985).

## 4.2 | Dispersification or key innovation? The role of heterostyly and dispersal in *Linum* evolution

The most important source of diversification in *Linum* s.l. is clearly within-area speciation, which is related to size of the areas we have used for our ancestral area reconstruction. Most species diversified in the Western Palearctic, which is the area of origin of the MRCA of the genus (Figure 1 and Figure S5 on Appendix S1). No events of subset within-area speciation were found, and vicariance and anagenetic dispersal constitute only isolated events in the evolution of *Linum* s.l. (Figure 1 and Figure S5 on Appendix S1). As mentioned above, most anagenetic dispersal events have not led to a diversification process in newly colonised areas (Figure 1). The exception was the clade constituted by the section *Linopsis* A. The MRCA of *Linum* section *Linopsis* dispersed south from the Western Palearctic to the Afrotropic; from there, the MRCA of *Linopsis* A dispersed to the Nearctic, then later from the Nearctic to the Neotropic (Figure 1).

Each of these colonisations, including the Afrotropic realm, was followed by a diversification processes (particularly in South Africa), although shifts in diversification rates did not couple with these biogeographic movements (Figure S8 in Appendix S1). Although diversification rates remained constant throughout the phylogeny (Figure S8 in Appendix S1), biogeographic changes due to dispersal into a new area have predated and led to diversification processes (appearance of new species on these areas) in section *Linopsis* A.

The term “dispersification” was defined by Moore and Donoghue (2007) as a shift in diversification rates caused by the colonisation of a given new area, rather than by the appearance of novel morphological traits, although it may be applied to any other phenotypic traits related to differentiation. This concept is based on the idea that new taxa can fill unexplored places with similar environmental conditions rapidly if colonising species are pre-adapted to those conditions (Donoghue, 2008). Dispersification implies (i) a change in the geographic range and (ii) a shift in diversification rates without the need to involve a phenotypic trait change (key innovation), assuming that species are pre-adapted to the new area, and can diversify rapidly (Moore & Donoghue, 2007). Heterostyly has been viewed as a mechanism that promotes outcrossing and thus could foster differentiation (de Vos, Hughes, et al., 2014), although the mechanism remains obscure (but see Haller et al., 2014) and is expected to depend strongly on the occurrence of specific pollinators which properly transfer pollen between morphs (Darwin, 1877; Lloyd & Webb, 1992), which could vary across regions. However, the only available study to date does not report an increase of speciation rates in polymorphic lineages compared to monomorphic ones, and only reduced extinction rate in the former was reported (de Vos, Hughes, et al., 2014). In the current study, no significant differences in diversification rates were found between monomorphic and polymorphic lineages or source versus sink geographic ranges (in relation to diversification, speciation, extinction and dispersal/transition events; Figure 3), which reject the hypotheses of dispersification and heterostyly as a key innovation.

It is remarkable that all lineages in our phylogeny (Figure 1) that have dispersed are monomorphic (Figure 4). Nevertheless, although they are minority, the *Linopsis* A clade in southern Africa includes two heterostylous species (*L. comptoni* and *L. heterostylum*; Rogers, 1981; Ruiz-Martín et al., 2018), which remains a noteworthy exception to the rule in the clade. In fact, we have inferred a strongly significant dependent model in which reproductive system depends on geographical distribution; in the Palearctic, species evolve towards a polymorphic reproductive system whereas in dispersed areas species evolve towards a monomorphic reproductive system. This might reflect two patterns that are indirectly related to heterostyly. First, monomorphism is evolutionarily associated with annual life cycle in *Linum* (Ruiz-Martín et al., 2018), and this life form (and in general short-lived plants) is more prone to colonisation, both on ecological (Baker, 1974) and evolutionary (Lavergne et al., 2013) time scales. Although we have inferred a marginally significant model in which reproductive system depends on life history trait, we have inferred an independent model for life history and geographical distribution (Tables S4–S6 in Appendix S1), which discards life history as

having a crucial role in the dependence of reproductive system on geographic distribution. Second, as in many other heterostylous groups, all tested polymorphic species in *Linum* show heteromorphic self-incompatibility, whereas self-compatible species are all monomorphic, although the species sample is very limited, particularly for American and African species (Ruiz-Martín et al., 2018). It would be very interesting to know if selfing is the dominant mating system amongst colonisers, as Baker's law predicts (Baker, 1955; Stebbins, 1957). At least, our results are consistent with this prediction if an increased selfing rate is expected in monomorphic species in comparison with obligate self-incompatible heterostylous species. Although not explored in our study, there is also wide variation in herkogamy of monomorphic species (Ruiz-Martín, 2017), from non-herkogamy (monomorphic homostyly) to several degrees of approach and reverse herkogamy (see description in Material and Methods section). This might have profound effects on the ability to self-pollinate amongst species that are self-compatible (Webb & Lloyd, 1986) and thus on colonisation capacity, at least with regards to pollination environment (Busch, 2011). This has been shown in *Primula*, where high selfing rates were found for homostylous and thus non-herkogamous populations (Zhong et al., 2019). There is a need to explore breeding systems and pollination biology on selected species in these colonising lineages to determine a closer fit to Baker's law. One remaining question is why five of these monomorphic lineages found only outside the area of origin (i.e. *L. lewisii*, *L. marginale*, *L. pallescens*, *L. stelleroides* and *L. volkensisii*) had stasis in diversification after colonisation but three of them (three main clades within sect. *Linopsis* A; Figure 1) diversified after colonisation.

What is clear is that heterostyly is mostly confined to the ancestral area, the Western Palearctic (particularly in the Mediterranean Basin), and is possible that the genetic system coding heterostyly was already present in the ancestor of *Linum*, as this breeding system is present in other Linaceae (see also Figure 4 for the ancestral state reconstruction of heterostyly in *Linum*). Diversification of current Mediterranean *Linum* could be fostered by a dynamic paleogeography during the Oligocene, when land fragmentation was the rule, compared with the rest of the world, parallel to what occurred in many other Mediterranean groups (e.g. Barres et al., 2013; Manafzadeh et al., 2014). In this sense, the date of dispersal could be a factor for diversification. Species or lineages which diversified more recently, might not have had enough time to further diversify (or available niche could have been already mostly occupied by species that colonised before). Nevertheless, although time could influence diversification, this possible explanation is not applicable in some cases of ancient dispersals, such as *L. stelleroides* or *L. volkensisii*. Whatever the causes of this diversification, it remains a mystery why two species of the coloniser Afrotropical clade (within *Linopsis* A; Figure 1) became polymorphic at the southern tip of the continent, within a clade that is otherwise monomorphic (Figure 4), in a period (Pliocene, Figure 1) when a truly Mediterranean-type climate was already fully established in the region (Rundel et al., 2016). Other transitions, like monomorphism acquired from style polymorphism,

are frequent in *Linopsis* B and *Linopsis* C in the western Palearctic, although derived monomorphism (particularly non-herkogamous homostyly) is a frequently reported case in heterostyly research (e.g. *Primula*; de Vos, Wüest, et al., 2014; Zhou et al., 2017).

## 5 | FINAL REMARKS

The origin of the MRCA of the genus *Linum* s.l. was clearly in the Western Palearctic from the late Eocene to mid-Oligocene. All studied species (or their MRCA) that dispersed outside the Western Palearctic were monomorphic, perhaps indicating the importance of self-pollination for their establishment in new places. The high incidence of heterostyly in the Mediterranean Basin (within the Western Palearctic) is apparently not related with the high regional diversification there but with the paleogeographic setting of the territory.

## ACKNOWLEDGEMENTS

We thank two reviewers and the chief and associate editors for comments and suggestions that significantly improved the original version of the manuscript, especially Dr. G. Schneeweiss for his strong interest and in depth revision of the manuscript. We thank Dr. C. Scotese for sending and allowing us to publish a modified version of the Oligocene world map as part of Figure 1 as well as Dr. M. Luceño and R. Turner for sharing pictures for their publication in this article. We thank CICA (Scientific Andalusian Informatics Center) for providing computational resources. No collecting permits were needed for this study. This work was funded through research grants from the Spanish Government and FEDER (EU) funds (CGL2013-45037-P and PGC2018-099608-B-100) granted to JA, ME and EM, a "Juan de la Cierva-Formación" fellowship to EM (Spanish Government; FJCI-2017-32314) and a grant FPI fellowship to JRM (BES-2008-003946).

## DATA AVAILABILITY STATEMENT

All sequence data were downloaded from NCBI GenBank database. Accession numbers can be found in Table S1 in Appendix S1.

## ORCID

Enrique Maguilla  <https://orcid.org/0000-0002-8245-3024>

## REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Armbruster, W. S., Pérez-Barrales, R., Arroyo, J., Edwards, M. E., & Vargas, P. (2006). Three-dimensional reciprocity of floral morphs in wild flax (*Linum suffruticosum*): A new twist on heterostyly. *New Phytologist*, *171*, 581–590. <https://doi.org/10.1111/j.1469-8137.2006.01749.x>
- Assefa, A., Ehrich, D., Taberlet, P., Nemomissa, S., & Brochmann, C. (2007). Pleistocene colonization of afro-alpine 'sky islands' by the arctic-alpine *Arabis alpina*. *Heredity*, *99*, 133–142. <https://doi.org/10.1038/sj.hdy.6800974>
- Avise, J. C. (2000). *Phylogeography: The history and formation of species*. Harvard University Press.



- Bacon, C. D., Silvestro, D., Jaramillo, C., Smith, B. T., Chakrabarty, P., & Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 6110–6115. <https://doi.org/10.1073/pnas.1423853112>
- Baker, H. G. (1955). Self-compatibility and establishment after "long-distance" dispersal. *Evolution*, 9, 347–348. <https://doi.org/10.2307/2405656>
- Baker, H. G. (1974). The evolution of weeds. *Annual Review of Ecology and Systematics*, 5, 1–24. <https://doi.org/10.1146/annurev.es.05.110174.000245>
- Barres, L., Sanmartín, I., Anderson, C. L., Susanna, A., Buerki, S., Galbany-Casals, M., & Vilatersana, R. (2013). Reconstructing the evolution and biogeographic history of tribe *Cardueae* (compositae). *American Journal of Botany*, 100, 867–882. <https://doi.org/10.3732/ajb.1200058>
- Barrett, S. C. H. (1992). *Evolution and function of heterostyly*. Springer.
- Barrett, S. C. H. (2019). 'A most complex marriage arrangement': Recent advances on heterostyly and unresolved questions. *New Phytologist*, 224, 1051–1067. <https://doi.org/10.1111/nph.16026>
- Barrett, S. C. H., & Shore, J. S. (2008). New insights on heterostyly: Comparative biology, ecology and genetics. In V. E. Franklin-Tong (Ed.), *Self-incompatibility in flowering plants* (pp. 3–32). Springer.
- Bell, C. D., Soltis, D. E., & Soltis, P. S. (2010). The age and diversification of the angiosperms re-visited. *American Journal of Botany*, 97, 1296–1303. <https://doi.org/10.3732/ajb.0900346>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A software platform for bayesian evolutionary analysis. *PLoS Computational Biology*, 10, 1–6.
- Busch, J. W. (2011). Demography, pollination, and Baker's law. *Evolution*, 65, 1511–1513. <https://doi.org/10.1111/j.1558-5646.2011.01224.x>
- Cavagnetto, C., & Anadón, P. (1996). Preliminary palynological data on floristic and climatic changes during the Middle Eocene–Early Oligocene of the eastern Ebro Basin, northeast Spain. *Review of Paleobotany and Palynology*, 92, 281–305. [https://doi.org/10.1016/0034-6667\(95\)00096-8](https://doi.org/10.1016/0034-6667(95)00096-8)
- Charlesworth, B., & Charlesworth, D. (1979). The maintenance and breakdown of distyly. *The American Naturalist*, 114, 499–513. <https://doi.org/10.1086/283497>
- Cheptou, P. O. (2012). Clarifying Baker's law. *Annals of Botany*, 109, 633–641. <https://doi.org/10.1093/aob/mcr127>
- Crawford, D. J., Anderson, G. J., Bernardello, G., Bramwell, D., & Caujape-castells, J. (2011). The reproductive biology of island plants. In D. Bramwell & J. Caujapé-Castells (Eds.), *The biology of Island Floras* (pp. 11–36). Cambridge University Press.
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9, 772. <https://doi.org/10.1038/nmeth.2109>
- Darwin, C. (1877). *The different forms of flowers on plants of the same species*. Cambridge University Press.
- Davies, T. J., Barraclough, T. G., Chase, M. W., Soltis, P. S., Soltis, D. E., & Savolainen, V. (2004). Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 1904–1909. <https://doi.org/10.1073/pnas.0308127100>
- de Vos, J. M., Hughes, C. E., Schneeweiss, G. M., Moore, B. R., & Conti, E. (2014). Heterostyly accelerates diversification via reduced extinction in primroses. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20140075. <https://doi.org/10.1098/rspb.2014.0075>
- de Vos, J. M., Wüest, R. O., & Conti, E. (2014). Small and ugly? Phylogenetic analyses of the "selfing syndrome" reveal complex evolutionary fates of monomorphic primrose flowers. *Evolution*, 68, 1042–1057. <https://doi.org/10.1111/evo.12331>
- Désamoré, A., Laenen, B., Devos, N., Popp, M., González-Mancebo, J. M., Carine, M. A., & Vanderpoorten, A. (2011). Out of Africa: North-westwards Pleistocene expansions of the heather *Erica arborea*. *Journal of Biogeography*, 38, 164–176. <https://doi.org/10.1111/j.1365-2699.2010.02387.x>
- Dodd, M. E., Silvertown, J., & Chase, M. W. (1999). Phylogenetic analysis of trait evolution and species diversity variation among angiosperm families. *Evolution*, 53, 732–744. <https://doi.org/10.1111/j.1558-5646.1999.tb05367.x>
- Donoghue, M. J. (2008). Colloquium paper: A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(Suppl), 11549–11555. <https://doi.org/10.1073/pnas.0801962105>
- Dressler, S., Reppinger, M., & Bayer, C. (2014). Linaceae. In K. Kubitzki (Ed.), *Flowering plants. Eudicots: Malpighiales* (pp. 237–246). Springer.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4, 699–710. <https://doi.org/10.1371/journal.pbio.0040088>
- Dupin, J., Matzke, N. J., Särkinen, T., Knapp, S., Olmstead, R. G., Bohs, L., & Smith, S. D. (2017). Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography*, 44, 887–899. <https://doi.org/10.1111/jbi.12898>
- Feliner, G. N. (2014). Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics*, 16, 265–278. <https://doi.org/10.1016/j.ppees.2014.07.002>
- Ferrero, V., Rojas, D., Vale, A., & Navarro, L. (2012). Delving into the loss of heterostyly in Rubiaceae: Is there a similar trend in tropical and non-tropical climate zones? *Perspectives in Plant Ecology, Evolution and Systematics*, 14, 161–167. <https://doi.org/10.1016/j.ppees.2011.11.005>
- Fitzjohn, R. G. (2012). Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, 3, 1084–1092. <https://doi.org/10.1111/j.2041-210X.2012.00234.x>
- Goldberg, E. E., Lancaster, L. T., & Ree, R. H. (2011). Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, 60, 451–465. <https://doi.org/10.1093/sysbio/syr046>
- Guggisberg, A., Mansion, G., Kelso, S., & Conti, E. (2006). Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid–polyploid species complex of *Primula*. *New Phytologist*, 171, 617–632. <https://doi.org/10.1111/j.1469-8137.2006.01722.x>
- Haller, B. C., de Vos, J. M., Keller, B., Hendry, A. P., & Conti, E. (2014). A tale of two morphs: Modeling pollen transfer, magic traits, and reproductive isolation in parapatry. *PLoS One*, 9, e106512. <https://doi.org/10.1371/journal.pone.0106512>
- Ilgic, B., & Busch, J. W. (2013). Is self-fertilization an evolutionary dead end? *New Phytologist*, 198, 386–397. <https://doi.org/10.1111/nph.12182>
- Koch, M. A., Kiefer, C., Ehrich, D., Vogel, J., Brochmann, C., & Mummehoff, K. (2006). Three times out of Asia Minor: The phylogeography of *Arabis alpina* L. (Brassicaceae). *Molecular Ecology*, 15, 825–839. <https://doi.org/10.1111/j.1365-294X.2005.02848.x>
- Kreitschitz, A., Kovalev, A., & Gorb, S. N. (2015). Slipping vs sticking: Water-dependent adhesive and frictional properties of *Linum usitatissimum* L. seed mucilaginous envelope and its biological significance. *Acta Biomaterialia*, 17, 152–159. <https://doi.org/10.1016/j.actbio.2015.01.042>
- Landis, M., Matzke, N., Mooer, B., & Huelsenbeck, J. (2013). Bayesian analysis of biogeography when the number of area is large. *Systematic Biology*, 62, 789–804. <https://doi.org/10.1093/sysbio/syt040>
- Lavergne, S., Hampe, A., & Arroyo, J. (2013). In and out of Africa: How did the Strait of Gibraltar affect plant species migration and local diversification? *Journal of Biogeography*, 40, 24–36. <https://doi.org/10.1111/j.1365-2699.2012.02769.x>
- Lloyd, D. G., & Webb, C. J. (1992). The evolution of heterostyly. In S. C. H. Barrett (Ed.), *Evolution and function of heterostyly* (pp. 151–178). Springer.



- Louca, S., & Pennell, M. W. (2020). Extant timetrees are consistent with a myriad of diversification histories. *Nature*, *580*, 502–505. <https://doi.org/10.1038/s41586-020-2176-1>
- Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, *56*, 701–710. <https://doi.org/10.1080/10635150701607033>
- Maguilla, E., Escudero, M., Hipp, A. L., & Luceño, M. (2017). Allopatric speciation despite historical gene flow: Divergence and hybridization in *Carex furva* and *C. lucenoiiberica* (Cyperaceae) inferred from plastid and nuclear RAD-seq data. *Molecular Ecology*, *26*, 5646–5662. <https://doi.org/10.1111/mec.14253>
- Manafzadeh, S., Salvo, G., & Conti, E. (2014). A tale of migrations from east to west: The Irano-Turanian floristic region as a source of Mediterranean xerophytes. *Journal of Biogeography*, *41*, 366–379. <https://doi.org/10.1111/jbi.12185>
- Marshall, L. (1985). *The Great American Biotic Interchange* (F. Stehli & S. Webb Ed., pp. 49–85). Plenum Press.
- 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. <https://doi.org/10.21425/F55419694>
- Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, *63*, 951–970. <https://doi.org/10.1093/sysbio/syu056>
- McDill, J., Repplinger, M., Simpson, B. B., & Kadereit, J. W. (2009). The Phylogeny of *Linum* and Linaceae subfamily Linoideae, with implications for their systematics, biogeography, and evolution of heterostyly. *Systematic Botany*, *34*, 386–405. <https://doi.org/10.1600/036364409788606244>
- McDill, J. R., & Simpson, B. (2011). Molecular phylogeny of Linaceae with complete generic sampling and data from two chloroplast genes. *Botanical Journal of the Linnean Society*, *165*, 64–83. <https://doi.org/10.1111/j.1095-8339.2010.01096.x>
- Meulenkaamp, J. E., & Sissingh, W. (2003). Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the African-Eurasian convergent plate boundary zone. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *196*, 209–228. [https://doi.org/10.1016/S0031-0182\(03\)00319-5](https://doi.org/10.1016/S0031-0182(03)00319-5)
- Moeller, D. A., Briscoe Runquist, R. D., Moe, A. M., Geber, M. A., Goodwillie, C., Cheptou, P.-O., Eckert, C. G., Elle, E., Johnston, M. O., Kalisz, S., Ree, R. H., Sargent, R. D., Vallejo-Marin, M., & Winn, A. A. (2017). Global biogeography of mating system variation in seed plants. *Ecology Letters*, *20*, 375–384. <https://doi.org/10.1111/ele.12738>
- Moore, B. R., & Donoghue, M. J. (2007). Correlates of diversification in the plant clade Dipsacales: Geographic movement and evolutionary innovations. *The American Naturalist*, *170*, S28–S55. <https://doi.org/10.1086/519460>
- Morlon, H., Hartig, F., & Robins, S. (2020). Prior hypotheses or regularization allow inference of diversification histories from extant timetrees. *bioRxiv*, 2020.07.03.185074.
- Myers, N., Mittermeier, R., Mittermeier, C., da Fonseca, G., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*, 853–858. <https://doi.org/10.1038/35002501>
- Ockendon, D. J., & Walters, S. M. (1968). *Linum* L. In T. G. Tutin, V. H. Heywood, N. A. Burges, D. M. Moore, D. H. Valentine, S. M. Walters, & D. A. Webb (Eds.), *Flora Europaea* (Vol. 2, pp. 206–211). Cambridge University Press.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., Amico, J. A. D., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wesley, W., Hedao, P., & Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience*, *51*, 933–938. [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
- Pagel, M. (1994). Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society: London B*, *255*, 37–45.
- Pannell, J. R., Auld, J. R., Brandvain, Y., Burd, M., Busch, J. W., Cheptou, P.-O., Conner, J. K., Goldberg, E. E., Grant, A.-G., Grossenbacher, D. L., Hovick, S. M., Igic, B., Kalisz, S., Petanidou, T., Randle, A. M., de Casas, R. R., Pauw, A., Vamosi, J. C., & Winn, A. A. (2015). The scope of Baker's law. *New Phytologist*, *208*, 656–667. <https://doi.org/10.1111/nph.13539>
- Peterson, A., Harpke, D., Peterson, J., Harpke, A., & Peruzzi, L. (2019). A pre-Miocene Irano-Turanian cradle: Origin and diversification of the species-rich monocot genus *Gagea* (Liliaceae). *Ecology and Evolution*, *9*, 5870–5890. <https://doi.org/10.1002/ece3.5170>
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*, 7–11.
- Pokorný, L., Riina, R., Mairal, M., Meseguer, A. S., Culshaw, V., Cendoya, J., Serrano, M., Carbajal, R., Ortiz, S., Heuertz, M., & Sanmartín, I. (2015). Living on the edge: Timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. *Frontiers in Genetics*, *6*, 154. <https://doi.org/10.3389/fgene.2015.00154>
- R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rabosky, D. L., Donnellan, S. C., Grundler, M., & Lovette, I. J. (2014). Analysis and visualization of complex macroevolutionary dynamics: An example from Australian Scincid lizards. *Systematic Biology*, *63*, 610–627. <https://doi.org/10.1093/sysbio/syu025>
- Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., Huang, H., & Larson, J. G. (2014). BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, *5*, 701–707. <https://doi.org/10.1111/2041-210X.12199>
- Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*, *4*, 1–8. <https://doi.org/10.1038/ncomm32958>
- Rambaut, A., & Drummond, A. J. (2013). Tracer v1.6. <http://beast.bio.ed.ac.uk/>
- Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography*, *45*, 741–749. <https://doi.org/10.1111/jbi.13173>
- Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, *57*, 4–14. <https://doi.org/10.1080/10635150701883881>
- Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*, 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Rodríguez-Sánchez, F., Pérez-Barrales, R., Ojeda, F., Vargas, P., & Arroyo, J. (2008). The Strait of Gibraltar as a melting pot for plant biodiversity. *Quaternary Science Reviews*, *27*, 2100–2117. <https://doi.org/10.1016/j.quascirev.2008.08.006>
- Rogers, C. M. (1981). A revision of the genus *Linum* in southern Africa. *Nordic Journal of Botany*, *1*, 711–722. <https://doi.org/10.1111/j.1756-1051.1981.tb01157.x>
- Ronquist, F. (1997). Dispersal-Vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, *46*, 195–203. <https://doi.org/10.1093/sysbio/46.1.195>
- Ronquist, F., & Sanmartín, I. (2011). Phylogenetic methods in biogeography. *Annual Review of Ecology, Evolution, and Systematics*, *42*, 441–464. <https://doi.org/10.1146/annurev-ecolsys-102209-144710>





- Rosenbaum, G., Lister, G. S., & Duboz, C. (2002). Reconstruction of the tectonic evolution of the western Mediterranean since the Oligocene. *Journal of the Virtual Explorer*, 8, 107–130. <https://doi.org/10.3809/jvirtex.2002.00053>
- Ruiz-Martín, J. (2017). *Evolución y función de la heterostilia en el género Linum (Linaceae)*. PhD thesis, Universidad de Sevilla, Sevilla, Spain.
- Ruiz-Martín, J., Santos-Gally, R., Escudero, M., Midgley, J. J., Pérez-Barrales, R., & Arroyo, J. (2018). Style polymorphism in *Linum* (Linaceae): A case of Mediterranean parallel evolution? *Plant Biology*, 20, 100–111. <https://doi.org/10.1111/plb.12670>
- Rundel, P. W., Arroyo, M. T., Cowling, R. M., Keeley, J. E., Lamont, B. B., & Vargas, P. (2016). Mediterranean biomes: Evolution of their vegetation, floras, and climate. *Annual Review of Ecology, Evolution, and Systematics*, 47, 383–407. <https://doi.org/10.1146/annurev-ecolsys-121415-032330>
- Santos-Gally, R., Vargas, P., & Arroyo, J. (2012). Insights into Neogene Mediterranean biogeography based on phylogenetic relationships of mountain and lowland lineages of *Narcissus* (Amaryllidaceae). *Journal of Biogeography*, 39, 782–798. <https://doi.org/10.1111/j.1365-2699.2011.02526.x>
- Schneeweiss, G. M., Winkler, M., & Schönswetter, P. (2017). Secondary contact after divergence in allopatry explains current lack of ecogeographical isolation in two hybridizing alpine plant species. *Journal of Biogeography*, 44, 2575–2584. <https://doi.org/10.1111/jbi.13071>
- Schneider, A. C., Freyman, W. A., Guilliams, C. M., Springer, Y. P., & Baldwin, B. G. (2016). Pleistocene radiation of the serpentine-adapted genus *Hesperolinon* and other divergence times in Linaceae (Malpighiales). *American Journal of Botany*, 103, 221–232. <https://doi.org/10.3732/ajb.1500379>
- Scotese, C. R. (2014). *Atlas of Paleogene paleogeographic maps (Mollweide projection), Maps 8–15* (Vol. 1). The Cenozoic, PALEOMAP Atlas for ArcGIS, PALEOMAP Project.
- Shi, J. J., & Rabosky, D. L. (2015). Speciation dynamics during the global radiation of extant bats. *Evolution*, 69, 1528–1545. <https://doi.org/10.1111/evo.12681>
- Sorensen, A. E. (1986). Seed dispersal by adhesion. *Annual Review of Ecology and Systematics*, 17, 443–463. <https://doi.org/10.1146/annurev.es.17.110186.002303>
- Stebbins, G. L. (1957). Self fertilization and population variability in the higher plants. *The American Naturalist*, 91, 337–354. <https://doi.org/10.1086/281999>
- Thompson, J. (2020). *A classification of the biogeographical provinces of the world* (p. 48). IUCN occasional paper. International Union for Conservation of Nature and Natural Resources.
- Udvardy, M. D. F. (1975). *A classification of the biogeographical provinces of the world* (p. 48). IUCN occasional paper. International Union for Conservation of Nature and Natural Resources.
- Vamosi, J. C., Magallón, S., Mayrose, I., Otto, S. P., & Sauquet, H. (2018). Macroevolutionary patterns of flowering plant speciation and extinction. *Annual Review of Plant Biology*, 69, 685–706. <https://doi.org/10.1146/annurev-arplant-042817-040348>
- Vamosi, J. C., & Vamosi, S. M. (2011). Factors influencing diversification in angiosperms: At the crossroads of intrinsic and extrinsic traits. *American Journal of Botany*, 98, 460–471. <https://doi.org/10.3732/ajb.1000311>
- Watanabe, K., & Sugawara, T. (2015). Is heterostyly rare on oceanic islands? *AoB PLANTS*, 7, plv087. <https://doi.org/10.1093/aobpla/plv087>
- Webb, C. J., & Lloyd, D. G. (1986). The avoidance of interference between the presentation of pollen and stigmas in angiosperms II. Herkogamy. *New Zealand Journal of Botany*, 24, 163–178. <https://doi.org/10.1080/0028825X.1986.10409726>
- Yang, X., Baskin, J. M., Baskin, C. C., & Huang, Z. (2012). More than just a coating: Ecological importance, taxonomic occurrence and phylogenetic relationships of seed coat mucilage. *Perspectives in Plant Ecology, Evolution and Systematics*, 14, 434–442. <https://doi.org/10.1016/j.ppees.2012.09.002>
- Zhang, Z., Ramstein, G., Schuster, M., Li, C., Contoux, C., & Yan, Q. (2014). Aridification of the Sahara desert caused by Tethys Sea shrinkage during the Late Miocene. *Nature*, 513, 401–404. <https://doi.org/10.1038/nature13705>
- Zhong, L., Barrett, S. C. H., Wang, X. J., Wu, Z. K., Sun, H. Y., Li, D. Z., Wang, H., & Zhou, W. (2019). Phylogenomic analysis reveals multiple evolutionary origins of selfing from outcrossing in a lineage of heterostylous plants. *New Phytologist*, 224, 1290–1303. <https://doi.org/10.1111/nph.15905>
- Zhou, W., Barrett, S. C., Li, H. D., Wu, Z. K., Wang, X. J., Wang, H., & Li, D. Z. (2017). Phylogeographic insights on the evolutionary breakdown of heterostyly. *New Phytologist*, 214, 1368–1380. <https://doi.org/10.1111/nph.14453>

**BIOSKETCH**

**Enrique Maguilla** is a postdoctoral researcher at the Universidad de Sevilla (Spain). His research is focused on the evolution, phylogenetics and biogeography of Angiosperms, with special emphasis on the cosmopolitan genus *Carex* (Cyperaceae) and some Mediterranean taxa.

Author contributions: Juan Arroyo and Marcial Escudero conceived the idea; Enrique Maguilla and Marcial Escudero analyzed the data; Enrique Maguilla drafted the manuscript; all authors revised and contributed to the preparation of the final version of the manuscript.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Maguilla E, Escudero M, Ruiz-Martín J, Arroyo J. Origin and diversification of flax and their relationship with heterostyly across the range. *J Biogeogr.* 2021;48:1994–2007. <https://doi.org/10.1111/jbi.14129>