RANGE-WIDE POPULATION GENETICS OF *LINUM SUFFRUTICOSUM* S.L., AN HETEROSTYLOUS POLYPLOID AND TAXONOMIC COMPLEX

Maria Antònia Vanrell Bosch

Tutores: Juan Arroyo Marín y Violeta Simón-Porcar

Máster Universitario en Biología Avanzada
Facultad de Biología
Noviembre de 2022

UNIVERSIDAD D SEVILLA

Abstract

Linum suffruticosum L. s.l. (Linaceae) is a taxonomic complex widespread in the western Mediterranean basin. It is characterized by its high cytogenetic diversity and its 3D heterostyly system, which makes it an obligate outcrosser. In this work we studied the patterns of genetic diversity and population structure throughout its distribution using neutral genetic data obtained from microsatellite markers. It is shown that there is no relationship between taxonomic treatment and genetic variability, so a revision of the taxonomy would be useful. *L. suffruticosum* s.l. has a high level of genetic diversity, promoted by polyploidization processes, and maintained by the existence of abundant and reproductively connected populations. The populations are genetically structured following a pattern of isolation by distance, with the Rif mountains as the most important genetic barrier. However, genetic diversity does not covary at the latitudinal or altitudinal level, nor is it related to population size or the proportion of stylar morphs. This suggests that variations in genetic diversity are not mostly determined by selective forces, but by stochastic processes.

Resumen

Linum suffruticosum L. s.l. (Linaceae) es un complejo taxonómico ampliamente extendido en la cuenca mediterránea occidental. Se caracteriza por su alta diversidad citogenética y su sistema de heterostília en 3D, lo que lo convierte en un outcrosser obligado. En este trabajo se estudian los patrones de diversidad genética y estructura poblacional a lo largo de su distribución utilizando datos genéticos neutros obtenidos a partir de marcadores de microsatélites. Se ha demostrado que no existe relación entre el tratamiento taxonómico y la variabilidad genética, por lo que sería útil una revisión de la taxonomía. *L. suffruticosum* s.l. presenta una elevada diversidad genética, promovida por procesos de poliploidización, y mantenida por la existencia de poblaciones abundantes y conectadas reproductivamente. Las poblaciones están estructuradas genéticamente siguiendo un patrón de aislamiento por distancia, con las montañas del Rif como la barrera genética más importante. Sin embargo, la diversidad genética no covaría con la latitud o la altitud, ni está relacionada con el tamaño de la población o la proporción de morfos estilares. Esto sugiere que las variaciones en la diversidad genética no están determinadas principalmente por fuerzas selectivas, sino por procesos estocásticos.

Introduction

The Mediterranean basin is a biodiversity hotspot with an exceptional concentration of plant species (Heywood, 1995; Médail & Quezel, 1997). There are approximately 25 000 vascular plant taxa, thus representing 4.3% of the total plant species on the planet (Myers et al., 2000). Remarkably, about half of them (about 13 000 taxa) are endemic to the area. This exceptional species richness and high percentage of endemic taxa has been favored by the exceptional habitat heterogeneity, the climatic history strongly influenced by Quaternary glaciations, and the intense habitat fragmentation in this area (Greuter, 2001). Many widely distributed plant species present high variation in the morphological, demographic, and biological traits of populations, which are determined by geographic, ecological, and historical factors (López-Goldar et al., 2020). This diversity and the underlying microevolutionary processes are reflected in the patterns of genetic diversity and structure of the populations and are key for the destiny of the species.

Linum suffruticosum L. s.l. (Linaceae) is a taxonomic complex of perennial woody plants with distinctive heterostylous white flowers widely distributed throughout the western Mediterranean basin (Martínez-Labarga & Muñoz -Garmendia, 2015). This species group has a particularly complex taxonomy due to its enormous morphological variability, which is probably increased by processes of hybridization, introgression between populations and polyploidy (Martínez-Labarga & Muñoz -Garmendia, 2015). In the absence of clear diagnostic characters, the definition of taxonomic entities has been made through morpho-geographical divisions. Historically, at least three taxa have been consensually accepted as distinct species: *L. salsoloides* Lam., *L. appressum* Caball. and *L. suffruticosum* L., (Jahandiez & Maire, 1932; López González, 1979; Ockendon & Walters, 1968), in addition to its different varieties. Nevertheless, the latest taxonomic treatment comprised more than 20 taxa for the Iberian Peninsula alone (Martínez-Labarga & Muñoz -Garmendia, 2015).

Linum suffruticosum s.l. also presents a remarkable cytogenetic diversity. The group has been described as a polyploid complex with a basic chromosome number of 8 or 9, displaying a polyploid series from diploidy to decaploidy. This variation in the basic number of chromosomes and in the level of ploidy suggests multiple past events of chromosomal rearrangements, whole genome duplications and hybridization events (Afonso et al., 2021). The various cytotypes of *L. suffruticosum* s.l. are spatially segregated across its distribution and their geographic patterns are complex. The different ploidy levels are distributed parapatrically, with several contact zones between cytotypes that are not reproductively isolated, although mixed-ploidy populations are rarely found. This is expected to influence the genetic diversity and structure of these populations (Afonso et al., 2021).

Polyploidization events can have a major effect on the genetic diversity of a lineage. In polyploidization, alteration of the genome triggers genetic and epigenetic responses, mostly of a stochastic nature (Madlung & Wendel, 2013). In the specific case of allopolyploidy, in which hybridization events have participated, complex gene expression combinations tend to be created, since new polyploid lineages are often formed from the hybridization of different populations, resulting in arrangements that incorporate greater genetic diversity (D. E. Soltis et al., 2009; P. S. Soltis & Soltis, 2000; van de Peer et al., 2021). Thus, polyploidy can promote diversity, but the formation of a new cytotype can cause a reproductive barrier to be created with the rest of the population. As a consequence, a bottleneck would be caused, and the isolation could be linked to a shift towards autogamy that prevents the exclusion of the minority cytotype (Levin, 1975; van Drunen & Husband, 2019). In that case there would be a reduction in genetic diversity. That said, polyploidy is considered a driver of speciation and rapid

diversification (van de Peer et al., 2017, 2021), if populations are isolated but viable, they can promote genetic differentiation (Chester et al., 2012; Levin, 2013). On the contrary, if gene flow is maintained, it may happen that the populations increase their genetic variability and success or that the recombination results in the homogenization of the populations, thus maintaining diversity but not differentiation (Meimberg et al., 2009; Modliszewski & Willis, 2012). The effects of polyploidization are complex and can have a variety of evolutionary consequences in lineages.

Linum suffruticosum s.l. presents an interesting reproductive strategy which includes floral polymorphism. Heterostyly - defined by reciprocal position of anthers and stigmas of opposite morphs within a population which improves pollen transfer between morphs (Barrett et al., 2000; Darwin, 1863, 1877) - is one of the most distinctive characteristics of *L. suffruticosum* (Armbruster et al., 2006; Martínez-Labarga & Muñoz -Garmendia, 2015). Among the flowers of *L. suffruticosum*, two different morphs can be found, but beyond the common distyly, in this case the stigmas and anthers of complementary morphs show full reciprocity in three dimensions: the anthers and stigmas are not reciprocal in height, but rather in how the stamens and styles twirl. In long-styled morph flowers (L-morph), styles form a column in the center of the flower with stigmas facing outward, and stamens spread from near the base, appressed to the corolla wall, with inwardly dehisced anthers. In short-styled morph flowers (S-morph), styles spread outward with the stigmas facing inwards, and the stamens are erect, forming a column in the center of the flower with their anthers rotated such that their dehiscence is outwards (**Figure 1**). The result of this polymorphism is that, depending on the morph, the pollen would be placed on the dorsal or ventral part of the pollinator; and, in the same way, the stigmas would come into contact only with the upper or underside of the pollinator. This mechanism ensures cross-pollination in a more effective way than with the common distyly, in which the pollen lands in a more frontal or more distal part of the body of the pollinator. To date, *Linum suffruticosum* is the only known species with this type of 3D heterostyly (Armbruster et al., 2006).

Figure 1. Flowers of *Linum suffruticosum* with petals removed, note the difference in the two morphs: The short-styled morph (left) and the long-styled morph (right). Font: Armbruster et al. (2006).

The morphological polymorphism in *L. suffruticosum* is associated to a heteromorphic self-incompatibility system that only allows fertilization between morphs, favoring disassortative mating (Afonso, 2022; Nicholls, 1985). Evolutionary models state that, in a population with disassortative mating in equilibrium, natural selection yields equal proportions of style morphs (Lloyd & Webb, 1992). Stochastic factors (e.g. habitat fragmentation) can deviate the morph ratio of populations. Unbalanced morph ratios reduce the abundance of compatible mates (i.e., the effective population size) (van Rossum et al., 2006a), and this can result in a loss of genetic diversity (Meeus et al., 2012). According to the observations of Afonso et al. (2021), most wild flax populations present an approximate 1:1 ratio of style morphs irrespective of ploidy and

population size, suggesting that this complex reproductive system maintains self-incompatibility. Even that, it is unknown yet to what extent the heteromorphic self-incompatibility system and the proportion of morphs may vary in populations throughout the distribution range.

Analyzing the genetic diversity and differentiation of *L. suffruticosum* s.l. populations appears as a key to understanding the evolutionary history and ecological forces driving the intra-specific variation in this taxonomic complex. The selective pressures faced by populations throughout the whole distribution, in addition to the stochastic processes that affect the differentiation between populations will be reflected in their neutral genetic variation (Barrett, 2010). Reproductive biology is also a relevant factor in explaining genetic diversity and population structure. The complex heterostylous system of the wild flax is expected to have important effects on the level of genetic differentiation of populations and will have different consequences than those found among self-fertilizing species (Crawford & Whitney, 2010).

The main aim of this study was to determine the patterns of genetic diversity and population structure of *L. suffruticosum* s.l. in populations of contrasted environmental conditions throughout the distribution range of the species, through the use of nuclear microsatellite markers. In addition, we aimed to determine the correlates of the genetic variation with various biological and ecological (morph-ratio, population size, latitude, and elevation, ploidy level and taxonomic entity as surrogate of morphological variation) traits of populations. Apart from possible taxonomic implications, these results will shed light on the microevolution of this group and, in general, on the genetic structuring patterns of Mediterranean outcrossing polyploid species complexes.

Materials and methods

Sample collection

We sampled a total of 32 *L. suffruticosum* s.l. populations throughout the entire species' range, from Morocco to the southern East coast of France, through the East half of the Iberian Peninsula, during the 2018 to 2021 spring flowering seasons. For each population, we estimated population size by counting plants across haphazardly chosen areas and extrapolating the number for the complete population. We estimated the morph ratio by visual determination of the floral morph of 60 randomly chosen flowers from different individuals (one flower per plant, total N = 1427 flowers). In populations smaller than 60, all plants were checked. We individually collected leaves from 12 individuals of known morph and conserved them in silica gel. For 21 (65%) of these sampled populations, we compiled available information on ploidy level (diploid to octoploid populations) and entities of putative taxonomic value (determined as *L. suffruticosum,* Moroccan *L. suffruticosum,* or *L. appressum-salsoloides*) from Afonso et al. (2021). We determined the taxonomic entity of the remaining populations following the same criteria as in Afonso et al. (2021) (**Table S1, Supplementary material**).

DNA isolation, SSR amplification and genotyping

Genomic DNA was extracted from the dried leaf samples (total N=382 individuals) in 2022 using the Invisorb® Spin Plant Mini Kit (Invitek Molecular GmbH, Germany). The concentration and quality of the extracted DNA were assessed with a Nanodrop DeNovix DS-11 Spectophotometer (DeNovix Inc., USA).

We used the specific *L. suffruticosum* SSR panel developed by Olmedo et al. (unpublished; **Table S2, Supplementary material**) to genotype the samples. Although this panel includes a total of 12 SSR markers, only six of these appeared suitable for this study after an exploratory probe in one individual from each of 16 populations. We performed simplex PCR for each of these six SSR in 20 µL of master mix including: 1x MyTaq Red Reaction Buffer (Bioline), 0.40 µM dye-labeled M13 primer, 0.40 µM PIG-tailed reverse primer, 0.04 µM M13-tailed forward primer, 0.01% bovine serum albumin (BSA, Promega), 0.5 µL MyTaqTM Red DNA Polymerase (Bioline), 50-70 ng gDNA and deionized water up to 20 µL. M13-tailed forward primers were labelled with either 6-FAM, VIC, NED, or PET fluorescent labels. The PCRs were performed in a Veriti™ 96-Well Thermal Cycler (Applied Biosystems™, USA) following the same touchdown procedure for all loci. This consisted of an initial denaturation for 2 min at 94°C; followed by 10 cycles of 92°C for 30 s, 30 s at 63°C with an increment of −1°C per cycle, and 30 s at 72°C; followed by 20 cycles of 94°C for 30 s, 30 s at 56°C, and 30 s at 72°C; and an extra extension of 5min at 72°C. Amplification products were genotyped by MACROGEN (Madrid, Spain) in an automatic ABI 3730 capillary DNA sequencer, using a GeneScan 500 LIZ internal size standard. The resulting electropherograms were analyzed for allele binning and calling using PeakScanner™ Software V2.0 (Applied Biosystems™, USA).

Data analysis

Genetic diversity

The retrieved genotypes were used to characterize the genetic diversity of *L. suffruticosum* s.l. populations. Given the autopolyploid nature of the studied complex, GenoDive V3.06 (Meirmans, 2020) was used to calculate the standard metrics of genetic diversity per population and per locus: number of alleles (N_a), effective number of alleles (N_e), expected heterozygosity (H_e) and total and corrected total heterozygosity (H_t, and H'_t), the last two only on a per locus basis. The F_{IS} statistic was used to test deviations from Hardy-Weinberg Equilibrium (HWE). Depending on the analysis, GenoDive corrects for the unknown dosage of alleles in polyploids in the estimates of genetic parameters by a maximum likelihood method based on random mating within populations or a method based on population allele frequencies. Since this correction is not possible in the calculation of observed heterozygosity (H_o) and inbreeding coefficient (G_{IS}) statistics, these parameters were not estimated.

Genetic structure

We evaluated the genetic differentiation among populations, ploidy levels and taxonomic entities by analyses of molecular variance (AMOVA) run in GenoDive (Meirmans, 2020). The isolation by distance was assessed by regressing population pairwise genetic distances (based in F_{ST}) against the geographic Euclidean distances between populations using a Mantel test as implemented in the vegan 2.6-4 R package (Oksanen et al., 2022). Given the genetic differentiation found among ploidy levels, we carried out a second Mantel test to examine correlation between F_{ST}-based genetic distances and ploidy level-based distances (calculated as the pairwise difference between population ploidy levels; e.g. a value of six for a diploid - octoploid population pair). A Bayesian clustering of individuals was performed using the STRUCTURE V2.3.4 software (Pritchard et al., 2000), which has been shown to be the most suitable clustering method in mixed-ploidy systems (Stift et al., 2019). First, we estimated the population structure for the complete dataset using the admixture model and the locations as prior with a $2x10^4$ burn-in period and subsequent $1x10^5$ MCMC steps. We evaluated $1 \leq K \leq 10$ potential genetic clusters (K) with 10 replicates for each K value. The optimal number of K was determined by the Evanno method (Evanno et al., 2005), checked with the aid of the Pophelper app (Francis, 2017). As in this first clustering the optimal K was found to be 2 (**Figure S1 A,**

Supplementary material), we created a subsample of the dataset including the individuals of the populations assigned to cluster 1 and repeated the analysis using the same parameters. The graphical representation of population genetic structure was generated with the Pophelper app. Last, a principal component analysis (PCA) was carried out based on a matrix of covariance of allele frequencies among populations with GenoDive and plotted using the ggplot2 R package (Wickham, 2016).

Ecological and biological correlates of genetic diversity

We tested whether the genetic diversity of populations differed among ploidy levels, taxonomic entities and genetic clusters retrieved by STRUCTURE analyses (see below) using ANOAs and Tukey HSD post hoc tests, using the stats R package (R Core Team, 2021). Three populations had unassigned taxonomic entities and one population had a mixture of diploid and triploid individuals. The analyses were repeated by assigning these populations to each putative group. Similar results were retrieved in all cases, and hence we report the results for the analyses classifying such populations as *L. suffruticosum* and as diploid. We also analyzed the possible correlation between genetic diversity and the following continuous ecological variables of populations: percentage of L-morph, population size, latitude and elevation. Given the spatial nature of our data, we first calculated the Moran's I statistic using the R package ape V5.6 (Paradis & Schliep, 2019) to evaluate the possible spatial autocorrelation of each variable. Given the lack of significant spatial autocorrelation for most of them **(Table S3**, **Supplementary material**), we performed linear regressions to analyze the association between each genetic diversity parameter and ecological variables.

Results

Characterization of SSR markers

Among the 6 SSR markers analyzed, a total of 84 different alleles were found, with an average of 14 alleles per locus (range 10-30). X337 was the locus with the greatest genetic diversity: besides having the highest number of alleles (N_a=30), it had the highest effective number of alleles (N_e = 3.883) and the highest expected, total and corrected total heterozygosity (H_e = 0.835, H_t = 0.924, H'_t = 0.927) (**Table 1**).

Genetic diversity of populations

The average number of alleles in a population (N_a) was 4.844, ranging from 2.167 (FLAX24) to 7.833 (FLAX78). Similarly, the average number of effective alleles in a population (N_e) was 3.05, ranging from 1.709 (FLAX24) to 4.86 (FLAX78). The expected heterozygosity (H_e) ranged from 0.381 (FLAX 61) to 0.744 (FLAX9), with an average of 0.584 (**Table 2**). Based on F_{IS} statistics, all the populations deviated from HWE (P <0.001).

Table 1. Indices of genetic diversity per locus. Abbreviations: N_a, Number of alleles per locus; N_e Effective number of alleles per locus; H_e, Expected Heterozygosity; Ht, Total Heterozygosity; H't, Corrected Total Heterozygosity.

Locus	N_a	N_e	H_e	H_t	H'_t
X114	12	1.971	0.515	0.671	0.676
X337	30	3.883	0.835	0.924	0.927
X373	10	2.743	0.690	0.852	0.857
X395	10	1.243	0.209	0.713	0.729
X421	11	3.026	0.698	0.843	0.848
X943	11	2.044	0.549	0.766	0.773
Overall	14 ± 7.188	2.485 ± 0.848	0.583 ± 0.197	0.795 ± 0.087	0.802 ± 0.084

Table 2. Indices of genetic diversity in each population. Abbreviations: N_a, Number of alleles per population; N_e, Effective number of alleles per population; H_e, Expected heterozygosity.

Population	N_a	N_e	H_e
FLAX2	4.333	2.709	0.640
FLAX5	4.000	2.494	0.520
FLAX9	3.833	2.883	0.744
FLAX13	4.167	2.413	0.602
FLAX24	2.167	1.709	0.398
FLAX29	5.000	2.926	0.546
FLAX30	4.833	3.371	0.601
FLAX33	7.167	4.533	0.736
FLAX36	2.833	2.026	0.459
FLAX37	5.167	3.282	0.531
FLAX39	4.833	3.075	0.518
FLAX45	5.500	2.841	0.581
FLAX46	3.667	2.622	0.539
FLAX53	4.500	3.050	0.583
FLAX56	5.167	3.373	0.646
FLAX59	5.333	3.590	0.603
FLAX61	2.333	1.840	0.381
FLAX62	5.167	3.332	0.567
FLAX69	4.000	2.797	0.661
FLAX71	7.333	4.200	0.680
FLAX73	3.833	2.934	0.508
FLAX75	7.333	4.266	0.689
FLAX77	4.500	2.518	0.560
FLAX78	7.833	4.860	0.717
FLAX79	6.500	3.675	0.544
FLAX80	6.833	3.448	0.611
FLAX81	5.333	2.814	0.497
AA100	5.000	2.988	0.669
AA105	3.667	2.550	0.647
AA107	3.167	2.117	0.515
AA111	6.000	3.419	0.690
AA112	3.667	2.955	0.510
Overall	4.844 ± 1.447	3.050 ± 0.731	0.584 ± 0.092

Genetic structure

The analyses of molecular variance (AMOVA) revealed a significant differentiation among populations and among ploidy levels (**Table 3 A**), but not among taxonomic entities (**Table 3 B**). First, 45.9% of total genetic variation was found within populations, followed by a 42.2% found among populations, and an 11.9% found among ploidy levels. Nevertheless, there was no correlation between the genetic distance and ploidy differentiation of populations, according to the corresponding Mantel test (**Table 4 A**). However, another Mantel test showed a significant correlation between the genetic distance and the geographical distance of populations, fitting the isolation by distance model (**Table 4 B**).

The Bayesian population analyses conducted with STRUCTURE found that the most likely ΔK was maximized at K=2 (**Figure S1 A, Supplementary material**), suggesting two distinct genetic clusters (**Figure 2 A**). Only three Moroccan populations were assigned to the cluster 2 (AA100, AA111 and AA112), and the other 29 populations were assigned to the cluster 1 (**Figure 2 B**). When removing the three populations of cluster 2 from the analysis, the ΔK was maximized in the value of K=7 (**Figure S1 B, Supplementary material**), which translates into seven differentiated genetic clusters (**Figure 2**).

Table 3. Analysis of molecular variance (AMOVA). Test of the variations within populations, among populations and among ploidy levels (A) or among taxonomic entities (B).

Source of variation	P-value	%Var
(A)		
Within Population	-	0.459
Among Population	$P < 0.001$ *	0.422
Among ploidy levels	$P < 0.001$ *	0.119
(B)		
Within Population	-	0.482
Among Population	$P < 0.001$ *	0.521
Among taxonomic entities	0.579	0

Table 4. Results of the Mantel's test showing the correlation between genetic differentiation (distance based on F_{ST}) and ploidy-based distance (A) or geographic Euclidean distance (B).

	Mantel statistic r	Simulated P-value
(A)		
F_{ST} distance – Ploidy-based distance	-0.165	0.956
(B)		
F_{ST} distance – Geographic distance	0.281	0.001 *

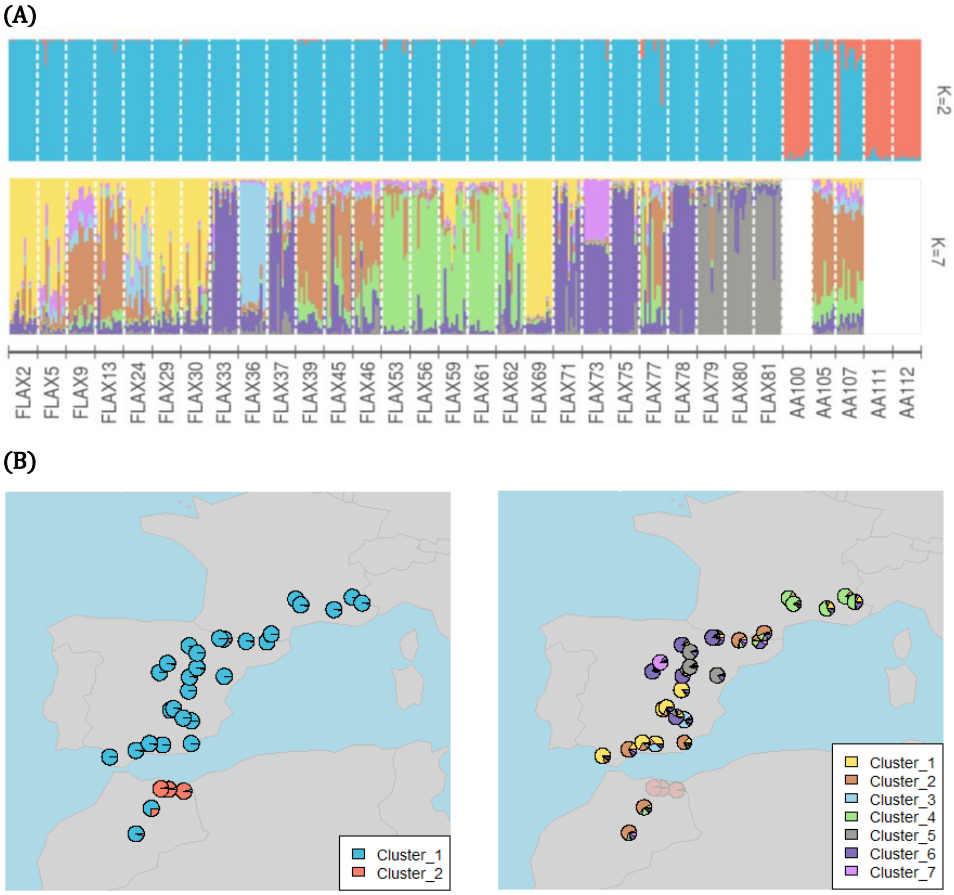

Figure 2. Genetic structure of the sampled populations of *Linum suffruticosum*.
(A) Results of the Bayesian analysis of population structure (with software STRUCTURE, see Material and Methods). Plots showing two genetic clusters(K=2) above, and seven genetic clusters (K=7) after removing three Moroccan populations from the analyses (below).
(B) Geographic distribution of the genetic clusters. Each population is represented by a pie chart that shows the percentage of membership in each genetic cluster for K=2 (left) and K=7 after removing three Moroccan populations (right).

The PCA analysis did show some diffuse trends of separation of populations along multivariate axes. Although with several intermediate points, there were some moderately dense population groups which corresponded the best with the genetic clusters assigned by STRUCTURE (**Figure 3 A**). Even those populations that remain isolated in the PCA space coincided with those assigned to a genetic cluster individually. The ploidy level of the populations also presented some correspondence with the PCA clustering (**Figure 3 B**) PCA1 differentiated diploid, tetraploid and hexaploid populations, although the ellipses around them. PCA2 separated some tetraploid populations as well. Finally, the taxonomic entities were not differentiated in the PCA space (**Figure 3 C**). It is observed that the ellipses that delimit the three defined entities overlap completely.

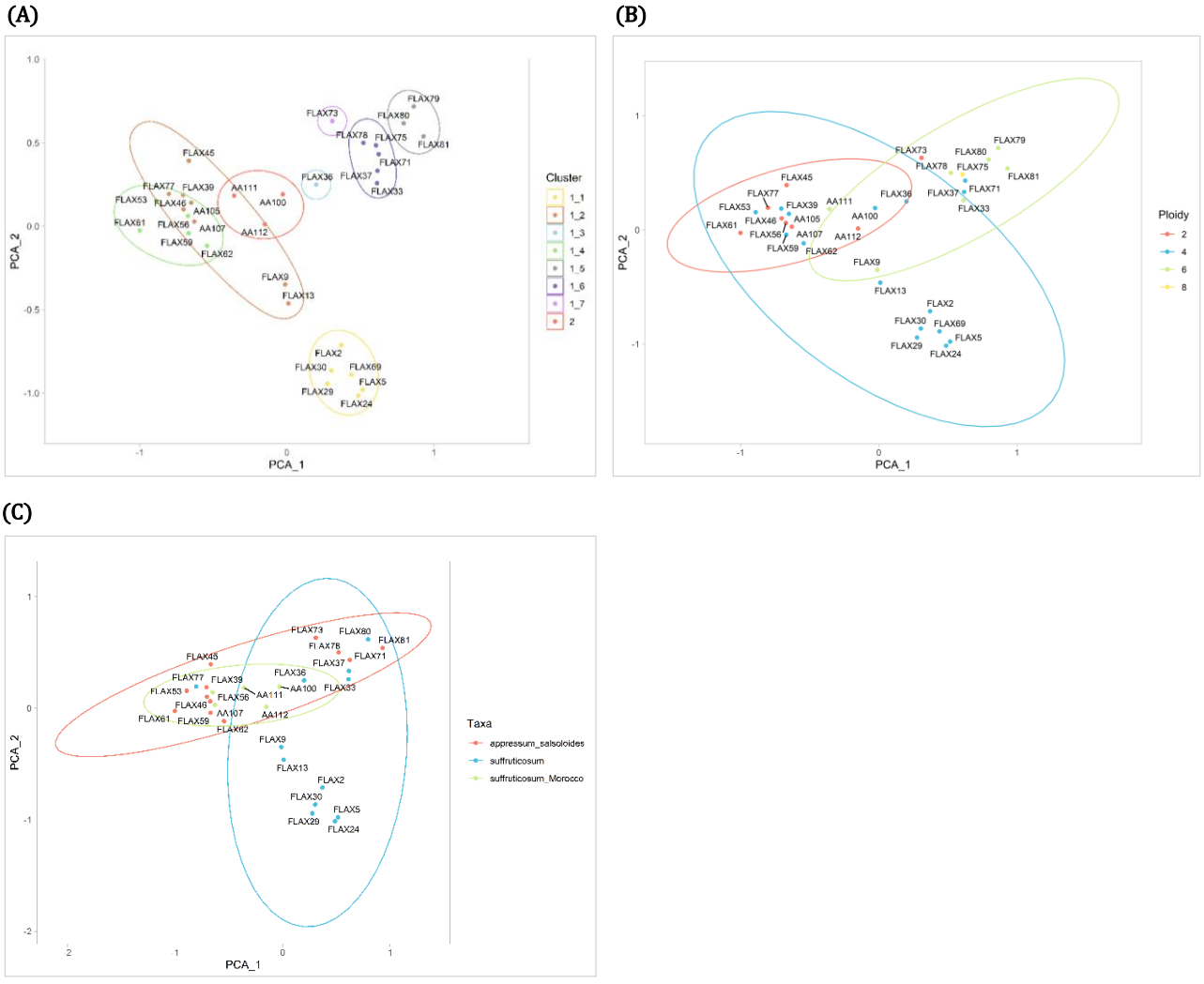

Figure 3. PCA scatter plots. The sampled populations grouped by genetic clusters (A), ploidy (B), ang taxonomic entities (C).

Ecological and biological correlates of genetic diversity

The ANOVA tests carried out showed significant between-population among ploidy differences in their genetic diversity for all analyzed parameters ($P < 0.04$; ANOVA, **Table 5 A**). The number of alleles, the effective number of alleles and the expected heterozygosity increased with the ploidy level of populations: diploids had the lowest genetic diversity (mean N_a =3.979,

Table 5. Results of the ANOVAs test on the differences in genetic diversity parameters among ploidy levels (A), taxonomic entities(B) and genetic clusters (C). Abbreviations: N_a, Number of alleles per population; N_e, Effective number of alleles per population; H_e, Expected heterozygosity.

(A)	Df	Sum Sq	Mean Sq	F value	Pr(>F)
N_a	1	24.350	24.345	17.990	$P < 0.001$ *
N_e	1	5.180	5.180	13.660	$P < 0.001$ *
H_e	1	0.063	0.063	9.770	0.041 .
(B)					
N_a	2	2.620	1.308	0.650	0.530
N_e	2	0.959	0.480	0.919	0.412
H_e	2	0.005	0.003	0.282	0.756
(C)					
N_a	7	41.290	5.898	5.985	$P < 0.001$ *
N_e	7	10.531	1.504	5.994	$P < 0.001$ *
H_e	7	0.074	0.011	1.394	0.253

mean N_e =2.65, mean H_e =0.53) and octoploids presented the highest values of genetic diversity (mean N_a =7.333, mean N_e =4.266, mean H_e =0.689) (**Figure 4 A; Table S4 A, Supplementary material**). The Tukey H_e D post hoc test showed significant differences between diploids and hexaploids, between tetraploids and hexaploids, and between diploids and octoploids for at least one genetic diversity parameter (**Figure 4 A; Table S5 A, supplementary material**).

There were not significant differences in the genetic diversity of populations from different taxonomic entities (P> 0.4; ANOVA, **Table 5 B**)**.** Finally, we found significant differences in N_a and N_e (but not in H_e) of populations between different genetic clusters (P < 0.001; ANOVA, **Table 5 C**). The cluster 1_3 showed the lowest values (mean N_a=2.833, mean N_e = 2.026) and the cluster 1_6 presented the higher values of genetic diversity (mean N_a =6.967, mean N_e = 4.228) (**Figure 4 B; Table S4 B, Supplementary material**). The 1_6 cluster's N_a and N_e significantly differed from 1_1, 1_2, 1_3 and 1_4, according to the post hoc test (**Figure 4 B; Table S5 B, Supplementary material**).

The genetic diversity of populations was not correlated with any ecological variable analyzed (percentage of L-morph, population size, latitude, and elevation) (P >0.05; **Figure 5; Table S6, Supplementary material**).

Discussion

The genetic characteristics of populations and their variation throughout the distribution of a species are defined from the joint effects of adaptive forces governed by selection, and by non-adaptive forces driven by stochastic processes like genetic drift, founder events, barriers in the gene flow, and bottlenecks. In the case of the latter, the microevolutionary processes that occur during the history of a lineage will leave a trace imprinted in the neutral genetic variation (Barrett et al., 2010; Pool et al., 2010). Intraspecific genetic variations originating from both adaptive and non-adaptive forces have an important value in the adaptive capacity of species. High levels of genetic diversity maintain the evolutionary potential of the species since they may be directly linked to a greater phenotypic diversity, which is critical in the adaptative processes and in the divergence of diverse populations in different environments, particularly if gene flow is limited (Frankham et al., 2002).

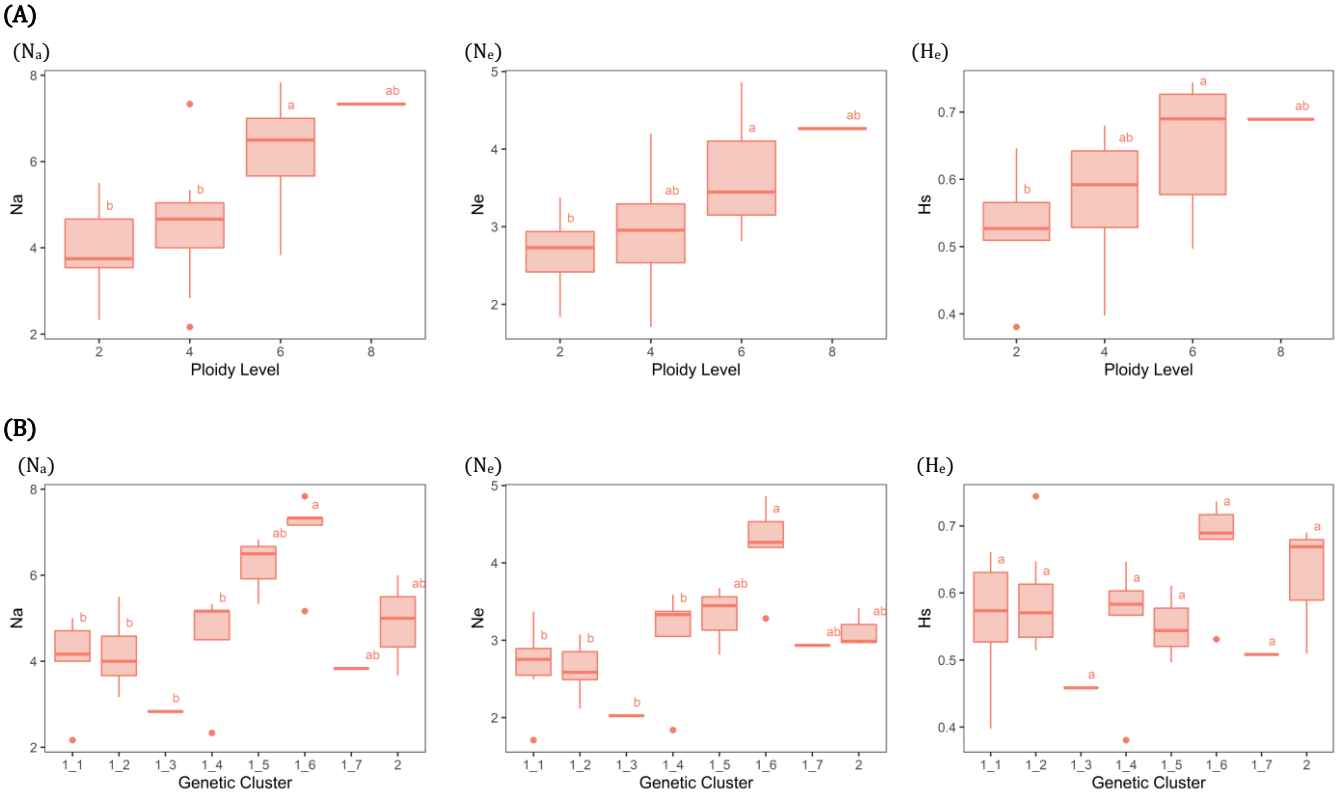

Figure 4. Boxplot representing the results of the ANOVA tests. Significant differences between groups according to the Tukey test are represented by different letters over each box in each plot:
(A) Relation between different levels of ploidy and genetic diversity parameters: (N_a) Number of alleles per population; (N_e) effective number of alleles per population, (H_e) expected heterozygosity.
(B) Relation between different genetic clusters and genetic diversity parameters: (N_a) Number of alleles per population; (N_e) effective number of alleles per population, (H_e) expected heterozygosity.

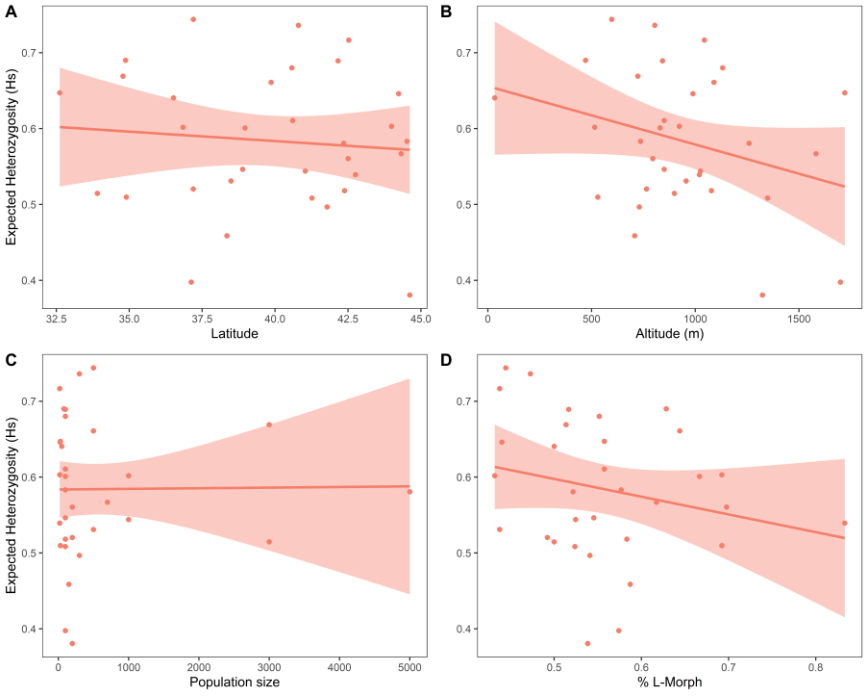

Figure 5. Scatter plots representing the linear regression between the expected heterozygosity (H_e) and latitude (A), altitude (m.a.s.l.) (B), population size (C) and % of long-styled (L-morph) morph individuals in the population.

Population genetic diversity

Our results show that *L. suffruticosum* s.l. shows a comparatively high level of genetic diversity in comparison to other widely distributed species of the Mediterranean Basin (Douaihy et al., 2011; González-Martínez et al., 2002; Jump et al., 2009; Martínez-López et al., 2020; Stöcklin et al., 2009). However, it seems that this genetic variation is more related to stochastic processes than to selective processes mediated by environmental cues which leave a phenotypic signal. In fact, the only biological process that is clearly correlated with genetic diversity is ploidy level. The wide distribution - from Morocco to the Maritime Alps and central France, with abundant and large populations without major disjunctions, may explain the high genetic diversity. In large populations, allele frequencies remain stable more easily than in small populations, especially if they fluctuate, since the former are less affected by drift (Frankham et al., 2002). In the theoretical framework, widely distributed species present center-periphery patterns of diversity, concentrating the highest values of diversity in the central populations, while genetic diversity is expected to be lower in populations at range limits, where the populations are smaller and more isolated since they are in an area where the abiotic and biotic factors that influence the survival and reproduction of individuals become less optimal (Eckert et al., 2008). This seems not to be the case for *L. suffruticosum* s.l.. Although *L. suffruticosum* s.l. populations span a wide latitudinal and altitudinal range, we did not find a significant relationship between these geographical factors and the level of genetic diversity of the populations, thus ruling out the center-periphery pattern. Likewise, no correlation was found with population size. The fact that there is not less genetic diversity in populations of reduced size indicates that they are not completely reproductively isolated from neighboring populations, thus maintaining gene flow (Loveless & Hamrick, 1984). Nevertheless, the existence of an isolation by distance pattern in genetic distance of populations may point out to some effect of stochastic forces plus limited gene flow if environmental differences does not affect. We can conclude that the balance selection vs stochasticity is biased to the latter in this species.

Genetic diversity and breeding system

The genetic variability of populations is also strongly influenced by the reproductive strategy of a species, since the outcrossing rate is key in the gene flow that maintains the exchange of genes and the balance in allele frequencies (Si et al., 2020). In the case of the wild flax, the peculiar 3D heterostyly system makes the species an obligate outcrosser (Afonso, 2022), which may also explain the high levels of genetic diversity (Yuan et al., 2017). In distylous species, stochastic processes and selection can also affect the morph ratio of populations (Meeus et al., 2012). In turn, this also entails changes in genetic diversity. For instance, a relaxation in the heteromorphic self-incompatibility system may cause populations with a higher proportion of L-morphs to have greater variability than those found at equilibrium (van Rossum et al., 2006b; van Rossum & Triest, 2007; Wedderburn & Richards, 1990). In the case of *L. suffruticosum* s.l., our results confirm a prior survey (Afonso et al., 2021) and indicate that most populations throughout the entire distribution range of the species show no large deviations from the 1:1 ratio (isoplethism). However, we also found populations with percentages of S-morph as low and high as 0.17 and 0.57, respectively. We are unaware of the strength of the heteromorphic self-incompatibility across morphs and populations of the species. The fact that the morph ratio is not related to the genetic diversity of populations suggests that the variations in the morph ratio are subtle and their possible causes do not counteract the effect of large population size and that they are not isolated.

Population genetic structure

It has been established that *L.suffruticosum* s.l. presents high levels of genetic diversity and, despite the lack of correlation between this and environmental and demographic variables, it does not necessarily mean that there is no genetic differentiation between populations (Stöcklin et al., 2009). Our results revealed a genetic structuration of populations driven by a pattern of isolation by distance. Precisely, the Rif functions as a significant genetic barrier for the species complex, as it separates the populations of the two main genetic clusters. Although the north of Morocco and the Iberian Peninsula share most of their flora (Valdés, 1991), the Strait of Gibraltar functions as a clear genetic barrier between populations of this region, at least for some linages (Arroyo et al., 2008; Escudero et al., 2008; Molina-Venegas et al., 2015). Despite that, two of the Moroccan populations located further west are more similar to several ones of the Iberian Peninsula than to the other three nearby northern Moroccan populations, demonstrating that the Rif has historically acted as a more important genetic barrier than the Strait (Molina-Venegas et al., 2015).

Genetic diversity and taxonomy of Linum suffruticosum group

Taxonomic entities within *l. suffruticosum* group present similar levels of genetic diversity and do not show any relationship with the differentiation among populations. Considering that the definition of taxonomic entities has been made through morpho-geographical criteria (Jahandiez & Maire, 1932; López González, 1979; Martínez-Labarga & Muñoz -Garmendia, 2015; Ockendon & Walters, 1968) and that it has been proven that there is no correlation between genetic diversity and these ecological variables, it is logical that there is no relationship between taxonomic treatment and genetic variability. This lack of congruence between the factors that genetically differentiate some populations from others and the factors that have been taken in the determination of taxa shows that the taxonomy of the group should be reviewed. Even though according to Afonso et al. (2021) *Linum salsoloides* and *L. appressum* are described as diploids, and *L. suffruticosum* s.str. as a polyploid complex including diploid tetraploid and octoploid individuals, there is also a lack of correspondence between species identity and ploidy level.

Genetic diversity and polyploidy

As expected, the cytogenetic diversity that characterizes the complex does have a direct effect on genetic diversity: the results show that polyploidization increases the genetic diversity levels of populations. The set of populations of the complex include polyploids of a parapatric nature that would have been recurrently formed from chromosomal rearrangements, whole genome duplications and hybridization events (Afonso et al., 2021). This mode of diversification allows the incorporation of genetic diversity from different parental populations, so the genetic diversity of polyploid populations will be greater than that of diploid parents and will also present less inbreeding depression (Soltis & Soltis, 2000). The role of ploidization in the population genetics of *L. suffruticosum* s.l. affects the differentiation of populations. It is evident that there is a significant differentiation among populations, but the results also indicated that there is a significant genetic differentiation of different ploidy levels. In general, polyploidization has consequences on the capacity to adapt to habitats, becoming different than those of their parents, so a niche segregation is produced. This, added to reproductive isolation, allows the establishment of polyploid differentiated populations (Baniaga et al., 2020; Comai, 2005; Moura et al., 2021). This would explain the scarcity of mixed polypoid populations in the *L. suffruticosum* complex. On the other hand, in the case of this wild flax, there is evidence that such reproductive isolation does not exist, on the contrary, populations throughout the distribution are in contact and hybridization is common, so that a process that could often carry out the

segregation of populations leads to a balance between genetic differentiation and the maintenance of genetic diversity (Balao et al., 2011, 2015).

The Bayesian analysis of populations' structure confirms that the structuring of genetic clusters has a clear geographical component. The populations of the same cluster appear close to each other, although without clear boundaries, suggesting that the genetic flow can be maintained, except for the genetic barrier that the Rif represents. Observing the PCA plots, we see that the genetic clusters reliably group the populations. However, it cannot be affirmed that the populations of the same ploidy level are grouped following the PCA. This indicates that the structuring of the population is more strongly influenced by isolation by distance than by the ploidy of the populations, although, as already mentioned, these explain part of the genetic differentiation. In fact, the cytogenetic diversity of *L. suffruticosum* s.l. itself, like that of most polyploid complexes, presents geographic segregation (Afonso et al., 2021; Nicholls, 1986; Rice et al., 2019; Ruiz-Martín et al., 2018). In the scenario of climate change, and the consequent changes in the distribution of species, it remains to be examined if the populations can maintain the population size and the genetic flow that keeps the levels of genetic diversity, if the disturbances will have an effect on the population segregation, and if this can compromise their viability.

Concluding remarks

The study of neutral genetic data has unraveled the processes that determine the patterns of diversity and genetic structure of *L. suffruticosum* s.l. across its whole range. We have verified that this complex presents a high level of genetic diversity, promoted by processes of polyploidization. Diversity is maintained thanks to the abundant and reproductively connected populations, which allow the maintenance of gen flow. The effects on genetic diversity that the characteristic heteromorphic self-incompatibility system (high within- and low among-population genetic diversity) and the segregation of polyploid populations can potentially cause are thus mitigated. The populations are structured following a pattern of isolation by distance with the Rif mountains as the most important genetic barrier. Lastly, the current taxonomic treatment does not fit the patterns of population differentiation and should be revisited. It is worth noting that the available evidence from this study and former phylogenetic studies where this species group has been included (Maguilla et al., 2021), which depicts a very recent (Pleistocene) origin and diversification, both point out to a very active group where range expansion has occurred fast, with low signature of adaptive processes linked to differentiation in particular habitats, being the most influential process that of polyploidy. Ploidy levels also show low differentiation in niche characteristics (Afonso, 2022). The fact that this species group has not been able to colonize islands in Western Mediterranean, despite the potential niche is widely present there, when these islands were already disconnected from the mainland, also support this interpretation. Only populations south of Rif mountains might be old enough to undergo colonization and differentiation process.

Acknowledgements

This work was funded by MSCA-IF-GF grant 897890 – FLAXMATE, and MICINN-FEDER grant PGC2018 099608 B 100 “REPROGRAD”. MAV was supported by a University of Seville Research Introduction grant VIIPPIT-2022-II.1.

References

Afonso, A. (2022). *The role of polyploidy in shaping the diversity within Linum suffruticosum s.l.(Linaceae)* [PhD Dissertation]. University of Coimbra.

Afonso, A., Loureiro, J., Arroyo, J., Olmedo-Vicente, E., & Castro, S. (2021). Cytogenetic diversity in the polyploid complex Linum suffruticosum s.l. (Linaceae). *Botanical Journal of the Linnean Society*, *195*(2), 216–232. https://doi.org/10.1093/BOTLINNEAN/BOAA060

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*(3), 581–590. https://doi.org/10.1111/J.1469-8137.2006.01749.X

Arroyo, J., Aparicio, A., Albaladejo, R. G., Muñoz, J., & Braza, R. (2008). Genetic structure and population differentiation of the Mediterranean pioneer spiny broom Calicotome villosa across the Strait of Gibraltar. *Biological Journal of the Linnean Society*, *93*(1), 39–51. https://doi.org/10.1111/J.1095-8312.2007.00916.X

Barrett, S. C. H. (2010). Darwin's legacy: the forms, function and sexual diversity of flowers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1539), 351–368. https://doi.org/10.1098/RSTB.2009.0212

Barrett, S. C. H., Jesson, L. K., & Baker, A. M. (2000). The Evolution and Function of Stylar Polymorphisms in Flowering Plants. *Annals of Botany*, *85*(suppl_1), 253–265. https://doi.org/10.1006/ANBO.1999.1067

Barrett, S. C. H., Yakimowski, S. B., Field, D. L., & Pickup, M. (2010). Ecological genetics of sex ratios in plant populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2549–2557. https://doi.org/10.1098/RSTB.2010.0002

Chester, M., Gallagher, J. P., Symonds, V. V., da Silva, A. V. C., Mavrodiev, E. v., Leitch, A. R., Soltis, P. S., & Soltis, D. E. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae). *Proceedings of the National Academy of Sciences of the United States of America*, *109*(4), 1176–1181. https://doi.org/10.1073/PNAS.1112041109/SUPPL_FILE/PNAS.201112041SI.PDF

Crawford, K. M., & Whitney, K. D. (2010). Population genetic diversity influences colonization success. *Molecular Ecology*, *19*(6), 1253–1263. https://doi.org/10.1111/J.1365-294X.2010.04550.X

Darwin, C. (1863). On the existence of two forms, and on their reciprocal sexual relation, in several species of the genus Linum. *Journal of the Proceedings of the Linnean Society of London. Botany*, *7*(26), 69–83. https://doi.org/10.1111/J.1095-8312.1863.TB01061A.X

Darwin, C. (1877). *The Different Forms of Flowers on Plants of the Same Species* (J. Muray, Ed.). http://www.public-library.uk/pdfs/5/205.pdf

Douaihy, B., Vendramin, G. G., Boratyński, A., Machon, N., & Dagher-Kharrat, M. B. (2011). High genetic diversity with moderate differentiation in Juniperus excelsa from Lebanon and the eastern Mediterranean region. *AoB PLANTS*, *2011*(1). https://doi.org/10.1093/AOBPLA/PLR003

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170–1188. https://doi.org/10.1111/J.1365-294X.2007.03659.X

Escudero, M., Vargas, P., Valcárcel, V., & Luceño, M. (2008). Strait of Gibraltar: an effective gene-flow barrier for wind-pollinated Carex helodes (Cyperaceae) as revealed by DNA sequences, AFLP, and cytogenetic variation. *American Journal of Botany*, *95*(6), 745–755. https://doi.org/10.3732/AJB.2007342

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/J.1365-294X.2005.02553.X

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Frankham, R., Ballou, J. D., Briscoe, D. A., & McInnes, K. H. (2002). Introduction to Conservation Genetics. *Introduction to Conservation Genetics*. https://doi.org/10.1017/CBO9780511808999

González-Martínez, S. C., Alía, R., & Gil, L. (2002). Population genetic structure in a Mediterranean pine (Pinus pinaster Ait.): a comparison of allozyme markers and quantitative traits. *Heredity 2002 89:3*, *89*(3), 199–206. https://doi.org/10.1038/sj.hdy.6800114

Greuter, W. (2001). Diversity of Mediterranean island floras. *Bocconea*, *13*, 55–64.

Heywood, V. H. (1995). The Mediterranean flora in the context of world biodiversity. *Ecologia Mediterranea*, *21*(1–2), 11–18.

Jahandiez, É., & Maire, R.-C.-J.-E. (1932). *Catalogue des plantes du Maroc. Tome deuxiéme. Dicotylédones Archichalamydées*. Alger : Imprimerie Minerva. En Vente à Paris, chez P. Lechevalier, libraire 12, Rue de Tournon (6º). https://doi.org/10.3/JQUERY-UI.JS

Jump, A. S., Rico, L., Lloret, F., & Peñuelas, J. (2009). Microspatial population genetic structure of the Mediterranean shrub Fumana thymifolia. *Plant Biology*, *11*(2), 152–160. https://doi.org/10.1111/J.1438-8677.2008.00109.X

Levin, D. A. (1975). Minority cytotipe exclusion in local plant populations. *Taxon*, *24*(1), 35–43. https://doi.org/10.2307/1218997

Levin, D. A. (2013). The timetable for allopolyploidy in flowering plants. *Annals of Botany*, *112*(7), 1201–1208. https://doi.org/10.1093/AOB/MCT194

Lloyd, D. G., & Webb, C. J. (1992). The Evolution of Heterostyly. In *Evolution and Function of Heterostyly* (pp. 151–178). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-86656-2_6

López González, G. (1979). Algunas consideraciones sobre los linos del grupo Linum tenuifolium L. en España. In *Memoires de la Société Botanique de Genéve* (Vol. 1, pp. 99–109). https://doi.org/10.1016/J.TREE.2008.10.008

López-Goldar, X., Zas, R., & Sampedro, L. (2020). Resource availability drives microevolutionary patterns of plant defences. *Functional Ecology*, *34*(8), 1640–1652. https://doi.org/10.1111/1365-2435.13610/SUPPINFO

Loveless, M. D., & Hamrick, J. L. (1984). Ecological Determinants of Genetic Structure in Plant Populations. *Annual Review of Ecology and Systematics*, *15*, 65–95. https://www.jstor.org/stable/2096943#metadata_info_tab_contents

Madlung, A., & Wendel, J. F. (2013). Genetic and Epigenetic Aspects of Polyploid Evolution in Plants. *Cytogenetic and Genome Research*, *140*(2–4), 270–285. https://doi.org/10.1159/000351430

Maguilla, E., Escudero, M., Jiménez-Lobato, V., Díaz-Lifante, Z., Andrés-Camacho, C., & Arroyo, J. (2021). Polyploidy Expands the Range of Centaurium (Gentianaceae). *Frontiers in Plant Science*, *12*, 373. https://doi.org/10.3389/FPLS.2021.650551/BIBTEX

Martínez-Labarga, J., & Muñoz -Garmendia, F. (2015). Linum L. In Castroviejo S, Muñoz Garmendia F, Navarro C, Quintanar A, & Buira A (Eds.), *Flora Ibérica. Vol. 9 Rhamnaceae-Polygalaceae* (Vol. 9, pp. 172–249). Real Jardín Botánico, CSIC.

Martínez-López, V., García, C., Zapata, V., Robledano, F., & de la Rúa, P. (2020). Intercontinental long-distance seed dispersal across the Mediterranean Basin explains population genetic structure of a bird-dispersed shrub. *Molecular Ecology*, *29*(8), 1408–1420. https://doi.org/10.1111/MEC.15413

Médail, F., & Quezel, P. (1997). Hot-spots analysis for conservation of plant biodiversity in the mediterranean basin. *Annals of the Missouri Botanical Garden*, *84*(1), 112–127.

Meeus, S., Honnay, O., Brys, R., & Jacquemyn, H. (2012). Biased morph ratios and skewed mating success contribute to loss of genetic diversity in the distylous Pulmonaria officinalis. *Annals of Botany*, *109*(1), 227–235. https://doi.org/10.1093/AOB/MCR272

Meimberg, H., Rice, K. J., Milan, N. F., Njoku, C. C., & McKay, J. K. (2009). Multiple origins promote the ecological amplitude of allopolyploid Aegilops (Poaceae). *American Journal of Botany*, *96*(7), 1262–1273. https://doi.org/10.3732/AJB.0800345

Meirmans, P. G. (2020). genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources*, *20*(4), 1126–1131. https://doi.org/10.1111/1755-0998.13145

Modliszewski, J. L., & Willis, J. H. (2012). Allotetraploid Mimulus sookensis are highly interfertile despite independent origins. *Molecular Ecology*, *21*(21), 5280–5298. https://doi.org/10.1111/J.1365-294X.2012.05706.X

Molina-Venegas, R., Aparicio, A., Lavergne, S., & Arroyo, J. (2015). The building of a biodiversity hotspot across a land-bridge in the Mediterranean. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1813). https://doi.org/10.1098/RSPB.2015.1116

Myers, N., Mittermeler, R. A., Mittermeler, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858.

Nicholls, M. S. (1985). Pollen flow, population composition, and the adaptive significance of distyly in Linum tenuifolium L. (Linaceae). *Biological Journal of the Linnean Society*, *25*(3), 235–242. https://doi.org/10.1111/J.1095-8312.1985.TB00395.X

Nicholls, M. S. (1986). Variation and evolution inLinum tenuifolium (Linaceae). *Plant Systematics and Evolution 1986 153:3*, *153*(3), 243–258. https://doi.org/10.1007/BF00983691

Ockendon, D. J., & Walters, S. M. (1968). Linaceae. In *Flora Europea* (Vol. 2, pp. 584–603).

Oksanen, J., Simpson, G. L., Blanchet, G., Kindt, R., Legendre, P., Minchin, P. R., & O'Hara, R. B. (2022). *vegan: Community Ecology Package. R package* (2.6-4). https://CRAN.R-project.org/package=vegan

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/BIOINFORMATICS/BTY633

Pool, J. E., Hellmann, I., Jensen, J. D., & Nielsen, R. (2010). Population genetic inference from genomic sequence variation. *Genome Research*, *20*(3), 291–300. https://doi.org/10.1101/GR.079509.108

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/GENETICS/155.2.945

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. In *ci.nii.ac.jp*. https://www.R-project.org/

Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., Meiri, S., Belmaker, J., & Mayrose, I. (2019). The global biogeography of polyploid plants. *Nature Ecology & Evolution*, *3*(2), 265–273. https://doi.org/10.1038/s41559-018-0787-9

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

Si, X., Gao, L., Song, Y., Khayatnezhad, M., & Minaeifar, A. A. (2020). Understanding population differentiation using geographical, morphological and genetic characterization in erodium cicunium. *Indian Journal of Genetics and Plant Breeding*, *80*(4), 459–467. https://doi.org/10.31742/IJGPB.80.4.12

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., Sankoff, D., DePamphilis, C. W., Wall, P. K., & Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany*, *96*(1), 336–348. https://doi.org/10.3732/AJB.0800079

Soltis, P. S., & Soltis, D. E. (2000). The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences*, *97*(13), 7051–7057. https://doi.org/10.1073/PNAS.97.13.7051

Stift, M., Kolář, F., & Meirmans, P. G. (2019). Structure is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity 2019 123:4*, *123*(4), 429–441. https://doi.org/10.1038/s41437-019-0247-6

Stöcklin, J., Kuss, P., & Pluess, A. R. (2009). Genetic diversity, phenotypic variation and local adaptation in the alpine landscape: case studies with alpine plant species. *Botanica Helvetica 2009 119:2*, *119*(2), 125–133. https://doi.org/10.1007/S00035-009-0065-1

Valdés, B. (1991). Andalucia and the Rif. Floristic links and common flora. *Chronica Botanica* , *10*, 117–124.

van de Peer, Y., Ashman, T. L., Soltis, P. S., & Soltis, D. E. (2021). Polyploidy: an evolutionary and ecological force in stressful times. *The Plant Cell*, *33*(1), 11–26. https://doi.org/10.1093/PLCELL/KOAA015

van de Peer, Y., Mizrachi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics 2017 18:7*, *18*(7), 411–424. https://doi.org/10.1038/nrg.2017.26

van Drunen, W. E., & Husband, B. C. (2019). Evolutionary associations between polyploidy, clonal reproduction, and perenniality in the angiosperms. *New Phytologist*, *224*(3), 1266–1277. https://doi.org/10.1111/NPH.15999

van Rossum, F., de Sousa, S. C., & Triest, L. (2006a). Morph-specific differences in reproductive success in the distylous Primula veris in a context of habitat fragmentation. *Acta Oecologica*, *30*(3), 426–433. https://doi.org/10.1016/J.ACTAO.2006.06.005

van Rossum, F., de Sousa, S. C., & Triest, L. (2006b). Morph-specific differences in reproductive success in the distylous Primula veris in a context of habitat fragmentation. *Acta Oecologica*, *30*(3), 426–433. https://doi.org/10.1016/J.ACTAO.2006.06.005

van Rossum, F., & Triest, L. (2007). Fine-scale spatial genetic structure of the distylous Primula veris in fragmented habitats. *Plant Biology*, *9*(3), 374–382. https://doi.org/10.1055/S-2006-924632/ID/60

Wedderburn, F., & Richards, A. J. (1990). Variation in within-morph incompatibility inhibition sites in heteromorphic Primula L. *New Phytologist*, *116*(1), 149–162. https://doi.org/10.1111/J.1469-8137.1990.TB00520.X

Wickham, H. (2016). ggplot2. Elegant Graphics for Data Analysis. In *ggplot2*. Springer-Verlag. https://doi.org/10.1007/978-0-387-98141-3

Yuan, S., Barrett, S. C. H., Duan, T., Qian, X., Shi, M., & Zhang, D. (2017). Ecological correlates and genetic consequences of evolutionary transitions from distyly to homostyly. *Annals of Botany*, *120*(5), 775–789. https://doi.org/10.1093/AOB/MCX098

Supplementary material

Table S1. Ecological and biological variables corresponding to each of the analyzed populations. Abbreviations: POP, Population; LAT, Latitude; LONG, Longitude; 2x, diploid; 3x, triploid; 4x, tetraploid; 6x, hexaploid; 8x, octoploid; % L, Proportion of long-styled morph individuals in the population; Cluster (K2), assigned cluster by STRUCTURE assuming two populations; Cluster (K8), assigned cluster by STRUCTURE assuming eight populations.

POP	LAT	LONG	Location	Ploidy	Taxa	% L	Altitude (m.a.s.l)	Population size	Cluster (K2)	Cluster (K8)
FLAX2	36.516	-6.138	Cádiz, España	4x	*L. suffruticosum*	0.5	33	50	1	1_1
FLAX5	37.194	-4.122	Granada, España	4x	*L. suffruticosum*	0.492	765	200	1	1_1
FLAX9	37.198	-1.999	Almería, España	6x	*L. suffruticosum*	0.444	597	500	1	1_2
FLAX13	36.846	-4.805	Málaga, España	4x	*L. suffruticosum*	0.432	515	1000	1	1_2
FLAX24	37.127	-3.451	Granada, España	4x	*L. suffruticosum*	0.574	1700	100	1	1_1
FLAX29	38.887	-3.053	Ciudad Real, España	4x	*L. suffruticosum*	0.545	850	100	1	1_1
FLAX30	38.971	-2.896	Ciudad Real, España	4x	*L. suffruticosum*	0.667	830	100	1	1_1
FLAX33	40.800	-3.613	Madrid, España	6x	*L. suffruticosum*	0.473	805	300	1	1_6
FLAX36	38.349	-1.993	Albacete, España	4x	*L. suffruticosum*	0.587	708	150	1	1_3
FLAX37	38.490	-2.412	Albacete, Spain	4x	*L. suffruticosum*	0.438	956	500	1	1_6
FLAX39	42.381	0.791	Lérida, Spain	4x	*L. appressum-salsoloides*	0.583	1078	100	1	1_2
FLAX45	42.353	1.842	Girona, Spain	2x-3x	*L. appressum-salsoloides*	0.522	1259	5000	1	1_2
FLAX46	42.757	2.057	Aude, France	2x	*L. appressum-salsoloides*	0.833	1020	20	1	1_2
FLAX53	44.523	3.307	Lozère, France	4x	*L. appressum-salsoloides*	0.577	736	100	1	1_4
FLAX56	44.236	3.554	Lozère, France	2x	*L. appressum-salsoloides*	0.44	989	25	1	1_4
FLAX59	43.992	5.242	Vaucluse, France	4x	*L. appressum-salsoloides*	0.692	923	20	1	1_4
FLAX61	44.619	6.174	Hautes- France	2x	*L. appressum-salsoloides*	0.538	1324	200	1	1_4
FLAX62	44.324	6.664	Alpes-de-Haute-Provence, France	4x	*L. appressum-salsoloides*	0.617	1582	700	1	1_4
FLAX69	39.863	-2.135	Cuenca, Spain	4x	NA	0.644	1090	500	1	1_1
FLAX71	40.573	-2.089	Cuenca, Spain	4x	*L. appressum-salsoloides*	0.552	1132	100	1	1_6
FLAX73	41.263	-3.210	Guadalajara, Spain	2x	*L. appressum-salsoloides*	0.524	1349	100	1	1_7
FLAX75	42.157	-2.109	La Rioja, Spain	8x	Intermediate entity	0.517	842	100	1	1_6
FLAX77	42.504	-0.349	Huesca, Spain	2x	*L. suffruticosum*	0.698	795	200	1	1_2
FLAX78	42.527	-0.556	Huesca, Spain	6x	*L. appressum-salsoloides*	0.438	1044	20	1	1_6
FLAX79	41.038	-1.709	Zaragoza, Spain	6x	NA	0.525	1025	1000	1	1_5
FLAX80	40.605	-0.327	Castellón, Spain	6x	*L. suffruticosum*	NA	850	100	1	1_5
FLAX81	41.787	-1.703	Zaragoza, Spain	6x	*L. appressum-salsoloides*	0.541	730	300	1	1_5
AA100	34.785	-2.383	Berka, Morocco	4x	*L. Suffruticosum* Morocco	0.514	723	3000	2	2
AA105	32.616	-4.799	Midelt, Morocco	4x	*L. Suffruticosum* Morocco	NA	1721	30	1	1_2
AA107	33.908	-4.031	Taza, Morocco	2x	*L. Suffruticosum* Morocco	0.5	899	3000	1	1_2
AA111	34.876	-3.149	Nador, Morocco	6x	*L. Suffruticosum* Morocco	0.629	471	80	2	2
AA112	34.901	-3.547	Driouch, Morocco	2x	*L. Suffruticosum* Morocco	0.692	530	30	2	2

Table S2. Partial characterization of the 12 SSR loci identified in *Linum suffruticosum* by Olmedo et al. (unpublished), and their inclusion in this study based on their performance with M13 labelling in a preliminary test with 16 individuals from 16 populations. Primer sequences are not printed here for the sake of respecting publication authority.

Locus	Repeat motif	Repeat length	Amplicon size (bp)	Included	M13- labelling
Ls_1145191	TGA	10	116	Yes	6-FAM
Ls_1169143	TG	18	158	No	
Ls_1178187	TTC	14	189	No	
Ls_144692	TTC	23	243	No	
Ls_246481	CCA	11	103	No	
Ls_337128	AG	25	250	Yes	PET
Ls_37372	AC	16	195	Yes	NED
Ls_395648	AC	13	243	Yes	6-FAM
Ls_421659	GGA	8	189	Yes	PET
Ls_807222	TTC	9	173	No	
Ls_889692	GAA	14	192	No	
Ls_9438	GAA	24	247	Yes	VIC

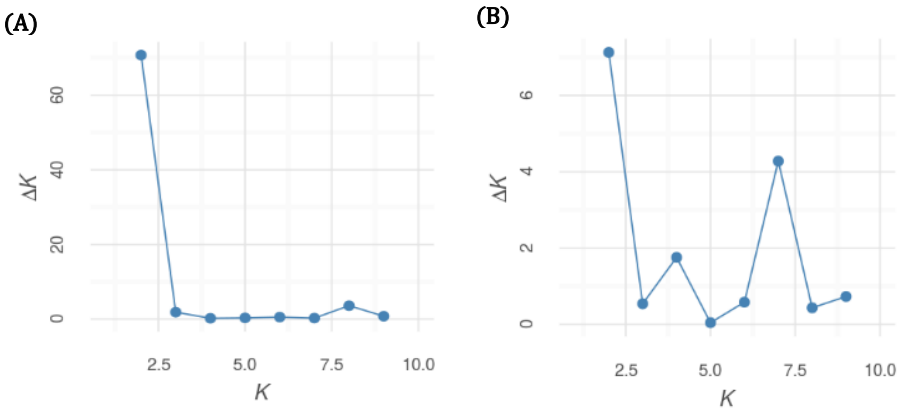

Figure S1. Delta-K plot, calculated based on Evanno et al. (2005) to estimate the optimal number of genetic clusters. (A) Delta-K was calculated from the output of the Bayesian analysis of the population structure (STRUCTURE software (Pritchard et al., 2000)), based on 6 SRRs and including the 32 populations involved in this study. (B) Delta-K was calculated from the output of the Bayesian analysis of the population structure (STRUCTURE software (Pritchard et al., 2000)), based on 6 SRRs and including 29 populations assignet to the Cluster 1 in the former analysis.

Table S3. Moran's I spatial correlation test results. Abbreviations: N_a, Number of alleles per population; N_e, Effective number of alleles per population; H_e, Expected heterozygosity; % L, Proportion of long-styled morph individuals in the population.

Parameter	Observed	Expected	Sd	P-value
N_a	0.034	-0.032	0.054	0.219
N_e	-0.01	-0.032	0.054	0.68
H_e	-0.07	-0.032	0.054	0.481
Altitude (m.a.s.l.)	0.032	-0.032	0.053	0.22
Ploidy level	-0.002	-0.032	0.054	0.58
Population size	-0.061	-0.032	0.045	0.514
% L	-0.083	-0.032	0.053	0.332

Table S4. Genetic diversity parameters' means for each ploidy level (A), and for each genetic cluster (B). Abbreviations: N_a, Number of alleles per population; N_e, Effective number of alleles per population; H_e, Expected heterozygosity.

(A)			
Ploidy level	Mean N_a	Mean N_e	Mean H_e
2	3.979	2.65	0.53
4	5.521	2.907	0.577
6	6.214	3.662	0.648
8	7.333	4.266	0.689
(B)			
Genetic cluster			
1_1	4.056	2.668	0.561
1_2	4.167	2.627	0.588
1_3	2.833	2.026	0.459
1_4	4.5	3.037	0.556
1_5	6.222	3.312	0.55
1_6	6.967	4.228	0.671
1_7	3.833	2.934	0.508
2	4.889	3.12	0.623

Table S5. Tukey HSD post hoc test results. Comparison of the differences in the Number of alleles (i), Effective number of alleles (ii) and Expected Heterozygosity (iii), between ploidy level pairs (A) and genetic cluster pairs (B).

(A)				
Ploidy level pairs	Difference	Lower	Upper	P adjusted
(i) Number of alleles				
4 – 2	0.541	-0.834	1.917	0.707
6 – 2	2.235	0.590	3.880	0.004 *
8 – 2	3.354	-0.017	6.725	0.051 .
6 - 4	1.693	0.253	3.133	0.016 *
8 - 4	2.812	-0.463	6.088	0.112
8 - 6	1.119	-2.278	4.516	0.805
(ii) Effective number of alleles				
4 – 2	0.256	-0.479	0.992	0.776
6 – 2	1.011	0.131	1.891	0.019 *
8 – 2	1.616	-0.186	3.419	0.091
6 - 4	0.754	-0.015	1.524	0.056 .
8 - 4	1.359	-0.392	3.111	0.171
8 - 6	0.604	-1.212	2.421	0.800
(iii) Expected heterozygosity				
4 – 2	0.046	-0.051	0.144	0.570
6 – 2	0.118	0.001	0.235	0.046 .
8 – 2	0.159	-0.080	0.399	0.288
6 - 4	0.071	-0.030	0.174	0.246
8 - 4	0.112	-0.120	0.345	0.557
8 - 6	0.041	-0.201	0.282	0.966

(B)

Genetic clusters pairs	Difference	Lower	Upper	P adjusted
(i) Number of alleles				
1_2-1_1	0.111	-1.665	1.887	1.000
1_3-1_1	-1.222	-4.773	2.329	0.941
1_4-1_1	0.444	-1.546	2.435	0.995
1_5-1_1	2.167	-0.158	4.491	0.080
1_6-1_1	2.911	0.920	4.902	0.001 *
1_7-1_1	-0.222	-3.773	3.329	1.000
2-1_1	0.833	-1.491	3.158	0.928
1_3-1_2	-1.333	-4.821	2.154	0.902
1_4-1_2	0.333	-1.541	2.208	0.999
1_5-1_2	2.056	-0.170	4.281	0.085
1_6-1_2	2.800	0.926	4.674	0.001 *
1_7-1_2	-0.333	-3.821	3.154	1.000
2-1_2	0.722	-1.504	2.948	0.956
1_4-1_3	1.667	-1.935	5.268	0.783
1_5-1_3	3.389	-0.408	7.185	0.105
1_6-1_3	4.133	0.532	7.735	0.017 *
1_7-1_3	1.000	-3.650	5.650	0.996
2-1_3	2.056	-1.741	5.852	0.630
1_5-1_4	1.722	-0.679	4.123	0.297
1_6-1_4	2.467	0.387	4.546	0.012 *
1_7-1_4	-0.667	-4.268	2.935	0.998
2-1_4	0.389	-2.012	2.790	0.999
1_6-1_5	0.744	-1.657	3.146	0.965
1_7-1_5	-2.389	-6.185	1.408	0.452
2-1_5	-1.333	-4.018	1.351	0.720
1_7-1_6	-3.133	-6.735	0.468	0.121
2-1_6	-2.078	-4.479	0.323	0.125
2-1_7	1.056	-2.741	4.852	0.981
(ii) Effective number of alleles				
1_2-1_1	-0.040	-0.936	0.856	1.000
1_3-1_1	-0.642	-2.434	1.150	0.928
1_4-1_1	0.369	-0.635	1.374	0.919
1_5-1_1	0.644	-0.529	1.818	0.614
1_6-1_1	1.560	0.556	2.565	0.001 *
1_7-1_1	0.266	-1.526	2.058	1.000
2-1_1	0.453	-0.720	1.626	0.898
1_3-1_2	-0.602	-2.362	1.158	0.943
1_4-1_2	0.410	-0.536	1.356	0.832
1_5-1_2	0.685	-0.439	1.808	0.491
1_6-1_2	1.601	0.655	2.547	0.001 *
1_7-1_2	0.306	-1.454	2.066	0.999
2-1_2	0.493	-0.630	1.617	0.823
1_4-1_3	1.012	-0.806	2.829	0.599
1_5-1_3	1.287	-0.629	3.203	0.373
1_6-1_3	2.203	0.385	4.020	0.010 *
1_7-1_3	0.908	-1.438	3.255	0.897
2-1_3	1.095	-0.821	3.011	0.568
1_5-1_4	0.275	-0.937	1.487	0.994
1_6-1_4	1.191	0.142	2.241	0.018 *
1_7-1_4	-0.103	-1.921	1.714	1.000
2-1_4	0.084	-1.128	1.295	1.000
1_6-1_5	0.916	-0.296	2.128	0.241
1_7-1_5	-0.379	-2.294	1.537	0.997
2-1_5	-0.191	-1.546	1.163	1.000
1_7-1_6	-1.295	-3.112	0.523	0.305
2-1_6	-1.107	-2.319	0.104	0.091
2-1_7	0.187	-1.729	2.103	1.000

Table S6. Results of the linear regressions testing differences in genetic diversity parameters among ecological and biological variables: Proportion of long-styled morph individuals in the population (i), Population size (ii), Latitude (iii) and Altitude (m.a.s.l.) (iiii). Abbreviations: N_a, Number of alleles per population; N_e, Effective number of alleles per population; H_e, Expected heterozygosity.

Genetic diversity parameter	Adjusted R-squared	P-value
(i) % L-morph individuals		
N_a	0.019	0.212
N_e	-0.002	0.337
H_e	0.023	0.195
(ii) Population size		
N_a	-0.033	0.939
N_e	-0.006	0.377
H_e	-0.033	0.956
(iii) Latitude		
N_a	0.057	0.098
N_e	0.067	0.081
H_e	0.06	0.094
(iiii) Altitude (m.a.s.l.)		
N_a	-0.012	0.443
N_e	-0.024	0.604
H_e	-0.014	0.463