PCP
PLANT & CELL PHYSIOLOGY

Special Issue - Regular Paper

Genetic Isolation among Four Lineages of *Silene nutans*

Zoé Postel[1,2,†,*], Hélène Martin[1,3,†], Camille Roux[1], Cécile Godé[1], Mathieu Genete[1], Éric Schmitt[1], François Monnet[1], Xavier Vekemans[1] and Pascal Touzet[1]

[1]University Lille, CNRS, UMR 8198—Evo-Eco-Paleo, Lille F-59000, France
[2]Department of Ecology, Environment and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden.
[3]Present address: Département des Sciences Animales, Université Laval, Québec, QC, Canada.
[†]These authors contributed equally to this work.

*Corresponding author: E-mail, zoe.postel@su.se

Received 7 April 2024; Accepted 24 September 2024

Speciation is the process leading to the emergence of new species. While being usually progressive, it can sometimes be fast with rapid emergence of reproductive barriers leading to high level of reproductive isolation. Some reproductive barriers might leave signatures in the genome, through elevated level of genetic differentiation at specific loci. Similar signatures might also be the results of linked selection acting in low recombination regions. Nottingham catchfly (*Silene nutans*) is a Caryophyllaceae species composed of four genetically differentiated lineages for which strong and asymmetric levels of reproductive isolation have been identified. Using population transcriptomic data from several individuals of the four lineages, we inferred the best evo-demographic scenario leading to the current reproductive isolation of these four lineages. We also tested whether loci exhibiting high level of genetic differentiation represented barrier loci or were located in low recombination regions, evolving under strong influence of linked selection. Overall, the four lineages of *S. nutans* have diverged in strict isolation, likely during the different glacial period, through migration in distinct glacial refugia. Speciation between these four lineages appeared to be particularly fast, likely due to fast evolving plastid genome accelerating plastid-nuclear co-evolution and the probability of plastid-nuclear incompatibilities in inter-lineage hybrids.

Keywords: Evo-demographic scenario • DILS • Plant speciation • Plastid-nuclear incompatibilities • Reproductive isolation

Introduction

Speciation is the process by which one species composed initially of different populations evolves into separate species entities, i.e. reproductively isolated to each other (Coyne and Orr 2004, Butlin and Stankowski 2020, Matute and Cooper 2021). Speciation is a gradual process (Seehausen et al. 2014, Stankowski and Ravinet 2021). Its establishment involves the emergence of successive reproductive barriers preventing gene flow between previously connected populations and ultimately leading to total reproductive isolation (RI) between these populations (Feder et al. 2012, Lowry 2012, Kulmuni et al. 2020). These barriers can act before (pre-zygotic) or after (post-zygotic) reproduction between individuals from different populations (Coughlan and Matute 2020). While alternative modes of speciation were classically identified based on the spatial context during the evolution of the new species (i.e. allopatric, sympatric, and parapatric) (Coyne and Orr 2004), the process of speciation is now rather considered as a continuum of divergence with inference of speciation scenarios addressing the following questions, integrating the spatial context rather than being defined by it: did the populations diverge while exchanging gene flow? Are there any secondary contacts? Are there any signs for local adaptation that might have driven divergence? What are the demographic events that might have enhanced or not speciation? (Butlin et al. 2008). These speciation scenarios should leave distinguishable signatures in the genome and can thus be inferred using molecular data (Feder et al. 2012, Seehausen et al. 2014, Roux et al. 2016, Ravinet et al. 2017). The reproductive barriers involved in the speciation process will also leave marks in the genomes. If speciation is primarily due to post-zygotic reproductive barrier, RI is thought to be the result of a mismatch between independently evolved allelic combinations in isolated lineages, called Bateson–Dobzhansky Muller incompatibilities (BDMIs) (Presgraves 2010, Fishman and Sweigart 2018). Because of that, if at some point these lineages cross together, while the rest of the genome will be a mix of the parental lineages, these loci will not and will thus remain highly differentiated (Wu 2001, Burri et al. 2015). This would give rise to 'speciation Islands' representing regions/loci in the genome where introgression does not occur and that can reach high levels of genetic differentiation (Turner et al. 2005, Feder et al. 2012, Seehausen et al. 2014). But recurrent selection in low recombination zones, either negative against deleterious

Plant And Cell Physiology, 2025, 66(4), 514–528
doi:https://doi.org/10.1093/pcp/pcae110
Advance Access publication on 25 September 2024

mutations, i.e. background selection, or positive for adaptive mutations, i.e. local adaptation, induces a reduction in polymorphism causing patterns of high genetic differentiation, and thus can be mistaken as barriers to introgression (Cruickshank and Hahn 2014, Shang et al. 2023). Distinguishing between these two processes is thus a key point when studying speciation.

Nottingham catchfly (*Silene nutans*) is a Caryophyllaceae species largely distributed in Europe. Previous studies identified two strongly differentiated evolutionary lineages within this species in relation to past climatic event and post-glacial recolonization: an eastern one (E1) widespread in the north of Europe (e.g. western Europe, Great Britain) and a western one, composed of three sub-lineages: W1 distributed in England/France and Belgium, W2 restricted to Spain and south-western France and W3 in the Alp and Italy (**Fig. 1**) (Martin et al. 2016, Van Rossum et al. 2018). These four lineages have specific plastid haplotypes. At secondary contact zones between lineages E1 and W1, in France, southern Belgium and in south England, no hybridization events were detected suggesting the absence of gene flow between these lineages (Van Rossum et al. 1997, 2018). These two lineages also show strong and complete post-zygotic reproductive isolation, expressed as high proportion of seedling chlorosis and reduced hybrid fitness. It could be the result of genetic incompatibilities accumulated in allopatry prior to the spread of lineages from their glacial refugia (Martin et al. 2017). Results from diallelic crosses between the four lineages (i.e. E1/W1/W2/W3) also suggested the presence of strong and asymmetric post-zygotic barriers not only between E1 and the Western sub-lineages but also between the western sub-lineages (Van Rossum et al., manuscript in preparation). The intensity of these barriers was found to depend on which lineage was the cytoplasm donor, with especially high strength when lineages E1 or W2 are the donors (Van Rossum et al., manuscript in preparation). Plastid-nuclear incompatibilities

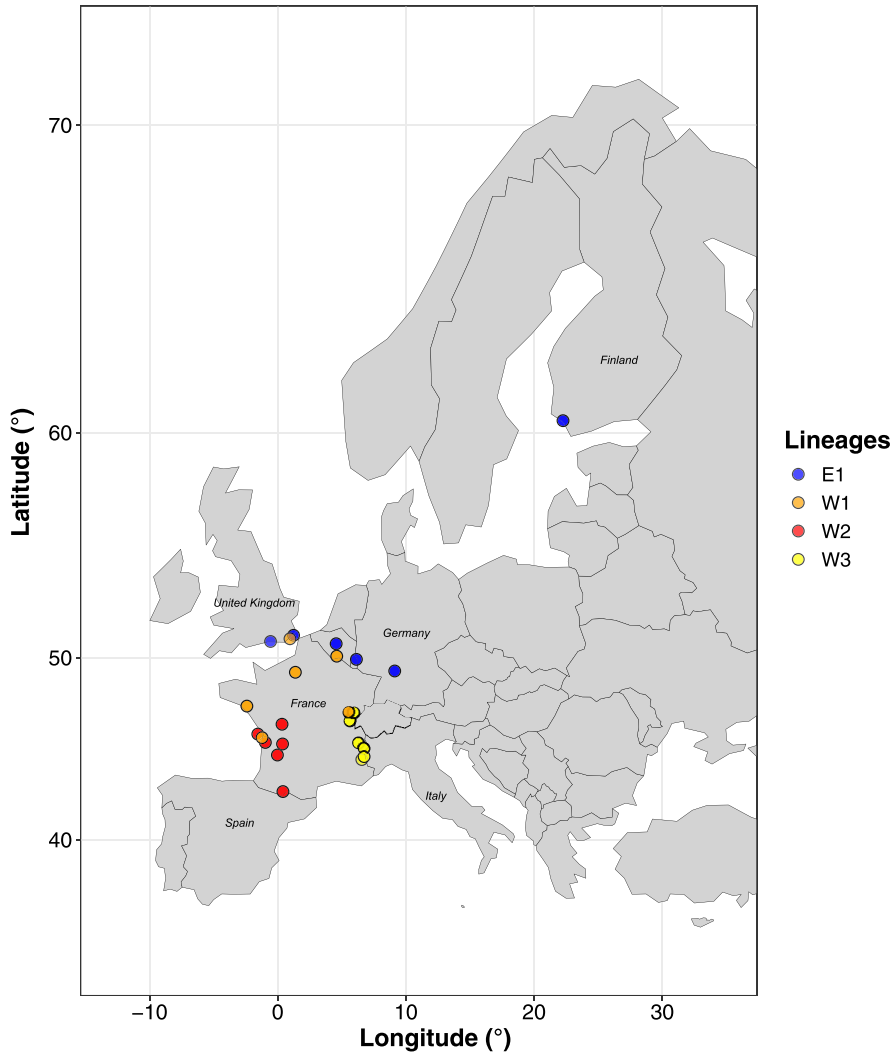

Fig. 1 Geographic locations of all sampled and sequenced individuals for the four lineages. *Silene nutans* is native in Europe and is widely distributed all over the European countries. On the countries represented on this figure, *S. nutans* is present in each of them except Ireland (Van Rossum et al. 2018).

(PNIs), a type of BDMIs, could be involved in the speciation process at stake between these four lineages (Postel et al. 2022). Overall, these studies suggest that *S. nutans* might be composed of four potentially cryptic species, but this should be confirmed by studying patterns of genomic differentiation among them.

In this study, using an Approximate Bayesian Computation (ABC) approach, we inferred the demographic history and patterns of differentiation among the four lineages of *S. nutans*. Specifically, we estimated the relative times of split of the lineages and tested for gene flow between pairs of lineages, despite the identification of strong reproductive barriers as described earlier. We also tried to discriminate among different demographic and speciation scenarios by analyzing patterns of covariation of different statistics among loci.

Results

We inferred evo-demographic scenarios for the four lineages of *Silene nutans* with DILS—Demographic Inference with Linked Selection, a recently developed software that implement ABC approaches (Fraïsse et al. 2021), using transcriptomic data from 11–12 individuals per lineage (**Fig. 1**, **Supplementary Table S1**). Very briefly, DILS analyzes a random sample of 1000 nuclear loci from the whole dataset and compute several summary statistics. After simulation of datasets under distinct evo-demographic scenarios, based on a set of prior, DILS compares summary statistics values between simulated and observed datasets to infer the most plausible scenario, under which model parameters such as population effective size (*Ne*) and time of split (T_{split}) between lineages, will be estimated. We ran in two different ways for two different purposes. First, we ran the regular version of DILS, with two populations, to estimate the summary statistics and the best speciation scenarios for each pair of lineages. Secondly, we ran the four populations version of DILS to estimate the model parameters under the best speciation scenario (**Fig. 2**). Using DILS-2 populations to find the best speciation scenarios, we reduced the number of models to compare and using DILS-4 populations to estimate

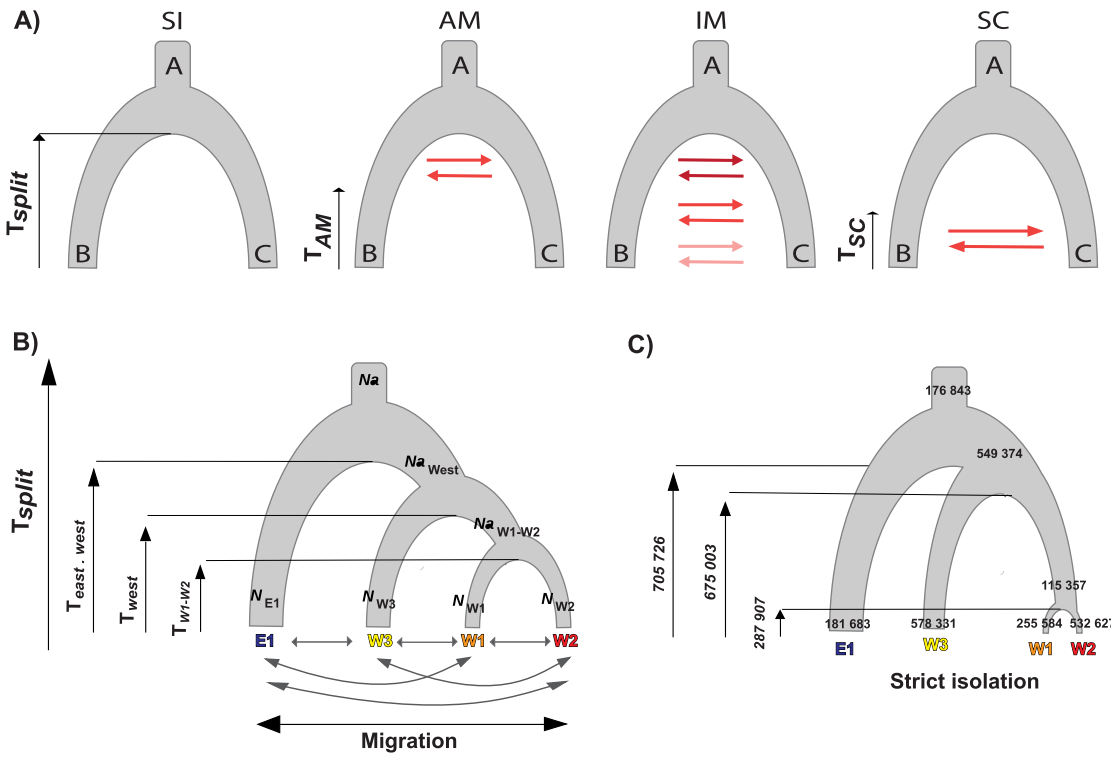

Fig. 2 Demographic model used in DILS 2 and 4 populations. (A) Alternative models for DILS 2 populations. SI—strict isolation: model of divergence without gene flow since time of split (T_{split}); AM—ancient migration: model of divergent with evidence for ancestral gene flow but none since their cessation (T_{AM}); IM—model of divergence with gene flow; SC—secondary contact: model of divergence without gene flow but current migration between populations through secondary contact (T_{SC}). Letters A represents the ancestral populations. The arrows represent the gene flow between extant populations B and C and the color gradient the level of gene flow (from dark red—high level—to light—low level). (B) Using a tree topology from a former study, four populations of the four lineages have been modeled here, each with independent current population sizes (annotated as *N*). The model describes three successive splitting events ($T_{W1–W2}$, T_{West} and $T_{East-West}$) up to the ancestral population. Each ancestral population is associated with an effective size independent of daughter populations (N_a). The six possible migration relationships are bidirectional and asymmetric secondary contacts (the gray arrows). (C) Estimation of the T_{split}, *N* and N_a under the best demographic scenario. Confidence interval for these estimations can be found in **Supplementary Table S6**. T_{split} is expressed in years (considering that one generation = 3 years). The best scenario was divergence in strict isolation for all four lineages so all of the migration's arrows are suppressed.

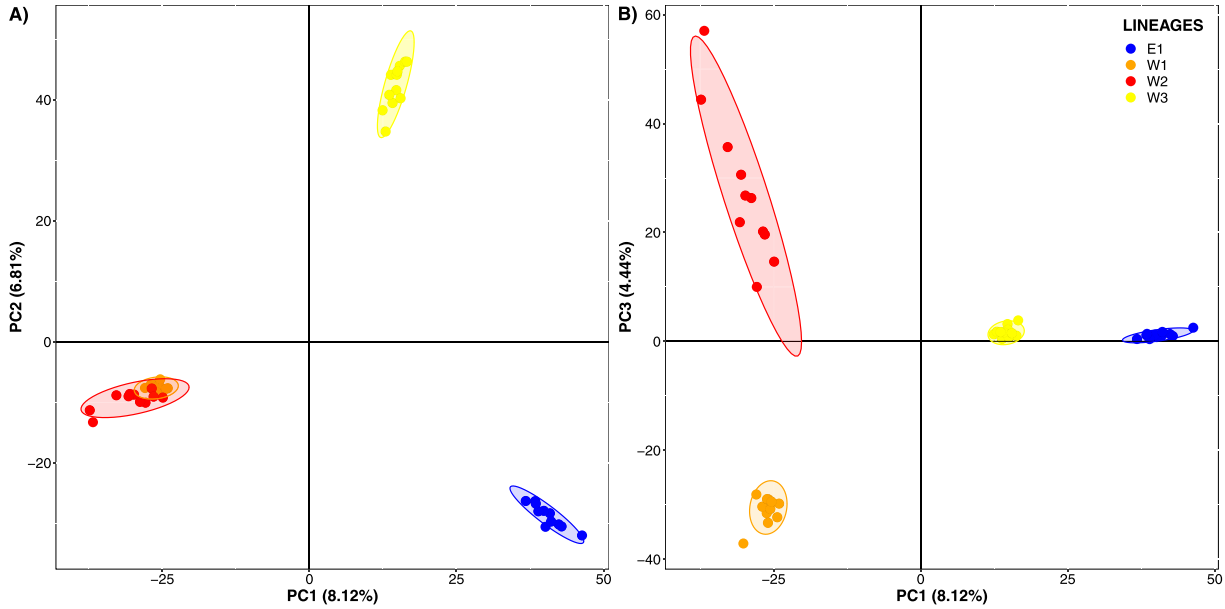

Fig. 3 Individuals coordinates on the PC1, PC2 and PC3. Eigen values for each PC axes are given.

the model parameters, we reduced the number of estimations to conduct.

Prior to run ABC using DILS (Fraïsse et al. 2021), we checked that the four lineages clustered separately in a Principal Component Analysis (PCA) with the 1,251,229 SNPs identified with reads2snps. Overall, lineages clustered separately along PC1, PC2 and PC3 (**Fig. 3**). Lineages W1 and W2 were undistinguishable in a PC1–PC2 plot, with W1 individuals clustering within W2, but they appeared well differentiated according to PC3 (**Fig. 3**).

Population genetic statistics

Summary statistics were estimated using DILS-2 populations for each lineage and pair of lineages, under the best speciation scenarios. Differences between lineages or comparison were tested using statistical Kruskal–Wallis test followed by Dunn test when the former was significant. Here after, we only present a subset of all summary statistics computed. For the rest of the summary statistics, results can be found in **Supplementary data**.

Lineage W3 significantly exhibited the highest level of genetic diversity (measured by the pairwise nucleotide diversity π) while lineage E1 the lowest (mean W3 = 0.0120 versus mean E1 = 0.0101, **Fig. 4**, **Supplementary Table S3**). Values of Tajima's D showed the exact opposite patterns, each difference between lineages being statistically significant (mean E1 = −0.23 versus mean W3 = −0.47) (**Fig. 4**, **Supplementary Table S3**). Levels of mean absolute divergence (D_{xy}) between pairs of lineages were the highest between lineages E1 and W1 (mean = 1.7e-2) and the lowest between lineages W1 and W2 (mean = 1.2e-2) (**Fig. 4**, **Supplementary Table S4**). Overall, lineage E1 was significantly the most divergent lineage and W3 had similar level of divergence with all the others, lower than E1 but higher than between W1 and W2. Genetic differentiation between lineages, F_{ST}, also followed the exact same trend, with all differences between pairs of lineages being significant (**Fig. 4**, **Supplementary Table S4**).

In order to assess the impact of linked selection on the speciation process between these four lineages we analyzed patterns of covariation of F_{ST}, D_{xy}, average Rho, average π among loci. Indeed, large values of F_{ST} are caused by the occurrence of strong local reproductive barriers in a scenario of divergence with migration, whereas they are caused by local reduction of polymorphism associated with low recombination in a scenario of linked selection in the absence of gene flow (Cruickshank and Hahn 2014, Shang et al. 2023). For the divergence between all western lineages, we found a positive significant relation between D_{xy} and F_{ST} ($mean\ tau = 0.32$) and between D_{xy} and π ($mean\ tau = 0.82$) while we detected a negative significant relation between F_{ST} and π ($mean\ tau = 0.14$) (**Fig. 5**, **Table 1**). A strong positive relationship between D_{xy} and π is expected in a scenario with linked selection in the absence of gene flow, whereas the reverse is expected under divergence with migration, suggesting that the former scenario is more likely. This scenario is also validated by the observation of a significant negative relation between F_{ST} and Rho ($mean\ tau = 0.07$), suggesting that local recombination rate is an important determinant of genetic differentiation through its effect on linked selection. Similar results were obtained between lineage E1 and the western lineages, though without statistical significance for the negative correlation between F_{ST} and Rho (**Fig. 5**, **Table 1**).

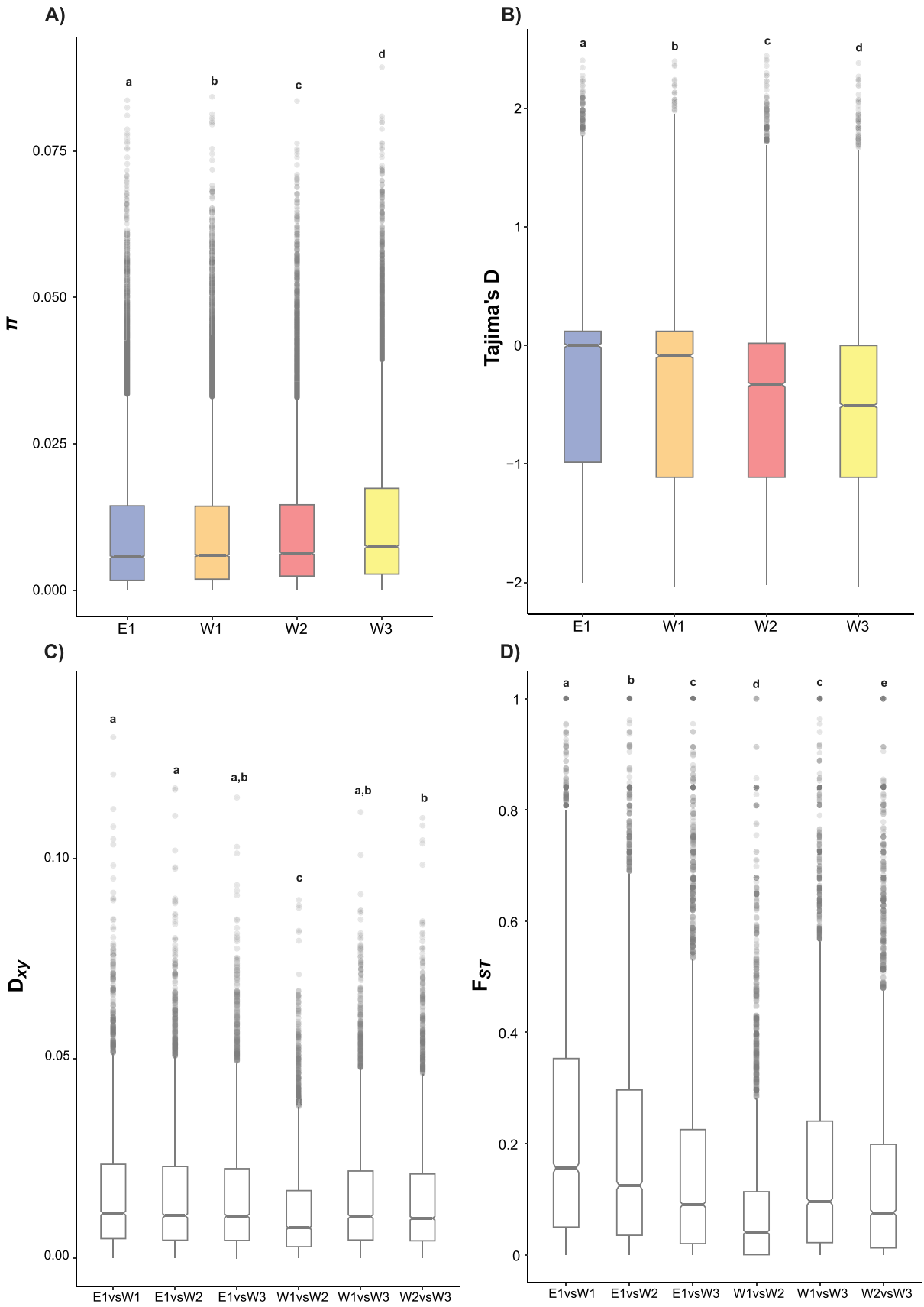

Fig. 4 Summary statistics values for each pairwise combinations and lineage. (A) Tajima D values per lineage. (B) Neutral genetic diversity, π, per lineage. (C) D_{xy} for each lineage's comparison. (D) F_{ST} for each lineage's comparison. Blue: lineage E1; orange = lineage W1; red = lineage W2; yellow = lineage W3. Results of the Dunn tests are given. Detailed result values and means for the other summary statistics can be found in **Supplementary Table S3 and S4**.

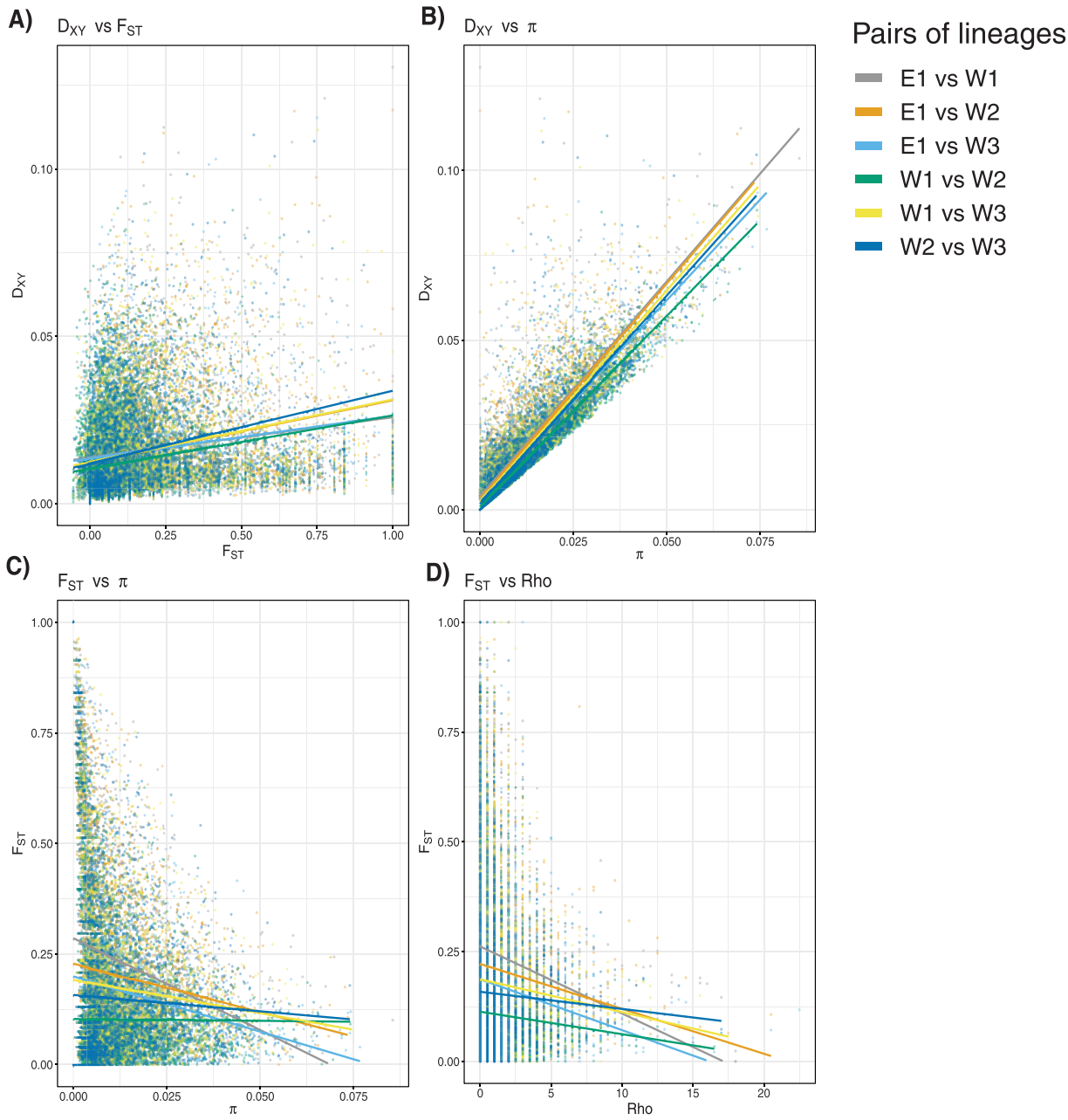

Fig. 5 Correlations between F_{ST}, D_{xy}, Rho, π. Detailed mean values for each of the comparison can be found in **Table 1**. (A) Correlation between F_{ST} and D_{xy}; (B) Correlation between D_{xy} and π; (C) Correlation between F_{ST} and π; (D) Correlation between F_{ST} and Rho.

The selfing rates of the different lineages were inferred from estimates of identity disequilibrium following the method proposed by David et al. (2007), E1 had the highest selfing rate (0.22), then W1 (0.1), W2 (0.06) and W3 (0.02).

Evo-demographic model selection

We ran DILS-2 populations for each pair of lineages ($n = 6$). For each, we selected the best evo-demographic model explaining the observed dataset looking at the posterior probabilities for each model and the goodness-offit produced as an output of DILS (**Fig. 2A**). For each of the six pairs, the best model was found to fit correctly to the data: on the PCA representing the goodness-of-fit for the best model, the observed dataset was found roughly in the middle of the optimized posterior distribution (**Figure S1**) when looking at PC1, 2 and 3, meaning that values of the summary statistics obtained in simulations of the best model were similar to those of the observed dataset. Equivalent results were obtained for the analysis using DILS with a model of four populations (**Figure S2**).

Table 1 Correlation test calculating Tau coefficient between D_{xy}, F_{ST}, average π and average Rho for each pair of lineages

		Pairs of lineages					
		E1 vs W1	E1 vs W2	E1 vs W3	W1 vs W2	W1 vs W3	W2 vs W3
D_{xy} vs F_{ST}	Tau coeff.	0.29	0.34	0.30	0.28	0.34	0.36
	Z-stats	25.49	29.91	26.34	24.23	29.8	31.25
	P-value	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
F_{ST} **vs** π	Tau coeff.	1.4e-3	0.09	0.08	0.15	0.12	0.16
	Z-stats	0.13	7.80	6.88	13.00	10.18	13.94
	P-value	0.90	0.00***	0.00***	0.00**	0.00***	0.00***
D_{xy} **vs** π	Tau coeff.	0.71	0.74	0.78	0.87	0.78	0.80
	Z-stats	62.62	65.85	68.52	76.57	68.30	70.74
	P-value	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
F_{ST} **vs** Rho	Tau coeff.	−1.09e-3	0.02	0.02	0.06	0.05	0.10
	Z-stats	−0.08	1.17	1.14	4.27	3.42	7.23
	P-value	0.93	0.24	0.25	0.00***	0.00***	0.00***

Average Rho and π were calculated as the mean of π or Rho estimated per lineage and per loci in each comparison. F_{ST}: genetic differentiation between lineages; D_{xy}: absolute divergence between lineages; π: mean genetic diversity of the two lineages compared; Rho: mean recombination rate of the two lineages compared.
***: *P*-value < 0.001.

Table 2 Best demographic model assessed with DILS using a two-step process

Pairs of lineages	Ongoing migration versus isolation		Ancient migration versus strict isolation	N-Homo versus N-Hetero
	DILS 2 pops	DILS 4 pops	DILS 2 pops	DILS 2 pops
E1 vs W1	Isolation	Isolation	Ancient migration	N-hetero
	0.885	0.898	0.711	0.997
E1 vs W2	Isolation	Isolation	Ancient migration	N-hetero
	0.912	0.929	0.764	0.997
E1 vs W3	Isolation	Isolation	Ancient migration	N-hetero
	0.915	0.940	0.859	0.992
W1 vs W2	Isolation	Isolation	Ancient migration	N-hetero
	0.832	0.857	0.784	0.989
W1 vs W3	Isolation	Isolation	Ancient migration	N-hetero
	0.892	0.894	0.750	0.979
W2 vs W3	Isolation	Isolation	Ancient migration	N-hetero
	0.876	0.890	0.800	0.979

Number under the name of the demographic model represents the posterior probabilities, from 0 to 1. Ancient Migration vs Strict Isolation was not tested with the DILS using four populations. N-homo/N-hetero: heterogeneous or homogenous population effective size.

Posterior probabilities for the best demographic model in all model comparisons and all pairs of lineages ranged from 0.711 to 0.997, indicating unambiguous support for the best model (**Table 2**). Following DILS hierarchical selection process, we first compared posterior probabilities between models with current isolation (SI—strict isolation + AM—ancient migration) versus those with ongoing migration (IM—isolation with migration + SC—secondary contact) (**Fig. 2A**). Overall, for all pairs of lineages, the best model was always SI + AM, supporting the absence of current gene flow among all four lineages. Secondly, between the two scenarios without current migration (SI versus AM) (**Fig. 2A**), the AM model, a two-step scenario with gene flow occurring during the first step after the split, followed by total isolation in the second phase, seemed to better explain the observed dataset (**Figure S1**). Finally, we tested for heterogeneity versus homogeneity of N_e among loci. Again, for all pairs of lineages, the best speciation models were the one including variation of N_e (i.e. model $N_e$hetero). Analyses using the DILS-4 populations produced the same overall result (though we could not test heterogeneity of N_e): a scenario with strict isolation among all four lineages was retained with posterior probabilities ranging from 0.86 to 0.94 (**Fig. 2**, **Table 2**).

Demographic models' parameter estimation

We estimated demographic parameters for each of the six pairwise analyses between E1, W1, W2 and W3 (**Supplementary Table S5**). This approach yields six values for ancestral population sizes at six different times. Additionally, we obtained three distinct estimates of effective population sizes, as each population was considered in three separate analyses. To simplify the interpretation of such estimates, we have estimated the demographic history of a scenario that directly includes these four populations using DILS-4 populations and under a speciation scenario of strict isolation since time of split. Lineage W3 seems to exhibit the highest N_e followed by lineages W2 > W1 > E1 (**Fig. 2**, **Supplementary Table S6**). The ancestral population of all lineages displays relatively low N_e compared to current lineages N_e, similarly to the N_e of the ancestral population of lineages W1 and W2 (**Fig. 2**, **Supplementary Table S6**). The ancestral population of the western lineages exhibits one

of the highest N_e (i.e. around 550 000 individuals). Regarding time of split (T_{split}), with a generation time of 3 years (Hepper 1956), eastern and western lineages split around 700000 years ago (IC95= 283,695–1,363,743) (**Fig. 2**, **Supplementary Table S6**). Lineage W3 separated from lineages W1 and W2 soon after [around 680,000 years ago (IC95= 202,092–1,204.050)] (**Fig. 2**, **Supplementary Table S6**). Finally, the split between lineages W1 and W2 occurred the most recently, around 300,000 years ago (IC95 = 82,248–782,862) (**Fig. 2**, **Supplementary Table S6**).

Discussion

Allopatric speciation for all lineages of *S. nutans*

Regarding DILS results with both 2 and 4 populations, highest posterior probabilities were observed for models of divergence without recent gene flow for all four lineages. Lineage E1 separated first, followed shortly after by W3 and then, more recently, by W1 and W2. The estimated small split time between E1 and the rest of the western lineages might suggest either incomplete lineage sorting or use of an incorrect topology between the four lineages. Yet, when randomly sampling nuclear loci or building a phylogeny with chloroplast data of the four lineages in former studies, the associated topology was correctly resolved: [E1, (W3, (W1, W2)] (Postel et al. 2022, 2023b). Estimated split times with DILS-4 populations likely match glacial periods of Quaternary glaciation cycles, indicating that divergence might have occurred as a consequence of geographical separation in distinct glacial refugia (Martin et al. 2016, Van Rossum et al. 2018) as found in many plant lineages (Kadereit and Abbott 2021). The time of split between lineages E1 and the rest of the western ones is compatible with the Günz glaciation, that started around 760,000 years ago and ended around 530,000 years ago. As derived from the current distribution of the lineages, and knowledge about pathways of post-glacial recolonization pathways in Europe (Kadereit and Abbott 2021) refugia for E1, W3 and the direct ancestor of W1-W2 would have been located, respectively, in Eastern Europe, in the Italian Peninsula, and Western Europe (Martin et al. 2016) (**Fig. 6**). Subsequent separation occurred then between lineages W1 and W2 around 300,000 years ago, which corresponds to the early stage of the Riss glacial period, possibly in association with geographical separation between W1, in south-central France (the French Massif Central area) and W2 in the Iberian Peninsula (Van Rossum et al. 2018) (**Fig. 6**). During the most recent postglacial recolonization period, lineages W2 and W3 recolonization seemed to have been restricted to south-western France for W2 and the Alps and south-eastern Europe for W3 (Van Rossum et al. 2018) (**Fig. 6**). W1 could have recolonized from the Massif Central area toward northern Europe, reaching its northern margins in southern England and Belgium (Van Rossum et al. 2018). Lastly, E1 followed a north-west expansion in Europe, having apparently the widest European geographical distribution as compared to the western sub-lineages (Van Rossum et al. 2018) (**Fig. 6**). Despite recent geographical contact between some of the lineages, we could not detect evidence of recent introgression, confirming the evolution of strong reproductive barriers among all lineages in allopatry during the glacial periods (Postel et al. 2024). Therefore, the four lineages of *Silene nutans* can be considered as genuine species.

When looking at the summary statistics results, Tajima's D values are negative but near zero for lineages E1 and W1 and the lowest for W2 and W3. Negative average values for Tajima's D usually reflect population expansion after colonization events. Moreover, levels of genetic diversity (π) and effective sizes (N_e) are higher for lineages W2 and W3 as compared to lineages E1 and W1 (W3 > W2 > W1 > E1). Both the signatures of population expansion and values of N_e are at odds with the currently known areas of geographical distribution of *S. nutans* lineages showing almost opposite patterns (E1 > W1 > W3 > W2). This could perhaps be due to different intensities of genetic bottlenecks during glaciation periods, or due to differences in mating systems, as such processes would leave similar signatures in the genomes (Wright et al. 2013, Burgarella and Glémin 2017, Hartfield et al. 2017). *Silene nutans* is mainly outcrossing, yet it is self-compatible with some occurrence of selfing (Van Rossum and Prentice 2004, Vanderplanck et al. 2020). Consistently, selfing rate estimations followed the exact opposite pattern of the N_e values above, suggesting that occasional selfing of *S. nutans* lineages could explain our estimations of N_e. The proportion of females in gynodioecious populations that favors outcrossing could also be variable among the lineages. Variation in our sampling strategy across lineages might also have influenced our estimates of N_e. It is well known for European plant lineages that current diversity in regional samples may depend on their geographical distance from the last glacial refugia (Hewitt 2000, Petit et al. 2003). In the present study, irrespective of the lineage, we sampled individuals in the marginal range area of the lineages. However, their distribution is wider and goes beyond those sampled areas (Van Rossum et al. 2018). For lineage W1, we sampled individuals near the glacial refugia and beyond, spanning almost the whole geographical distribution of this lineage (Martin et al. 2016, Van Rossum et al. 2018). However, it is not the case for E1 as we only sampled populations at the margin of its geographical repartition and far from its glacial refugia, as highlighted with the relatively lower allelic richness of individuals from our sampled populations (Martin et al. 2016, Van Rossum et al. 2018). Hence, our estimate of N_e in E1 could have been biased downwards, due to restricted sampling. For W3, its geographical distribution extends in south eastern Europe and we only sampled individuals close to its glacial refugia in the Alps, with high allelic richness of the W3 populations (Martin et al. 2016, Van Rossum et al. 2018). This might be also true for W2 though we also sampled individuals more distant from the W2 glacial refugia in the Pyrenees (Van Rossum et al. 2018). Alternatively, the high diversity observed for lineage W3 might suggest that its *census* size is higher than for the other lineages, or local population structure is higher, which is a feature often seen in current populations close to their glacial refugia

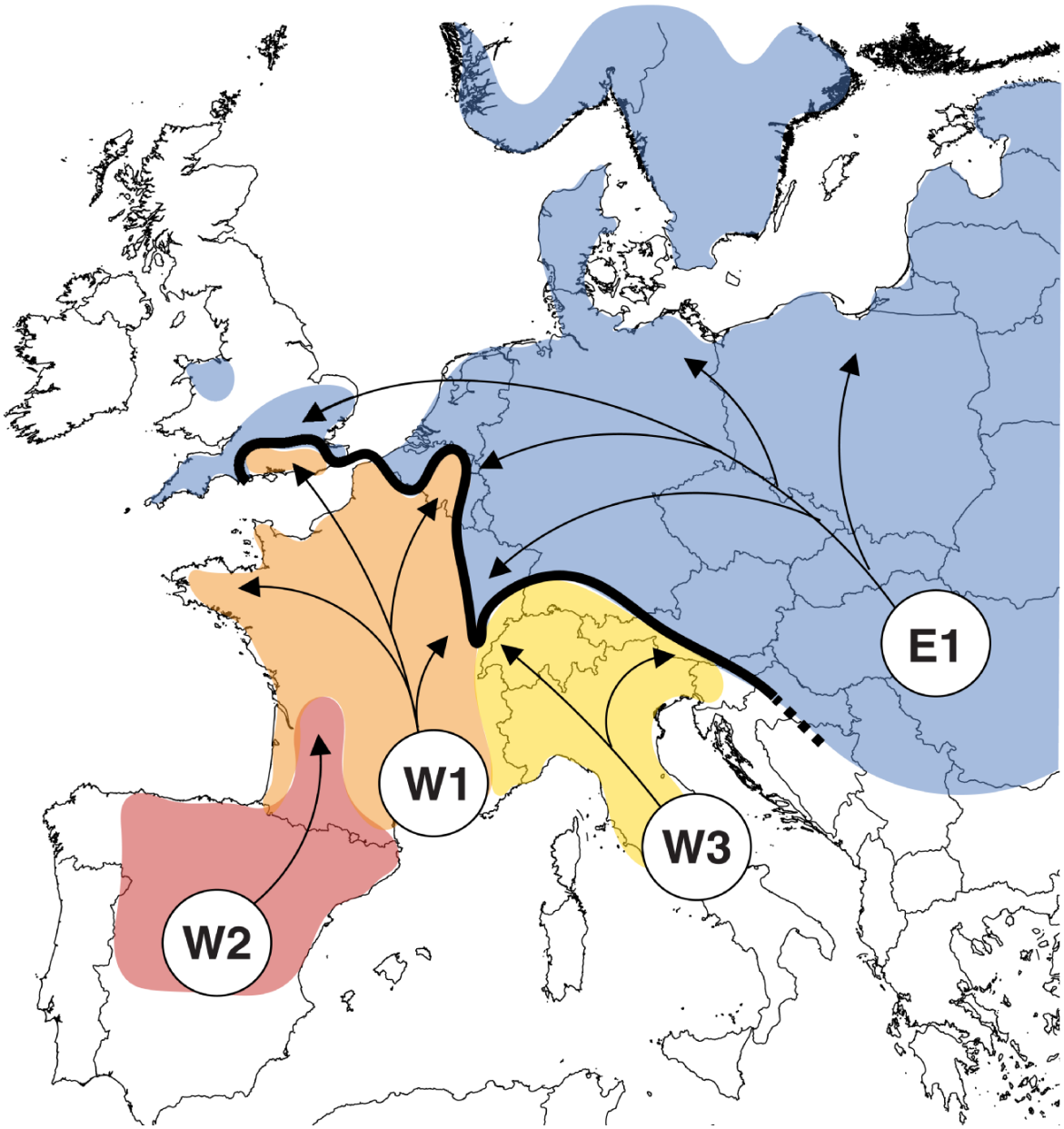

Fig. 6 Evo-demographic scenarios for the four lineages of *S. nutans*. The different colors correspond to the different lineages (blue = E1; orange = W1; red = W2; yellow = W3). The arrows represent the post-glacial expansion of each lineage from their glacial refugia (circles) to their extrapolated distribution (colored areas). The thick black line represents the contact zone between lineages E1, W1 and W3.

(Heuertz et al. 2004). The latter would be consistent with the geographical distribution of this lineage in mountain habitat which might increase local differentiation between populations.

Rapid speciation between lineages of *S. nutans*

The current distribution of the four lineages of *S. nutans* shows secondary contact zones at least between lineages E1 and W1 in south of England and south of Belgium, as well as potential overlap between geographical ranges of lineages W1 and W2 in central-western France (**Fig. 1**). Yet, at these contact zones and overlapping areas no hybridizing events are detected from our analyses, which is in agreement with the strong post-zygotic reproductive isolation demonstrated from the diallelic crosses (Postel et al. 2024). Several reproductive isolation mechanisms could have played a role in the overall pattern of ongoing speciation in the lineages of *S. nutans*. Regarding the pre-zygotic barriers, no evidence for pollinator isolation (premating barrier) was identified between lineages E1 and W1 at secondary contact zone in south of Belgium: specialized pollinators were found on individuals of both lineages and pollen flow between/within lineages were similar (Cornet et al. 2022). Pollen-stigma incompatibilities were identified between these

Belgian lineages, potentially representing a post-pollination pre-zygotic reproductive barrier with pollen tubes having less probability to develop when resulting from inter-lineages pollination (Van Rossum et al. 1996). Lineages E1 and W1 also exhibit morphological differences and different flowering time (De Bilde 1973, Van Rossum 2000), which might also generate pre-zygotic reproductive barrier (Baack et al. 2015). These two lineages represent distinct edaphic ecotypes in Belgium, lineage E1 being specialized on calcicolous soil and W1 on siliceous ones (De Bilde 1973, Van Rossum et al. 1996, Van Rossum 2000). This ecotypic specialization seems to have occurred following the divergence of both lineages in allopatry (De Bilde 1973, Martin et al. 2016, Van Rossum et al. 2018). Soil specialization and habitat isolation might represent a pre-zygotic barrier, with local adaptation to different soil type playing a role in the creation of genetic incompatibilities between maladapted alleles in the hybrids, or alternatively, association of soil specialization alleles with other genetic incompatibilities could have resulted from a coupling effect (Bierne et al. 2011). Finally, post-zygotic reproductive isolation might also be the result of PNIs, with disruption of lineage-specific plastid-nuclear co-adaptation in hybrids (Postel et al. 2022). Even though divergence between lineages of *S. nutans* seems to be recent given the split times, repeated bottlenecks or founding events experienced during post-glacial recolonization could have shaped the plastid genetic diversity independently in each lineage, increasing probability to observe lineage-specific plastid-nuclear co-adaptation and PNIs in hybrids (Postel et al. 2022).

Speciation is supposed to be gradual but can sometimes be fast (Nosil et al. 2017, Stankowski and Ravinet 2021). Rapid speciation may result from various processes such as adaptation to rapid environmental turnover, genome rearrangement generating genetic incompatibilities, polyploidization, founder events... (Nosil et al. 2017). A survey across many pairs of animals' species highlighted the presence of a gray zone of speciation along the speciation continuum, where for a specific level of net divergence $D_a = 10^{-2}$, levels of gene flow between pairs of populations could span the entire range from 0—no gene flow, to 1—gene flow (Roux et al. 2016). Within this gray zone, defining whether pairs of populations belong or not to the same species is challenging. For lower levels of net divergence, pairs of populations belong to the same species and for higher levels, to different species. When mapping our pairs of *S. nutans* lineages on the gray zone figure (**Fig. 7**), for overall low levels of D_a, probability of ongoing migration was quite low and lower than for similar level of net divergence in animal pairs. Except for comparison between W1/W2 and E1/W1, where distinction between 'true' species and semi-isolated one is ambiguous, lineages can be considered as distinct species (**Fig. 7**). This might indicate a rapid speciation in *S. nutans*. Consistently, level of RI does not seem to reflect levels of genetic divergence between lineages. When crossed in both directions, these lineages exhibit various levels of hybrid mortality and chlorosis (Postel et al. 2024): except with E1 being the most divergent and leading to highest levels of RI when crossed, mortality is higher between W1 and W2 that are less divergent from one another than when crossed with W3 which is more genetically dissimilar. This highlights rapid evolution of reproductive barriers between lineages as for very low levels of divergences, estimated level of RI is strong. This is also striking when comparing with level of divergence and migration between two distinct species of *Silene*, *Silene latifolia* and *Silene dioica*. These two species have separated around 120,000 years ago and under a scenario of speciation with gene flow (Liu et al. 2020). Between

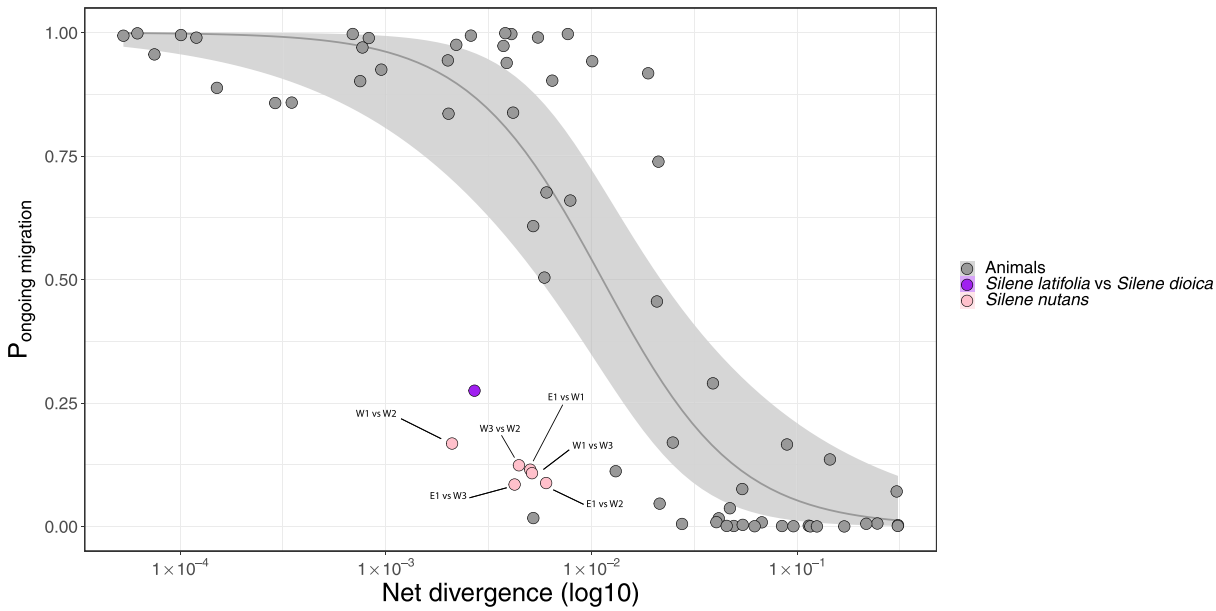

Fig. 7 Gray zone of speciation. Probability of ongoing migration estimated using DILS-2 populations against the net divergence of neutral sites (D_a). Data for animals come from Roux et al. (2016). Data for *Silene latifolia* versus *Silene dioica* come from Liu et al. (2020).

this pair of species, values of D_a are similar to those estimated between lineages of *S. nutans* (i.e. $D_a = 2.7 \times 10^{-3}$ between *S. latifolia* and *S. dioica* and on average $= 4.44 \times 10^{-3}$ between *S. nutans* lineages). For similar levels of absolute genetic divergence, probability of ongoing migration between lineages of *S. nutans* is lower. Yet, a recent study seems to highlight that speciation could be overall faster in the plant kingdom compare to animal (Monnet et al. 2023).

As discussed earlier, former studies using these lineages identified potential involvement of PNIs reducing hybrid viability and fertility (Postel et al. 2022, 2023a, 2024). There is growing evidence for involvement of such PNIs in RI, especially as they might be one of the first post-zygotic reproductive barriers (Barnard-Kubow et al. 2016, Postel and Touzet 2020). Their study might be complicated by the fact that plastid genes are generally strongly conserved (Jansen et al. 2007), so lineage-specific co-adaptation between plastid and nuclear genes might require time of isolation long enough to accumulate divergence between lineages. But in some angiosperm lineages, and especially in the genus *Silene*, acceleration of the rate of evolution of the plastid genome has been identified (Barnard-Kubow et al. 2014, Sloan et al. 2014, Ruhlman and Jansen 2018, Shrestha et al. 2019). This acceleration might speed up plastid–nuclear co-adaptation and divergence at these nuclear and plastid genes in isolated lineages: if evolution of the plastid genes is higher, the interacting nuclear loci will follow up, increasing divergence between lineages but potentially only at these nuclear and plastid genes. In lineages of *S. nutans*, we observed a pattern of high diversity of the plastid genes, with numerous lineage specific mutations and potentially accelerated rates of plastid genome evolution (Postel et al. 2022, 2023a). If PNIs are one of the main barriers inducing RIx between the lineages of *S. nutans*, then rapid evolution of reproductive barriers and strong RI despite low levels of divergence at nuclear loci might be driven by rapid evolution of the plastid genome in this species. In another angiosperm species, *Campanulastrum americanum*, similar levels of RI were observed between isolated clades of this species, with similar acceleration of the evolution (Barnard-Kubow et al. 2014, 2016, Barnard-Kubow and Galloway 2017). Inferences of evo-demographic scenarios with the different clades of this species have not been done yet but we might expect a similar relationship between genetic divergence at nuclear loci and levels of RI between them. For other angiosperm species with reported high rates of plastid genome evolution, independent lineages were not identified so far and levels of RI not assessed, and it would be interesting to investigate those patterns in a wider set of taxa, including control lineages that do not present such elevated levels of plastid genome evolution, in order to set up a more formal test of this hypothesis.

The main driver of genetic differentiation of lineages is linked selection

By assessing the correlations between summary statistics of genetic diversity of lineages and differentiation between lineages, linked selection appears to be the main driver of the genomic landscape of *S. nutans* in regions with low rate of recombination. Indeed, for every pair of lineages analyzed, we observed a strong positive correlation between D_{xy} and π which would not be expected under the alternative scenario where differentiation is driven by the heterogeneity of the rate of introgression due to genetic reproductive barriers (Shang et al. 2023). This scenario is also compatible with the observation of a negative correlation between genetic diversity and local recombination rate. Moreover, our ABC approach indicated that the best evo-demographic scenario of speciation was strict isolation with heterogeneity of N_e among loci, as expected under the effect of linked selection (Cruickshank and Hahn 2014). Recombination rate is known to be heterogeneous along the genome and especially low near chromosomes centromeres that are poor of gene content (Brazier et al. 2022) . *Silene latifolia* that shares the same number of chromosomes as *S. nutans* is no exception: gene density as well as rate of recombination are reduced in the pericentromeric regions (Moraga et al. 2023, Yue et al. 2023). Consequently, the occurrence of outliers of differentiation traditionally used to pinpoint potential reproductive barriers cannot be used in our case. However the potential involvement of such low recombining regions in speciation through divergent selection remains an open question (Schluter and Rieseberg 2022).

Materials and Methods

Transcriptomic data

A transcriptome assembly was already available as well as transcriptomic data for 22 individuals from the four genetic lineages of *S. nutans* (**Supplementary Table S1**) [(Muyle et al. 2021);PRJEB39526]; but the number of individuals from each lineage was uneven: 11 individuals from E1, 7 from W1, 2 for W2 and 2 for W3. We increased the sampling effort to get a total of 11 individuals per lineage (i.e. +4 individuals for W1, 9 for W2 and W3). We followed the same sampling strategy (i.e. two individuals per population) (**Supplementary Table 1**, **Fig. 1**). To test for potential gene flow within the western lineages, we sampled, when possible, individuals from population in close geographical proximity (**Fig. 1**). RNAs were extracted from flower buds using NucleoSpin RNA plus kit from Macherey Nagel. Libraries were constructed using NextFlex Rapid RNAseq kit. Sequencing for the 24 samples was done in paired-end 2x100bp. Libraries were sequenced on Illumina NextSeq500 HO in paired-end (2 × 100) at the LIGAN platform (UMR 8199 LIGAN-PM Genomics platform—Lille, France), resulting in a total of 44,436 GB and 21,654 GB after demultiplexing.

Newly acquired reads were aligned on the previously assembled transcriptome (Muyle et al. 2021) using Bowtie v2.4.1 (Langmead and Salzberg 2012), and duplicates were eliminated using MarkDuplicate implemented in Picard v2.21.4 ('Picard Toolkit', 2019. Broad Institute, https://broadinstitute.github.io/picard/; Broad Institute). Reads were then cleaned, sorted and indexed using samtools v1.10 (Li et al. 2009). To call the variants and produce the output format used for further analyses, we ran reads2snp v2.0 on these reads using the following options: -nbth 20 as a threshold for sites and reads mapping; -min 8 for minimal coverage; -aeb to control for allelic bias expression suited for RNAseq data (Tsagkogeorga et al. 2012, Gayral et al. 2013). The output multifasta format contains all biallelic nuclear loci for each individual.

The transcriptomic data can sometimes contain organellar gene sequences. The reference transcriptome was annotated using TAIR identifier, so we could easily find the organellar loci. Using seqkit (Shen et al. 2016), we specifically removed those loci. Before running the ABC analysis using DILS (Fraïsse et al.

2021), we checked whether individuals of the four lineages were indeed genetically differentiated. Using the fasta output of reads2snp, we constructed a PCA with popPhyl_PCA (https://github.com/popgenomics/popPhyl_PCA) (Fraïsse et al. 2021).

The selfing rates of the different lineages were inferred from estimates of identity disequilibrium using a method proposed by David et al. (2007). In a nutshell, partial selfing in a population induces an excess of multihomozygous or multiheterozygous genotypes, i.e. a stronger correlation in heterozygosity among loci within individuals, which can be used to estimate the selfing rate of the population. The Identity disequilibrium (estimated through the $\hat{g}2$ statistic, see David et al. 2007) was calculated with the Rpackage *InbreedR* (Stoffel et al. 2016) as input with a fasta file converted to a format suitable for *InbreedR* (the scripts are available at https://github.com/popgenomics/inbreeding). This input file, a table with the information of heterozygosity per site and per individual coded as 0 for homozygotes and 1 for heterozygotes, was obtained with the script *fasta2dos.py* and *hierf2inbreed.py*. Once the $\hat{g}_2$ was estimated, the selfing rates *s* were estimated through the formula (1) (David et al. 2007) with the R script *inbred_stats.R* (from the same repository).

$$\hat{S}_{g_2} = \frac{1 + 5\hat{g}_2 - \sqrt{10\hat{g}_2 + 9\hat{g}_2^2}}{2\hat{g}_2} \quad (1)$$

Demographic inference using DILS ABC framework

Model comparison: 2 populations DILS analyzes a random sample of 1000 nuclear loci, from the overall dataset, to compute several summary statistics: a joint site frequency spectrum (jSFS), the number of private and shared polymorphic sites (S_X and S_S respectively), the number of fixed differences between pairs of lineages (S_F), polymorphism statistics such as pairwise nucleotide diversity (π), Watterson's θ, Tajima's D, population divergence statistics D_{xy} (absolute divergence) and D_a (net divergence), and F_{ST} values of differentiation under different demographic scenarios. DILS also estimates the number of recombination events per locus running a four-gamete test. Giving a set of prior, simulated datasets are computed according to distinct alternative scenarios of speciation and divergence among the four lineages of *S. nutans*, including scenarios with or without current gene flow. In the end, DILS compares values of the summary statistics from simulated and observed datasets to infer the most plausible scenario. Once selected, parameters such as N_e and the time of split (T_{split}) will be estimated under this scenario. We ran DILS for the six pairwise combinations of lineages, using the same priors (**Supplementary Table S2**) for each combination and tested different demographic models of speciation: strict isolation (SI), ancient migration (AM), isolation with migration (IM) and secondary contact (SC) (**Fig. 2A**). When searching for the best model, DILS also considered changes in population size along the genome (variation of the N_e among loci—genome homogeneous *versus* heterogeneous *Ne*) and semipermeable barriers to gene flow (variation of the migration rate m_e among loci due to linkage to barriers). So, for each demographic scenario, fluctuating m_e and N_e were also testing through heterogeneity of these parameters along the genome. Because estimates of T_{split} can be tedious using only pairwise combinations, we also used a modified version of DILS performing model comparisons and parameters estimates on four-population models.

Model comparison: 4 populations The four-population model explored assumes a topology separating the eastern lineage E1 from the western lineages [W3, (W1, W2)] based on previous studies (Martin et al. 2016, Postel et al. 2022) (**Fig. 2**). This topology was simulated using MSMS (Ewing and Hermisson 2010) as given further, where tbs means a value to be specified: msms tbs 10,000 -s tbs -r tbs tbs -I 4 tbs tbs tbs tbs 0 -n 1 tbs -n 2 tbs -n 3 tbs -n 4 tbs -m 1 3 tbs -m 3 1 tbs -m 2 4 tbs -m 4 2 tbs -m 1 2 tbs -m 2 1 tbs -m 2 3 tbs -m 3 2 tbs -m 1 4 tbs -m 4 1 tbs -m 3 4 tbs -m 4 3 tbs -em tbs 3 4 0 -em tbs 4 3 0 -em tbs 3 2 0 -em tbs 2 3 0 -em tbs 3 1 0 -em tbs 1 3 0 -em tbs 2 4 0 -em tbs 4 2 0 -em tbs 2 1 0 -em tbs 1 2 0 -em tbs 4 1 0 -em tbs 1 4 0 -ej tbs 3 4 -en tbs 4 tbs -ej tbs 4 2 -en tbs 2 tbs -ej tbs 2 1 -eN tbs tbs

In this model, the four current lineages E1, W1, W2 and W3 have the possibility to exchange alleles during six independent secondary contacts. These secondary contacts involve six possible lineage pairs: E1-W3 (at time $TSC_{E1-W3} < T_{west}$), E1–W1 (at time $TSC_{E1-W1} < T_{W1-W2}$), E1–W2 (at time $TSC_{E1-W2} < T_{W1-W2}$), W3-W1 (at time $TSC_{W3-W1} < T_{W1-W2}$), W3–W2 (at time $TSC_{W3-W2} < T_{W1-W2}$) and W1–W2 (at time $TSC_{W1-W2} < T_{W1-W2}$). These secondary contact constraints on split times are intended to test only recent gene flow involving only pairs in independent combinations. At each population split, the new ancestral population (backward in times) has an independent size from the daughter populations (N_{anc}, N_{west} and N_{W1-W2}). Under this model, we randomly simulated multi-locus datasets with properties corresponding to the observed sampling (number of loci, locus length, number of gametes per locus and per sequenced population). For each simulation, coalescent trees of loci were obtained for parameter values randomly drawn from prior distributions. The exact number of mutations corresponding to that observed for each locus was then randomly placed in the simulated tree, thus not following a molecular clock according to an assumed mutation rate. With this mutation model, the parameters are expressed in coalescent units and not demographic units, allowing only relative parameter estimates to be obtained.

Current and ancestral population sizes are randomly drawn in a uniform distribution between 0 and 10.N_e; T_{split} is drawn in a uniform distribution between 0 and 20.N_e; T_{West} is drawn between 0 and T_{split}; T_{W1-W2} is drawn between 0 and T_{West}; effective migration $N_e.m$ is drawn in a uniform distribution between 0 and 10; migration is assumed to be heterogeneously distributed in the genome according to a beta distribution of alpha and beta shape parameters drawn in a uniform distribution between 0 and 20.

Introgression between lineages was tested independently for each of the six possible pairs. For this, 64 alternative models were simulated, depending on whether the migration for a given pair is null ($N_e.m = 0$ for all loci) or non-null ($N_e.m$ drawn between 0 and 10). Each of the 64 alternative models was simulated 10,000 times. Model comparisons were then performed individually for each pairwise relationship. To test whether recent gene flow has shaped the genetic patterns for a pair involving E1 and W3, we label as 'isolation' the 32 submodels for which there is no migration between E1 and W3 (but with or without migration between the other population combinations) and 'migration' the 32 submodels for which there is migration. These six model comparisons were performed with the R package abcrf (Pudlo et al. 2016) and using a forest of 1,000 trained trees. The trainings were conducted using the summary statistics described for the two-population models.

The parameters of the best-supported model among the 64 tested were also inferred by an ABC approach. Simulations similar to the previous step were performed, but using a molecular clock assuming the mutation rate $\mu = 7.31 \times 10^{-9}$ (Krasovec et al. 2018). We calculated the split time in year considering that one generation is equivalent to 3 years. Summary of parameters and prior use for this analysis can be found in **Supplementary Table S2**.

Statistical analysis on summary statistics

We ran statistical tests to assess the significance of the different values of the summary statistics. As DILS randomly samples 1,000 loci in the whole nuclear datasets before the data filtering procedures, we did not use the same number of loci for each pairwise combination nor the exact same loci. So, we first filtered the common loci to all pairwise combinations ($n = 3{,}460$). Statistical tests were run on R v. 1.4.1717. We first assessed normality of the data for each of summary statistics. As data were not normally distributed (Schapiro test *P*-value < 2.2e-16), we ran Kruskal–Wallis tests, followed by Dunn tests when the Kruskal–Wallis results were significant, to test for differences between comparisons (e.g. E1–W1 level of genetic differentiation versus E1–W2) and between lineages (e.g. E1 level of genetic diversity versus W1).

In order to help discriminate among different demographic and speciation scenarios and assess the impact of linked selection on divergence between these lineages, we analyzed patterns of covariation of different statistics (F_{ST}, D_{xy}, average Rho and average π) among loci. To do so, we first estimated the mean Rho and mean π between pairs of lineages for each locus. Then, using R v.

1.4.1717 and the R package *stats* v.3.6.2, we first assessed data normality conducting Shapiro test. Because data were not normally distributed and because our data contained ties, we then assess correlations between (i) F_{ST} and D_{xy}, (ii) D_{xy} and π, (iii) F_{ST} and π, (iv) F_{ST} and Rho using correlation test and estimating tau correlation's coefficient.

Supplementary Data

Supplementary data are available at *PCP* online.

Data Availability

Raw reads have been deposited to NCBI. SRA numbers can be found in **Supplementary Table S1**. Scripts used to run DILS-4 populations can be found in the following GitHub: https://github.com/popgenomics/DILS_4pop.

Funding

Agence Nationale de la Recherche (ANR-11-BSV7-013-03 TRANS to P.T.), the Région Hauts-de-France, the Ministère de l'Enseignement Supérieur et de la Recherche (CPER Climibio), and the European Fund for Regional Economic Development, the Ministère de l'Enseignement Supérieur et de la Recherche for HM's PhD grant and the Région Hauts-de-France and the Ministère de l'Enseignement Supérieur et de la Recherche for Z.P.'s PhD grant.

Acknowledgments

We would like to thank Alexander Papadopulos, Sophie Karrenberg, Jacqui Shykoff and Fabienne Van Rossum for their constructive remarks on the discussions and comments on the early version of the manuscript. We would also like to thank the members of the E.E.P. facilities for their help and support.

Author Contributions

Z.P. was involved in Investigation, Data curation, Formal analysis, Visualization and Writing—original draft, H.M. contributed in Investigation, Data curation and Formal analysis, C.R. was involved in Methodology, Software, Supervision and Writing—review &editing, M.G. did the Data curation, C.G. and collected the Resources F.M. was involved in Investigation, X. V. contributed in the Methodology, Supervision, Formal analysis and Writing—review &editing and P.T. was involved in Conceptualization, Funding acquisition, Project administration, Supervision and Writing—original draft.

Disclosures

The authors have no conflicts of interest to declare.

References

Baack, E., Melo, M.C., Rieseberg, L.H. and Ortiz-Barrientos, D. (2015) The origins of reproductive isolation in plants. *New Phytol*. 207: 968–984.

Barnard-Kubow, K.B. and Galloway, L.F. (2017) Variation in reproductive isolation across a species range. *Ecol. Evol*. 7: 9347–9357.

Barnard-Kubow, K.B., Sloan, D.B. and Galloway, L.F. (2014) Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evol. Biol*. 14: 268.

Barnard-Kubow, K.B., So, N. and Galloway, L.F. (2016) Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* 70: 2752–2766.

Bierne, N., Welch, J., Loire, E., Bonhomme, F. and David, P. (2011) The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Mol. Ecol*. 20: 2044–2072.

Brazier, T., Glémin, S. and Henderson, I.R. (2022) Diversity and determinants of recombination landscapes in flowering plants. *PLoS Genet*. 18: e1010141.

Burgarella, C. and Glémin, S. (2017) Population Genetics and Genome Evolution of Selfing Species. ELS. John Wiley & Sons. Ltd, Chichester.

Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L., et al. (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res*. 25: 1656–1665.

Butlin, R.K., Galindo, J. and Grahame, J.W. (2008) Sympatric, parapatric or allopatric: The most important way to classify speciation?. *Philos. Trans. R Soc. B* 363: 2997–3007.

Butlin, R. and Stankowski, S. (2020) Is it time to abandon the biological species concept? No. *Natl. Sci. Rev*. 7: 1398–1400.

Cornet, C., Noret, N. and Van Rossum, F. (2022) Pollinator sharing between reproductively isolated genetic lineages of *Silene nutans*. *Front. Plant Sci*. 13: 927498.

Coughlan, J.M. and Matute, D.R. (2020) The importance of intrinsic postzygotic barriers throughout the speciation process. *Philos. Trans. R Soc. B* 375: 20190533.

Coyne, J.A. and Orr, H.A. (2004) Speciation. Oxford University Press, Sinauer, Sunderland.

Cruickshank, T.E. and Hahn, M.W. (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol*. 23: 3133–3157.

David, P., Pujol, B., Viard, F., Castella, V. and Goudet, J. (2007) Reliable selfing rate estimates from imperfect population genetic data. *Mol. Ecol*. 16: 2474–2487.

De Bilde, J. (1973) Etude genecologique du *Silene nutans* L. en Belgique: Populations du *Silene nutans* L. sur substrat siliceux et calcaires. *Revue Generale de Botanique* 80: 161–176.

Ewing, G. and Hermisson, J. (2010) MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 26: 2064–2065.

Feder, J.L., Egan, S.P. and Nosil, P. (2012) The genomics of speciation-with-gene-flow. *Trends Genet*. 28: 342–350.

Fishman, L. and Sweigart, A.L. (2018) When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. *Annu. Rev. Plant Biol*. 69: 707–731.

Fraïsse, C., Popovic, I., Mazoyer, C., Spataro, B., Delmotte, S., Romiguier, J., et al. (2021) DILS: Demographic inferences with linked selection by using ABC. *Mol. Ecol. Resour*. 21: 2629–2644.

Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., et al. (2013) Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet*. 9: e1003457.

Hartfield, M., Bataillon, T. and Glémin, S. (2017) The evolutionary interplay between adaptation and self-fertilization. *Trends Genet*. 33: 420–431.

Hepper, F.N. (1956) *Silene nutans* L. *J. Ecol*. 44: 693–700.

Heuertz, M., Hausman, J.-F., Hardy, O.J., Vendramin, G.G., Frascaria-Lacoste, N. and Vekemans, X. (2004) Nuclear microsatellites reveal contrasting patterns of genetic structure between western and southeastern European populations of the Common Ash (*Fraxinus Excelsior* L.). *Evolution* 58: 976–988.

Hewitt, G. (2000) The genetic legacy of the quaternary ice ages. *Nature* 405: 907–913.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., dePamphilis, C.W., Leebens-Mack, J., et al. (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *PNAS* 104: 19369–19374.

Kadereit, J.W. and Abbott, R.J. (2021) Plant speciation in the quaternary. *Plant Ecol. Divers* 14: 105–142.

Krasovec, M., Chester, M., Ridout, K. and Filatov, D.A. (2018) The mutation rate and the age of the sex chromosomes in *silene latifolia*. *Curr. Biol*. 28: 1832–1838.

Kulmuni, J., Butlin, R.K., Lucek, K., Savolainen, V. and Westram, A.M. (2020) Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers: progress towards complete speciation. *Philos. Trans. R Soc. B* 375: 20190528.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Liu, X., Glémin, S. and Karrenberg, S. (2020) Evolution of putative barrier loci at an intermediate stage of speciation with gene flow in campions (*Silene*). *Mol. Ecol*. 29: 3511–3525.

Lowry, D.B. (2012) Ecotypes and the controversy over stages in the formation of new species. *Biol. J. Linn. Soc*. 106: 241–257.

Martin, H., Touzet, P., Dufay, M., Godé, C., Schmitt, E., Lahiani, E., et al. (2017) Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution (N Y)* 71: 1519–1531.

Martin, H., Touzet, P., Van Rossum, F., Delalande, D. and Arnaud, J.F. (2016) Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity* 116: 286–294.

Matute, D.R. and Cooper, B.S. (2021) Comparative studies on speciation: 30 years since Coyne and Orr. *Evolution* 75: 764–778.

Monnet, F., Postel, Z., Touzet, P., Fraïsse, C., Van De Peer, Y., Vekemans, X., et al. (2023) Rapid establishment of species barriers in plants compared to animals. *bioRxiv* 1–51.

Moraga, C., Branco, C., Rougemont, Q., Veltsos, P., Jedlička, P., Muyle, A., et al. (2023) The *Silene latifolia* genome and its giant Y chromosome. *bioRxiv* 558754.

Muyle, A., Martin, H., Zemp, N., Mollion, M., Gallina, S., Tavares, R., et al. (2021) Dioecy is associated with high genetic diversity and adaptation rates in the plant genus *silene*. *Mol. Biol. Evol*. 38: 805–818.

Nosil, P., Feder, J.L., Flaxman, S.M. and Gompert, Z. (2017) Tipping points in the dynamics of speciation. *Nat. Ecol. Evol*. 1: 0001.

Petit, R.J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., et al. (2003) Glacial Refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563–1565.

Postel, Z., Mauri, T., Lensink, M.F. and Touzet, P. (2023a) What is the potential impact of genetic divergence of plastid ribosomal genes between *Silene nutans* lineages in hybrids? An *in silico* approach using the 3D structure of the plastid ribosome. *Front. Plant Sci*. 14: 1167478.

Postel, Z., Poux, C., Gallina, S., Varré, J.S., Godé, C., Schmitt, E., et al. (2022) Reproductive isolation among lineages of *Silene nutans* (Caryophyllaceae): a potential involvement of plastid-nuclear incompatibilities. *Mol. Phylogenet Evol*. 169: 107436.

Postel, Z., Sloan, D.B., Gallina, S., Godé, C., Schmitt, E., Mangenot, S., et al. (2023b) The decoupled evolution of the organellar genomes of *Silene nutans* leads to distinct roles in the speciation process. *New Phytol*. 239: 766–777.

Postel, Z. and Touzet, P. (2020) Cytonuclear genetic incompatibilities in plant speciation. *Plants* 9: 487.

Postel, Z., Van Rossum, F., Godé, C., Schmitt, E. and Touzet, P. (2024) Paternal leakage of plastids rescues inter-lineage hybrids in *Silene nutans*. *Ann. Bot*. 133: 427–434.

Presgraves, D.C. (2010) The molecular evolutionary basis of species formation. *Nat. Rev. Genet*. 11: 175–180.

Pudlo, P., Marin, J., Estoup, A., Cornuet, J., Gautier, M. and Robert, C.P. (2016) Reliable ABC model choice via random forests. *Bioinformatics* 32: 859–866.

Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., et al. (2017) Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol*. 30: 1450–1477.

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N. and Bierne, N. (2016) Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol*. 14: e2000234.

Ruhlman, T.A. and Jansen, R.K. (2018) Aberration or analogy? The atypical plastomes of geraniaceae. *Adv. Bot. Res*. 85: 223–262.

Schluter, D. and Rieseberg, L.H. (2022) Three problems in the genetics of speciation by selection. *Proc. Natl. Acad. Sci. U.S.A*. 119: e2122153119.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., et al. (2014) Genomics and the origin of species. *Nat. Rev. Genet*. 15: 176–192.

Shang, H., Field, D.L., Paun, O., Rendón-Anaya, M., Hess, J., Vogl, C., et al. (2023) Drivers of genomic landscapes of differentiation across a *Populus* divergence gradient. *Mol. Ecol*. 32: 4348–4361.

Shen, W., Le, S., Li, Y., Hu, F. and Zou, Q. (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11: e0163962.

Shrestha, B., Weng, M., Theriot, E.C., Gilbert, L.E., Ruhlman, T.A., Krosnick, S.E., et al. (2019) Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. *Mol. Phylogenet Evol*. 138: 53–64.

Sloan, D.B., Triant, D.A., Forrester, N.J., Bergner, L.M., Wu, M. and Taylor, D.R. (2014) A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Mol. Phylogenet Evol*. 72: 82–89.

Stankowski, S. and Ravinet, M. (2021) Defining the speciation continuum. *Evolution* 75: 1256–1273.

Stoffel, M.A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., et al. (2016) inbreedR: An R package for the analysis of inbreeding based on genetic markers. *Methods Ecol. Evol*. 7: 1331–9.

Tsagkogeorga, G., Cahais, V. and Galtier, N. (2012) The population genomics of a fast evolver : high levels of diversity, functional constraint, and molecular adaptation in the tunicate *ciona intestinalis*. *Genome Biol. Evol*. 4: 852–861.

Turner, T.L., Hahn, M.W., Nuzhdin, S.V. and Barton, N. (2005) Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biol*. 3: e285.

Vanderplanck, M., Touzet, P., Rossum Van, F., Lahiani, E., Cauwer, De, I. and Dufaÿ, M. (2020) Does pollination syndrome reflect pollinator efficiency in *Silene nutans*?. *Acta Oecol*. 105: 103557.

Van Rossum, F. (2000) Amplitude synécologique de *Silene nutans* L. (Caryophyllacae) en Belgique. *Dumortiera* 75: 11–24.

Van Rossum, F., De Bilde, J. and Lefèbvre, C. (1996) Barriers to hybridization in calcicolous and silicicolous populations of *Silene nutans* from Belgium. *R. Bot. Soc. Belgium* 129: 13–18.

Van Rossum, F., Martin, H., Le Cadre, S., Brachi, B., Christenhusz, M.J.M. and Touzet, P. (2018) Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspect. Plant Ecol. Evol. Syst*. 35: 44–51.

Van Rossum, F. and Prentice, H.C. (2004) Structure of allozyme variation in Nordic *Silene nutans* (Caryophyllaceae): population size, geographical position and immigration history. *Biol. J. Linn. Soc*. 81: 357–371.

Van Rossum, F., Vekemans, X., Meerts, P., Gratia, E. and Lefèbvre, C. (1997) Allozyme variation in relation to ecotypic differentiation and population size in marginal populations of *Silene nutans*. *Heredity* 78: 552–560.

Wright, S.I., Kalisz, S. and Slotte, T. (2013) Evolutionary consequences of self-fertilization in plants. *Proc. R. Soc. B* 280: 20120133.

Wu, C. (2001) The genic view of the process of speciation. *J. Evol. Biol*. 14: 851–865.

Yue, J., Krasovec, M., Kazama, Y., Zhang, X., Xie, W., Zhang, S., et al. (2023) The origin and evolution of sex chromosomes, revealed by sequencing of the *Silene latifolia* female genome. *Curr. Biol*. 33: 2504–2514.e3.