

SCIENTIFIC REPORTS



OPEN

Analysis of spatial genetic variation reveals genetic divergence among populations of *Primula veris* associated to contrasting habitats

Pablo Deschepper¹, Rein Brys², Miguel A. Fortuna³ & Hans Jacquemyn¹

Genetic divergence by environment is a process whereby selection causes the formation of gene flow barriers between populations adapting to contrasting environments and is often considered to be the onset of speciation. Nevertheless, the extent to which genetic differentiation by environment on small spatial scales can be detected by means of neutral markers is still subject to debate. Previous research on the perennial herb *Primula veris* has shown that plants from grassland and forest habitats showed pronounced differences in phenology and flower morphology, suggesting limited gene flow between habitats. To test this hypothesis, we sampled 33 populations of *P. veris* consisting of forest and grassland patches and used clustering techniques and network analyses to identify sets of populations that are more connected to each other than to other sets of populations and estimated the timing of divergence. Our results showed that spatial genetic variation had a significantly modular structure and consisted of four well-defined modules that almost perfectly coincided with habitat features. Genetic divergence was estimated to have occurred about 114 generations ago, coinciding with historic major changes in the landscape. Overall, these results illustrate how populations adapting to different environments become structured genetically within landscapes on small spatial scales.

Understanding how the environment shapes species is a central question in evolutionary biology^{1–4}. When species are distributed across large areas, geographical isolation can limit gene flow and cause strong genetic differentiation, ultimately leading to ecotype and even species formation. This process can take place either along an ecological gradient or across isolated, well-defined habitat entities^{1, 5, 6}. Strong genetic differentiation can also occur at much smaller local scales as a result of contrasting environmental features that affect gene flow and population genetic structure^{7–10}. It has for instance been observed that recently diverged sister species originate from habitat-specific adaptations following colonization events in new environments^{11–14}. Similarly, landscape heterogeneity can lead to formation of locally adapted populations with reduced gene flow between habitats.

Several reproductive barriers that prevent gene flow can arise when a species occupies contrasting environments within the same geographical region, which in turn can elicit genetic structuring^{15, 16}. Prezygotic barriers such as differences in flowering time significantly reduce gene transfer between neighboring populations growing under contrasting environmental conditions or in different habitats^{17–21}. Additionally, differences in pollinator communities between habitats can shape floral morphology and therefore significantly impact on gene flow by pollen transfer^{22–25}. Different habitats also demand different adaptations in a way that the right alleles are required to thrive in a specific environment²⁶. This restricts certain genotypes to settle or flourish in a given environment characterized by specific soil characteristics, predators, or light availability^{4, 21, 27, 28}. Selection against immigrants due to maladaptation hampers integration of genes of less fit genotypes into the native genotype and can for this reason be considered as an additional legitimate barrier against gene flow and thus ultimately a driver for genetic divergence among populations^{29–32}.

However, how these aspects affect gene flow at neutral loci between diverged populations has only been studied recently^{3, 19–21}. Because ecological divergent selection can give rise to barriers to gene flow³³, which in turn

¹Division of Plant Ecology and Systematics, Biology Department, University of Leuven, Leuven, Belgium. ²Research Institute for Forest and Nature, Gaverstraat 4, B-9500, Geraardsbergen, Belgium. ³Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland. Correspondence and requests for materials should be addressed to P.D. (email: pablo.deschepper@kuleuven.be)

may reduce the homogenizing effect of gene flow among habitats, this can cause genetic divergence at neutral loci between habitats due to genetic drift^{10,33}. Demonstrating isolation by adaptation with neutral markers is not always possible since this process largely depends on the stochastic nature of genetic drift or the presence of weakly linked genomic regions under divergent selection³³. Nonetheless, several studies have recently used neutral markers to assess population genetic structure and its relation to natural selection^{10,32,34,35}. Direct comparison between neutral microsatellite markers and a genetic marker directly influenced by natural selection has shown that microsatellites are better suited to determine population genetic structure as a consequence of neutral processes such as gene flow³⁶. Furthermore, a simulation study showed that ecological differences can cause a reduced gene flow at neutral markers resulting in genetic differentiation among populations³⁷.

Forests and grasslands represent two contrasting habitats that largely differ in several environmental conditions, such as light availability, humidity and temperature, which in turn may select for different traits and result in ecological divergence. Recent research on the distylous *Primula veris* has shown clear evidence for habitat-specific differentiation in phenology and flower characteristics between neighboring populations growing in forests and grasslands³⁸. Due to increasing shade during the growing season, plants in the forest habitat flower about three weeks earlier compared to neighboring populations growing in open grasslands, restricting the chances of gene flow between populations of both habitats³⁸. The size of the flowers and positioning of the anthers and stigma also differed significantly between grassland and forest plants. In particular, the L-morph flower of forest plants showed strong deviation from anther-stigma separation³⁸. Deviations from reciprocal placement of sexual organs can hamper pollen deposition on compatible stigmas²⁵ and therefore restrict pollen flow between grassland and forest populations. Due to these differences, we hypothesized that restricted gene flow between forest and grassland populations has created genetically structured groups within a mosaic landscape consisting of both grassland and forest habitats and therefore may present the very first signs of ecological speciation.

To test this hypothesis, we investigated patterns of genetic variation and structure in a large set of populations of *P. veris* using 12 polymorphic microsatellite loci. We used Bayesian clustering techniques and a landscape genetic analysis based on network theory to test whether grassland and forest populations can be divided into clusters that act as independent evolutionary units in the landscape and to see whether these clusters can be brought back to specific characteristics of the habitats they were sampled from. Additionally, coalescent-based approximate Bayesian computation (ABC) was used to make inferences about population history and to estimate the timing of divergence between populations growing in grassland and forest populations.

Results

Analysis of genetic diversity. Population sizes ranged from 60 to approximately 12 000 flowering individuals (Table 1). There was no significant ($P > 0.05$) difference in population size between grassland (average size: 345 ± 26.35) and forest populations (207 ± 167.81) after omitting one outlier. In total, 119 different alleles were found across 12 microsatellite marker loci in 792 sampled plants. The number of alleles per locus ranged from four to 23 (mean: 9.92). Measures concerning genetic diversity are summarized in Table 1 with mean values given for both habitats. The number of alleles ($P = 0.09$, $t = 1.73$, $df = 31$) and allelic richness ($P = 0.17$, $t = 1.41$, $df = 31$) were not significantly different between both habitats. Expected (H_e) and observed (H_o) heterozygosity ranged from 0.325 to 0.606 and from 0.385 to 0.589 in grassland and forest populations respectively and did not differ significantly between habitats (mean H_e grassland: 0.535 and mean H_e forest: 0.529, $P = 0.67$, $t = 0.43$, $df = 31$; mean H_o grassland: 0.475 and mean H_o forest: 0.466, $P = 0.70$, $t = 0.38$, $df = 31$). The mean values for F_{is} were positive for both habitats (mean grassland: 0.136; mean forest: 0.119), indicating slight inbreeding. The inbreeding coefficient did not differ significantly between habitats ($P = 0.60$, $t = 0.63$, $df = 31$). All but two loci showed significant heterozygote deficiency across all populations. Linkage disequilibrium (LD) was found in 7 out of 66 possible locus pairs following Bonferroni correction. Significant linkage for at least one locus pair was present in three grassland and three forest populations (populations 1, 2, 11, 19, 27 and 30).

Genetic structure and isolation by distance. The AMOVA analysis revealed that most of the genetic variation was found between individuals within populations (81%), whereas 14% of variation was due to differences between populations and 4% due to differences between habitats. All values were highly significant with $P < 0.001$ (Table 2). Overall genetic differentiation was moderate ($F_{ST} = 0.069$, $G_{ST} = 0.263$, Jost's $D = 0.204$) (Table 3). The ANOVA-like Mantel test indicated that genetic differentiation was significantly larger between habitats than within habitats for every parameter of genetic differentiation ($P < 0.01$) (Table 3). Inspecting the values of the genetic differentiation measures within and between habitats, this result was most likely caused by the higher gene flow between grassland populations (Table 3). Indeed, genetic differentiation was significantly higher between forest populations than between populations of grassland habitats ($F_{ST} = 0.087$ and $F_{ST} = 0.052$ respectively with $P < 0.001$). Additionally, traditional Mantel tests showed a significant positive relationship between genetic and geographic distance for forest populations ($R^2 = 0.126$; $P < 0.01$), whereas no such relationship was detected for grassland populations ($R^2 = 0.008$; $P = 0.225$) (Fig. 1).

Bayesian clustering (Fig. 2) revealed two distinct genetic clusters inferred by Evanno's ΔK ³⁹. The inferred clusters strongly corresponded with the habitats from which individuals were sampled, confirming differentiation by environment ($\Delta K = 140.8$ for $K = 2$, second largest $\Delta K = 11.5$ for $K = 3$, third largest $\Delta K = 5.9$ for $K = 4$). The average Q values for grassland populations were 0.705 and 0.295 for the red and blue cluster respectively, and 0.321 and 0.679 for the forest populations. The value for $LnP(D)$ reached a plateau at $K = 4$ and corresponding cluster plots are shown in Fig. 2. We did not find a different optimal number of clusters when populations that showed linkage at one or more loci were omitted from the clustering analysis ($\Delta K = 26.2$ for $K = 2$, second largest $\Delta K = 10.1$ for $K = 4$, third largest $\Delta K = 5.8$ for $K = 3$).

Habitat	ID	Size	N	A	A _r	H _o	H _e	F _{is}
Grassland	1	210	24	4.7	2.80	0.538	0.590	0.083
Grassland	2	490	24	4.3	2.46	0.454	0.492	0.101
Grassland	3	85	24	4.7	2.68	0.494	0.532	0.092
Grassland	4	100	24	4.6	2.63	0.525	0.543	0.053
Grassland	5	100	24	5.3	2.86	0.576	0.574	0.007
Grassland	6	100	24	5.2	2.71	0.486	0.541	0.132
Grassland	7	70	24	5.6	2.88	0.606	0.585	0.031
Grassland	8	200	24	4.3	2.59	0.325	0.500	0.414
Grassland	9	750	24	4.9	2.79	0.521	0.569	0.107
Grassland	10	550	24	4.5	2.46	0.444	0.467	0.037
Grassland	11	1250	24	5.2	2.69	0.434	0.518	0.170
Grassland	12	12000	24	5.0	2.60	0.453	0.512	0.133
Grassland	13	220	24	4.5	2.54	0.440	0.481	0.195
Grassland	14	420	24	4.9	2.83	0.494	0.576	0.176
Grassland	15	160	24	5.3	2.82	0.439	0.555	0.269
Grassland	16	90	24	5.0	2.76	0.538	0.558	0.050
Grassland	17	860	24	4.6	2.59	0.388	0.526	0.162
Grassland	18	210	24	4.8	2.61	0.391	0.519	0.291
Grassland habitat mean				4.9	2.68	0.475	0.535	0.136
Forest	19	320	24	4.4	2.80	0.548	0.615	0.091
Forest	20	400	24	4.0	2.75	0.537	0.623	0.104
Forest	21	110	24	5.1	2.96	0.589	0.633	0.027
Forest	22	210	24	5.2	2.44	0.410	0.463	0.186
Forest	23	190	20	3.9	2.51	0.429	0.496	0.160
Forest	24	90	20	4.1	2.58	0.385	0.478	0.170
Forest	25	60	24	5.0	2.74	0.420	0.546	0.225
Forest	26	90	24	4.5	2.66	0.439	0.549	0.190
Forest	27	70	24	4.4	2.50	0.471	0.497	0.047
Forest	28	210	23	4.4	2.55	0.451	0.496	0.124
Forest	29	380	19	4.8	2.58	0.450	0.520	0.117
Forest	30	60	24	4.3	2.48	0.462	0.492	0.055
Forest	31	680	24	5.5	2.67	0.477	0.538	0.105
Forest	32	160	24	4.7	2.42	0.452	0.472	0.087
Forest	33	75	24	4.8	2.58	0.474	0.510	0.100
Forest habitat mean				4.6	2.61	0.466	0.529	0.119

Table 1. Genetic diversity measures for all sampled *P. veris* populations among the two habitats. Size, number of flowering individuals; N, number of sampled individuals; A, number of alleles per locus; A_r, allelic richness; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{is}, inbreeding coefficient. The mean value of every genetic measure is given for each habitat.

Source of variation	d.f.	Sum of Squares	Phi-statistics	% of total variance	P-value
Between groups	1	244.111	0.043	4.30	< 0.01
Among populations within groups	31	1523.110	0.150	14.35	< 0.01
Among individuals within populations	759	7127.583	0.186	81.35	< 0.01
Total	791	8894.804			

Table 2. Hierarchical analysis of molecular variance based on 12 microsatellite loci and 33 populations.

	all populations	within grassland habitat	within forest habitat	between habitats
F_{ST}	0.069	0.052	0.087	0.084
G_{ST}	0.263	0.189	0.337	0.326
D	0.204	0.125	0.242	0.234

Table 3. Different parameters for pairwise genetic differentiation given for the two habitats. Values for all genetic differentiation parameters between populations of different habitats were significantly different from values between populations of the same habitat as reported in the ANOVA-like mantel test ($P < 0.01$). Genetic differentiation was also significantly greater among populations within the forest habitat than between grassland populations for all parameters ($P < 0.001$).

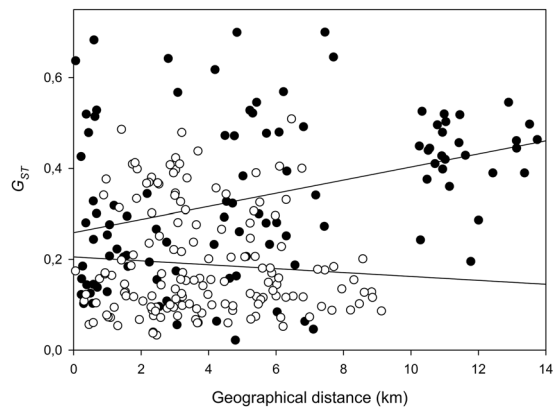


Figure 1. Relationship between geographic distance and G_{ST} for grassland (white dots) and forest (black dots) populations.

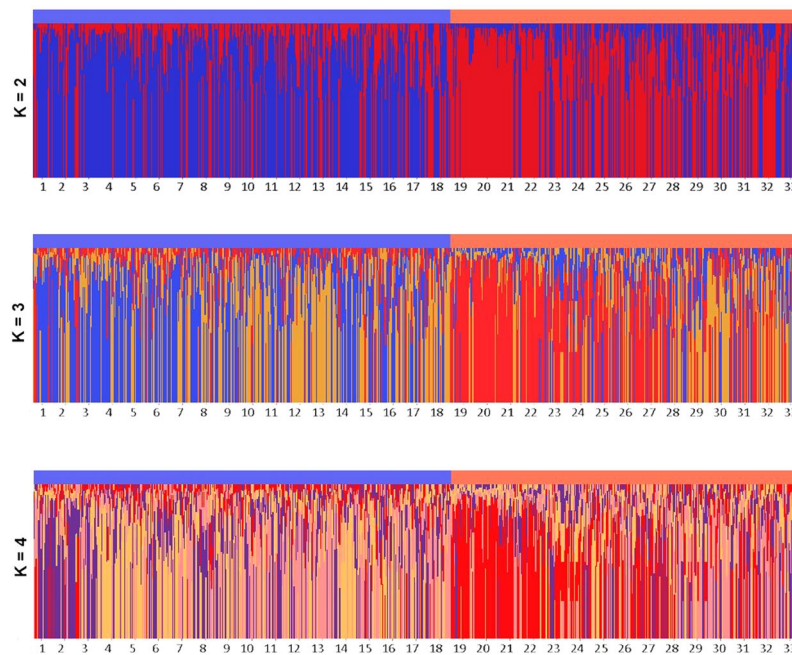


Figure 2. Bayesian cluster analysis (from $K = 2$ to $K = 4$) for all *P. veris* populations with a color indication for the respective habitat for each of the 33 populations visualized with Structure Plot shiny web application⁹⁶ (blue for the grassland habitat and red for the forest habitat).

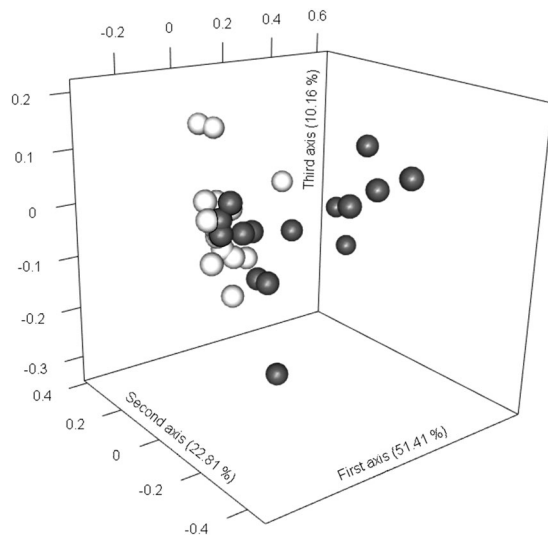


Figure 3. PCoA of the 33 *P. veris* populations with grassland and forest populations indicated in light grey and black respectively. The percentage of explained variation by each axis is shown between the brackets.

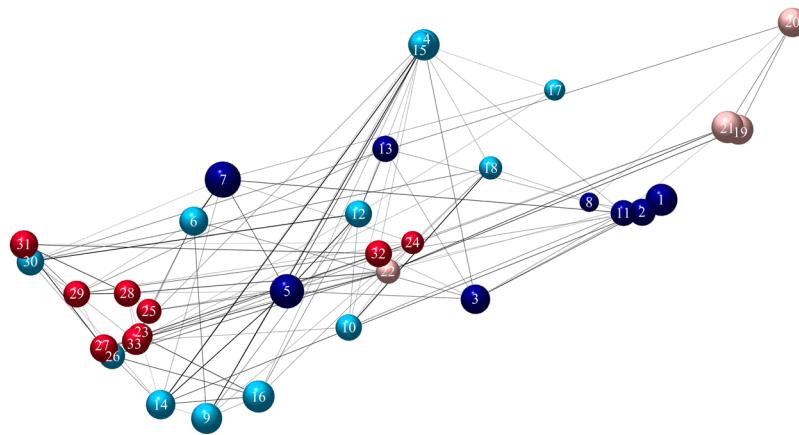


Figure 4. Spatial network of genetic variation with four detected modules (different colors). Links represent significant genetic similarity with line thickness positively correlated to the level of similarity. Node size shows population heterozygosity and node position reflects the geographic position. Population 1 to 18 and 19 to 33 are grassland and forest populations respectively.

Clustering of the populations belonging to the two habitats was also visually confirmed by the PCoA analysis (Fig. 3). Furthermore, distances between forest populations in the PCoA were larger than distances between grassland populations, supporting our previous finding that forest populations were less closely related to each other than grassland populations.

Networks of spatial genetic differentiation and identification of modules. The network of spatial genetic variation of *P. veris* populations (Fig. 4) showed 33 nodes connected with 107 links. 43% of the links were established between grassland populations (average number of links per population: 2.56) and 31% between forest populations (average: 2.20). 26% of the links were established between nodes of a different habitat. Overall network connectance (number of remaining links over all possible links) was 0.203. Network connectance was markedly lower between (0.104) than within habitats (0.301 for the grassland habitat and 0.314 for the forest habitat). The overall network showed a significant modular structure (modularity $M = 0.32$; $P = 0.013$). Four different modules were detected (Fig. 4). Almost all the populations within one module corresponded to one habitat type. Only one module contained a mixture of grassland and forest populations (light blue module) with two out of 12 populations belonging to the forest instead of the grassland habitat. The two forest modules (red and light red) have their core on separate sides of the study area, suggesting stronger isolation by distance than the grassland modules.

Time of divergence and historical gene flow. Results from ABC showed that grassland and forest populations diverged ca. 114 generations ago (95% CI: 126 y - 1030 y). This is 228 y before present assuming a generation time of 2 years. Model checking indicated that summary statistics from the posterior predictive distribution of the simulated datasets correctly estimated parameters. More specifically, none of the summary statistics of the 1% closest simulated datasets deviated significantly from those of the observed dataset ($P = 0.288-0.925$). The MIGRATE-N analysis showed that the mutation-scaled immigration rate M was highly directional with migration occurring mainly from the grassland into the forest habitat ($M = 1043$) and far less from the forest into the grassland habitat ($M = 278$).

Discussion

Identifying the factors that promote genetic differentiation is of interest for understanding the processes initiating the early stages of speciation. Gene flow among populations inhabiting different environments can be reduced by geographical distance or by divergent selection resulting from local adaptation^{10, 27, 35, 40, 41}. However, disentangling the relative role of spatial and environmental factors in shaping patterns of population differentiation is not straightforward as both are likely to be intertwined. Several studies have addressed genetic divergence that is attributable to environmental adaptation to the environment^{10, 19, 35, 41, 42}. Here, we used microsatellite markers to test the hypothesis that differences in phenology, flower morphology and potentially other adaptations to contrasting habitats (e.g. shade and drought tolerance) between grassland and forest populations of the perennial herb *P. veris* translated into genetic structuring and the occurrence of well-defined genetic entities in the landscape.

Genetic differentiation across all populations was moderate ($F_{ST} = 0.069$), with forest populations showing the largest differentiation. The analysis of molecular variance showed that 4% of the total genetic variation could be attributed to differences between habitats, indicating that environmental factors significantly contributed to partitioning of genetic variation and that gene flow seems to be restricted even if populations are located close to each other. In the related *Primula merrilliana*, Shao *et al.*⁴³ documented a much higher value of variation that could be attributed to different habitats (13.30%). In this study, however, both contrasting habitats (foothill versus mountain habitat) were clearly spatially separated with extreme differences in growing conditions, both of which can be expected to have a considerable impact on genetic variation. On the other hand, Jacquemyn *et al.*⁴⁴ found lower amounts of variation (<4%) that were associated with patch age in the related forest herb *Primula elatior*. A very similar study on the epiphytic orchid *Fumellea rossi* that grew in different forest types reported a value of 2.1% for the part of the total genetic variation that could be ascribed to habitat⁴⁰. A comparable value of 2.3% was reported by Andrew *et al.*³⁵ in the sunflower *Helianthus petiolaris* occurring in dune and non-dune populations, which they denoted as different ecotypes.

Our results further showed a higher genetic differentiation between forest populations than between grassland populations, suggesting that grassland populations are stronger connected to each other than forest populations. Indeed, isolation by distance is causing partitioning of genetic variation in the forest habitat and not in grasslands probably because of the lack of connecting suitable habitat patches in the landscape. This result is not surprising given that gene flow by seeds is restricted in *P. veris*^{45, 46} and that most gene flow therefore is the result of pollen flow. Within the study area, populations of *P. veris* can often be found along grassy road verges, which provide corridors for gene flow and essentially connect different grassland populations. Open old growth forests, on the other hand, are not as widely or evenly distributed across the study area and are often surrounded by denser younger forest and separated by agricultural fields or intensively managed grasslands. Moreover, *P. veris* shows reduced performance under shade and is therefore not likely to be present in forests with a dense shrub layer⁴⁶, further restricting the presence of *P. veris* to very specific forest conditions. In our case, forest populations clustered within three large forest patches embedded within agricultural fields, which may to some extent have contributed to the isolation-by-distance pattern observed for forest populations. As a result, pollinators have to fly large distances to spread genes from one forest population patch to another. Additionally, it has been observed that bumblebees, the main pollinators of *P. veris* in forests, tend to fly over forests rather than fly through them to be cost efficient⁴⁷, further restricting pollen flow among forest patches.

Differentiation between habitats was considerable ($F_{ST} = 0.084$) and was visualized by a Bayesian cluster analysis and a PCoA, which revealed two distinct groups of populations largely overlapping with the two *P. veris* habitats, indicating a strong concordance between the type of habitat and the partitioning of gene diversity in the landscape. This result is similar with findings of Mallet *et al.*⁴⁰, where the number of clusters equaled the number of different habitats. The network analysis of spatial genetic variation largely confirmed these results and showed that genetic variation was spatially distributed across four modules that almost entirely overlapped with populations belonging to the grassland or the forest habitat. Populations within habitats were also more connected to each other than to populations of the other habitat, suggesting that gene flow between forest and grassland populations was restricted. Similar results were shown by Lowry *et al.*¹⁵, who showed significant genetic clustering between coastal and inland races of *M. guttatus*. However, in contrast with this study, the forest and grassland populations investigated here occurred interspersed within the landscape. The detection of four different evolutionary units by our module finding algorithm therefore also suggests exploration for clustering at a deeper level than the cluster analysis implemented in STRUCTURE⁴⁸.

The presence of significant modularity points to the existence of independent evolutionary units that may form the basis for further adaptation and ultimately ecotype formation. Several adaptations to specific environmental features linked to the habitats in which *P. veris* grows can be expected. First, the observed differences in flowering time may limit the chances for gene flow between forest and grassland populations and can cause genetic differentiation^{17, 20, 27}. Differences in flowering time are most likely an adaptation to different light conditions encountered in forest and grassland habitats⁴⁹⁻⁵¹. Previous research has shown that the peak in the light saturated rate of CO₂ assimilation (Amax) in *P. veris* occurs at the beginning of April, when plants start flowering, and declined during the growing season⁵². *Primula veris* thus shows an increased carbon gain early in spring

and exploits the high spring light phase before expansion of the vegetation canopy. This high light phase largely coincides with flower and fruit production, and a substantial carbon gain in spring could guarantee the initiation of flower primordia and/or increase the proportion of dormant flower buds developing⁵². Flowering earlier in the year in forest habitats, before canopy closure, can therefore be considered as a reproductive strategy of *P. veris* to escape competition for light.

Field observations have also indicated that pollinator communities differ between grassland and forest populations, with little overlap between them⁵³. Whereas pollinator communities in grasslands are dominated by the hairy-footed flower bee *Anthophora plumipes*, this species was absent in forest populations. Here, bumblebees (*Bombus terrestris*, *B. lapidarius*) and several *Lasioglossum* species were the most frequent insects visiting *P. veris* flowers. Separation of flowering in time and pronounced differences in pollinator communities therefore most likely act as reproductive barriers that limit gene flow between habitats^{21, 27, 54}.

Furthermore, successful establishment after immigration to the opposing habitat could be hampered by immigrant unviability because of maladaptation^{19, 35, 55}. Grasslands and forest are strongly different habitats, demanding different sets of adaptations which can directly compromise the ability of a plant to germinate, grow or reproduce in a non-native environment^{26, 55, 56}. For example, previous research has shown that compared to the related *Primula elatior* and *P. vulgaris*, *P. veris* is least well-adapted to survive moderate shade, as a consequence of its relatively low quantum efficiency, high light saturation point and high dark respiration⁵². The pronounced differences in specific leaf area and stomatal density point to habitat-specific adaptations to light conditions⁵⁷ or water status⁵⁸, which may restrict establishment of grassland plants in forest habitats and vice versa.

The time of divergence between grassland and forest populations was estimated to occur at the end of the 18th century. This period coincides with large alterations of the landscape^{59, 60}. In 1775, the date of first map of the study area, the entire area was widely covered with calcareous grasslands and old-growth forest was limited. These grasslands are known for their high diversity in flowering plants⁶¹ and attract a wider variety of insects because they can supply copious nectar for pollinators⁶². It is therefore reasonable to assume that pollinator diversity was highest in calcareous grasslands and that gene flow occurred predominantly from grassland to forest. This is in line with our results that showed that historical migration rates were about 4 times larger from grassland to forest populations than from forest populations to grassland populations. However, in the 18th and especially the 19th century, many calcareous grasslands were abandoned⁶⁰ or planted with high-productivity forests that mainly consisted of *Pinus nigra*. These secondary forests are unsuitable to support *P. veris* populations and this may have initiated increased isolation between forest and grassland populations. Assuming that, due to the increased isolation, less pollen was dispersed from grasslands into forest populations, forest and grassland populations gradually started to diverge and phenotypic and genotypic differences between forest and grassland populations started to increase.

To conclude, our results demonstrated that populations of *P. veris* formed clear genetic entities in the landscape that were related to the habitat from which they were sampled. Historic changes in landscape configuration and spatial isolation and the associated changes in gene flow probably have gradually induced phenotypic and genotypic differences between plants from grassland and forest populations. Future research is needed to examine whether the observed phenotypic differences, such as specific leaf area or stomatal density, are the result of plasticity or are in fact evolutionary adaptations⁶³ and to identify genomic regions that are subject to adaptation.

Materials and Methods

Species. *Primula veris* L. (cowslip) is a herbaceous, spring flowering perennial plant species that usually grows in calcareous grasslands, but that can also be found in old growth forest and hedgerows⁴⁶. This rosette forming hemicryptophyte can be found throughout most of temperate Europe and the British Isles until the western Russian border. In early spring, *P. veris* forms a rosette of leaves and produces flowers that grow in umbels. Since *P. veris* prefers a warm microclimate, it has a preference for sunny slopes and open forest patches. Selfing is prevented by a diallelic self-incompatibility system and heterostyly with two reciprocal flower morphs (pin and thrum)^{38, 64}. Pollen flow is mainly accomplished by early Hymenoptera (mostly bumble bees and bees) that are able to reach the nectar⁴⁶. Seed dispersal is restricted to only a few meters from the maternal plant since there is no mechanism for long-distance seed dispersal⁴⁶.

Previous research has shown some remarkable differences between forest and grassland plants³⁸. Cowslip flowers about three weeks earlier in forest habitats than in grasslands⁴⁶, possibly as a way to take advantage from the higher light availability in spring before canopy closure⁴⁹. Additionally, *P. veris* shows habitat specific variation in flower morphology, with forest populations generally producing larger flowers that show strong deviations in stigma-anther separation, particularly in the L-morph³⁸. This deviation is mainly driven by variation in stigma height, resulting in high and asymmetric reciprocity indices and the occurrence of several short-styled homostylous plants³⁸. In contrast, flowers of grassland plants show clear distyly with low and symmetric reciprocity indices at both the lower and upper level³⁸. Analysis of vegetative traits further shows that plants in grasslands produce more, but shorter flowering stalks. Interestingly, grassland plants have a lower specific leaf area and a higher stomatal density than forest plants (Deschepper, unpublished data), also suggesting habitat specific adaptations to cope with the different environmental conditions imposed by forest and grassland habitats^{58, 65}.

Study area and sampling. The study area is located in the river valley of the Viroin River in the south of Belgium, in the Namur province located in the Calestienne region. The study area consists of a mosaic landscape of forests, calcareous grasslands and agricultural fields and covers a total surface area of 40 km² (Fig. 5). Topographic altitudes within the study area range from 170 m on foothills to 230 m on hilltop locations. Mean annual temperature is 9.8 °C and the average annual precipitation is 780 mm (Royal Meteorological Institute Belgium). More detailed information on the history of the region can be found in Adriaens *et al.*⁵⁹.

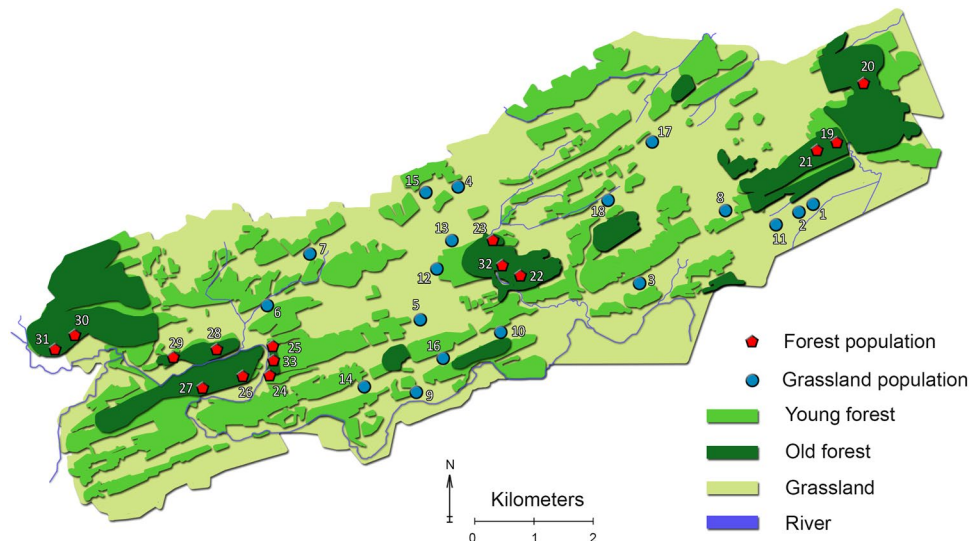


Figure 5. Map of the study area with forest populations indicated by red pentagons and grassland populations by blue circles. Old forest are forests that were present at least from 1775 until today. Map modified with QGIS (www.qgis.org) from OpenStreetMap.org. OpenStreetMap is made available under the Open Database License: <http://opendatacommons.org/licenses/odbl/1.0/>. Any rights in individual contents of the database are licensed under the Database Contents License: <http://opendatacommons.org/licenses/dbcl/1.0/>.

In spring 2015, a total of 33 populations on the northern side of the Viroin River were sampled (Fig. 5). Because it is crucial to separate ecological from spatial factors in determining genetic differentiation¹⁰, populations were sampled for each habitat as uniformly as possible over the whole study area. The sampled populations showed a patchy distribution separated by (intensely) managed farmlands throughout the region (Fig. 5). Forest populations ($n = 15$) were located in open, old growth forests, whereas the grassland populations ($n = 18$) were selected in species-rich calcareous grasslands or grassy road edges. Grazing was common practice in the management of calcareous grasslands in the past and is still applied today⁶⁰. Patches containing *P. veris* populations ranged from approximately 40 m² in size for roadside populations up to several hundreds of square meters in grassland populations. For each population, population size was assessed by counting the total number of flowering plants. Additionally, a large healthy leaf was taken from 24 plants per population and was immediately stored within silica gel until further analyses in the lab.

DNA extraction and genotyping. 20 mg of dry leaf material was weighted and homogenized with the use of zirconia beads and a FastPrep-24 Instrument (MP Biomedicals, USA), after which DNA was isolated using the NucleoSpin[®] Plant II DNA extraction kit (Macherey Nagel, Germany). DNA quality and concentration were evaluated using a NanoDrop ND-2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). For genotyping, 12 primer pairs were used (developed by Bickler *et al.*⁶⁶ and Seino *et al.*⁶⁷ in 3 multiplex reactions for each locus containing 1 μ l DNA sample, 5 μ l of Qiagen Multiplex PCR Master Mix, 2 μ l of one of the multiplexed primer combinations (1 μ M for both the forward and reverse primer) and 2 μ l of RNase-free water. Forward primers were fluorescently labeled with one of the four available colours (FAM, VIC, PET and NED). A 2720 Thermal Cycler (Applied Biosystems, CA, USA) was used for the polymerase chain reaction with the following program: an initial denaturation at 94 °C for 3 min, then 10 cycles of 30 s at 94 °C, 60 s at 57 °C and 30 s at 72 °C. This was followed by a further 28 cycles of 94 °C for 30 s, 55 °C for 60 s and 72 °C for 30 s, finishing with a final extension phase at 72 °C for 7 min. After the fragment elongation, capillary electrophoresis was performed using an ABI3130 instrument (Applied Biosystems, USA) using a GeneScan 500 LIZ (Life Technologies, USA) size standard. Microsatellite peaks were visualized and scored using GeneMapper[®] Software v4.0 (Applied Biosystems).

Analysis of genetic diversity. Micro-Checker was used to check for null-alleles, stutter and large allele dropout⁶⁸. We used R packages 'hierfstat'⁶⁹ and 'adegenet'⁷⁰ in R 3.3.0⁷¹ to estimate the average number of alleles per locus (A), allelic richness (A_r), observed (H_o), expected heterozygosity (H_e) and the inbreeding coefficient (F_{is}) for each population. Linkage disequilibrium was checked with GENEPOP 4.0⁷². Conformance to Hardy-Weinberg equilibrium was determined by assessing the significance of the F_{is} values by means of Fisher's exact tests implemented in GENEPOP 4.0, with specified Markov chain parameters of 5,000 dememorization steps, followed by 1,000 batches of 5,000 iterations per batch. The sequential Bonferroni correction was applied to obtain critical confidence limits for multiple comparisons, with an initial α of 0.05. To test for significant differences in population genetic parameters between forest and grassland populations, a two-tailed t -test or Mann-Whitney U test was performed in R in accordance to assumption fulfillment.

Genetic structure and isolation by distance. Assessment of total genetic diversity partitioned among both habitats (forest vs. grassland), among populations within one habitat, and within populations was done by carrying out a hierarchical analysis of molecular variance (AMOVA) on Euclidean pairwise distances among

individuals implemented in GENALEX⁷³. Individuals were assigned to two groups reflecting the two habitats. Significances were determined using permutation tests.

Parameters for pairwise genetic differentiation F_{ST} (adegenet⁷⁰), Hedrick's standardized measure for G_{ST} and Jost's D (both calculated with the 'mmod' package⁷⁴) were calculated (5000 permutations). A mantel test was used as a nonparametric equivalent of ANOVA to test the hypothesis that genetic differentiation among populations within one habitat was significantly different from differentiation among populations from different habitats⁷⁵. Therefore we constructed a dissimilarity matrix with zeroes in the within group submatrices and ones in the between groups submatrices. If differences between habitats are greater, then the ones in the design matrix will be associated with larger differences. This ANOVA-like Mantel test was performed for the three different measures for differentiation and 5000 random permutations were used. Furthermore, a Mann-Whitney U test was performed to test for significant differences for all three differentiation parameters among habitats. The combined information of different differentiation measures enables a more profound and reliable understanding of the relations between populations^{76–81}.

Evidence of isolation by distance among populations was obtained by examining correlations between matrices of genetic distances (G_{ST}) and geographical distances⁸². Significance of the observed relationships was obtained by using a Mantel test^{75, 83, 84}. A total of 5000 random permutations were performed. The same test was also used to look for isolation by distance separately among populations in grasslands and among populations in forests.

We also investigated whether the two habitats translated into distinct genetic clusters using a Bayesian clustering method applying Markov Chain Monte Carlo estimation implemented in STRUCTURE 2.3.4 software⁴⁸. K was set to range from 1 to 10 with a Markov Chain Monte Carlo of 500 000 iterations following a burn-in of 100 000 iterations. The best K value was inferred from the modal value of the run with the highest log likelihood³⁹. In addition, genetic clustering of populations was visualized by subjecting the G_{ST} matrix to a principal coordinates analysis (PCoA) using the *ape* package in R⁸⁵.

Network analysis of spatial genetic variation. We applied a network approach to spatial genetic variation known as population graphs⁸⁶. Population graphs go beyond describing the traditional F -statistics and provide better insights into the overall population genetic structure by visualizing links of significant genetic similarity between populations^{86, 87} and providing quantitative measurements of link density and strength between and within habitats as estimates of population connectivity. Each node represents a population and can be connected to several other populations by edges or links. In this way, a graphic representation of the identified modules in the network can provide insights into the factors (e.g. geography, habitat differences) that define the assignment of a population to a cluster of genetically similar populations. To perform the network analysis, first a genetic distance matrix was constructed and converted into a correlation matrix in several steps following Fortuna *et al.*⁸⁷.

Some populations are only poorly connected to others and therefore several links can be excluded in order to simplify the model without losing information about the genetic covariance structure among sites. Based on conditional independence it is possible to detect these redundant edges. Edge exclusion deviance is a theoretic measure to calculate whether an edge should be excluded from a fully saturated population graph^{88, 89}. Edge exclusion was calculated as:

$$\tau = -N \ln \left[1 - (r_{ij})^2 \right] \quad (1)$$

where N is the number of individuals in the entire data set, and r_{ij} is the partial correlation coefficient between sites i and j . This statistic asymptotically follows a chi-square distribution. A link among populations i and j is removed if the value of its deviance (theta) is less than 3.84 (the 5% threshold of the chi-square distribution with $df=1$). For further details on graph theory of genetic structures see Dyer and Nason⁸⁶ and Fortuna *et al.*⁸⁷

Modularity analysis. We applied a modularity analysis on the network of spatial genetic variation to detect modules or sets of populations that are more genetically related to each other than to populations belonging to other modules. Newman's modularity⁹⁰, M , can be written as:

$$M = \sum_{\text{all modules } i} \left(\frac{e_i}{L} - \sum_{\text{species in } i} \frac{k_m k_n}{2L \cdot 2L} \right) = \sum_{\text{all modules } i} \left(\frac{e_i}{L} - \frac{d_i d_i}{2L \cdot 2L} \right) \quad (2)$$

where e_i is the number of edges or links within module i , k_m (k_n) is the degree of node m (n) and d_i is the sum of the degrees of all nodes or patches in module i .

Multiple techniques consist of maximizing modularity of a complex network by exploring different possible states of the population structure. Simulated annealing is an optimization techniques that explores low cost configurations of community structure⁹¹ to find high level modules. In this study, we ran 100 replicates of the Guimera and Amaral algorithm⁹¹ for M and then obtained the maximum value of M . In order to test to what extent the value of modularity departs significantly from random expectation we ran 1000 randomizations of the network of genetic variation keeping exactly the same number of links per node, but reshuffling them randomly using a local rewiring algorithm⁹². The P -value was estimated as the fraction of random networks with a modularity value equal to, or higher than, the value obtained for the empirical network. The final population graph displaying the different modules was constructed using the *igraph* package⁹³ in R 3.3.0⁷¹.

Genetic inference of divergence time and historical gene flow. To estimate the timing of divergence between populations growing in grassland and forest patches, we used the coalescent-based approximate

Bayesian computation (ABC) algorithm implemented in the program DIY-ABC version 2.0⁹⁴. This software simulates one or several user specified scenarios of historic and/or demographic events and compares the observed data with summary statistics of the simulated data to calculate posterior distribution of demographic parameters⁹⁴. Our goal was to determine the time of divergence between grassland and forest populations. Therefore, we constructed a scenario with a bifurcation event between both groups and compared it to our data. We set wide priors (10–10000) for all parameters and used 300000 simulated dataset to infer parameters. The 1% datasets with summary statistics closest to the observed data were used for approximate Bayesian computation of parameters by using regression rejection steps of the algorithm following logistic regression of the parameter values. We evaluated the goodness of fit of the model-parameter posterior distribution with the model checking function implemented in DIY-ABC version 2.0⁹⁴. Each of the 14 summary statistics of the observed dataset is ranked against the distribution of the corresponding summary statistics from the posterior predictive distribution and deviations are checked using the estimated *P*-value.

In addition, past migration rates were estimated using MIGRATE-N 3.6⁹⁵. MIGRATE-N uses maximum likelihood or Bayesian inference to jointly estimate historical migration rates and effective population sizes. To compensate for fluctuating population sizes at each site, we ran a simplified migration model with populations grouped per habitat type. Bayesian inference was used with one long chain of 100000 steps and a sampling increment of 1000. The default static heating scheme was used. The burn-in was set to 100000.

References

1. Mayr, E. Ecological factors in speciation. *Evolution* **1**, 263–288 (1947).
2. Rundle, H. D. & Nosil, P. Ecological speciation. *Ecol. Lett.* **8**, 336–352 (2005).
3. Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421**, 259–264 (2003).
4. Rieseberg, L. H. & Willis, J. H. Plant speciation. *Science* **317**, 910–4 (2007).
5. Hoskin, C. J., Higgie, M., McDonald, K. R. & Moritz, C. Reinforcement drives rapid allopatric speciation. *Nature* **437**, 1353–1356 (2005).
6. Salomon, M. A revised cline theory that can be used for quantified analyses of evolutionary processes without parapatric speciation. *J. Biogeogr.* **29**, 509–517 (2002).
7. Wang, I. J., Glor, R. E. & Losos, J. B. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecol. Lett.* **16**, 175–182 (2013).
8. Wang, I. J. & Bradburd, G. S. Isolation by environment. *Mol. Ecol.* **23**, 5649–5662 (2014).
9. Sexton, J. P., Hangartner, S. B. & Hoffmann, A. A. Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* **68**, 1–15 (2014).
10. Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J. & De Meester, L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol. Ecol.* **22**, 5983–5999 (2013).
11. Webb, C. O., Ackerly, D. D., McPeck, M. A. & Donoghue, M. J. Phylogenies and Community Ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505 (2002).
12. Carvalho, P., Diniz-Filho, J. A. F. & Bini, L. M. The impact of Felsenstein's 'Phylogenies and the comparative method' on evolutionary biology. *Scientometrics* **62**, 53–66 (2005).
13. Harvey, P. H. & Pagel, M. D. The comparative method in evolutionary biology. *Oxford Series in Ecology and Evolution* **1**, 239 (1991).
14. Ackerly, D. D., Schilck, D. W. & Webb, C. O. Niche evolution and adaptive radiation: testing the order of trait divergence. *Ecology* **87**, S1 (2006).
15. Lowry, D. B. Ecotypes and the controversy over stages in the formation of new species. *Biol. J. Linn. Soc.* **106**, 241–257 (2012).
16. Fishman, L., Stathos, A., Beardsley, P. M., Williams, C. F. & Hill, J. P. Chromosomal rearrangements and the genetics of reproductive barriers in *Mimulus* (monkey flowers). *Evolution* **67**, 2547–2560 (2013).
17. Grant, V. Plant speciation, 2nd edition. (1981).
18. Ellstrand, N. C. Is gene flow the most important evolutionary force in plants? *Am. J. Bot.* **101**, 737–753 (2014).
19. Lowry, D. B., Rockwood, R. C. & Willis, J. H. Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution* **62**, 2196–2214 (2008).
20. Melo, M. C., Grealy, A., Brittain, B., Walter, G. M. & Ortiz-Barrientos, D. Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsel. *New Phytol.* **203**, 323–334 (2014).
21. Baack, E., Melo, M. C., Rieseberg, L. H. & Ortiz-Barrientos, D. The origins of reproductive isolation in plants. *New Phytologist* **207**, 968–984 (2015).
22. Martin, N. H. & Willis, J. H. Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* **61**, 68–82 (2007).
23. Anderson, B., Alexandersson, R. & Johnson, S. D. Evolution and coexistence of pollination ecotypes in an african *Gladiolus* (iridaceae). *Evolution* **64**, 960–972 (2010).
24. Kay, K. M. & Sargent, R. D. The role of animal pollination in plant speciation: integrating ecology, geography, and genetics. *Annu. Rev. Ecol. Syst.* **40**, 637–656 (2009).
25. Keller, B., Thomson, J. D. & Conti, E. Heterostyly promotes disassortative pollination and reduces sexual interference in Darwin's primroses: evidence from experimental studies. *Funct. Ecol.* **28**, 1413–1425 (2014).
26. Anderson, J. T., Willis, J. H. & Mitchell-Olds, T. Evolutionary genetics of plant adaptation. *Trends in Genetics* **27**, 258–266 (2011).
27. Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A. & Willis, J. H. Review. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 3009–3021 (2008).
28. Nosil, P. Reproductive isolation caused by visual predation on migrants between divergent environments. *Proc. R. Soc. B Biol. Sci.* **271**, 1521–1528 (2004).
29. Montalvo, A. M. & Ellstrand, N. C. Transplantation of the subscrub *Lotus scoparius*: testing the home-site advantage hypothesis. *Conserv. Biol.* **14**, 1034–1045 (2014).
30. Plath, M. *et al.* Genetic differentiation and selection against migrants in evolutionarily replicated extreme environments. *Evolution* **67**, 2647–2661 (2013).
31. Via, S., Bouck, aC. & Skillman, S. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* **54**, 1626–1637 (2000).
32. Hendry, A. P. Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evol. Ecol. Res.* **6**, 1219–1236 (2004).
33. Nosil, P., Egan, S. P. & Funk, D. J. Heterogeneous genomic differentiation between walking-stick ecotypes: 'Isolation by adaptation' and multiple roles for divergent selection. *Evolution* **62**, 316–336 (2008).
34. Lowry, D. B. *et al.* Adaptations between ecotypes and along environmental gradients in *Panicum virgatum*. *Am. Nat.* **183**, 682–692 (2014).

35. Andrew, R. L., Ostevik, K. L., Ebert, D. P. & Rieseberg, L. H. Adaptation with gene flow across the landscape in a dune sunflower. *Mol. Ecol.* **21**, 2078–2091 (2012).
36. André, C. *et al.* Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity* **106**, 270–280 (2011).
37. Thibert-Plante, X. & Hendry, A. P. When can ecological speciation be detected with neutral loci? *Mol. Ecol.* **19**, 2301–2314 (2010).
38. Brys, R. & Jacquemyn, H. Disruption of the distylous syndrome in *Primula veris*. *Ann. Bot.* **115**, 27–39 (2015).
39. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
40. Mallet, B., Martos, F., Blambert, L., Pailler, T. & Humeau, L. Evidence for isolation-by-habitat among populations of an epiphytic orchid species on a small oceanic island. *PLoS One* **9**, (2014).
41. Spurgin, L. G., Illera, J. C., Jorgensen, T. H., Dawson, D. A. & Richardson, D. S. Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Mol. Ecol.* **23**, 1028–1039 (2014).
42. Galloway, L. F. & Fenster, C. B. Population differentiation in an annual legume: local adaptation. *Evolution* **54**, 1173–1181 (2000).
43. Shao, J. *et al.* Genetic diversity and gene flow within and between two different habitats of *Primula merrilliana* (Primulaceae), an endangered distylous forest herb in eastern China. *Bot. J. Linn. Soc.* **179**, 172–189 (2015).
44. Jacquemyn, H., Honnay, O., Galbusera, P. & Roldán-Ruiz, I. Genetic structure of the forest herb *Primula elatior* in a changing landscape. *Mol. Ecol.* **13**, 211–219 (2004).
45. Van Rossum, F. & Triest, L. Fine-scale spatial genetic structure of the distylous *Primula veris* in fragmented habitats. *Plant Biol.* **9**, 374–382 (2007).
46. Brys, R. & Jacquemyn, H. Biological Flora of the British Isles: *Primula veris* L. *J. Ecol.* **97**, 581–600 (2009).
47. Kreyer, D., Oed, A., Walther-Hellwig, K. & Frankl, R. Are forests potential landscape barriers for foraging bumblebees? Landscape scale experiments with *Bombus terrestris* agg. and *Bombus pascuorum* (Hymenoptera, Apidae). *Biol. Conserv.* **116**, 111–118 (2004).
48. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
49. Kudo, G., Ida, T. Y. & Tani, T. Linkages between phenology, pollination, photosynthesis, and reproduction in deciduous forest understory plants. *Ecology* **89**, 321–331 (2008).
50. Campbell, D. R. Pollinator Sharing and Seed Set of *Stellaria pubera*: competition for Pollination. *Ecology* **66**, 544–553 (1985).
51. Heinrich, B. Flowering Phenologies: bog, woodland, and disturbed habitats. *Ecology* **57**, 890–899 (1976).
52. Whale, D. M. A physiological study of habitat preference in *Primula* species. (1982).
53. Sanders, R. Pollinatoren en reproductieve isolatie van de gulden sleutelbloem (*Primula veris*), unpublished master's thesis (2014).
54. Van Der Niet, T., Peakall, R. & Johnson, S. D. Pollinator-driven ecological speciation in plants: new evidence and future perspectives. *Ann. Bot.* **113**, 199–211 (2014).
55. Nosil, P., Vines, T. H. & Funk, D. J. Immigrants from divergent habitats. *Evolution* **59**, 705–719 (2005).
56. Hufford, K. M. & Mazer, S. J. Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends Ecol. Evol.* **18**, 147–155 (2003).
57. Mott, K. A. & Michaelson, O. Amphistomy as an adaptation to high light intensity in *Ambrosia cordifolia* (Compositae). *American Journal of Botany* **78**, 76–79 (1991).
58. Xu, Z. & Zhou, G. Responses of leaf stomatal density to water status and its relationship with photosynthesis in a grass. *J. Exp. Bot.* **59**, 3317–3325 (2008).
59. Adriaens, D., Honnay, O. & Hermy, M. No evidence of a plant extinction debt in highly fragmented calcareous grasslands in Belgium. *Biol. Conserv.* **133**, 212–224 (2006).
60. Poschlod, P. & WallisDeVries, M. F. The historical and socioeconomic perspective of calcareous grasslands - lessons from the distant and recent past. *Biol. Conserv.* **104**, 361–376 (2002).
61. Butaye, J., Honnay, O., Adriaens, D., Delescaille, L. & Hermy, M. Phytosociology and phytogeography of the calcareous grasslands on Devonian limestone in Southwest Belgium. *Belgian J. Bot.* **138**, 24–38 (2005).
62. Hegland, S. J. & Boeke, L. Relationships between the density and diversity of floral resources and flower visitor activity in a temperate grassland community. *Ecol. Entomol.* **31**, 532–538 (2006).
63. Knight, C. A. & Ackerly, D. D. Evolution and plasticity of photosynthetic thermal tolerance, specific leaf area and leaf size: congeneric species from desert and coastal environments. *New Phytol.* **160**, 337–347 (2003).
64. Van Rossum, F., De Sousa, S. C. & Triest, L. Morph-specific differences in reproductive success in the distylous *Primula veris* in a context of habitat fragmentation. *Acta Oecologica* **30**, 426–433 (2006).
65. Abrams, M. D., Klooppel, B. D. & Kubiske, M. E. Ecophysiological and morphological responses to shade and drought in two contrasting ecotypes of *Prunus serotina*. *Tree Physiol.* **10**, 343–355 (1992).
66. Bickler, C., A'Hara, S., Cottrell, J., Rogers, L. & Bridle, J. Characterisation of thirteen polymorphic microsatellite markers for cowslip (*Primula veris* L.) developed using a 454 sequencing approach. *Conserv. Genet. Resour.* **5**, 1185–1187 (2013).
67. Seino, M. M., Vega, C., de Bazaga, P., Jacquemyn, H. & Herrera, C. M. Development and characterization of microsatellite loci for the primrose *Primula vulgaris* and successful cross-amplification in the congeneric *P. elatior* and *P. veris*. *Conserv. Genet. Resour.* **6**, 653–655 (2014).
68. Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538 (2004).
69. Goudet, J. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
70. Jombart, T. ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
71. R Core team. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, **55**, 275–286 (2015).
72. Rousset, F. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106 (2008).
73. Peakall, R. & Smouse, P. E. GenALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539 (2012).
74. Winter, D. J. MMOD: An R library for the calculation of population differentiation statistics. *Mol. Ecol. Resour.* **12**, 1158–1160 (2012).
75. Sokal, R. R. & Rohlf, F. J. *Biometry: the principles and practice of statistics in biological research*. New York **3**, (1995).
76. Gerlach, G., Jueterbock, A., Kraemer, P., Deppermann, J. & Harmand, P. Calculations of population differentiation based on G_{ST} and D : Forget G_{ST} but not all of statistics. *Mol. Ecol.* **19**, 3845–3852 (2010).
77. Ryman, N. & Leimar, O. G_{ST} is still a useful measure of genetic differentiation - A comment on Jost's D . *Mol. Ecol.* **18**, 2084–2087 (2009).
78. Jost, L. G_{ST} and its relatives do not measure differentiation. *Mol. Ecol.* **17**, 4015–4026 (2008).
79. Verity, R. & Nichols, R. A. What is genetic differentiation, and how should we measure it - G_{ST} , D , neither or both? *Mol. Ecol.* **23**, 4216–4225 (2014).
80. Hedrick, P. W. A standardized genetic differentiation measure. *Evolution* **59**, 1633–1638 (2005).
81. Meirmans, P. G. & Hedrick, P. W. Assessing population structure: F_{ST} and related measures. *Mol. Ecol. Resour.* **11**, 5–18 (2011).
82. Slatkin, M. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264–279 (1993).

83. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
84. Legendre, P. & Fortin, M. J. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol. Ecol. Resour.* **10**, 831–844 (2010).
85. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
86. Dyer, R. J. & Nason, J. D. Population graphs: The graph theoretic shape of genetic structure. *Mol. Ecol.* **13**, 1713–1727 (2004).
87. Fortuna, M. A., Albaladejo, R. G., Aparicio, A. & Bascompte, J. Networks of spatial genetic variation across species. *Proc. Natl. Acad. Sci.* **106**, 19044–19049 (2009).
88. Magwene, P. M. New tools for studying integration and modularity. *Evolution* **55**, 1734–45 (2001).
89. Whittaker, J. Graphical Methods in Applied Multivariate Statistics. (1990).
90. Newman, M. E. J. Fast algorithm for detecting community structure in networks. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **69**, 1–5 (2004).
91. Guimerà, R. & Amaral, L. A. N. Cartography of complex networks: modules and universal roles. *J. Stat. Mech.* **2005** nihpa35573 (2005).
92. Gale, D. A theorem on flows in networks. *Pacific J. Math.* **7**, 1073–1082 (1957).
93. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* **1695**, 1695 (2006).
94. Cornuet, J. M. *et al.* DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30**, 1187–1189 (2014).
95. Beerli, P. & Felsenstein, J. Maximum-likelihood estimation of migration rates and effective population numbers. *Genetics* **152**, 763–773 (1999).
96. Ramasamy, R. K., Ramasamy, S., Bindroo, B. B. & Naik, V. G. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springerplus* **3**, 431 (2014).

Acknowledgements

This research was funded by a grant from the Flemish Fund for Scientific Research (FWO). Many thanks to Thomas Sebrechts for his help in the laboratory procedures.

Author Contributions

Conceived and designed the study: B.J., J.H., D.P. Collected tissue samples: D.P. Analysed the data: D.P., F.M. Wrote the paper: D.P. All authors contributed to and approved the final manuscript version.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017