Ecology and life history affect different aspects of the population structure of 27 high-alpine plants

PATRICK G. MEIRMANS,*† JEROME GOUDET,† IntraBioDiv Consortium[1] and OSCAR E. GAGGIOTTI*

Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, 38041 Grenoble, France, †Department of Ecology & Evolution, Biophore, University of Lausanne, CH-1015 Lausanne, Switzerland

Abstract

A plant species' genetic population structure is the result of a complex combination of its life history, ecological preferences, position in the ecosystem and historical factors. As a result, many different statistical methods exist that measure different aspects of species' genetic structure. However, little is known about how these methods are interrelated and how they are related to a species' ecology and life history. In this study, we used the IntraBioDiv amplified fragment length polymorphisms data set from 27 high-alpine species to calculate eight genetic summary statistics that we jointly correlate to a set of six ecological and life-history traits. We found that there is a large amount of redundancy among the calculated summary statistics and that there is a significant association with the matrix of species traits. In a multivariate analysis, two main aspects of population structure were visible among the 27 species. The first aspect is related to the species' dispersal capacities and the second is most likely related to the species' postglacial recolonization of the Alps. Furthermore, we found that some summary statistics, most importantly Mantel's r and Jost's D, show different behaviour than expected based on theory. We therefore advise caution in drawing too strong conclusions from these statistics.

Keywords: AFLP, Alps, dispersal, genetic structure, glacial refugia, redundancy analysis

Received 5 February 2010; revision received 9 May 2011; accepted 13 May 2011

Introduction

It is well known that a plant species' life-history traits influence both its genetic diversity and the way this diversity is distributed over populations (Hamrick & Godt 1989, 1996; Nybom & Bartish 2000; Nybom 2004; Thiel-Egenter *et al.* 2009). The most important observations that have emerged from review studies on this subject are that outcrossing, long-lived, late successional and animal-dispersed species have high genetic diversity and most of their genetic variation is found within populations (Thiel-Egenter *et al.* 2009). Conversely, in selfing, annual and early successional species most of

Correspondence (Present address): Patrick Meirmans, Fax: +31 20 525 7832;

E-mail: p.g.meirmans@uva.nl

[1]http://intrabiodiv.vitamib.com/: Nadir Alvarez, Felix Gugerli, Rolf Holderegger, Riccardo Negrini, Peter Schönswetter, Pierre Taberlet, Conny Thiel-Egenter, Andreas Tribsch (see supplementary data for addresses)

the variation is found between populations. In addition, species' life-history traits have been shown to have a significant effect on the degree of spatial autocorrelation in the distribution of genetic variation (Vekemans & Hardy 2004). These patterns are independent of the type of genetic marker that is used (Nybom 2004) as they have been shown for allozymes (Hamrick & Godt 1989, 1996), RAPDs (Nybom & Bartish 2000), microsatellites (Nybom 2004) and amplified fragment length polymorphisms (AFLPs, Thiel-Egenter *et al.* 2009).

It is also well known that historical events can have profound effects on the distribution of genetic variation. One important example is the pleistocene ice ages that caused dramatic fluctuations in species' ranges and population sizes (Taberlet *et al.* 1998; Hewitt 2000). A common observation is that genetic variation is structured as a result of the postglacial recolonization from multiple refugia, often resulting in a hierarchical population structure and a lower diversity in the once-glaciated part of the range (Schönswetter *et al.* 2005). Here,

the ecological niche and life history of species also play a role as they determine the species' ability to respond to climate changes. For example, using the IntraBioDiv data set of 27 alpine plant species, Alvarez *et al.* (2009) found that the spatial arrangement of population clusters, that presumably resulted from postglacial recolonization, was related to the species' soil preference.

From the above, we see that there are different ways in which a species' ecology and life-history traits affect the distribution of genetic variation. Determining the relative importance of these factors is difficult as they can all influence the amount of differentiation among populations, the strength of the spatial genetic structure, or the genetic diversity within populations. Nowadays, many different methods are used to test different hypotheses regarding the genetic structure of species. *F*-statistics are used to test population differentiation, assignment tests are used to detect migrants, while Mantel tests are used to analyse isolation by distance. Even though several studies have shown the link between species' genetic structure and life-history traits (Hamrick & Godt 1989, 1996; Nybom & Bartish 2000; Nybom 2004; Vekemans & Hardy 2004), we still lack an overview of how different genetic summary statistics are correlated to each other, how they are jointly correlated to species' ecology and life-history traits, and whether different statistics are correlated to different traits. For example, a new statistic, *D*, has been developed for quantifying the amount of differentiation among populations (Jost 2008), which circumvents some statistical problems associated with F_{ST}. However, *D* has been criticized since its value is independent of one of the most important demographic parameters, the population size, and is therefore less suited for inferences of the effect of demography on the genetic structure (Ryman & Leimar 2009; Meirmans & Hedrick 2011). It would therefore be interesting to include both statistics into a single analysis to test how they are correlated with each other and with species' ecological and life-history traits. However, a joint analysis of multiple genetic summary statistics and species traits is difficult to perform by reviewing the existing literature due to the different sampling strategies, laboratory techniques and different statistical methods used by different studies.

Here, we take advantage of the extraordinarily large IntraBioDiv data set, already used by Gugerli *et al.* (2008) and Alvarez *et al.* (2009), to investigate how different aspects of the species' genetic population structure are related to each other and to the species' ecological preferences and life-history traits. The data set comprises 27 high-alpine species sampled using the same protocol and analysed with the same laboratory technique (AFLP) and statistical methods. We used this data set to calculate for each species a set of eight genetic summary statistics, which we jointly correlated to a set of six species traits. We are especially interested in looking whether the ecological traits and the life-history traits are differently related to the species' genetic structure. On a more technical note, we also explore how the different summary statistics relate to each other and to these traits.

Materials and methods

Sampling and genotyping

We used the IntraBioDiv AFLP-data set of 27 high-alpine plant species (Alvarez *et al.* 2009). The included species are all widespread and abundant and represent a wide range of life-histories and ecological characteristics. A detailed description of the sampling and genotyping protocols can be found in Alvarez *et al.* (2009) and Gugerli *et al.* (2008). In short, sampling was performed across the European Alps on a grid with a cell size of 20′ longitude by 12′ latitude (~20 × 22.5 km). Plant material was sampled from every second cell of the grid by taking leaves from three individuals at a single location within the cell, when a species was present in the cell. The number of locations sampled per species ranged from 44 to 137, and the maximum distance between sampling locations for the species ranged from 559 to 906 km (see Table 1). For genotyping, the species were distributed over five different laboratories, with every laboratory processing all individuals for five or six species (Table 1). DNA extraction was performed using a CTAB protocol (lab E), or the DNeasy 96 Plant Kit (QIAGEN, labs A–D). The sampled individuals were genotyped using AFLP (Vos *et al.* 1995) with three primer/enzyme combinations, yielding between 58 and 252 marker loci. Markers were separated using either electrophoresis on 8% polyacrylamide gels (lab E) or on automatic capillary sequencers (labs A–D). Scoring the markers as presence–absence was done by hand (lab E) or automatically (labs A–D). This data set has already been analysed with respect to the spatial distribution of population clusters (Alvarez *et al.* 2009) and the within-population heterozygosity (Thiel-Egenter *et al.* 2009).

Life history and ecological characters

We selected six life-history traits and ecological characteristics that are either directly related to the species' mode of dispersal or thought to affect the species' genetic structure in other ways: soil substrate affinity, altitude, ecological dominance, successional status, mode of seed dispersal, and mode of pollination (Table 1). The soil substrate affinity was classified into three categories based on the classification of Landolt (1977): (i) calcicolous, growing on substrate from alka-

Table 1 Plant species, abbreviations, number of sampled populations, sampling range (maximum distance between sampling locations), number of used AFLP markers and values for six different species traits. See the Materials and methods section for information on the classification of the species traits

Species	Abbreviation	Lab	Populations	Range (km)	Markers	Soil	Succession	Dominance	Pollination	Seed	Altitude
Androsace obtusifolia	Aob	B	45	814	138	3	3	1	1	1	2
Arabis alpina	Aal	A	129	902	151	1	1	1	1	1	3
Campanula barbata	Cba	D	104	799	114	3	3	1	1	1	2.5
Carex firma	Cfi	E	76	730	58	1	1	2	2	3	3
Carex sempervirens	Cse	C	137	898	125	2	3	2	2	2	2.5
Cerastium uniflorum	Cun	D	44	641	93	3	1	2	1	1	2.5
Cirsium spinosissimum	Csp	D	110	774	95	2	1	1	1	3	2
Dryas octopetala	Doc	A	124	861	101	1	1	2	1	3	3
Gentiana nivalis	Gni	D	74	825	166	2	3	1	1	1	2.5
Geum montanum	Gmo	C	122	862	93	3	3	1	1	3	2.5
Geum reptans	Gre	C	51	708	61	3	1	1	1	3	2.5
Gypsophila repens	Gyr	C	107	794	94	1	1	1	1	1	2.5
Hedysarum hedysaroides	Hhe	E	76	807	123	1	3	1	1	3	2
Hornungia alpina	Hal	B	97	853	225	1	1	1	1	1	3
Hypochaeris uniflora	Hun	A	59	683	102	3	3	1	1	3	3
Juncus trifidus	Jtr	C	91	813	88	3	2	2	2	3	1.5
Ligusticum mutellinoides	Lmu	E	56	808	97	2	2	1	1	3	1
Loiseleuria procumbens	Lpr	A	90	776	121	3	3	1	1	1	2
Luzula alpinopilosa	Lal	D	82	725	234	3	2	1	2	2	2
Peucedanum ostruthium	Pos	E	117	783	113	2	2	1	1	3	2
Phyteuma betonicifolium	Pbt	B	104	743	165	3	3	1	1	1	2
Phyteuma hemisphaericum	Phm	B	76	662	234	3	2	1	1	1	2.5
Ranunculus alpestris	Ral	B	79	841	252	1	2	1	1	3	2.5
Rhododendron ferrugineum	Rfe	A	126	850	119	3	3	2	1	1	2
Saxifraga stellaris	Sst	B	101	806	199	2	1	1	1	1	3
Sesleria caerulea	Sca	E	137	906	70	1	2	2	2	3	3.5
Trifolium alpinum	Tal	E	64	559	98	3	3	2	1	3	2

line limestone bedrock; (ii) intermediate, growing on either crystalline or limestone bedrock; (iii) silicicolous, growing on acidic, crystalline bedrock. The range of the altitudinal preference for a species was calculated as the number of vegetation belts (colline, montane, subalpine, alpine and nival belt) in which it occurs, using data from the *Flora alpina* (Aeschimann *et al.* 2004). Vegetation belts where a species occurs only sparsely were counted as 0.5, whereas vegetation belts where the species occurs more frequently were counted as 1.0. So for example, *Carex sempervirens*, which occurs at low frequencies in the montane belt, but mostly in the subalpine and alpine belts, was given an altitudinal range score of (0.5 + 1.0 + 1.0) = 2.5. We also calculated the average altitude for species based on the vegetation belts but we found that these values were correlated with the altitudinal range, with species with a preference for high altitudes having a smaller altitudinal range.

All other traits were obtained from local flora and experts' knowledge of the species (A. Tribsch, M. Ronikier, S. Ertl & T. Englisch, unpublished, also see Thiel-Egenter *et al.* 2009). The ecological dominance of species in their main distribution range was classified as:

(i) nondominant; (ii) dominant. The successional stage of the species was divided into three categories: (i) early successional; (ii) mid-successional; (iii) late successional. The mode of seed dispersal was classified into three groups (i) gravity dispersal (boleochory); (ii) animal dispersal (zoochory); (iii) wind dispersal (anemochory). The mode of pollination was classified into two groups (i) animal pollination; (ii) wind pollination. No strict selfers were included among the 27 species.

Genetic summary statistics

We used the AFLP-data to calculate a set of eight genetic summary statistics, selecting statistics that are generally thought to be related to a species' dispersal capacities and its population structure: Mantel's r; the genetic neighbourhood Nb; the within-population diversity H_S; the total diversity H_T; the fixation index F_{ST}; the differentiation measure D; the average dispersal distance from an assignment test; and the kurtosis of the distribution of dispersal distances from an assignment test. For the calculation of these summary statistics, every sampling location was treated as a separate popu-

lation. Since only three individuals were sampled per location, the estimates of the population allele frequencies are necessarily poor, which may lead to lower H_S estimates and inflated F_{ST}. However, we believe that this does not greatly affect our analysis, since the sampling strategy was very uniform among the species and we are primarily interested in comparing the summary statistics among the 27 included species and not in the values per se. Furthermore, because of the large number of loci and the large number of sampling locations, the standard errors of the estimated F-statistics were generally very low, at a value of 0.02 averaged over all species, for an average F_{ST} of 0.34.

We used Mantel's r (Mantel 1967) to describe the strength of the correlation between the genetic and geographical distance. A matrix of F_{ST} values between all pairs of populations was calculated, as well as a matrix of geographical distances. The Mantel test was then performed using the matrix of $F_{ST}/(1 - F_{ST})$ and the logarithm of the geographic distance, as the relationship between these two transformed matrices is expected to be linear in a two-dimensional habitat (Rousset 1997). These same two matrices were also used to calculate Rousset's Nb, which is a product of the population density and the mean squared distance of gene movements (Rousset 1997), though it is also interpreted as the size of the genetic neighbourhood. Nb can be estimated as $Nb = 1/b_{log}$, where b_{log} is the slope of the regression between $F_{ST}/(1 - F_{ST})$ and the logarithm of the geographical distance. Calculation and transformation of the distance matrices, the Mantel tests and the regression analyses were performed using the program GENO-DIVE v. 2.0 (Meirmans & Van Tienderen 2004, available from http://www.patrickmeirmans.com).

The within-population diversity H_S and the total diversity H_T were estimated from the AFLP data following the method of Lynch & Milligan (1994), using the software AFLP-SURV (Vekemans 2002, available from http://www.ulb.ac.be/sciences/lagev/aflp-surv.html). Estimates of F_{ST} were obtained from an Analysis of Molecular Variance (Excoffier et al. 1992), based on pairwise differences between the AFLP profiles using GENODIVE. The estimates of the differentiation statistic D were calculated from Jost's (2008) equation 11 using estimates of H_S and H_T obtained with AFLP-SURV. We did not include the results of a clustering analysis of the data to detect higher-level population structure (Pritchard et al. 2000; Jombart et al. 2010), since we were unable to obtain satisfactory estimates of the number of clusters for all species.

To estimate dispersal, we performed for every species an assignment test in which we assigned all individuals to their most likely location of origin. We used a distinctly spatial approach for the assignment test, where

the marker frequencies of unsampled grid cells were estimated by spatial interpolation using the geostatistical technique of Ordinary kriging (Cressie 1993). We then inferred for every individual its most likely location of origin on the grid based on the interpolated allele frequencies, using the likelihood method of Paetkau et al. (1995) in combination with the leave-one-out strategy (Paetkau et al. 1998). Interpolated marker frequencies lower than 0.005 were replaced with a frequency of 0.005 (Paetkau et al. 1998). We then calculated for every individual the inferred dispersal distance, i.e. the distance between the sampling location and the inferred location of origin. The inferred dispersal distances were then used to calculate the last two summary statistics: the average dispersal distance of putative migrants (individuals with nonzero distances) and the kurtosis of the distribution of the inferred dispersal distances. We did not use an exclusion method to filter putative migrants based on their likelihood values (Rannala & Mountain 1997; Cornuet et al. 1999; Paetkau et al. 2004), as we noticed that those methods tend to exclude mostly individuals with very short inferred dispersal distances. The spatial assignment test was performed in R (R Development Core Team 2010) using a custom script, which is available upon request. We used the *fitvario* and *Kriging* functions from the *RandomFields* package (Schlater 2001) for maximum likelihood fitting of an exponential variogram and subsequent kriging of the marker frequencies over the landscape.

Analysis of congruence between data sets

We used a redundancy analysis (RDA, Rao 1964) to perform a direct comparison between the matrix with genetic summary statistics and the matrix of species traits. RDA is a canonical ordination technique where the ordination of a matrix of variables Y is constrained to be maximally related to combinations of variables in a second matrix X. In practice, RDA is a combination of a Principal Components Analysis (PCA) and multiple regression (Legendre & Legendre 1998). First, a multiple regression is performed between every variable in matrix Y (the genetic summary statistics) and the explanatory variables in matrix X (life-history and ecological traits), and a matrix of fitted values of Y is produced. A PCA is then performed on the matrix of fitted values. This PCA returns a number of independent axes that describe how the variation in Y can be explained by the correlations with the variables in matrix X. We performed the RDA with the matrix of genetic summary statistics as the matrix Y of dependent variables, and the matrix of species traits as the matrix X of independent variables. Both matrices were standardized,

and the kurtosis summary statistic was log-transformed before performing the RDA.

An earlier analysis of the within-species analysis using the same data set found that there was a strong difference between the values for lab E and the other four labs (fig. 2 in Thiel-Egenter *et al.* 2009). The reasons for this difference were not exactly clear, but Thiel-Egenter *et al.* (2009) gave two possible explanations. First, labs A–D used capillary sequencers and automated scoring of bands, whereas lab E used gel electrophoresis and manual scoring of the bands. Second, the species in lab E are all wind-dispersed and are therefore expected to have a higher genetic diversity (Hamrick & Godt 1996). Since this lab-effect can lead to spurious correlations in our analyses, we accounted for it in two different ways. In our main analysis, we used Lab E as a covariate in the RDA, so that the effect of the lab was 'partialled out' before the analysis. In addition, we performed the RDA without the species analysed in lab E, so with only the subset of 21 species that were analysed in labs A–D.

We used permutations to test the overall significance of the RDA as well as the significance of individual RDA-axes. As a test statistic for the overall test, we used the amount of inertia in matrix Y that is explained by the constrained analysis (i.e. the sum of all canonical eigenvalues). For testing the significance of individual RDA-axes, we used the axis-specific eigenvalue to calculate Ter Braak's (1990) F-statistic, using the approach outlined in Legendre & Legendre (1998). We performed the permutation test only for RDA-axes that explained more than 10% of the variation in the matrix of fitted values. The RDA was performed in R using the 'vegan' package (Oksanen *et al.* 2009) with the 'rda' command to calculate the RDA and the 'anova.cca' command to perform the permutations. The permutation test was run with a maximum of 99 999 random permutations. The results from the permutation test were verified using a series of univariate ANOVAs. In addition to the RDA, we performed standard PCAs on both matrices individually to see how their variables were related to each other outside of the constraints of the RDA. These PCAs were performed using the 'princomp' command in R.

Results

Table 2 shows that there is a wide range of population structure patterns present among the 27 high-alpine species, though in general all species had moderate to strong population differentiation, which was significant for all species ($P = 0.001$, 9999 permutations). F_{ST}-estimates ranged from 0.14 for *Ligusticum mutellinoides* to 0.68 for *Arabis alpina*, with an average of 0.34. In contrast,

estimates of D were much lower with an average of only 0.09, with a minimum of 0.04 for *Hypochaeris uniflora* and a maximum of 0.18 for *Hedysarum hedysaroides*. In all species except *Hornungia alpina*, a significant pattern of isolation by distance was present. The values of Mantel's r ranged from 0.04 for *H. alpina* to 0.65 for *Phyteuma hemisphaericum*, with an average of 0.29.

The RDA revealed a significant relationship between the matrix of genetic summary statistics and the matrix of species traits (Table 3). The RDA incorporating the lab-effect as a covariate explained about one-third of the variation in the matrix of summary statistics (constrained inertia = 29.8%), which proved strongly significant in the permutation test ($P = 0.0091$). This means that there is a significant correlation between the traits and the genetic structure of the species, as described by the selected summary statistics. The first two RDA-axes together explained more than 90% of the constrained inertia (49.1% and 41.2% for the first and second RDA-axis, respectively), but neither axis was significant in the permutation test ($P = 0.09$ and $P = 0.064$). The third RDA-axis only explained 6% of the constrained inertia, so we discarded the third and higher axes for all interpretations below. When we removed all species genotyped in lab E (Fig. S1, Supporting information), the percentage of explained variance was higher (50.7%) than when lab E was used as a covariate (Fig. 1). Though it explained less of the total variance, we chose to use the analysis with lab E as a covariate for our main results below, as it is the most conservative approach and it allowed us to include all 27 species.

Figure 1 provides a triplot of the first two axes of the RDA. Here, the positions of the names of the genetic summary statistics indicate their placement on the two axes. The arrows show the influence of the species traits on the ordination of the genetic summary statistics. Of the six life-history traits, there were four that showed a strong correlation with the first two RDA-axes: the mode of seed dispersal, the ecological dominance, the mode of pollination and the soil preference. These four arrows roughly form two groups that are more or less perpendicular, and therefore independent. For the successional stage and the altitudinal range, the correlations with the first two axes were much less pronounced. The lesser role of the altitudinal range was not due to the way we coded the altitudinal preference of the species. When we used the average altitude rather than the altitudinal range, this lead to a decrease in the percentage of explained variance in the RDA (see Fig. S2, Supporting information). When we restricted the analysis to those species that were not genotyped in Lab E (see Fig. S1, Supporting information), the same distribution in two groups of arrows was visible, though the influence of the altitudinal range was

Table 2 Values for eight genetic summary statistics calculated for all 27 species. See the Materials and methods section for information on how the statistics were calculated
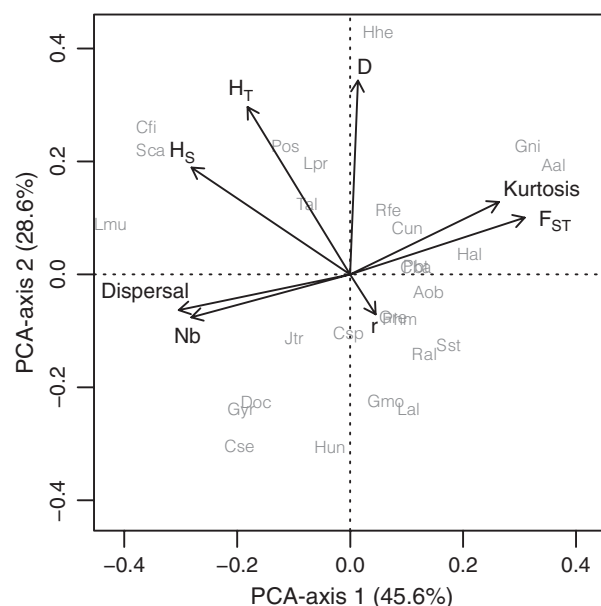
Species	r	Nb	H_S	H_T	F_{ST}	D	Dispersal	Kurtosis
Androsace obtusifolia	0.31	5.87	0.08	0.16	0.42	0.08	2.60	6.20
Arabis alpina	0.07	1.37	0.05	0.17	0.68	0.13	2.14	21.21
Campanula barbata	0.36	4.48	0.09	0.17	0.39	0.09	2.39	5.40
Carex firma	0.16	10.57	0.21	0.31	0.17	0.13	6.39	1.43
Carex sempervirens	0.14	13.29	0.07	0.12	0.33	0.05	7.98	0.64
Cerastium uniflorum	0.38	3.95	0.10	0.20	0.40	0.11	2.51	4.52
Cirsium spinosissimum	0.28	7.62	0.10	0.16	0.31	0.07	3.98	3.26
Dryas octopetala	0.28	12.47	0.11	0.15	0.20	0.05	5.33	1.62
Gentiana nivalis	0.17	2.05	0.07	0.19	0.60	0.13	1.92	25.25
Geum montanum	0.39	4.64	0.07	0.12	0.31	0.05	3.58	2.45
Geum reptans	0.60	2.47	0.09	0.17	0.38	0.08	3.17	1.74
Gypsophila repens	0.17	17.59	0.10	0.14	0.24	0.05	5.80	1.84
Hedysarum hedysaroides	0.32	2.27	0.13	0.29	0.49	0.18	2.64	2.66
Hornungia alpina	0.04	7.97	0.07	0.15	0.48	0.09	1.95	17.88
Hypochaeris uniflora	0.57	5.54	0.09	0.13	0.20	0.04	3.31	0.80
Juncus trifidus	0.21	6.52	0.11	0.17	0.29	0.07	4.10	0.62
Ligusticum mutellinoides	0.18	22.04	0.19	0.27	0.14	0.10	6.18	1.15
Loiseleuria procumbens	0.39	4.47	0.16	0.25	0.29	0.11	2.98	2.74
Luzula alpinopilosa	0.27	7.77	0.07	0.11	0.29	0.04	2.69	5.72
Peucedanum ostruthium	0.22	10.15	0.17	0.26	0.28	0.12	3.01	4.62
Phyteuma betonicifolium	0.57	2.68	0.10	0.18	0.40	0.10	2.33	2.98
Phyteuma hemisphaericum	0.65	3.49	0.10	0.17	0.34	0.08	1.99	1.99
Ranunculus alpestris	0.21	6.52	0.07	0.13	0.38	0.06	2.22	4.43
Rhododendron ferrugineum	0.22	3.14	0.11	0.21	0.38	0.10	2.63	4.07
Saxifraga stellaris	0.21	5.81	0.06	0.12	0.43	0.07	2.21	4.93
Sesleria caerulea	0.15	12.57	0.20	0.30	0.20	0.12	6.47	1.87
Trifolium alpinum	0.40	5.81	0.15	0.24	0.27	0.11	3.31	2.56

Table 3 Results of the redundancy analysis performed on the matrix of summary statistics constrained by the matrix of species traits, using laboratory as a covariate. Permutation tests were used to test for overall significance and the significance of the axes that explained >10% of the inertia

Axis	Eigenvalue	% Inertia	P-value
1	1.17	49%	0.090
2	0.98	41%	0.064
3	0.14	6%	—
4	0.07	3%	—
5	0.02	1%	—
6	0.00	<1%	—
All constrained	2.38	30%	0.0091
All unconstrained	5.62	70%	—
Total	8.00	100%	—

somewhat stronger and that of the ecological dominance was weaker. In addition, the same two groups of arrows were visible when the RDA analysis was performed on all 27 species without any correction for the lab-effect (not shown), indicating that the effect of Lab E changes the strength of the statistical support, but not the main results of the analysis.

The first group of arrows consists of the two dispersal related traits, namely the mode of seed dispersal and the mode of pollination, in combination with the ecological dominance. The fact that these arrows point in the same general direction indicates that they are similarly correlated with the species' genetic structure. Wind dispersal, wind pollination, and a dominant position in the vegetation all corresponded to high estimates of the average dispersal distances, a more platykurtic distribution of the dispersal distances, low D, and low F_{ST}. However, the effects of these three species traits were different enough to have slightly different correlations with different genetic summary statistics. The mode of pollination and the ecological dominance were most strongly correlated with the average dispersal distance, while the mode of seed dispersal was most strongly and negatively correlated with F_{ST} and the kurtosis. Surprisingly, however, this latter trait was only weakly correlated with Rousset's (1997) Nb. The second group of arrows was dominated by the soil type, with only a very minor effect of the successional stage. The soil preference was mainly correlated to the spatial genetic structure as described by Mantel's r, with silicicolous species having higher values for Mantel's r than

Fig. 1 Triplot representation of the results of the RDA performed on the matrix of summary statistics constrained by the matrix of species traits, using laboratory as a covariate. Bold grey text indicates the placement of the genetic summary statistics on the two axes; small grey text indicates the placement of the species. Arrows indicate the effect of the species traits on the axes.

Fig. 2 Biplot of the results of the PCA performed on the matrix with genetic summary statistics. Grey text indicates the placement of the species on the two axes. Arrows indicate the effect of the summary statistics on the axes.

calcicolous species. This group of arrows was fairly independent of all other summary statistics with the exception of *Nb*, to which it was weakly and negatively correlated.

When looking at the placement of the species on the RDA triplot (Fig. 1), roughly three groups of species can be recognized that are placed away from the centre. The most obvious cluster of species consists of *Gentiana nivalis*, *A. alpina* and *H. alpina*. These species all have gravity-dispersed seeds, insect-pollinated flowers, a nondominant position in the vegetation and a relatively large altitudinal range. They are also characterized by very high F_{ST} values, low dispersal and low genetic diversity. A second cluster consists of *Geum reptans*, *H. uniflora*, *Phyteuma betonicifolium*, *P. hemisphaericum* and *Trifolium alpinum*. These all have insect-pollinated flowers, a preference for calcareous soils, and with the exception of *T. alpinum*, a nondominant position in the vegetation. The most important genetic characteristic that they share is a very high value of Mantel's *r*. They all have intermediate F_{ST} values and below-average genetic diversity. A third group consists of the species *Carex sempervirens*, *Gypsophila repens*, *L. mutellinoides* and *Dryas octopetala*. This group presents a mixture of life-history traits, containing both insect and wind-pollinated species, species with gravity, animal, and wind dispersal of seeds, and species with a small and with a

large altitudinal range. With regards to the genetic structure, the species in this cluster were very similar with high dispersal, high diversity and low F_{ST}.

The main trends in the RDA were corroborated in a series of univariate ANOVAs (see Table S1, Supporting information): there were significant associations for the same four life-history traits that were most important in the RDA. In the ANOVAs, the strongest associations found were between the soil preference and Mantel's *r*, the ecological dominance and the average dispersal distance, the mode of pollination and the average dispersal distance, and between the mode of seed dispersal and the kurtosis of the dispersal distances. Remarkably, there was no single significant association with *D* for any of the six life-history traits.

It is of interest to compare the results of the RDA with those of standard PCAs performed on the two matrices separately. For the PCA performed on the genetic matrix, the first two axes together explained 74.2% of the variation (45.6% and 28.6%, respectively). The relationships between the genetic summary statistics in the PCA (Fig. 2) showed a pattern that is strikingly different from the one that was present in the RDA (Fig. 1). In the RDA, H_S and H_T had very central placements, while in the PCA they both have a very strong effect. The reason for this difference is probably that both statistics were much higher for the species genotyped in lab E than for the other species (Thiel-Egenter *et al.* 2009), and this effect has been partialled out in the RDA. Another striking difference is that Mantel's
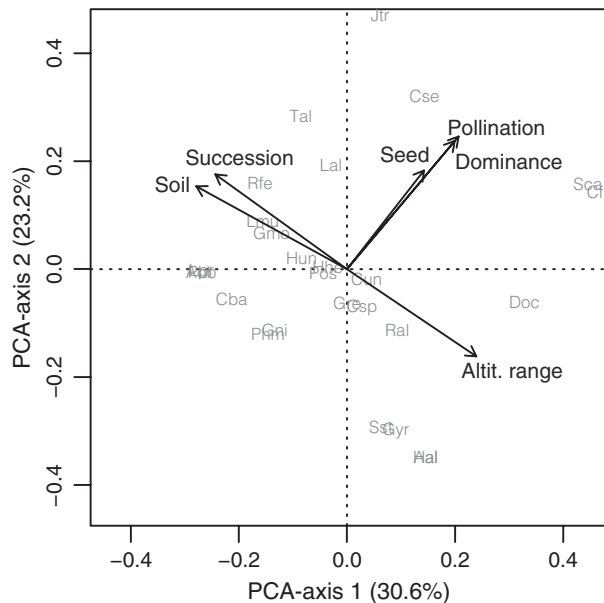
Fig. 3 Biplot of the results of the PCA performed on the matrix with species traits. Grey text indicates the placement of the species on the two axes. Arrows indicate the effect of the species traits on the axes.

r was much more prominent in the RDA than in the PCA, where it mainly affected the third axis.

There was less correlation among the species' life-history traits, so that only 59.4% of the variation was explained by the first two PCA axes (33.4% and 26.0%, respectively, with 16% explained by the third axis). Here, the main relationships between the species on the first two axes (Fig. 3) were largely the same as on the first two RDA axes (Fig. 1), showing two major groups of arrows. However, there were also some differences. For example, the placement of the altitudinal range was very different on the two figures, and this trait was much more important in the PCA than in the RDA. Also in the PCA, the influence of the successional stage was much stronger than it was in the RDA.

Discussion

Life history, history and genetic structure

We found a strong correlation between the plant species' ecological and life-history characters and their population structure. This is shown by the strong overall significance of the RDA, as well as the significance of the first axis. The six species traits together explained 30% of the variation in the matrix of genetic summary statistics. This percentage is lower than the result of the literature study of Hamrick & Godt (1989) who found that 47% of the among-species variation in G_{st} could be

explained by a data set of eight life-history traits, a difference that is due to our use of a covariate to account for the lab-effect. In this light, the percentage of variance explained in our study is remarkably high as there are other life-history and ecological traits that can affect the genetic structure of a species, apart from the six traits we included in our analysis. Furthermore, the genetic structure of a species is not only influenced by its life history and ecology but also by many stochastic factors such as fluctuations in population size, population extinction and other evolutionary processes.

In the RDA, the six life-history traits fell into two nearly orthogonal groups, indicating that they represent two independent aspects of the species' genetic population structure. The first group consisted mainly of dispersal related traits — both pollen and seed dispersal — and was correlated with several of the genetic summary statistics. Species with a strong population structure, that is with high F_{ST} and small inferred dispersal distances all had gravity-dispersed seeds, insect-pollinated flowers and a nondominant position in the vegetation. On the other hand, there were no traits shared among the species with weak population structure. This indicates that limited dispersal through one vector (e.g. seeds) can be compensated by high dispersal through another vector (pollen), and in that way still lead to high rates of gene flow. The higher F_{ST} values for species with insect-pollination and for species with gravity-dispersed seeds coincide with the findings of several previous studies (Hamrick & Godt 1989; Nybom & Bartish 2000; Nybom 2004). However, we found that the dispersal traits were more strongly related to the average dispersal distance estimated from the assignment test. In our data set, the relationship between F_{ST} and the mode of dispersal is nicely illustrated by the observation that four species that have obvious adaptations to long-distance seed dispersal by wind — *Cirsium spinosissimum*, *Dryas octopetala*, *Geum montanum* and *Hypochaeris radicata* — all had below-average F_{ST} values. *Geum reptans* also has hairy appendices on its seeds that aid in wind dispersal, but its life cycle is mostly dominated by clonal reproduction through stolons (Pluess & Stöcklin 2005), resulting in an above-average F_{ST}. The seeds of *Arabis alpina* — the species with the highest F_{ST} value — do not have any obvious adaptations to dispersal, but neither do the seeds of *Carex firma* — a species with one of the lowest F_{ST} values. So we see that there is a significant, but certainly not a strict, association between adaptation to dispersal and genetic structure. The contrast between the dispersal capacities of *D. octopetala* and *A. alpina* has been found before in a study of the postglacial recolonization of the arctic archipelago of Svalbard. Out of nine species, *D. octopetala* was the one with the highest estimate for the number of propagules that colonized Svalbard, whereas *A. alpina* was the one

with the lowest estimate (Alsos *et al.* 2007). Besides its lack of any adaptations to dispersal, the high F_{ST} value for *A. alpina* may also be partly explained by inbreeding. In a previous study, alpine populations were found to be significantly inbred, while Italian populations were in Hardy–Weinberg equilibrium (Ansell *et al.* 2008).

The second main trend that we observed in the RDA was a relationship between three ecological characters, most notably the soil preference, and the value of Mantel's *r*. Silicicolous species showed generally higher values of Mantel's *r* than calcicolous species. The nine species with the highest values for Mantel's *r* all had a preference for alkaline soils, most importantly *G. reptans*, *Hypochaeris uniflora*, *Phyteuma betonicifolium* and *Phyteuma hemisphaericum*. In population genetic studies, Mantel's *r* is generally interpreted to represent the strength of the pattern of isolation by distance, resulting from spatially restricted dispersal. In this light, it is surprising to see that in the RDA, Mantel's *r* and the soil preference are placed perpendicularly to the species' dispersal traits and the associated genetic summary statistics. It is difficult to explain why isolation by distance should be affected by the species' soil preference independently of differences in dispersal abilities. One explanation could be that in the Alps the geographical distance is not a good indicator of the ecological distance, which could be different for the two soil types. Another explanation may be that the two soil types differ in how they are distributed, with more patchy distributions leading to less connectivity among populations.

However, a historical explanation for the observed association between Mantel's *r* and the soil preference may be more likely. Using the same data set of 27 species, Alvarez *et al.* (2009) showed that the soil preference of the species was an important driver for the higher-level population structure. When regarding the geographic distribution of genetic clusters, they found that silicicolous species were predominantly arranged along the East-West axis ('horizontal' bands), while calcicolous species were arranged along the North-South axis ('vertical' bands) in the Alps. This difference in banding was explained using geographical records of glacier and snow cover during the last glaciation that showed that in the Northern Alps, there were only ice-free refugia available for calcicolous species, while in the Southern Alps there were refugia available for both calcicolous and silicicolous species (Schönswetter *et al.* 2005). Thus, in the case of calcicolous species it is possible to distinguish several genetic clusters corresponding to the different refugia that served as sources for todays populations. Although it may be possible to detect isolation by distance at a small spatial scale (within each cluster), at a large scale the pattern is broken by the existence of distinct genetic clusters, which affects

Mantel's *r* (Legendre & Legendre 1998). We think that it is the difference in the geographical distribution of the population clusters that causes the difference in Mantel's *r* between the calcicolous and silicicolous species. This would make the association between Mantel's *r* and the soil preference not strictly ecological but rather due to a mixture of ecological and historical factors.

Methodological considerations

As expected, we found strong relationships among the genetic summary statistics themselves; more than 70% of the variance in the matrix of summary statistics was explained by the first two PCA-axes. Thus, a few summary statistics suffice to describe the main patterns of genetic structure for this group of species. The two most important summary statistics in both the RDA and the ANOVAS were the average dispersal distance and the kurtosis estimated from the assignment test. This result is somewhat surprising as we did not use any exclusion method (Cornuet *et al.* 1999; Paetkau *et al.* 2004) to filter putative migrants based on their likelihood score, and therefore expect a high false positive rate. However, simulations have shown that despite this high false positive rate, the inferred average dispersal distance from an unfiltered assignment test can give a good indication of the actual dispersal distances (P.G. Meirmans, unpublished). Furthermore, we found that the estimated average dispersal distance was strongly correlated with the mode of pollen dispersal, while the kurtosis of the distribution of dispersal distances was strongly correlated with the mode of seed dispersal. This indicates that the results of assignment tests can capture different aspects of dispersal. However, we regard the estimates of the average dispersal distance simply as a useful tool for comparing overall dispersal strategies among species, and we caution against taking the distance estimates, the shape of the distribution, and the individual assignments too literally.

The observed association between Mantel's *r* and the soil preference means that one should be careful in drawing conclusions about isolation by distance from Mantel's *r*. The pattern of isolation by distance may be obfuscated by higher-level population structure that is a result of historical events and the species ecology. It is therefore advised to always perform a test for isolation by distance in combination with an analysis of higher-level population structure. For example a partial Mantel test could be used to correct for the effects of population structure on isolation by distance (Landguth *et al.* 2010). Alternatively, a standard Mantel test may be performed in each of the clusters separately. Unfortunately, we were unable to obtain satisfactory results from pop-

ulation clustering methods (Pritchard *et al.* 2000; Jombart *et al.* 2010) for all species, and therefore could not include such an analysis in our study.

Recently, F_{ST} has been criticized for being dependent on the amount of within-population diversity H_S (Hedrick 2005; Jost 2008). Jost (2008) therefore developed the D statistic that does not have this dependency and should therefore be more suitable than F_{ST} for comparisons among species. However, in our data, this dependency is not very pronounced and F_{ST} and H_S were placed almost orthogonally in the PCA of the genetic summary statistics. The placement of D relative to the other summary statistics changed rather drastically among our different analyses, so that its relationship to the species traits remains unclear. Indeed, of the eight summary statistics, D was the only one that did not show any significant association with any of the traits in our series of ANOVAS. F_{ST} on the other hand was significantly correlated with the mode of seed dispersal. The reason for the bad results for D is probably that, theoretically, the value of D is expected to be independent of the species' effective population size (Jost 2008). Our results confirm that despite its shortcoming, F_{ST} is still a better statistic for making demographic inferences than D (Meirmans & Hedrick 2011; Whitlock 2011). Another statistic that was developed to circumvent the dependency of F_{ST} on H_S, Hedrick's (2005) G'_{ST}, occupied the same position as F_{ST} when included in the RDA, and was therefore not included to prevent problems of multicollinearity.

The used data set is remarkable in its scope and in the thoroughness of the sampling and genotyping. Earlier reviews on plant population structure have noted large differences in geographical sampling range, genotyping strategies and statistical methods, which may influence the results (Hamrick & Godt 1996; Thiel-Egenter *et al.* 2009). Because of its uniformity in sampling, genotyping and analysis, the present data set is particularly suited to this type of analysis. Nevertheless, there are some drawbacks to the data set. First, despite an effort to standardize the genotyping process, one of the involved laboratories used different equipment for the AFLPs. Unfortunately, this same laboratory was appointed only wind-dispersed species. As a result, it is impossible to determine whether the higher diversity in these species is a result of their mode of dispersal or of the lab. We therefore took a conservative approach and removed the effect of this laboratory.

Another drawback of the data set is that it contains only high-alpine plant species, creating a lack of variation in certain life-history traits. For example, in the study of Hamrick & Godt (1989) the most important life-history traits were breeding system and life-form. We did not include these two traits as there was too little variation among the 27 species that we used; there simply are few strict inbreeders and no trees in high-alpine habitats. In addition, all included wind-pollinated species were graminoids, most of which occupy a dominant position in the vegetation (Thiel-Egenter *et al.* 2009). In fact, this explains the relationship found between dominance and the two modes of dispersal.

Conclusions

In summary, we find that the genetic population structure of the 27 studied species is indeed the result of the combined effects of history and the species' ecology and life history. As expected, the most important aspect of the species' genetic structure was determined by their dispersal strategies. However, very strong population structure was only found when both seed and pollen dispersal was restricted, while weaker population structure was found for several different life-history strategies. A second important aspect was related to the species soil preference, whose effect on the genetic structure was due to differences in the geographic distribution of these two soil types across the Alps, which determined the distribution of refugia during the last glaciation.

On a more technical note, we also found that the new differentiation statistic D provided very little information when analysed together with the species traits, confirming its limited usefulness for demographic studies. Unexpectedly, we found that Mantel's r was not related to the species' dispersal modes, but instead was strongly affected by the spatial distribution of the higher-level population structure that resulted from postglacial recolonization. This is a reminder that even nonparametric multivariate approaches involve underlying assumptions (in this case the absence of distinct clusters) and require careful interpretation.

Acknowledgements

References

Aeschimann D, Lauber K, Moser DM (2004). *Flora Alpina*. Hauptverlag, Bern.

Alsos IG, Eidesen P, Ehrich D *et al.* (2007) Frequent long-distance plant colonization in the changing Arctic. *Science*, **316**, 1606–1609.

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of patial genetic structure in alpine plants. *Ecology Letters*, **12**, 632–640.

Ansell S, Grundmann M, Russell S, Schneider H, Vogel J (2008) Genetic discontinuity, breeding-system change and population history of *Arabis alpina* in the Italian Peninsula and adjacent Alps. *Molecular Ecology*, **17**, 2245–2257.

Cornuet J, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, **153**, 1989–2000.

Cressie N (1993). *Statistics for Spatial Data*. Wiley, New York.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes — application to human mitochondrial-DNA restriction data. *Genetics*, **131**, 479–491.

Gugerli F, Englisch T, Niklfeld H *et al.* (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation — a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 259–281.

Hamrick J, Godt M (1989). Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding and Genetic Resources* (eds Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43–63. Sinauer, Sunderland, MA.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **351**, 1291–1298.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

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

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Landguth E, Cushman S, Schwartz M, Mckelvey K, Murphy M, Luikart G (2010) Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, **19**, 4179–4191.

Landolt E (1977). *Ökologische Zeigerwerte zur Schweizer Flora*. Veröffentlichungen des Geobotanischen Institutes der ETH, Stiftung Rübel, Zürich.

Legendre P, Legendre L (1998). *Numerical Ecology*. Elsevier, Amsterdam.

Lynch M, Milligan B (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Meirmans PG, Hedrick P (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, **11**, 5–18.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, **13**, 1143–1155.

Nybom H, Bartish I (2000) Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspectives in Plant Ecology, Evolution and Systematics*, **3**, 93–114.

Oksanen J, Kindt R, Legendre P, O'Hara B, Simpson G, Solymos P *et al.* (2009). *Vegan: Community Ecology Package* (R Package Version 1.15-3). http://CRAN.R-project.org/package=vegan.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, **4**, 347–357.

Paetkau D, Shields G, Strobeck C (1998) Gene flow between insular, coastal and interior populations of brown bears in Alaska. *Molecular Ecology*, **7**, 1283–1292.

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Pluess AR, Stöcklin J (2005) The importance of population origin and environment on clonal and sexual reproduction in the alpine plant *Geum reptans*. *Functional Ecology*, **19**, 228–237.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences, USA*, **94**, 9197–9201.

Rao C (1964) The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, **26**, 329–358.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Ryman N, Leimar O (2009) Gst is still a useful measure of genetic differentiation — a comment on Jost's D. *Molecular Ecology*, **18**, 2084–2087.

Schlater M (2001) Simulation of stationary and isotropic random fields. *R-News*, **1**, 18–20.

Schönswetter P, Stehlik I, Holderegger R, Tribsch A (2005) Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology*, **14**, 3547–3555.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Ter Braak C (1990). *Update Notes: CANOCO Version 3.10*. Agricultural Mathematics Groups, Wageningen.

Thiel-Egenter C, Gugerli F, Alvarez N *et al.* (2009) Effects of species traits on the genetic diversity of high-mountain plants: a multi-species study across the Alps and the Carpathians. *Global Ecology and Biogeography*, **18**, 78–87.

Vekemans X (2002). *AFLP-SURV Version 1.0*. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA-fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Whitlock MC (2011). G'_{ST} and D do not replace F_{ST}. *Molecular Ecology*, **20**, 1083–1091.

P.M. is a postdoctoral researcher interested in various aspects of plant ecology and population genetics. J.G. is a population geneticist interested in the consequences of population structure and dynamics for neutral and adaptive traits. O.G.'s research focuses on developing theory and statistical methods aimed at bridging the gap between population ecology, population genetics and evolution. Much of his research is applied to the study of metapopulations.

Data accessibility

Sampling locations and AFLP data for all 27 species: DRYAD entry doi:10.5061/dryad.f3rk4.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Results from a series of univariate ANOVAs using the genetic summary statistics as a dependent and the life-history traits (excluding the continuously distributed altitudinal range) as independent variables. The cell values represent the ANOVA F-statistics, with the degrees of freedom given in the column headers. The significance is indicated with asterixes (*$P < 0.05$, **$P < 0.01$). No correction for multiple testing was applied as the results are only used as an illustration next to the more powerful redundancy analysis.

Fig. S1 Redundancy analysis (RDA) triplot for 21 species, excluding the species genotyped in lab E. This RDA explained 50.7% of the variation in the matrix of genetic summary statistics ($P = 0.0069$). Bold grey text indicates the placement of the genetic summary statistics on the two axes; small grey text indicates the placement of the species. Arrows indicate the effect of the species traits on the axes.

Fig. S2 Redundancy analysis (RDA) triplot, using the average altitude preference for the 27 species, rather than the altitudinal range. This RDA explained 28.3% of the variation in the matrix of genetic summary statistics ($P = 0.0187$). Note that compared with the RDA using the altitudinal range in the main text (Fig. 1), the percentage of variation and significance of the RDA have decreased. However, the main trend visible in the data remains practically unchanged. Bold grey text indicates the placement of the genetic summary statistics on the two axes; small grey text indicates the placement of the species. Arrows indicate the effect of the species traits on the axes.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.