Molecular Ecology (2009) **18**, 2602–2614 doi: 10.1111/j.1365-294X.2009.04229.x

Quantifying the genetic component of phenotypic variation in unpedigreed wild plants: tailoring genomic scan for within-population use

C. M. HERRERA and P. BAZAGA
Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, Américo Vespuccio s/n, E-41092 Sevilla, Spain

Abstract

The study of adaptive genetic variation in natural populations is central to evolutionary biology. Quantitative genetics methods, however, are hardly applicable to long-lived organisms, and current knowledge on adaptive genetic variation in wild plants mostly refers to annuals and short-lived perennials. Studies on long-lived species are essential to explore possible life-history correlates of genetic variation, selection, and trait heritability. In this paper, we propose a method based on molecular markers to quantify the genetic basis of individual phenotypic differences in wild plants under natural conditions. Rather than focusing on inferring individual relatedness to estimate the heritability of phenotypic traits, we directly estimate the proportion of observed phenotypic variance that is statistically accounted for by genotypic differences between individuals. This is achieved by (i) identifying loci that are correlated across individuals with the phenotypic trait of interest by means of an amplified fragment length polymorphism (AFLP)-based explorative genomic scan, and (ii) fitting multiple regression and linear random effect models to estimate the effects of genotype, environment and genotype × environment on phenotypes. We apply this method to estimate genotypic and environmental effects on cumulative maternal fecundity in a wild population of the long-lived *Viola cazorlensis* monitored for 20 years. Results show that between 56–63% (depending on estimation method) of phenotypic variance in fecundity is accounted for by genotypic differences in 11 AFLP loci that are significantly related to fecundity. Genotype × environment effects accounted for 38% of fecundity variance, which may help to explain the unexpectedly high levels of genetic variance for fecundity found.

Keywords: ecological genomics, environmental variance, fecundity hierarchy, genetic variance, genomic scan, genotype-by-environment interaction, *Viola cazorlensis*

Received 2 February 2009; revision received 30 March 2009; accepted 9 April 2009

Introduction

Evolutionary processes are largely driven by the tension between the factors that originate or maintain heritable variation of fitness-related traits and the opposing forces that erode such variation, particularly natural selection. Evaluating the magnitude and distribution of adaptive (i.e. non-neutral, related to fitness) genetic variation within and between populations, its relationship to phenotypic differences, and the factors contributing to maintain it, are thus central themes in evolutionary biology. The study of genetic variation and its relationship with phenotypic differences is relatively straightforward in artificially assembled populations made up of captive individuals of known pedigree and kept under controlled environmental conditions. Practical and conceptual complications, however, become legion when the same issues are to be addressed in natural populations (Roff 1997; Lynch & Walsh 1998; Merilä & Sheldon 1999; Leinonen *et al*. 2008; Pujol *et al*. 2008), mainly because of the difficulties inherent to genotyping unpedigreed wild individuals. Such difficulties have been traditionally overcome by the use of quantitative genetics methods, which typically rely on artificially breeding individuals of known pedigree through controlled crosses between field-collected progenitors of known phenotypes, and exposing experimental populations to controlled environments (Roff 1997; Lynch & Walsh 1998).

Correspondence: Carlos M. Herrera, Fax: +34 954 62 11 25; E-mail: herrera@cica.es

There appears to be no substitute for pedigree-based, classical quantitative genetics methods in the foreseeable future as the best way of obtaining precise estimates of the magnitude and sources of genetic variation that underly phenotypic traits of interest in wild nonmodel organisms. Unfortunately, however, not all organisms are equally amenable to these methods. Among plants, long-lived species that cannot be cloned or propagated vegetatively, such as many trees, shrubs and herbaceous perennials, stand out as a prominent group for which quantitative genetics methods based on experimental crossing schemes are impractical. Consequently, efforts to elucidate natural patterns of adaptive genetic variation in wild plants have inevitably concentrated on annuals and short-lived perennials, and relatively little is known for long-lived species (e.g. Geber & Griffen 2003). An important implication of this life-history bias is that current knowledge on patterns of adaptive genetic variation in wild plants might itself be biased if the features that render species amenable to quantitative genetics methods are themselves correlated with the patterns and magnitude of adaptive genetic variation. This is not an unplausible possibility given, for example, the mounting evidence that evolutionary rates in plants are linked to life history, and particularly to lifespan (Andreasen & Baldwin 2001; Smith & Donoghue 2008). Studies on a broader range of long-lived species are therefore essential to explore possible life-history effects on patterns of adaptive genetic variation, selection, and heritability (Geber & Griffen 2003).

In parallel to the increased availability of molecular markers, increasing attention has been paid in recent years to the development of alternative methods that can be applied to the analysis of adaptive genetic variation in unpedigreed wild populations (Garant & Kruuk 2005). These methods have mostly focused on estimating the heritability of phenotypic traits by assessing the strength of the statistical relationship between phenotypic similarity and estimates of genetic relatedness between individuals inferred using genetic markers (Ritland 1996, 2000, 2005; Fernández & Toro 2006), although few studies have so far applied them to real situations (Ritland & Ritland 1996; Andrew *et al*. 2005; van Kleunen & Ritland 2005). In this paper, we propose a different approach to quantify the genetic basis of individual phenotypic variation in wild plants under natural conditions. Rather than targetting at estimating the heritability of the phenotypic trait of interest by inferring the relatedness of individuals and then relating it to their phenotypic similarity (Garant & Kruuk 2005), we will directly proceed to estimate the proportion of observed phenotypic variance that can be statistically accounted for by adaptive genetic differences between individuals, elucidated through a variant of explorative genomic scan tailored for within-population use. The method will be applied to estimate the magnitude of genetic and environmental effects on the long-term, cumulative maternal fecundity of individuals of the long-lived violet *Viola cazorlensis* growing under natural conditions and monitored for 20 years.

Genomic scans are useful to identify functional genetic polymorphisms that are related to phenotypic traits of interest (Luikart *et al*. 2003; Beaumont 2005; Storz 2005; Vasemägi & Primmer 2005; Bonin *et al*. 2006). Although these methods provide a strategy in which the genetic basis of adaptation can be considered directly at the genome level without prior knowledge about the phenotypes, the incorporation of phenotypic information makes them particularly powerful (Herrera & Bazaga 2008). In this study, we perform a genome screening to identify genomic regions (amplified fragment length polymorphism, AFLP fragments or 'loci') whose variation across individuals of the same population is significantly related to their phenotypic differences in long-term cumulative fecundity. These loci can be assumed to have a direct functional effect on fecundity, or to be closely linked to loci having such effects, and for convenience will be termed here 'adaptive loci' because of their association with variation in fecundity, a phenoypic trait closely related to fitness. Multiple linear regression and mixed linear models will then be applied to (i) test for the additive nature of the effects of adaptive AFLP loci on fecundity, and (ii) assess the proportion of total phenotypic variance in fecundity that is (statistically) accounted for by individual genotypic differences as measured using adaptive AFLP loci. *Viola cazorlensis* plants may occur in several substrate types that differ in important ecological features (Herrera 1989, 1993; see Study area and field methods below). Taking advantage of this fact, we will use substrate type as a surrogate for the environment faced by individual plants, which will allow us to evaluate the effects of genotype, environment, and genotype × environment on fecundity under natural field conditions.

Materials and methods

Study species

Viola cazorlensis (Violaceae) is a perennial, suffruticose violet endemic to a few contiguous limestone mountain ranges in southeastern Spain, where it occurs as small, discrete populations associated with rocky outcrops, cliffs, and 'islands' of sandy soils originating from heavily weathered dolomitic limestone between 900–2100-m elevation. In the Sierra de Cazorla, where the present study was conducted (see below), the species occurs as widely disjunct, discrete local populations generally separated by a few kilometers of unsuitable habitat (various types of coniferous and mixed forest). Flowers have pinkish-purple corollas, a long and thin spur, and are pollinated by the day-flying hawkmoth *Macroglossum stellatarum* (Lepidoptera, Sphingidae). Adult

plants are generally small (mean of longest diameter = 13.7 cm), and produce very few flowers and fruits annually (mean = 3.6 flowers and 1.7 fruits per plant·year). All reproduction takes place by seeds, which lack special dispersal mechanisms. Direct data on the longevity of *V. cazorlensis* plants are not available. In the set of 75 adult plants marked at the beginning of this study (see Study area and field methods below), mean survival time (± SE) over 1988–2008 was 19.7 ± 0.5 year (Kaplan–Meier survival estimation, analyses not shown). Since all plants were already adults when marked, their lifespans will exceed that average. It seems safe to assume, however, that our 20-year study encompassed a substantial portion of the lifetimes of marked plants, and that cumulative fecundity estimates for individual plants over the study period are correlated with lifetime fecundities. Additional details on the ecology and genetics of *V. cazorlensis* can be found in Herrera (1989, 1990a, b, 1993) and Herrera & Bazaga (2008).

Study area and field methods

This study was conducted during 1988–2008 in a large population of *V. cazorlensis* located at 1290-m elevation in La Cruz de Quique, Sierra de Cazorla (Jaén province, southeastern Spain). The vegetation of the area is open pine (*Pinus nigra* and *Pinus pinaster*) forest with an understorey of *Juniperus phoenicea* treelets and a sparse xerophytic shrub layer dominated by *Echinospartum boissieri* and *Rosmarinus officinalis*. Changes in the vegetation of the study area during the course of this study were nearly imperceptible, and mainly involved a weak reduction in percentage area of bare sandy soil and a slight increase in percentage cover of the shrub layer. The substrate of the site is a complex mosaic made up of large bare rocks at ground level, small cliffs (2–4 m in height), and pockets of loose, coarse-grained, limestone-derived sandy soils. *Viola cazorlensis* plants occur on all these substrate types. Three substrate categories were recognized for this study, namely pockets of sandy soil ('ground' hereafter), crevices of bare rocks at ground level ('rocks'), and crevices in vertical or overhanging cliffs > 2 m in height ('cliffs'). Substrates differ in ecological features including timing and duration of insolation, overtopping by competing or protective shrubs, severity of water stress, and incidence of vertebrate and invertebrate herbivores (Herrera 1989, 1993). This provides biological justification for using substrate type as a proxy to investigate environmental effects on fecundity.

In April 1988, a random sample of 75 reproductive individuals of *V. cazorlensis* was chosen and marked with permanent tags before the start of flowering in the population. Marked plants were distributed over an area of approximately 1300 m^2, and grew on all substrate types available at the site. These individuals were subsequently surveyed during the flowering and fruiting season (April–June) of every year until 2008, and the number of flowers and fruits produced by each plant were recorded. A total of 23 marked plants (30.7%) died from natural, albeit unknown causes over the 20-year study period. Information on the long-term fecundity and genetic characteristics of the 52 individuals that remained alive in the spring of 2008 provides the basis for the present study. Fecundity data consist of a 20-year-long series of yearly fruit production for each of the 52 study plants. During 1988–1991, the mean number of seeds per capsule was estimated every reproductive season from subsamples of capsules collected shortly before dehiscence and examined under a dissecting microscope. A close correlation existed across marked plants between the number of fruits and seeds produced ($r = 0.953$, $N = 75$ plants, $P < 0.0001$). Only the number of fruits produced per plant in each season was subsequently recorded during 1992–2008, and the total number of fruits produced over 1988–2008 by each plant was used as an estimate of its long-term, cumulative individual fecundity. All plants were measured (length of the longest canopy axis) at the beginning of the study, and again in 1991, 1999 and 2003. Average figures will be used to characterize the size of individual plants.

AFLP analyses

In April 2008, fresh leaf material was collected from each of the 52 study plants and dried immediately at ambient temperature.Total genomic DNA was extracted from approximately 35 mg of ground leaf material using DNeasy Plant Mini Kit (QIAGEN) and the manufacturer protocol. The AFLP analysis was performed essentially as originally described by Vos *et al*. (1995), with modifications involving the use of fluorescent dye-labelled selective primers following Applied Biosystems (2005). We used eight *Eco*RI + 3/*Mse*I + 3 and eight *Pst*I + 2/*Mse*I + 3 primer combinations, chosen from a broader sample of combinations previously assayed in a pilot study. For each combination of restriction enzymes, the primer pairs providing the most reliable, consistently scorable results were chosen (Table 1). Mounting evidence indicates that AFLP markers obtained using *Eco*RI and *Pst*I as rare cutters may differ widely in distribution and degree of clustering across plant genomes (Castiglioni *et al*. 1999; Vuylsteke *et al*. 1999; Young *et al*. 1999; Yuan *et al*. 2004; Ojha 2005). We used these two restriction enzymes to increase the chances of hitting genome segments functionally related to fecundity.

Each of the 52 study plants was fingerprinted using the 16 combinations chosen. Fragment separation and detection was made using an ABI PRISM 3100 DNA sequencer. The presence or absence of each marker in each individual plant was scored manually by visualizing electrophoregrams with GeneMapper 3.7 software. All scoring was blindly carried out by the same person (P.B.), who during the

Table 1 Primer combinations used, number of markers (loci) obtained in the size range 150–500 bp, observed polymorphism level, and estimates of scoring error rates, in the AFLP analysis of the 52 individual plants of *Viola cazorlensis* considered in this study

Primer combination	No. of markers	Percentage of polymorphic*	Scoring error rate (Percentage)†
1. *Pst*-AG/*Mse* CTT	52	59.6	3.7
2. *Pst*-AG/*Mse* CAT	34	50.0	0.9
3. *Pst*-AC/*Mse* CTT	34	64.7	1.9
4. *Pst*-AC/*Mse* CAT	47	80.8	1.6
5. *Pst*-AA/*Mse* CGT	22	86.4	0.7
6. *Pst*-AA/*Mse* CCC	28	82.1	0.5
7. *Pst*-AT/*Mse* CAC	34	82.4	2.7
8. *Pst*-AT/*Mse* CCT	48	68.7	0.7
9. *Eco*-ACA/*Mse*-CTT	41	70.7	7.4
10. *Eco*-ACT/*Mse*-CTA	33	72.7	3.3
11. *Eco*-AGG/*Mse*-CAA	33	72.7	2.4
12. *Eco*-AGA/*Mse*-CTT	51	84.3	0.3
13. *Eco*-AGC/*Mse*-CTC	37	78.4	2.2
14. *Eco*-AGC/*Mse*-CTT	38	78.9	5.0
15. *Eco*-AGG/*Mse*-CAT	31	77.4	4.9
16. *Eco*-AGG/*Mse*-CTG	36	66.7	2.7

*A locus was considered polymorphic if at least one individual showed a variant score.
†Calculated from the 28 individuals that were re-assayed as 100× (number of discordant scores on two independent analyses)/(number of scored markers × number of individuals).

process lacked any information on location, size, or any other characteristic of individual plants. Only fragments ≥ 150 bp in size were considered, as a way of reducing the potential impact of size homoplasy (Vekemans *et al*. 2002). Genotyping error rates were determined for each primer combination by running repeated, independent analyses for a total of 28 individual plants, and estimated as the ratio of the total number of loci with discordant scores (all individuals combined) to the product of the number of individuals by the total number of scored loci (Bonin *et al*. 2004). Results are shown in Table 1. Error rates varied among primer combinations, ranging between 0.5–3.7% for individual *Pst*I combinations (mean ± SE = 1.59 ± 0.40%), and between 0.3–7.4% for *Eco*RI combinations (3.52 ± 0.77%). The mean value (± SE) for the whole set of 16 combinations was 2.55 ± 0.49%. These figures are similar to recent estimates of genotyping error rates in AFLP studies of plants (Bonin *et al*. 2004, 2007; Jump *et al*. 2008; Parisod & Christin 2008).

Data analyses

The simplest way for evaluating possible links between genotype and quantitative traits is to search for significant associations across individuals between molecular markers and phenotypic values, an approach generally known as linkage disequilibrium mapping (Mackay 2001; Olsen & Purugganan 2004). Although this method has been generally used to infer associations betwen quantitative trait loci (QTL) and phenotypic traits, the same underlying principles can motivate genome screenings aimed at identifying putative functional associations between individual variation in anonymous DNA markers and phenotypic traits. In natural populations, however, significant marker-trait associations are not necessarily attributable to linkage, but can also arise from admixture between subpopulations that have different gene frequencies at marker loci and different trait values (Mackay 2001). Before searching for relationships between fecundity and individual genotypes, a preliminary analysis was undertaken to examine the possibility that the *V. cazorlensis* plants studied were a genetically heterogeneous mixture. This was explored by applying to the data the Structure 2.2 program (Falush *et al*. 2007; Pritchard *et al*. 2007). We performed 10 replicates of each simulation from $K = 1$ to 5, with a burn-in of 50 000 and 100 000 post burn-in Markov chain Monte Carlo (MCMC) iterations, assuming admixture and correlated allele frequencies. The most likely number of populations represented in our sample was determined by examination of the estimates of model log likelihood [$\log P(X \mid K)$]. These analyses showed that marked plants formed a single panmictic unit without detectable genetic substructuring (results not shown).

We searched for AFLP loci associated with individual differences in cumulative fecundity over the 20-year study period by running separate logistic regressions for each locus, using band presence as the dependent binary variable, and the total number of fruits produced during 1988–2008 ('fecundity' hereafter) as the independent one, log transformed to achieve normality ($P = 0.38$, Shapiro–Wilk normality test for log-transformed fecundity data). Analyses were conducted with SAS procedure Logistic (SAS Institute 2004), and the P values obtained from likelihood ratio tests

were used to identify statistically significant locus-fecundity relationships after properly accounting for the large number of simultaneous significance tests performed (see below). Likelihood-ratio tests, that are more powerful and reliable for moderate-to-small sample sizes as those of the present study (Agresti 2007), were used to assess the statistical significance of logistic regressions. In addition to obtaining maximum likelihood regression parameter estimates and standard errors with procedure LOGISTIC, significant loci-fecundity relationships were corroborated with Bayesian statistics. Logistic regression parameter estimates and 95% credible intervals were obtained using the Gibbs sampling implemented in SAS procedure BGENMOD (SAS Institute 2006), with burn-in = 50 000, MCMC = 500 000 and thinning interval = 1000.

Given the large number of logistic regressions involved in the screening for AFLP score-fecundity relationships, a distinct possibility exists of obtaining by chance an unknown number of false significant regressions (i.e. committing type I errors). We applied Storey & Tibshirani's (2003) *q*-value method for estimation of false discovery rates to the set of *P* values for individual loci obtained from likelihood-ratio tests of logistic regressions. The *q*-value obtained for a given locus is the expected proportion of false-positives incurred for the whole set of regressions when the logistic regression for that particular locus is considered as statistically significant. Using the Qvalue package (Storey & Tibshirani 2003), we calculated the *q*-values for all the locus-fecundity regressions, ranked them, and found the largest *q*-value leading to an expectation of less than one falsely significant regression [i.e. *q*-value × (number of regressions accepted as significant) < 1]. In this way, we enhanced the statistical power for detecting as many significant regressions as possible while keeping with the deliberately conservative constraint of avoiding any false-positive.

Two complementary statistical approaches were adopted to evaluate the effects of AFLP loci significantly related to fecundity ('adaptive loci' hereafter). A multiple linear regression was fitted to individual plant data using SAS procedure REG, with fecundity as the dependent variable and all adaptive loci as independent ones, coded as binary scores. In multiple regression, a response variable (y) is predicted on the basis of an assumed linear relationship with several independent variables (x_i), and the extent of the relationship between y and the x_i variables is assessed with the multiple correlation coefficient R. Adjusted R-square from the regression was interpreted as a direct estimate of the proportion of phenotypic variance in fecundity jointly explained by the additive, linear effects of adaptive loci scores. In addition, a random effect model was fitted to individual plant data to simultaneously estimate the genotype (as described by the set of adaptive loci), environment (substrate type), and genotype × environment effects on fecundity. To make the data amenable to this analysis, the dimensionality of the multilocus genotype of individual plants based on adaptive loci was reduced to a continuous two-dimensional representation. A matrix of pairwise linear genetic distances between individuals based on adaptive loci scores was computed with Genalex 6.1 (Peakall & Smouse 2006) using a band-based, simple-matching coefficient (Bonin *et al*. 2007). A principal coordinate analysis (PCO) was then performed on this distance matrix, and the coordinates of individual plants on the first two axes used as continuous descriptors of genotypes with respect to the adaptive loci related to fecundity. Variance components were estimated by fitting a linear model to individual fecundity data using SAS procedure MIXED with REML estimation and substrate, genotype, and genotype × substrate as independent variables, all treated as random effects. The covariance structure used to fit the model (VC or 'variance components' type in SAS terminology) assumed a different variance component for each random effect in the model. Heteroscedasticity in fecundity across substrates can bias asymptotic significance tests of variance components. Standard errors and confidence limits for estimates of percentage variance components were obtained by bootstrapping, with 10^5 generations and the standard percentile method. Our analytical approach to test genotype × environment interactions was conceptually similar to the two-way ANOVA method commonly used in experimental quantitative genetics studies (Lynch & Walsh 1998; see also Littell *et al*. 2006, p. 88). Nevertheless, given the mixture of quantitative (genotypes) and qualitative (substrate) independent variables in the random effects model, our assessment of the interaction effect relied on a homogeneity-of-slopes test in an ANCOVA framework. Because we are interested in evaluating sources of naturally occurring variance in fecundity in the original measurement scale (i.e. number of fruits), and to avoid distortions of the genotype × environment effect arising from data transformation (Stanton & Thiede 2005), the random effects model was fitted to untransformed fecundity data.

Results

Levels of polymorphism

The 16 primer combinations assayed produced a total of 599 AFLP fragments ('loci' hereafter) in the range 150–500 bp that could be unambiguously scored for the 52 plants studied. Both the number of scorable loci (range = 22–52) and levels of polymorphism (range = 50.0–86.4%) varied among primer combinations (Table 1). *Eco*RI- and *Pst*I-based combinations did not differ significantly in either number of loci or percentage of polymorphism per combination ($P = 0.94$ and 0.86, respectively; Wilcoxon two-sample tests). Loci that were monomorphic for the entire data set, and those present in < 2% or > 98% of individuals, were

Table 2 Results of separate logistic regressions relating scores for AFLP loci (1–0, presence–absence) to the log-transformed cumulative number of fruits produced by the $N = 52$ study plants over 1988–2008. Only results for the 11 loci exhibiting statistically significant relationships according to maximum likelihood tests are shown. Regressions model the probability of marker band presence (i.e. score = 1) as a function of fecundity, hence positive regression parameters denote a positive effect of the dominant allele, and vice versa. See examples in Fig. 2. Primer combination codes as in Table 1

Locus	Primer combination; AFLP marker size, base pairs	Maximum likelihood estimation		Bayesian estimation	
		Regression parameter (SE)	*P* value	Regression parameter	95% credible interval
56	3; 316	–2.15 (0.68)	0.0001	–2.35	–3.79– –1.10
64	4; 161	–3.46 (1.49)	0.0019	–4.35	–8.49– –1.58
72	4; 271	–2.29 (0.83)	0.0008	–2.53	–4.47– –1.09
143	7; 159	4.96 (2.40)	0.0026	6.18	1.99–12.42
151	7; 277	1.89 (0.64)	0.0008	2.05	0.93–3.53
152	7; 279	1.89 (0.64)	0.0008	2.05	0.93–3.53
162	8; 203	–1.79 (0.63)	0.0011	–1.92	–3.47– –0.72
180	8; 399	–1.54 (0.57)	0.0026	–1.66	–2.92– –0.56
200	9; 315	1.63 (0.58)	0.0018	1.73	0.64–2.96
268	12; 368	5.13 (2.48)	0.0022	6.23	2.00–12.32
323	14; 391	5.16 (2.49)	0.0021	6.76	1.89–13.09

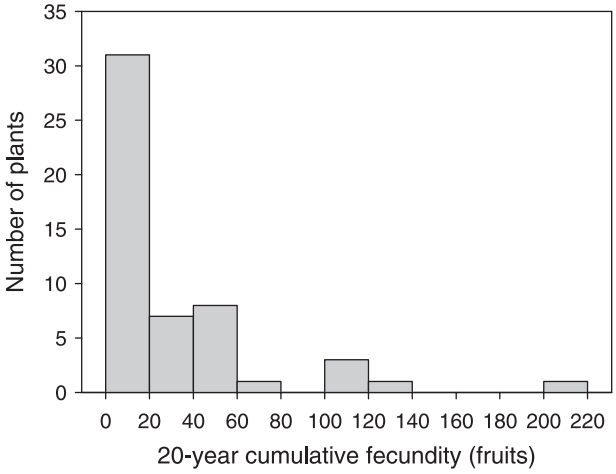

Fig. 1 Frequency distribution of individual cumulative fecundity (total number of fruits produced) of the $N = 52$ marked *Viola cazorlensis* plants over the 1988–2008 study period.

excluded from all subsequent analyses. Our final data set consisted of a total of 365 polymorphic loci scored for each of the 52 individual plants. All individuals exhibited unique AFLP profiles.

Individual fecundity

Study plants differed widely in cumulative fecundity over the period 1988–2008, ranging between 0–209 fruits (median = 13.5 fruits, mean ± SE = 29.0 ± 5.6 fruits, $N = 52$ plants). As documented in detail for this population by Herrera (1989, 1993) for the period 1987–1991, individual variation in fruit production over 1988–2008 depended closely on differences in flower production ($R^2 = 0.679$, $F_{1,50} = 105.7$, $P < 0.0001$), which were in turn directly related to mean plant size ($R^2 = 0.200$, $F_{1,50} = 12.5$, $P = 0.0009$). There was significant heterogeneity between substrate types in mean fecundity ($\chi^2 = 12.50$, d.f. = 2, $P = 0.0019$; Kruskal–Wallis analysis of variance), which increased from rock- (mean ± SE = 10.7 ± 2.7 fruits/plant) through ground- (23.7 ± 6.0 fruits/plant) to cliff-growing plants (69.9 ± 18.1 fruits/plant).

The frequency distribution of individual 20-year cumulative fecundity was strongly skewed to the right, with 88% of plants producing ≤ 60 fruits, and only 9% of individuals producing > 100 fruits (Fig. 1). These few, highly fecund individuals alone accounted collectively for nearly half (44.6%) of all the fruits produced by marked plants over the two decades encompassed by this study ($N = 1486$ fruits).

AFLP loci related to fecundity

A total of 11 logistic regressions relating presence–absence of individual AFLP markers in a plant genotype to its (log-transformed) 20-year fecundity were statistically significant according to our criterion of keeping the expected number of false-positives < 1 ($\chi^2 \geq 9.04$, $P \leq 0.0026$, q-value ≤ 0.063; expected number of false-positives = $0.063 \times 11 = 0.69$). Characteristics of the 11 AFLP fragments involved ('adaptive loci' hereafter), estimates and standard errors of regression parameters, and Bayesian credible intervals, are shown in Table 2. Adaptive loci represented 3.0% of the total of 365 loci assayed, and were about three times more frequent among *Pst*I (8 out of 183 total polymorphic loci, or 4.4%) than among *Eco*RI combinations (3 out of 182 polymorphic loci, or 1.6%). Regression parameter estimates obtained by maximum likelihood and

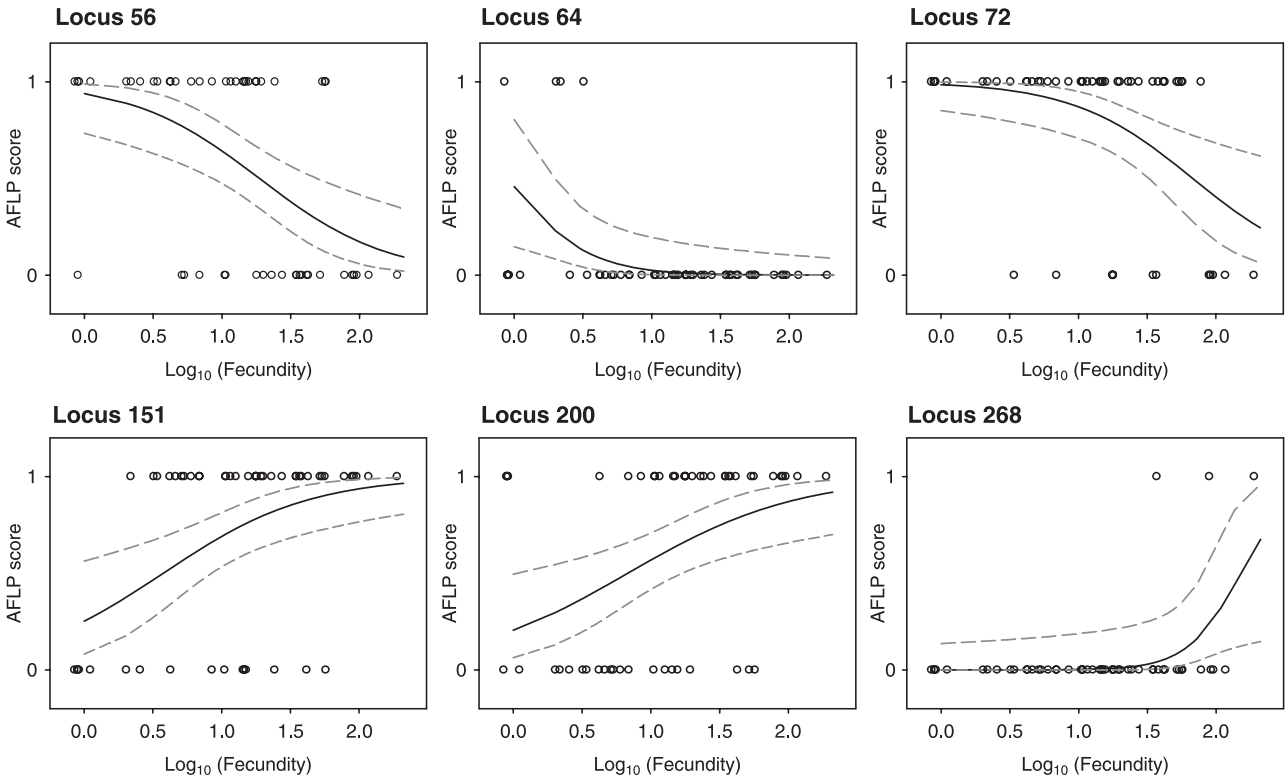

Fig. 2 Representative examples of inverse (upper row) and direct (lower row) relationships between AFLP scores (1, marker present; 0, marker absent) and 20-year cumulative fecundity of individual *Viola cazorlensis* plants, tested by means of logistic regressions (Table 2). Each dot corresponds to a plant, and lines are the fitted logistic regression (continuous line) and 95% confidence limits (dashed lines). A small amount of uniform random noise was added to *x*-coordinates to reduce point overlap.

Bayesian inference were closely coincident. There were both positive (regression parameter > 0; loci 143, 151, 152, 200, 268, 323) and negative (regression parameter < 0; loci 56, 64, 72, 162, 180) associations across individuals between fecundity and the presence of particular markers, as illustrated by the shapes and slopes of representative regressions depicted in Fig. 2.

Dissecting variance in individual fecundity

Multiple linear regressions of individual fecundity (dependent variable) against the scores of the 11 adaptive loci (independent variables) were highly significant for both raw ($F_{10,41} = 7.00$, $P < 0.0001$) and log-transformed ($F_{10,41} = 5.93$, $P < 0.0001$) fecundity data. Coefficients of determination (R^2) were 0.631 and 0.591 for regressions on raw and log-transformed fecundities, respectively. On either the original or log-transformed fecundity scale, therefore, the linear combination of scores of the loci identified as adaptive accounted for almost two-thirds of individual variance in cumulative fecundity, a finding that clearly supports the additivity of their effects on fecundity.

The relative contribution of genotype, environment and genotype × environment interaction to individual variance in fecundity was assessed by fitting a random effect model to the data. Individual plant genotypes were described by their scores on PC1 and PC2, the first two axes from a principal coordinates analysis of the pairwise genetic distance matrix based on the 11 adaptive AFLP loci, which accounted altogether for 61% of variance. Substrate type (ground, rock, or cliff) was used as a descriptor of environment. Results are summarized in Table 3. Genotype as described by PC1 had the greatest contribution to variance in individual fecundity (56%), followed closely by the PC1 × Substrate interaction (38%). Despite the broad confidence intervals of these two estimates, a comparison with the confidence intervals for other effects in the model indicates that most phenotypic variance in long-term fecundity in our study plants was attributable to the joint effects of genotype and the genotype × substrate interaction (Table 3). The negligible importance of the residual component of variance (6%) is a remarkable result, for it denotes that the fitted model captured all major sources of variance in fecundity.

The strong genotypic effect on fecundity is due to the close direct relationship between fecundity and PC1 (Fig. 3). The genotype × environment interaction mainly arises from the significant heterogeneity among substrate types

Table 3 Relative contribution of genotype (PC1 and PC2), environment (substrate type) and genotype × environment (PC1 × substrate and PC2 × substrate) effects to total phenotypic variance in cumulative 20-year fecundity of marked *Viola cazorlensis* plants. PC1 and PC2 are coordinates on the first two axes from a principal coordinates analysis of the pairwise genetic distance matrix between individuals based on the 11 adaptive AFLP loci shown in Table 2. Substrate refers to the type of substrate in which each plant was growing (ground, rock, or cliff). Variance components were obtained by fitting a random effect model to the data using restricted maximum likelihood estimation (REML). Standard errors and confidence intervals obtained by bootstrapping

	Percentage of individual variance in fecundity	
Effect	REML estimate (SE)	95% confidence interval
PC1	56.0 (17.1)	10.2–72.3
PC2	0 (4.3)	0–15.0
Substrate	0 (2.5)	0–4.5
PC1 × Substrate	38.2 (18.5)	9.3–79.2
PC2 × Substrate	0 (14.1)	0–49.5
Residual	5.8 (2.8)	0–10.4

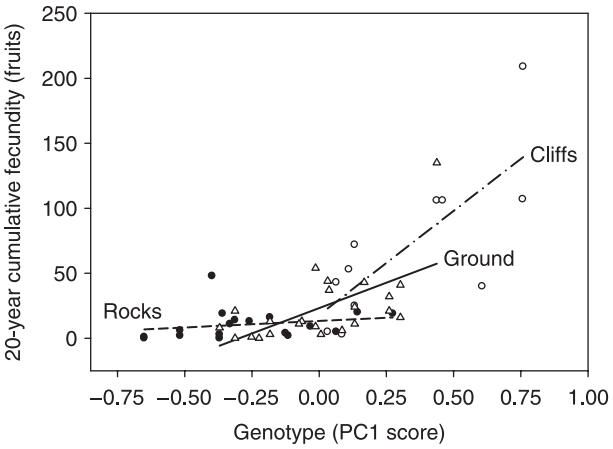

Fig. 3 Relationship between individual cumulative fecundity (total fruits produced over 1988–2008) and the coordinate on the first axis (PC1) from the principal coordinates analysis of the genetic distance matrix between individuals based on the 11 AFLP adaptive loci. Each symbol corresponds to a different plant, coded according to substrate type: black dots, rocks; triangles, ground; open dots, cliffs. The three lines are least-squares regressions separately fitted to the three groups of plants as defined by substrate.

in the slope of the fecundity–PC1 relationship. Model-adjusted, substrate-specific slopes obtained using the ESTIMATE statement in the MIXED procedure increased steadily from the rock (slope ± SE = 49 ± 19 fruits/PC1 unit, $P = 0.13$) through the ground (98 ± 23 fruits/PC1 unit, $P = 0.049$) to the cliff substrate (169 ± 17 fruits/PC1 unit, $P = 0.010$). As shown in Fig. 3, the genotype × environment effect is therefore the consequence of substrate-dependent phenotypic expression of the adaptive AFLP loci that contribute to the PC1 axis. Variation in these adaptive loci has little consequences for fecundity in plants that grow on rocks. In contrast, individual variation in the same loci has moderate to high phenotypic consequences in plants that grow in soil and cliffs, respectively. This reveals considerable phenotypic plasticity for fecundity, with genotypes having predictably different fecundities depending on substrate type.

Discussion

Results of this study have shown that genome scans, if used in combination with individual phenotypic data, can be profitably used to quantify the genetic component of phenotypic variation in populations of unpedigreed wild plants growing in the wild. Application of the method to long-term fecundity data for *Viola cazorlensis* has revealed that the extensive variation in fecundity exhibited by individuals of this long-lived species can almost entirely be statistically accounted for by genetic differences plus genotype × environment effects. We will discuss these two main aspects of our results in the following sections.

Fecundity hierarchies and genetic variance

The highly skewed fecundity distribution found here for *V. cazorlensis* is similar to that commonly reported by plant demographic studies of shorter duration (Solbrig & Solbrig 1984; Weiner & Thomas 1986). 'Fecundity hierarchies', characterized by a mixture of a few exceptionally fecund individuals and many producing moderate-to-low numbers of fruits are closely related to 'size hierarchies', and have been considered a 'universal characteristic of all populations of plants' (Solbrig & Solbrig 1984). A variety of ecological and physiological factors are known to contribute to them, including herbivory, exploitative and interference competition, growth rate, size of the originating seed, microhabitat heterogeneity and, in the case of perennials, age heterogeneity (Burdon & Harper 1980; Antlfinger *et al*. 1985; Stanton 1985; Waller 1985; Gange & Brown 1989; Herrera 1991).

The role of genetic differences as determinants of fecundity hierarchies has received comparatively less attention than ecological factors. This may be partly attributed to the discouraging effect of the negative results obtained by a pioneering study by Gottlieb (1977), who found no relationship between plant size and allozyme genotypes, and interpreted the result as validating 'the ecological practice of ignoring the genotypes of individual plants when accounting for the reproductive activity of a population, even when individuals make very different contributions

to the seed pool.' Further contributing to the relative neglect of genetic effects in studies of fecundity hierarchies is the widely held theoretical expectation that selection should quickly deplete genetic variance for traits closely correlated with fitness and, consequently, that fecundity differences should be mostly environmental in origin and weakly related to genotypic variation (Falconer 1989; Mousseau & Roff 1987). Despite this theoretical expectation, however, the few studies that have used quantitative genetics methods to look for the genetic basis of individual differences in size/fecundity in wild plants have often found significant amounts of additive genetic variance for these traits (Antlfinger *et al*. 1985; Thomas & Bazzaz 1993). In the forest herb *Viola sororia*, for example, narrow-sense heritabilities (h^2) of plant size, biomass and seed production ranged between 0.28–0.38 (Antlfinger *et al*. 1985). Assuming that our estimates of genetic effects on individual variation in fecundity for *V. cazorlensis* predominantly reflect additive genetic variance (Hill *et al*. 2008), then narrow-sense heritability for fecundity would equal 0.56–0.63 (depending on method), estimated as the proportion of phenotypic variance attributable to variance of genotypic values (Visscher *et al*. 2008). Our results thus confirm earlier studies by showing that, when a sufficiently broad range of genotypes are considered, genetic effects often emerge as important determinants of fecundity differences in wild plant populations. Failures to detect this relationship in other studies may perhaps be due to insufficient statistical power due to the small number of genotypes assayed (e.g. Solbrig 1981).

The persistence in natural populations of significant amounts of additive genetic variance for traits that, like fecundity, are closely related to fitness remains a challenge in evolutionary genetics, and a number of hypotheses have been advanced to explain it (Visscher *et al*. 2008). One of these hypotheses interprets the maintenance of additive genetic variance for fitness-related traits as the outcome of spatial or temporal variation in the fitness effects of alleles that contribute to the trait, that is, to genotype × environment interactions on the trait (Gillespie & Turelli 1989; Turelli & Barton 2004). Empirical support for this hypothesis is contradictory (Heywood & Levin 1984; Mazer & Wolfe 1992; Young *et al*. 1994; Stratton & Bennington 1998; Prati & Schmid 2000), possibly because it is not easy to estimate genotype × environment effects for a sufficiently representative number of genotypes across a reasonably broad environmental range by means of conventional quantitative genetics methods. One strength of the method presented here is precisely that it affords tests of genotype × environment effects on phenotypic traits over roughly the entire range of genotypes and environments occurring naturally in a population, rather than in a limited subsample thereof as usually imposed on quantitative genetics studies by practical limitations. Our results reveal that when a biologically realistic, broad genotype × environment sampling space is considered, strong genotype × environment effects on fecundity can arise, which in the case of the *V. cazorlensis* population studied here account for as much as 38% of phenotypic variance in fecundity. It is not possible at present to know whether similar results would be obtained for other *V. cazorlensis* populations.

In our study, genotype × environment effects arose as a consequence of differences between substrate types in the slope of the fecundity–genotype relationship, which was steepest for plants in cliffs, intermediate for those in the ground, and flat and nonsignificant in rocks. This gradient parallels one of ecological 'favourableness' for *V. cazorlensis*, running from cliffs (the most favourable microsite) through ground (intermediate) to rocks (the most adverse microsite) (Herrera 1989, 1993). Our results suggest that the extensive genetic variance for fecundity existing in the *V. cazorlensis* population studied may be explained by the contrasting relationship between genotype and fecundity in different microhabitats, and by the fact that there is not a single genotype that is consistently the best at all substrates (Gillespie & Turelli 1989; Turelli & Barton 2004). Since the three microhabitats are closely intermingled in space and presumably connected by extensive gene exchange via pollen and seeds, we hypothesize that the lack of a relationship between genotype and fecundity among rock-growing plants may contribute to maintain in the population alleles that would eventually be eliminated by selection at the other microhabitats, where fecundity and genotypes are significantly related. Rock-growing plants, representing about 40% of individuals at the study population (Herrera 1989), might thus be playing the role of local reservoirs of genetic diversity where unfavourable alleles remain 'hidden' to the eroding action of selection. Even though rock plants have lower maternal fecundities (Herrera 1993), they are expected to sire part of the seeds produced by the more fecund individuals in the other two substrates, which would contribute to maintain unfavourable alleles in the latter's progeny as envisaged by the hypothesis.

Genome scan at the within-population level

Most recent genome scans on wild plants have adopted a population-genomic approach, using DNA markers to screen the genomes of individuals from different populations (Black *et al*. 2001). In that implementation of genomic scan, loci subject to selection are sought ('outlier loci') that exhibit distinguishing signatures of selection in comparison to neutral ones. The identification of such outlier loci depends heavily on particular genetic models (Beaumont & Nichols 1996; Luikart *et al*. 2003; Beaumont & Balding 2004). The potential usefulness of genomic scan in evolutionary studies of wild plants is not limited, however, to questions framed at the between-population level. The method can

also be used to address important evolutionary questions at the within-population level. In particular, we expect genomic scan of wild individuals of known phenotypes to be helpful to unravel evolutionary processes currently taking place within populations like selection and local adaptation, as well as to estimate important genetic parameters when other methods are not applicable, as shown in this study. Like between-populations implementations, within-population applications of genomic scan also rely on some assumptions. The most important of these is that the set of loci found to be statistically related to individual variation in phenotypic traits (named adaptive loci here) are functionally related to such variation (or closely linked to loci functionally affecting such variation) and can be used as valid descriptors of individual genotypes. In the case of AFLP markers, several lines of evidence show that this assumption will hold true in many instances: (i) high-density linkage maps often reveal a close proximity of AFLP markers to quantitative trait loci (QTL) or genes with known functionality (Raman *et al*. 2002; Herselman *et al*. 2004; Yuan *et al*. 2004; Papa *et al*. 2007); (ii) AFLP allelic frequencies are responsive to artificial divergent selection on quantitative traits (Cameron *et al*. 2003; Jump *et al*. 2008); and (iii) studies on populations of the same species combining phenotypic information with population-genomic scans often reveal concordant phenotypic and AFLP variation (Klappert *et al*. 2007; Herrera & Bazaga 2008).

We found 11 AFLP loci significantly related to long-term fecundity of individual *V. cazorlensis* plants. Genotypic variation involving these loci accounted for nearly two-thirds of individual variance in fecundity, which denotes a strong genetic effect on fecundity. The genotype × environment effect accounted for an additional 38% of fecundity variance. The model fitted to the data accounted for as much as 94% of total variance in fecundity occurring in the sample, which is a remarkable finding in view of the low levels of explained variance commonly obtained in ecological and evolutionary studies (Møller & Jennions 2002). These results attest to the explanatory capacity of the variables included in the model and the validity of our method in the case of *V. cazorlensis*. In addition, the finding that a few loci account for a large proportion of phenotypic variance is consistent with expectations from theoretical models that predict an exponential distribution of effect sizes of individual loci (Orr 1998; Farrall 2004). Under this 'exponential model' of locus effect distribution, quantitative variation is determined by a few quantitative trait loci of relatively large effects and an increasing number of genes of progressively smaller effects. This predicted unequality of individual loci effects is apparent even in the restricted set of 11 adaptive loci found in this study. The three loci with the strongest effects on fecundity as judged from their logistic regression parameters (loci 143, 268 and 323; Table 2), accounted collectively for 47.6% of total variance in fecundity, while the remaining eight loci accounted altogether for only 15.5%. Each locus in the large-effect group thus explains, on average, 15.9% of phenotypic variance in fecundity, while each locus in the small-effect group explains only 1.9%, a difference of roughly one order of magnitude.

The ability of within-population genomic scan to produce reliable estimates of genetic and environmental variance components of phenotypic traits will depend on *both* the accuracy of the phenotypic representation of individuals and the number of adaptive loci identified. In long-lived plants, phenotypic traits of individuals frequently fluctuate between years, as found for fecundity and morphometric traits in *V. cazorlensis* (Herrera 1993). In these cases, using long-term average phenotypes to characterize individuals may prove essential for identifying phenotype–genotype relationships. The fact that fecundity estimates used in this study are based on 20 years of data has probably enhanced the likelihood of identifying fecundity-related markers. It must be noted, however, that similar results involving essentially the same AFLP markers were obtained when analyses were conducted separately on two subsets of fecundity data each lasting for 10 years (results not shown), which points to the robustness of the method even for more usual data sets spanning shorter periods. The greater the number of adaptive loci identified, the more accurate the genotypic characterization of individuals that can be achieved and hence the likelihood of detecting phenotype-genotype relationships. In this study, adaptive loci represented 3% of total AFLP fragments assayed. This proportion is similar to that of outlier loci reported in population-genomic scans based on AFLP markers (Herrera & Bazaga 2008, and references therein), which suggests the possible rule of thumb that at least several hundred polymorphic AFLP markers should be used in genome scans of wild plants. It may be intuited, however, that everything else being equal the expected number of adaptive loci will be inversely related to genome size. No information is available on the genome size of *V. cazorlensis*, but it is probably of small size in view of the low *C*-value (1.35 pg) of the congeneric *Viola riviniana* (Bennett & Leitch 2004). Larger genomes may require a larger number of polymorphic markers than used here.

The number of primer combinations and particular restriction enzymes used can also influence results of genomic scans. In this study, the proportion of adaptive loci varied between primer combinations, a finding suggesting that better results are to be expected with diverse arrays of primer combinations (Herrera & Bazaga 2008, and references therein). We found differences between combinations having *Pst* and *Eco* as rare-cutter enzymes (eight and three adaptive loci, representing 4.4% and 1.6% of total loci in each group, respectively). At least in *V. cazorlensis*, combinations using *Pst* should therefore be preferred over those using *Eco* to maximize the likelihood of identifying adaptive loci

with significant effects on quantitative traits. Observed differences between the two restriction enzymes in the likelihood of identifying adaptive loci are probably related to the fact that *Pst* is sensitive to DNA methylation while *Eco* is not. Consequently, *Pst*-based markers are expected to concentrate in gene-rich, transcriptionally active genomic regions, thus enhancing the likelihood of close linkage to quantitative trait loci (Vuylsteke *et al*. 1999; Yuan *et al*. 2004).

Concluding remarks

This paper has shown that explorative genomic scans conducted within populations can be used to quantify the genetic component of individual phenotypic differences and estimate trait heritabilities under natural conditions in species that are not amenable to conventional quantitative genetics methods. In this respect, our method provides an alternative to procedures that estimate trait heritability in the field by inferring pedigrees through estimation of relatedness (Ritland 1996, 2000, 2005; Garant & Kruuk 2005; Fernández & Toro 2006). By exclusively focusing on the strength of the statistical relationship between putative adaptive loci and phenotype, the method implemented in this study is free from the assumptions that may limit the application of procedures based on estimating relatedness between individuals. Interestingly, however, the heritability estimate for fecundity obtained from our data with Ritland's relatedness estimator for dominant markers implemented in program Mark (Ritland 2005) was 0.47, which falls close to values obtained with our method (0.56–0.63).

Although these aspects have been not considered explicitly in this paper, some of our results are relevant to the study of selection in natural populations. Population genomic approaches applied to multiple populations are useful to infer adaptive divergence by identifying the signatures left by selection on the allelic frequencies of adaptive loci. Within-population genomic scans may complement these studies by providing direct evidence of the shape and strength of local selection currently acting on quantitative traits and their associated adaptive loci. Assuming a diallelic nature for AFLP markers, relationships between adaptive loci and individual fecundity found in this study suggest that some alleles will be over-represented, and others under-represented, in the seed bank of the local *V. cazorlensis* population, and presumably also in the next generation, in comparison with their current allelic frequencies in the adult population. This falls close to providing a parsimonious, direct proof of ongoing local selection. If relationships could be identified linking fecundity-related loci with some phenotypic traits of interest (e.g. floral morphology, physiological traits), then identifying instances of current selection on such traits would become relatively straightforward. Future studies will examine these possibilities for the *V. cazorlensis* population studied here.

Acknowledgements

We are grateful to Rafael G. Albaladejo, José L. Garrido, Roger Jovani, Mónica Medrano, Clara de Vega and three reviewers for useful comments on the manuscript. Computer-intensive analyses were carried out at the Finis Terrae supercomputer of the Centro de Supercomputación de Galicia (CESGA). Permission to work in the Sierra de Cazorla was provided by the Consejería de Medio Ambiente, Junta de Andalucía. This work was supported by grants 2005-RNM-156 (Consejería de Innovación, Ciencia y Empresa, Junta de Andalucía), and CGL2006-01355 (Ministerio de Educación y Ciencia, Gobierno de España).

References

Agresti A (2007) *An Introduction to Categorical Data Analysis*, 2nd edn. Wiley, Hoboken, New Jersey.

Andreasen K, Baldwin BG (2001) Unequal evolutionary rates between annual and perennial lineages of checker mallows (*Sidalcea*, Malvaceae): evidence from 18S-26S rDNA internal and external transcribed spacers. *Molecular Biology and Evolution*, **18**, 936–944.

Andrew RL, Peakall R, Wallis IR *et al*. (2005) Marker-based quantitative genetics in the wild?: the heritability and genetic correlation of chemical defenses in *Eucalyptus*. *Genetics*, **171**, 1989–1998.

Antlfinger AE, Curtis WF, Solbrig OT (1985) Environmental and genetic determinants of plant size in *Viola sororia*. *Evolution*, **39**, 1053–1064.

Applied Biosystems (2005) *AFLP Plant Mapping Protocol*. Applied Biosystems, Foster City, California.

Beaumont MA (2005) Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Bennett MD, Leitch IJ (2004) *Angiosperm DNA C-Values Database* (release 5.0. Dec. 2004). Available from URL: http://www.kew.org/cvalues/homepage.html.

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, **46**, 441–469.

Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci por adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773–783.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

Burdon JJ, Harper JL (1980) Relative growth rates of individual members of a plant population. *Journal of Ecology*, **68**, 953–957.

Cameron ND, van Eijk MJT, Brugmans B, Peleman J (2003) Discrimination between selected lines of pigs using AFLP markers. *Heredity*, **91**, 494–501.

Castiglioni P, Ajmone-Marsan P, van Wijk R, Motto M (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. *Theoretical and Applied Genetics*, **99**, 425–431.

Falconer DS (1989) *Introduction to Quantitative Genetics*, 3rd edn. Longman, Harlow, Essex, UK.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Farrall M (2004) Quantitative genetic variation: a post-modern view. *Human Molecular Genetics*, **13**, R1–R7.

Fernández J, Toro MA (2006) A new method to estimate relatedness from molecular markers. *Molecular Ecology*, **15**, 1657–1667.

Gange AC, Brown VK (1989) Insect herbivory affects size variability in plant populations. *Oikos*, **56**, 351–356.

Garant D, Kruuk LEB (2005) How to use molecular marker data to measure evolutionary parameters in wild populations. *Molecular Ecology*, **14**, 1843–1859.

Geber MA, Griffen LR (2003) Inheritance and natural selection on functional traits. *International Journal of Plant Sciences*, **164**, S21–S42.

Gillespie JH, Turelli M (1989) Genotype-environment interactions and the maintenance of polygenic variation. *Genetics*, **121**, 129–138.

Gottlieb LD (1977) Genotypic similarity of large and small individuals in a natural population of the annual plant *Stephanomeria exigua* ssp. *coronaria* (Compositae). *Journal of Ecology*, **65**, 127–134.

Herrera CM (1989) Biología y ecología de *Viola cazorlensis*. II. Uso de sustratos, reproducción y consumo por los herbívoros. *Anales del Jardín Botánico de Madrid*, **47**, 125–138.

Herrera CM (1990a) The adaptedness of the floral phenotype in a relict endemic, hawkmoth-pollinated violet. 1. Reproductive correlates of floral variation. *Biological Journal of the Linnean Society*, **40**, 263–274.

Herrera CM (1990b) The adaptedness of the floral phenotype in a relict endemic, hawkmoth-pollinated violet. 2. Patterns of variation among disjunct populations. *Biological Journal of the Linnean Society*, **40**, 275–291.

Herrera CM (1991) Dissecting factors responsible for individual variation in plant fecundity. *Ecology*, **72**, 1436–1448.

Herrera CM (1993) Selection on floral morphology and environmental determinants of fecundity in a hawk moth-pollinated violet. *Ecological Monographs*, **63**, 251–275.

Herrera CM, Bazaga P (2008) Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. *Molecular Ecology*, **17**, 5378–5390.

Herselman L, Thwaites R, Kimmins FM *et al*. (2004) Identification and mapping of AFLP markers linked to peanut (*Arachis hypogaea* L.) resistance to the aphid vector of groundnut rosette disease. *Theoretical and Applied Genetics*, **109**, 1426–1433.

Heywood JS, Levin DA (1984) Genotype-environment interactions in determining fitness in dense, artificial populations of *Phlox drummondii*. *Oecologia*, **61**, 363–371.

Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, **4**, e1000008. doi. 10.1371/journal.pgen.1000008.

Jump AS, Peñuelas J, Rico L *et al*. (2008) Simulated climate change provokes rapid genetic change in the Mediterranean shrub *Fumana thymifolia*. *Global Change Biology*, **14**, 637–643.

Klappert K, Butlin RK, Reinhold K (2007) The attractiveness fragment—AFLP analysis of local adaptation and sexual selection in a caeliferan grasshopper, *Chorthippus biguttulus*. *Naturwissenschaften*, **94**, 667–674.

van Kleunen M, Ritland K (2005) Estimating heritabilities and genetic correlations with marker-based methods: an experimental test in *Mimulus guttatus*. *Journal of Heredity*, **96**, 368–375.

Leinonen T, O'Hara RB, Cano JM, Merilä J (2008) Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, **21**, 1–17.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006) *SAS for Mixed Models*, 2nd edn. SAS Institute, Cary, North Carolina.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer & Associates, Sunderland, Masachusetts.

Mackay TFC (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics*, **35**, 303–339.

Mazer SJ, Wolfe LM (1992) Planting density influences the expression of genetic variation in seed mass in wild radish (*Raphanus sativus* L. Brassicaceae). *American Journal of Botany*, **79**, 1185–1193.

Merilä J, Sheldon BC (1999) Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity*, **83**, 103–109.

Møller AP, Jennions MD (2002) How much variance can be explained by ecologists and evolutionary biologists? *Oecologia*, **132**, 492–500.

Mousseau TA, Roff DA (1987) Natural selection and the heritability of fitness components. *Heredity*, **59**, 181–197.

Ojha BR (2005) Comparison of Eco-Mse and Pst-Mse primer combinations in generating AFLP map of tomato. *Journal of the Institute of Agriculture and Animal Sciences*, **26**, 27–35.

Olsen KM, Purugganan MD (2004) Plant population genomics, linkage disequilibrium mapping, and the genetics of adaptation. In: *Plant Adaptation: Molecular Genetics and Ecology* (eds Cronk QCB, Whitton J, Ree RH, Taylor IEP), pp. 45–52. National Research Council, Ottawa, Canada.

Orr HA (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, **52**, 935–949.

Papa R, Bellucci E, Rossi M *et al*. (2007) Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Annals of Botany*, **100**, 1039–1051.

Parisod C, Christin PA (2008) Genome-wide association to fine-scale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist*, **178**, 436–447.

Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Prati D, Schmid B (2000) Genetic differentiation of life-history traits within populations of the clonal plant *Ranunculus reptans*. *Oikos*, **90**, 442–456.

Pritchard JK, Wen X, Falush D (2007) *Documentation for Structure Software: Version 2.2*. Department of Human Genetics, University of Chicago, Chicago, Illinois. Available at http://pritch.bsd.uchicago.edu/structure.html.

Pujol B, Wilson AJ, Ross RIC, Pannell JR (2008) Are Q_{ST}–F_{ST} comparisons for natural populations meaningful? *Molecular Ecology*, **17**, 4782–4785.

Raman H, Moroni JS, Sato K, Read BJ, Scott BJ (2002) Identification

of AFLP and microsatellite markers linked with an aluminium tolerance gene in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, **105**, 458–464.

Ritland K (1996) A marker-based method for inferences about quantitative inheritance in natural populations. *Evolution*, **50**, 1062–1073.

Ritland K (2000) Detecting inheritance with inferred relatedness in nature. In: *Adaptive Genetic Variation in the Wild* (eds Mousseau TA, Sinervo B, Endler J), pp. 187–199. Oxford University Press, Oxford, UK.

Ritland K (2005) Multilocus estimation of pairwise relatedness with dominant markers. *Molecular Ecology*, **14**, 3157–3165.

Ritland K, Ritland C (1996) Inferences about quantitative inheritance based on natural population structure in the yellow monkeyflower, *Mimulus guttatus*. *Evolution*, **50**, 1074–1082.

Roff DA (1997) *Evolutionary Quantitative Genetics*. Chapman & Hall, New York.

SAS Institute (2004) *SAS/STAT 9.1. User's Guide*. SAS Institute, Cary, North Carolina.

SAS Institute (2006) *Preliminary Capabilities for Bayesian Analysis in SAS/STAT Software*. SAS Institute, Cary, North Carolina.

Smith SA, Donoghue MJ (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science*, **322**, 86–89.

Solbrig OT (1981) Studies on the population biology of the genus *Viola*. 2. The effect of plant size on fitness in *Viola sororia*. *Evolution*, **35**, 1080–1093.

Solbrig OT, Solbrig DJ (1984) Size inequalities and fitness in plant populations. *Oxford Surveys in Evolutionary Biology*, **1**, 141–159.

Stanton ML (1985) Seed size and emergence time within a stand of wild radish (*Raphanus raphanistrum* L.): the establishment of a fitness hierarchy. *Oecologia*, **67**, 524–531.

Stanton ML, Thiede DA (2005) Statistical convenience vs biological insight: consequences of data transformation for the analysis of fitness variation in heterogeneous environments. *New Phytologist*, **166**, 319–337.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences, USA*, **100**, 9440–9445.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Stratton DA, Bennington CC (1998) Fine-grained spatial and temporal variation in selection does not maintain genetic variation in *Erigeron annuus*. *Evolution*, **52**, 678–691.

Thomas SC, Bazzaz FA (1993) The genetic component in plant size hierarchies: norms of reaction to density in a *Polygonum* species. *Ecological Monographs*, **63**, 231–249.

Turelli M, Barton NH (2004) Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and GxE interactions. *Genetics*, **166**, 1053–1079.

Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics*, **9**, 255–266.

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

Vuylsteke M, Mank R, Antonise R *et al*. (1999) Two high-density AFLP linkage maps of *Zea mays* L. analysis of distribution of AFLP markers. *Theoretical and Applied Genetics*, **99**, 921–935.

Waller DM (1985) The genesis of size hierarchies in seedling populations of *Impatiens capensis* Meerb. *New Phytologist*, **100**, 243–260.

Weiner J, Thomas SC (1986) Size variability and competition in plant monocultures. *Oikos*, **47**, 211–222.

Young HJ, Stanton ML, Ellstrand NC, Clegg JM (1994) Temporal and spatial variation in heritability and genetic correlations among floral traits in *Raphanus sativus*, wild radish. *Heredity*, **73**, 298–308.

Young WP, Schupp JM, Keim P (1999) DNA methylation and AFLP marker distribution in the soybean genome. *Theoretical and Applied Genetics*, **99**, 785–792.

Yuan L, Dussle CM, Muminovic J, Melchinger AE, Lubberstedt T (2004) Targeted BSA mapping of *Scmv1* and *Scmv2* conferring resistance to SCMV using *Pst*I/*Mse*I compared with *Eco*RI/*Mse*I AFLP markers. *Plant Breeding*, **123**, 434–437.