

Genomic Scan as a Tool for Assessing the Genetic Component of Phenotypic Variance in Wild Populations

Carlos M. Herrera

Abstract

Methods for estimating quantitative trait heritability in wild populations have been developed in recent years which take advantage of the increased availability of genetic markers to reconstruct pedigrees or estimate relatedness between individuals, but their application to real-world data is not exempt from difficulties. This chapter describes a recent marker-based technique which, by adopting a genomic scan approach and focusing on the relationship between phenotypes and genotypes at the individual level, avoids the problems inherent to marker-based estimators of relatedness. This method allows the quantification of the genetic component of phenotypic variance (“degree of genetic determination” or “heritability in the broad sense”) in wild populations and is applicable whenever phenotypic trait values and multilocus data for a large number of genetic markers (e.g., amplified fragment length polymorphisms, AFLPs) are simultaneously available for a sample of individuals from the same population. The method proceeds by first identifying those markers whose variation across individuals is significantly correlated with individual phenotypic differences (“adaptive loci”). The proportion of phenotypic variance in the sample that is statistically accounted for by individual differences in adaptive loci is then estimated by fitting a linear model to the data, with trait value as the dependent variable and scores of adaptive loci as independent ones. The method can be easily extended to accommodate quantitative or qualitative information on biologically relevant features of the environment experienced by each sampled individual, in which case estimates of the environmental and genotype \times environment components of phenotypic variance can also be obtained.

Key words: Amplified fragment length polymorphism, Genetic determination, Genomic scan, Heritability, Multiple linear regression, Phenotypic variance components

1. Introduction

The study of contemporary evolution in wild populations poses two major challenges, namely, quantifying selection on the phenotypic traits of interest (phenotypic selection) and evaluating the extent to which individual variation in these traits is due to genetic differences (heritability) (1). Phenotypic selection occurs when individuals with different phenotypic characteristics differ

in their fitness because of differences in viability and/or fecundity. Following the development of a theoretical framework and associated analytical toolbox for estimating the shape and strength of selection on quantitative traits (1–5), studies of phenotypic selection on wild populations have proliferated in the past 30 years, and have shown that phenotypic selection is widespread in nature (reviews in refs. 6–8). Nevertheless, “natural selection is not evolution” ((9), p. vii), and changes in phenotypic value distributions arising from selection can be constrained or altered by the pattern of inheritance of the traits of interest. Phenotypic selection studies on wild populations usually fall short of demonstrating ongoing evolutionary change because of the absence of information on the genetic basis of variation in the phenotypic traits of interest ((7); but see, e.g., refs. 10–12 for some exceptions), a limitation that stems from the problems involved in estimating heritability and genetic correlations under natural field conditions (13).

Traditional quantitative genetics techniques to estimate heritability rely on information about the relationships among individuals, generally in the form of a pedigree structure or a matrix of pairwise individual relatedness (14, 15). Except for some special cases, as when pedigrees can be inferred from detailed observations of matings (16), neither pedigrees nor information on individual relatedness are available for natural populations. This difficulty has been customarily circumvented by artificially breeding individuals of known pedigree through controlled crosses between field-collected progenitors of known phenotypes, and keeping the experimental populations in controlled artificial environments. Many organisms, however, are not amenable to these methods. Conventional quantitative genetics methods based on experimental crossing schemes are impractical or impossible with long-lived species that cannot be artificially crossed, propagated vegetatively or kept in captivity, or with those that reproduce irregularly or have long periods until first reproduction. In addition, the expression of quantitative traits usually differ between environments, and heritability estimates may differ widely between wild and artificial populations of the same species (17), hence field-based estimates are essential to investigate the possibility for evolution in the settings where the organisms naturally occur (18).

Alternative methods for estimating trait heritability in wild populations have been developed in recent years which take advantage of the increased availability of genetic markers to reconstruct pedigrees or to estimate relatedness between individuals (reviewed in refs. 19–21). These methods, however, are not a panacea, as their application also faces a number of difficulties. Most of these procedures either make assumptions on the structure of relatedness

in the population (e.g., assuming known, discrete classes of relatedness) or require additional information on parentage of individuals, which greatly limits their application to complex natural populations where individuals may differ widely in age and relatedness, as in uneven-aged populations of long-lived organisms. Ritland regression procedure (18, 22, 23), which is based on assessing the strength of the relationship between pairwise phenotypic similarity and relatedness for the individuals in a population, stands apart from the rest of marker-based methods in that it allows direct estimation of quantitative genetic parameters without requiring specification of an explicit pedigree or prior knowledge of population structure (19).

At first view, Ritland method appears widely applicable, since it is based on the assessment of the simple relationship between phenotype and genotype, and it has been used to estimate genetic parameters for wild, unpedigreed populations (24–27). Nevertheless, this method has two intrinsic limitations (see ref. 19 for a detailed discussion). First, since it depends heavily on the efficiency of the estimation of pairwise relatedness, the large statistical errors associated with these estimators will produce broad standard errors around genetic parameter estimates. Second, the method requires the existence of significant variance in relatedness in the sample analyzed (18, 22, 23). The limited number of applications of Ritland method and their modest success are perhaps attributable to the joint influence of these intrinsic limitations, and despite its appeal the utility of the method may be problematic (19, 25, 28–30). In the same pessimistic vein, it has been also suggested that, despite the increased availability of genetic markers, marker-based methods in general might not be as useful for genetic parameters estimation as previously thought (13, 19).

A new marker-based technique has been recently proposed by Herrera and Bazaga (31) which, by focusing on individual genotypes and adopting a genomic scan approach, avoids the problems associated with estimating pedigrees or individual relatedness. Although this method is not necessarily aimed at the direct estimation of quantitative genetic parameters, it does provide an alternative, relatively easy way for addressing one of the key aspects in evolutionary studies, namely, the quantitative assessment of the genetic component of phenotypic variance in wild populations. As stressed by van Kleunen and Ritland (26), “even the possibility of determining the presence of heritable genetic variation is a big improvement for studies testing for the potential for evolution in natural populations.” The motivation and a step-by-step description of the genomic scan-based method of Herrera and Bazaga are presented in this chapter, followed by an example of application and a comparison with results from the Ritland procedure for the same data.

2. Overview and Motivation

A trait can be “hereditary” in the sense of being transmitted from parent to offspring or in the sense of being determined by the genotype. Under the first meaning, the heritability of a character is “the relative importance of heredity in determining phenotypic values,” and is known as heritability in the narrow sense (14). Marker-based methods for estimating heritability mentioned in the preceding section are framed in this specific context, hence their reliance on pedigree reconstruction or relatedness estimation. Under the second meaning, a trait’s heritability is simply “a quantity defined as the proportion of phenotypic variance that is genetically determined” (32). This is the extent to which individual phenotypes are determined by the genotypes, and is known as heritability in the broad sense or degree of genetic determination (14). The method described in this chapter is framed in this second context, as it focuses on quantifying the fraction of population-wide phenotypic variance that is attributable to genetic differences between individuals. These genetic differences are evaluated directly through the application of a genomic scan approach at the within-population scale. Rather than targeting the narrow sense heritability of the phenotypic trait of interest by inferring the relatedness of individuals and then relating it to their phenotypic similarity, as done with the Ritland method, the present method directly proceeds 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.

Consider a population for which a sample of individuals are simultaneously characterized with regard to some quantitative phenotypic trait and a large number of anonymous, putatively neutral genetic markers. If the number of markers is large and are randomly distributed across the whole genome, a certain proportion of the markers is expected to be in linkage disequilibrium with non-neutral, adaptive loci having some causative effect on the phenotypic value of the trait of interest. It should therefore be possible to single out these non-neutral markers by searching for significant statistical associations between markers and phenotypic values across individuals, an approach usually known as association mapping or linkage disequilibrium mapping (33–35). This method has been used, for example, to infer associations between quantitative trait loci (QTLs) and phenotypic traits, and the same underlying principles can motivate genome screenings aimed at identifying functional associations between anonymous markers and phenotypic traits. Once a set of non-neutral (or linked to non-neutral) markers has been identified which are significantly correlated with phenotypic values, statistical models can be fitted to the data which

estimate the proportion of phenotypic variance in the sample statistically accounted for by the multilocus differences between individuals in these non-neutral loci.

The success of the method will depend on whether non-neutral genomic regions that are causatively associated with the phenotypic trait of interest, or adjacent regions in linkage disequilibrium, are “hit” by some of the anonymous markers chosen, so that genotype–phenotype correlations across individuals can be revealed. Selection of markers is thus critical. The larger the number of markers and the more thorough its distribution across the genome, the higher the expected probability of identifying non-neutral genomic regions causatively associated with the phenotypic trait of interest. The amplified fragment length polymorphism (AFLP) technique (36) allows for a virtually unlimited number of markers without prior sequence knowledge. Detailed linkage maps have also revealed that AFLP markers are thoroughly distributed over the whole genome (37–40). These features make AFLP markers particularly suited to genomic scan and linkage disequilibrium mapping approaches with non-model organisms (35, 41, 42), and also render them the markers of choice for the method described in this chapter. In the immediate future, large single-nucleotide polymorphism (SNP) datasets made available for non-model species by next-generation sequencing will provide additional opportunities for the application of the method described here, with the added advantage that SNPs could be easily selected in or near coding regions of the genome.

The method described here rests on the assumption that the set of loci found to be statistically related to individual variation in phenotypic traits are functionally, causatively related to such variation, or linked to loci functionally affecting such variation, and can therefore be used as valid descriptors of individual genotypes. In the case of AFLP markers, and here lies yet another reason supporting their use, three lines of evidence show that this assumption will probably hold true in most instances: (a) high-density linkage maps often reveal a close proximity of AFLP markers to QTLs or genes with known functionality (43–46); (b) AFLP allelic frequencies are responsive to artificial selection on quantitative traits (47–49); and (c) studies on populations of the same species combining phenotypic information with population-genomic scans often reveal concordant phenotypic and AFLP variation (50, 51).

3. Methods

In the following, it is assumed that the analysis focuses on a single quantitative phenotypic trait that has been measured in a randomly chosen sample of individuals from a single study population.

The same individuals are also fingerprinted using AFLP markers, so that phenotypic trait values and multilocus AFLP data are simultaneously available.

3.1. AFLP Fingerprinting

As noted above, using a large number of markers is essential to increase the chances of hitting at or near (linked) genomic regions causatively associated with the phenotypic trait under consideration. Proportions of non-neutral loci reported by previous AFLP-based genomic scans of natural populations of non-model organisms mostly fall between 1 and 5% of the polymorphic markers assayed ((35, 51) and references therein), which suggests the rule of thumb that individuals should be fingerprinted using a minimum of several hundred polymorphic AFLP markers. The probability of success will be enhanced if this large number of markers is achieved by selecting a variety of primer pairs for each of several combinations of restriction enzymes, rather than by selecting a single enzyme combination and a few primer pairs each yielding many markers. This recommendation is motivated by the observation that non-neutral AFLP markers often are nonrandomly distributed and clustered on particular primer pairs (51–53) or restriction enzyme combinations (31). AFLP markers obtained using *EcoRI* and *PstI* as rare cutter enzymes often differ widely in their distribution across genomes (38, 45, 54, 55), thus using a combination of *EcoRI*–*MseI* and *PstI*–*MseI* primer pairs will generally be a better starting off strategy than using primer pairs based on only one rare cutter. This is illustrated by results of the study by Herrera and Bazaga (31) on the wild violet *Viola cazorlensis*, where AFLP markers significantly associated with individual fecundity were about three times more frequent among *PstI*-based (4.4% of polymorphic loci) than among *EcoRI*-based combinations (1.6% of polymorphic loci).

3.2. Checking for Cryptic Population Substructuring

In natural populations, marker–phenotype associations may not reflect an underlying causal relationship. Significant marker–phenotype associations can be caused by admixture between subpopulations that have different allele frequencies at marker loci and different values of the trait (33, 56). Such situation might occur, for example, if the study population comprises several genetically distinct subgroups that originated from independent colonization events and differ in average phenotypes. The possibility of spurious marker–phenotype associations due to such cryptic population substructuring should therefore be confidently ruled out prior to conducting any further analyses. This can be accomplished with any of the model-based Bayesian assignment methods currently available for the identification of admixture events using multilocus molecular markers, as implemented, for example, in the programs structure (57–59) or baps (60, 61).

If results of these preliminary analyses do not suggest any cryptic substructuring, then it would be justified to proceed to the analytical step in Subheading 3.3. Alternatively, if there are reasons to suspect that the study population is cryptically substructured, one should further check whether genetic subgroups arising from the Bayesian analysis are phenotypically homogeneous (e.g., by comparing trait means across subgroups). If subgroups are phenotypically similar, despite genetic distinctiveness, it would be safe to proceed to the next step. If subgroups were found to differ phenotypically, it would still be possible to proceed by splitting the sample and conducting separate analyses on each subgroup. This would allow for testing the intriguing possibility that genetically distinct subgroups differ in the degree of genetic determination of the phenotypic trait under study.

3.3. Identifying Phenotype-Related Markers

To identify the AFLP markers associated with the phenotypic trait under consideration, separate logistic regressions will be run for each polymorphic marker, using band presence as the dependent binary variable and the phenotypic trait as the continuous, independent one. The statistical significance of each marker–trait relationship will be assessed by consideration of the p -value obtained from a likelihood ratio test. Although several tests are available for assessing the significance of logistic regressions, likelihood ratio tests are more powerful and reliable for sample sizes often used in practice (62). Given the large number of logistic regressions involved (as many as polymorphic AFLP markers), precautions must be taken to account for the possibility of obtaining by chance alone an unknown number of false significant regressions (see Note 1).

3.4. Estimating the Genetic Component of Variance

In this step, the combined phenotypic effects of all the AFLP markers found in Subheading 3.3 to be significantly related to the trait under consideration (“adaptive loci” hereafter) is estimated by fitting a multiple linear regression to the individual data, using trait value as the dependent variable (y) and all adaptive loci as independent ones (x_i ; $i = 1$ to k significant markers), coded as binary scores. In multiple linear regression, a response variable is predicted on the basis of an assumed linear relationship with several independent variables. The combined explanatory value of the x_i variables for predicting y is assessed with the multiple correlation coefficient R , and R^2 is known as the coefficient of determination or the squared multiple correlation (63). The R^2 from the fitted multiple regression linking the phenotypic trait with the set of adaptive loci can therefore be properly interpreted as an estimate of the proportion of phenotypic variance in the trait that is jointly explained by the additive, linear effects of adaptive loci scores. In other words, the regression R^2 provides an estimate of the degree of genetic determination, heritability in the broad sense, or genetic component

of variance, of the focal phenotypic trait. The overall statistical significance of the multiple linear regression can be determined using an ordinary F test.

Three aspects must be noted at this point. First, it is extremely unlikely that most or even a sizable fraction of the genomic regions influencing a given phenotypic trait (or closely linked to such regions) are hit in a genomic screening involving only several hundred markers. Estimates of the degree of genetic determination obtained with this method are thus expected to be biased downwards, because there will always be an unknown number of genomic regions whose effects on the phenotype are left out from the regression, and because adding independent variables x_i to a linear model always increases, never decreases, the value of R^2 (63).

Second, it follows from linear models' properties that, unless all the x 's are mutually orthogonal, the R^2 of the multiple regression cannot be partitioned into k components each uniquely attributable to an x_i (63). In the present context, this would imply that the allelic states of adaptive loci entering the model are uncorrelated (i.e., completely unlinked). If this condition is proven for the particular set of individuals sampled, then the contributions of individual loci to phenotypic variance could be inferred from the multiple regression, for example, by using some stepwise fitting procedure. Alternatively, if the x 's are statistically non-independent (i.e., in linkage disequilibrium), this property of linear models leads one to predict that the regression R^2 will be robust to the multicollinearity of the independent variables.

And third, if the number of individuals sampled is about the same order of magnitude than the number k of adaptive loci included as independent variables in the regression, then it is possible to obtain a large value of R^2 that is not meaningful. In this case, some x 's that do not contribute independently to predicting y may appear to do so, and the regression may provide a spurious estimate of the degree of genetic determination. To correct at least in part for this tendency, the adjusted R^2 should be used instead of the ordinary uncorrected R^2 (63). An additional way to circumvent the problem of a large number of adaptive loci relative to sample size is to reduce the dimensionality of the (adaptive) multilocus genotype of individuals. This can be achieved by obtaining individual scores on the first few axes of a principal coordinates analysis of the pairwise genetic distance matrix (e.g., using the program *genalex*; (64)), and then using these scores as independent variables in the multiple regression rather than the original binary scores of adaptive loci. Reducing the dimensionality of the multilocus genotypes of individuals based on adaptive loci can also be useful for other purposes, as shown in the next section.

**3.5. Possible
Enhancements:
Environmental
and Genotype ×
Environment
Components of
Phenotypic Variance**

The multiple regression method described in Subheading 3.4 will estimate the genetic component of phenotypic variance using exclusively information on the phenotypes and marker-defined genotypes of sampled individuals. But some study systems allow gathering information on features of the environment naturally experienced by the individuals sampled. In these cases, some enhancements of the method are possible which permit to evaluate also the environmental and genotype × environment components of phenotypic variance. Consider, for example, a wild-growing plant population where individuals differ in the microclimate (e.g., temperature, humidity) and soil physicochemical properties of growing sites. Because of the individuals' phenotypic plasticity (the ability of a given genotype to produce different phenotypes when exposed to different environments; (65)), environmental heterogeneity of the individuals' growing sites will generate a certain amount of phenotypic variance, which clearly corresponds to an environmental variance component. In addition, a genotype × environment variance component will also arise whenever individual plants (genotypes) respond differently to variations in environmental factors, such as temperature or soil fertility, because of differences in reaction norms (65).

The proportions of genetic, environmental, and genotype × environment components of phenotypic variance can be simultaneously assessed if information on the environment experienced by individuals is available in addition to their phenotypes and AFLP fingerprints. This can be achieved by fitting a random effect linear model to the individual data, using the focal phenotypic trait as the response variable, and genotype (as described by the set of adaptive loci), environment (as described by biologically relevant parameters), and genotype × environment as independent variables, all treated as random effects. By using a restricted maximum likelihood method (REML), estimates of the genetic, environmental, and genotype × environment components of phenotypic variance of sampled individuals under natural field conditions can be obtained from this model. Computations can be performed using the procedure MIXED in the SAS statistical package, specifying a VC or "variance components" type of covariance structure, which assumes a different variance component for each random effect in the model.

Unless the number of individuals sampled is really large, which is an unlikely situation in most field studies, the dimensionality of the genetic and environmental data should be reduced prior to fitting the mixed model, since otherwise a large number of variance components would be non-estimable. This would result from the large number of model parameters that should be estimated if separate genotype × environment effects are modeled for each adaptive locus. Dimensionality of the multilocus genotype of individuals

can be reduced by obtaining individual scores on the first few axes of a principal coordinates analysis on the genetic distance matrix, as described in Subheading 3.4 above. The method for reducing the dimensionality of environmental variables will depend on the number and nature (discrete vs. continuous variables) of the environmental parameters associated with each individual. For sets of quantitative environmental data (e.g., microclimate parameters), scores on the first axes from a principal component analysis could be used as descriptors of individual environments. For sets of qualitative environmental data (e.g., compass orientation, substrate type), individual scores on the axes from a nonmetric multidimensional scaling ordination could provide reduced-dimensionality descriptors of individual environments (see ref. 66 for a review of multivariate applications to genetic marker data).

4. Application to a Case Study

The methods described above were applied by Herrera and Bazaga (31) to investigate the genetic basis of individual variation in long-term cumulative fecundity in a sample of 52 wild-growing individuals of the long-lived Spanish violet *Viola cazorlensis* monitored for 20 years. Plants were fingerprinted using eight *EcoRI* + 3/*MseI* + 3 and eight *PstI* + 2/*MseI* + 3 primer combinations, which yielded a total of 365 polymorphic AFLP markers. The same plants were also characterized for quantitative floral traits including length of the floral peduncle, floral spur, and upper, middle, and lower petal blades, and an earlier study revealed significant phenotypic selection on most of these floral features (67). Application of the genomic scan method to evaluate the genetic basis of individual variation in these traits will provide insights on the possibility of evolutionary change in floral characteristics in that population.

Results are shown in Table 1, where the genetic determination estimate for 20-year cumulative fecundity for the same individuals is also included (31). Between 2 and 4 AFLP loci were significantly associated with individual variation in floral traits, and multiple regressions between floral traits and their associated AFLP loci were all statistically significant, which reveals that individual variation in all floral traits had a significant genetic basis. The number of statistically significant AFLP loci was lower for floral traits than for fecundity. This difference may reflect a real disparity between the two types of phenotypic traits in the number of genomic regions underlying their variation, but could also be due to a lower statistical power for detecting significant loci in the case of floral traits, e.g., because of lower phenotypic variance. The greater phenotypic variability of individual fecundity (coefficient of variation, CV = 138%) in comparison to floral traits (CV range = 8–23%) is

Table 1
Application of the genomic scan and Ritland methods for evaluating the genetic component of phenotypic trait variance in a wild population of the violet *Viola cazorlensis*

Phenotypic trait	Genomic scan method (regression between trait and significant loci)				
	Significant AFLP loci	R^2	F -value	p -Value	Ritland method (h^2) ^a
Floral peduncle	2	0.383	16.85	<0.0001	-0.042 (-0.416)
Floral spur	4	0.353	7.96	<0.0001	0.064 (-0.204)
Upper petal	3	0.279	7.59	0.0003	0.178 (-0.380)
Middle petal	2	0.190	6.99	0.0021	0.102 (-0.477)
Lower petal	2	0.199	7.33	0.0016	0.110 (-0.470)
20-year fecundity	11	0.631	7.00	< 0.0001	0.471 (-0.076)

^aThe lower 5% quantile of the bootstrap distribution is given in parentheses. For the heritability value to be significant at $p < 0.05$, this value must be larger than zero (26)

compatible with the latter interpretation. Anyway, these results for *V. cazorlensis* suggest that even modest numbers of significant AFLP loci can be sufficient to estimate the genetic component of phenotypic variance. Furthermore, the demonstration that individual variation in floral traits has a genetic basis provides the necessary complement to earlier results showing phenotypic selection on these traits (67), and suggests that ongoing selection is likely to promote evolutionary change of the average floral phenotype at the study population.

Estimates of the genetic component of phenotypic variation obtained with the genomic scan and Ritland methods are not strictly comparable, because they correspond to two different concepts of heritability as noted earlier. Keeping this caveat in mind, it is still interesting to compare Ritland h^2 and genomic scan-based R^2 figures for the same dataset and traits, to see whether the general conclusions reached by the two methods are similar. Ritland h^2 estimates for the six traits studied in the *Viola cazorlensis* dataset obtained with the program mark (68) are shown in Table 1. Estimates vary widely among traits, but none of them reaches statistical significance (i.e., $h^2 > 0$) at the 0.05 level. Ritland h^2 and genomic-scan R^2 values are correlated across traits, although the relationship does not reach statistical significance ($r = 0.655$, $N = 6$, $p = 0.16$). For this dataset, therefore, Ritland and genomic scan methods provide contrasting conclusions regarding the genetic component of phenotypic variation. While Ritland estimates would point to the absence of a significant additive genetic component

underlying the variation of floral traits, the genomic scan method does provide evidence of a significant genetic component in all cases, as generally found in other species (7). In addition, genomic-scan R^2 values for *Viola cazorlensis* in Table 1 are comparable to broad-sense heritability estimates obtained for morphological and life history traits in other species of *Viola* using quantitative genetics methods (69).

5. Note

1. The possibility of obtaining by chance alone an unknown number of false significant phenotype-markers logistic regressions (false positives or type I errors) could be controlled by using some simple correction to the p -value threshold required for significance, such as Bonferroni-style corrections (70, 71). Nevertheless, Bonferroni-style corrections can be overly conservative, and the more-sophisticated approach proposed by Storey and Tibshirani (72) is better suited to genome-wide studies, where the number of simultaneous tests can be very large. In this method, a measure of statistical significance called the q -value is associated with each separate test. The q -value is similar to the well-known p -value, except it is a measure of significance in terms of the false-discovery rate rather than the false-positive rate. In the present context, the q -value obtained for a given marker is the expected proportion of false positives incurred for the whole set of regressions when the logistic regression for that particular marker is considered as statistically significant. Using the `qvalue` package (72) for the *R* environment (73), q -values for all the marker–trait regressions can be computed, ranked, and the largest q -value leading to an expectation of less than one falsely significant regression [i.e., $q\text{-value} \times (\text{number of regressions accepted as significant}) < 1$] used as the threshold for determining the statistical significance of marker–trait associations. In this way, the statistical power for detecting as many significant regressions as possible is enhanced while keeping with the conservative constraint of avoiding any false positive.

Acknowledgments

I am grateful to Pilar Bazaga for her invaluable contribution to the development of the method described here; Antonio Castilla for discussion and comments on the manuscript; and Aurélie Bonin

and François Pompanon for the opportunity to contribute to this volume. Work supported by grant CGL2006-01355 (Ministerio de Educación y Ciencia, Gobierno de España).

References

1. Endler JA (1986) Natural selection in the wild. Princeton University Press, Princeton
2. Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution* 37:1210–1226
3. Arnold SJ, Wade MJ (1984) On the measurement of natural and sexual selection: theory. *Evolution* 38:709–719
4. Arnold SJ, Wade MJ (1984) On the measurement of natural and sexual selection: applications. *Evolution* 38:720–734
5. Manly BF (1985) The statistics of natural selection. Chapman and Hall, London
6. Kingsolver JG, Hoekstra HE, Hoekstra JM et al (2001) The strength of phenotypic selection in natural populations. *Am Nat* 157:245–261
7. Geber MA, Griffen LR (2003) Inheritance and natural selection on functional traits. *Int J Plant Sci* 164:S21–S42
8. Kingsolver JG, Pfennig DW (2007) Patterns and power of phenotypic selection in nature. *Bioscience* 57:561–572
9. Fisher RA (1958) The genetical theory of natural selection, 2nd edn. Dover, New York
10. Andersson S (1996) Floral variation in *Saxifraga granulata*: phenotypic selection, quantitative genetics and predicted response to selection. *Heredity* 77:217–223
11. Campbell DR (1996) Evolution of floral traits in a hermaphroditic plant: field measurements of heritabilities and genetic correlations. *Evolution* 50:1442–1453
12. Conner JK (1997) Floral evolution in wild radish: the roles of pollinators, natural selection, and genetic correlations among traits. *Int J Plant Sci* 158:S108–S120
13. Moore AJ, Kulk PF (2002) Quantitative genetic analysis of natural populations. *Nat Rev Genet* 3:971–978
14. Falconer DS, MacKay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Essex
15. Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland
16. Garant D, Kruuk LEB, McCleery RH et al (2004) Evolution in a changing environment: a case study with great tit fledging mass. *Am Nat* 164:E115–E129
17. Roff DA (1997) Evolutionary quantitative genetics. Chapman and Hall, New York
18. Ritland K (1996) A marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* 50:1062–1073
19. Garant D, Kruuk LEB (2005) How to use molecular marker data to measure evolutionary parameters in wild populations. *Mol Ecol* 14:1843–1859
20. Fernández J, Toro MA (2006) A new method to estimate relatedness from molecular markers. *Mol Ecol* 15:1657–1667
21. Pemberton JM (2008) Wild pedigrees: the way forward. *Proc R Soc B* 275:613–621
22. Ritland K (2000) Marker-inferred relatedness as a tool for detecting heritability in nature. *Mol Ecol* 9:1195–1204
23. Ritland K (2000) Detecting inheritance with inferred relatedness in nature. In: Mousseau TA, Sinervo B, Endler J (eds) Adaptive genetic variation in the wild. Oxford University Press, Oxford
24. Ritland K, Ritland C (1996) Inferences about quantitative inheritance based on natural population structure in the yellow monkeyflower, *Mimulus guttatus*. *Evolution* 50: 1074–1082
25. Klaper R, Ritland K, Mousseau TA, Hunter MD (2001) Heritability of phenolics in *Quercus laevis* inferred using molecular markers. *J Hered* 92:421–426
26. van Kleunen M, Ritland K (2005) Estimating heritabilities and genetic correlations with marker-based methods: an experimental test in *Mimulus guttatus*. *J Hered* 96:368–375
27. 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
28. Thomas SC, Coltman DW, Pemberton JM (2002) The use of marker-based relationship information to estimate the heritability of body weight in a natural population: a cautionary tale. *J Evol Biol* 15:92–99
29. Wilson AJ, McDonald G, Moghadam HK et al (2003) Marker-assisted estimation of quantitative genetic parameters in rainbow trout, *Oncorhynchus mykiss*. *Genet Res* 81:145–156
30. Rodríguez-Ramilo ST, Toro MA, Caballero A, Fernández J (2007) The accuracy of a heritabil-

- ity estimator using molecular information. *Conserv Genet* 8:1189–1198
31. Herrera CM, Bazaga P (2009) Quantifying the genetic component of phenotypic variation in unpedigreed wild plants: tailoring genomic scan for within-population use. *Mol Ecol* 18: 2602–2614
 32. Hedrick PW (2005) *Genetics of populations*, 3rd edn. Jones and Bartlett, London
 33. Mackay TFC (2001) The genetic architecture of quantitative traits. *Ann Rev Gen* 35:303–339
 34. Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature* 452:169–175
 35. Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100: 158–170
 36. Vos P, Hogers R, Bleeker M et al (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407–4414
 37. Van Eck HJ, Van der Voort JR, Draaistra J et al (1995) The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. *Mol Breeding* 1:397–410
 38. Castiglioni P, Ajmone-Marsan P, van Wijk R et al (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. *Theor Appl Genet* 99:425–431
 39. Peng J, Korol AB, Fahima T et al (2000) Molecular genetic maps in wild emmer wheat, *Triticum dicoccoides*: genome-wide coverage, massive negative interference, and putative quasi-linkage. *Genome Res* 10:1509–1531
 40. Rogers SM, Isabel N, Bernatchez L (2007) Linkage maps of the dwarf and normal lake whitefish (*Coregonus clupeaformis*) species complex and their hybrids reveal the genetic architecture of population divergence. *Genetics* 175:375–398
 41. Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Mol Ecol* 16:3737–3758
 42. Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends Plant Sci* 12:106–117
 43. Raman H, Moroni JS, Sato K et al (2002) Identification of AFLP and microsatellite markers linked with an aluminium tolerance gene in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 105:458–464
 44. 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. *Theor Appl Genet* 109:1426–1433
 45. Yuan L, Dussle CM, Muminovic J et al (2004) Targeted BSA mapping of *Scmv1* and *Scmv2* conferring resistance to SCMV using *PstI/MseI* compared with *EcoRI/MseI* AFLP markers. *Plant Breeding* 123:434–437
 46. 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. *Ann Bot* 100:1039–1051
 47. Cameron ND, van Eijk MJT, Brugmans B et al (2003) Discrimination between selected lines of pigs using AFLP markers. *Heredity* 91:494–501
 48. Jump AS, Peñuelas J, Rico L et al (2008) Simulated climate change provokes rapid genetic change in the Mediterranean shrub *Fumana thymifolia*. *Global Change Biol* 14:637–643
 49. Freeland JR, Biss P, Conrad KF, Silvertown J (2010) Selection pressures have caused genome-wide population differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow. *J Evol Biol* 23:776–782
 50. 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
 51. Herrera CM, Bazaga P (2008) Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. *Mol Ecol* 17:5378–5390
 52. Campbell D, Bernatchez L (2004) Genomic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol Biol Evol* 21:945–956
 53. Scotti-Saintagne C, Mariette S, Porth I et al (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* 168:1615–1626
 54. 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. *Theor Appl Genet* 99:921–935
 55. Ojha BR (2005) Comparison of Eco-Mse and Pst-Mse primer combinations in generating AFLP map of tomato. *J Inst Agric Anim Sci* 26:27–35
 56. Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. *Lancet* 361:598–604
 57. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959
 58. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals

- using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
59. Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7:574–578
60. Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* 163:367–374
61. Corander J, Marttinen P (2006) Bayesian identification of admixture events using multilocus molecular markers. *Mol Ecol* 15:2833–2843
62. Agresti A (2007) An introduction to categorical data analysis, 2nd edn. Wiley, Hoboken
63. Rencher AC, Schaalje GB (2008) Linear models in statistics, 2nd edn. Wiley, Hoboken
64. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295
65. Pigliucci M (2001) Phenotypic plasticity: beyond nature and nurture. Johns Hopkins University Press, London
66. Jombart T, Pontier D, Dufour AB (2009) Genetic markers in the playground of multivariate analysis. *Heredity* 102:330–341
67. Herrera CM (1993) Selection on floral morphology and environmental determinants of fecundity in a hawk moth-pollinated violet. *Ecol Monogr* 63:251–275
68. Ritland K (2005) Multilocus estimation of pairwise relatedness with dominant markers. *Mol Ecol* 14:3157–3165
69. Antlfinger AE, Curtis WF, Solbrig OT (1985) Environmental and genetic determinants of plant size in *Viola sororia*. *Evolution* 39:1053–1064
70. Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43:223–225
71. Narum SR (2006) Beyond Bonferroni: less conservative analyses for conservation genetics. *Conserv Genet* 7:783–787
72. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100:9440–9445
73. R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna