Spatial autocorrelation of allozyme and quantitative markers within a natural population of *Centaurea jacea* (Asteraceae)

O. J. HARDY, S. VANDERHOEVEN, P. MEERTS & X. VEKEMANS

Laboratoire de Génétique et Ecologie Végétales, Université Libre de Bruxelles, Chaussée de Wavre 1850, 1160 Bruxelles, Belgium

Keywords:

allozymes;
Centaurea;
genetic drift;
Moran's *I*;
quantitative genetics;
selection;
simulation;
spatial autocorrelation.

Abstract

This paper compares the fine-scale genetic structure of quantitative traits and allozyme markers within a natural population of *Centaurea jacea* s.l. To that end, a spatial autocorrelation approach is developed based on pairwise correlation coefficients between individuals and using sib families. Statistical properties of the proposed statistics are investigated with numerical simulations. Our results show that most quantitative traits have a significant spatial structure for their genetic component. On average, allozyme markers and the genetic component of quantitative traits have similar patterns of spatial autocorrelation that are consistent with a neutral model of isolation by distance. We also show evidence that environmental heterogeneity generates a spatial structure for the environmental component of quantitative traits. Results are discussed in terms of mechanisms generating spatial structure and are compared with those obtained on a large geographical scale.

Introduction

Spatial genetic structure can result from two major evolutionary mechanisms, i.e. random genetic drift and natural selection. The relative importance of natural selection and genetic drift in shaping the pattern of spatial genetic structure is a central question in population genetics. One way to investigate the effect of selection relative to drift for a given character in natural populations is to compare its genetic structure with that of neutral genes (polymorphism at molecular markers, including allozymes, is often recognized as neutral or near neutrality). One advantage of this approach is that estimates of demographic parameters and heritability are unnecessary (Spitze, 1993), contrary to many other tests for neutral evolution (e.g. Lande, 1977). There has been some debate about the validity of comparing polygenic characters with monogenic markers (e.g. Lewontin, 1984). However, theoretical studies have shown that, under a neutral hypothesis, the number of loci controlling a character does not affect the degree of genetic structure (e.g. Rogers & Harpending, 1983; Lande, 1991, 1992). Hence, comparison of genetic structure between monogenic (molecular markers) and polygenic characters (quantitative ones) is meaningful (Rogers, 1986; Felsenstein, 1986). To assess broad-scale differentiation among populations, the genetic variance can be partitioned within and among populations, using *F*-statistics. Estimators of F_{ST} adapted to quantitative characters have been developed and used to compare molecular and quantitative markers in natural populations (Prout & Barker, 1993; Spitze, 1993; Long & Singh, 1995; Podolsky & Holtsford, 1995; Bonnin *et al.*, 1996; Yang *et al.*, 1996; Kremer *et al.*, 1997; Kuittinen *et al.*, 1997; Lynch *et al.*, 1999). The overall trend revealed by those studies is that quantitative traits often show a higher level of population differentiation than molecular markers, suggesting that diversifying selection is commonly operating on the former, at least on a broad-scale (i.e. among populations).

How important the effect of selection is in shaping the genetic structure of quantitative traits at a small-scale (i.e. within populations) is still little known. In a recent review, Linhart & Grant (1996) argue that natural selection is an important force shaping the fine-scale differentiation within natural plant populations. They show evidence from several studies that local adaptation

Correspondence: Dr Olivier J. Hardy, Laboratoire de Génétique et Ecologie Végétales, Université Libre de Bruxelles, Chaussée de Wavre 1850, 1160 Bruxelles, Belgium.
Tel.: +32 2675 96 03; fax: +32 2672 02 84; e-mail: ohardy@ulb.ac.be

to different environmental factors (e.g. soil characteristics) can occur on a scale of a few metres, or even centimetres. In most studies reviewed by Linhart & Grant (1996) detection of selection depended on the prior identification of the environmental factors of potential adaptive significance. One advantage of the methods involving comparison of differentiation at quantitative traits and neutral genetic markers is that no *a priori* knowledge of the factors that may drive selection is required. However, such comparisons are still lacking at a microgeographical scale (but see Bonnin *et al.*, 1996).

Estimators of F_{ST} may not be ideal to assess the fine-scale genetic structure of a continuous population because of the need to define arbitrarily discrete subpopulations and the risk of losing information about processes occurring within subpopulations. An alternative is to assess the genetic structure using spatial autocorrelation techniques that describe genetic relatedness between pairs of individuals according to distance. Those techniques have sometimes been criticised on the ground that they lack a sound foundation in population genetics theory. However, one of those, Moran's *I* statistic, can be used to estimate meaningful parameters from the theory of population genetics because it can be related to probabilities of identity of pairs of genes (Hardy & Vekemans, 1999). Using a spatial autocorrelation approach, Argyres & Schmitt (1991) demonstrated a significant microgeographical structure of genetic origin for a set of quantitative traits in a natural plant population. In their study, it was, however, not possible to test whether the observed genetic structure resulted from local adaptation in a heterogeneous environment or local random genetic drift.

In this paper we use a spatial autocorrelation method to assess the genetic structure of a set of allozyme markers and a set of quantitative traits (morphological and phenological ones) within a small plant population of knapweeds. Our main objective is to compare the genetic structure observed for allozyme markers and quantitative traits when the latter are measured both *in situ* and in controlled conditions. More specifically, we address the following questions: (1) Is the pattern of spatial structure of allozyme markers consistent with an isolation by distance process? (2) Is there evidence of a spatial structure of the environmental component of the quantitative traits? (3) Does the spatial structure of the genetic component of quantitative traits differ from that of allozyme markers, suggesting different evolutionary forces?

We show that useful statistics to compare genetic markers and quantitative characters using natural sib families are based on the ratio of correlation coefficients between pairs of individuals belonging to different families over an intraclass correlation coefficient (see Appendix). Those correlation coefficients can be plotted as a function of the spatial distance to characterize the spatial genetic structure (Rousset, 2000). We use numerical simulations to check that these statistics have identical expectations for genetic markers and quantitative traits under a neutral hypothesis, and to compare their stochastic variances. We show that the microgeographical differentiation of quantitative traits in our study was consistent with neutral expectations, indicating the absence of selection or that the balance selection – gene flow is in favour of the latter. Finally, we discuss the power of the approach to detect selection.

Materials and methods

Study organism and experimental setup

Centaurea jacea L. *sensu lato* (Asteraceae) is an extremely polymorphic complex whose taxonomic treatment is still controversial (e.g. Gardou, 1972). In Belgium, this comprises several microspecies, with the microspecies *C. thuillieri* (Dostál) J. Duvigneaud and Lambinon being most widespread (Lambinon *et al.*, 1992). Only tetraploid *C. thuillieri* plants have been found in Belgium. They show a tetrasomic pattern of allozyme segregation (Hardy, 2000; Hardy *et al.*, 2000) indicating an autopolyploid origins. The high polymorphism of quantitative traits that occurs in tetraploid knapweeds from western Europe (Gardou, 1972; Hardy, 2000) makes them suitable for a spatial structure study. It is a self-incompatible species and, as is typically observed with outcrossing species, much of the genetic variability is expressed within populations (Hamrick & Godt, 1989).

In a preliminary survey made in Belgium on four natural populations of knapweeds belonging to the microspecies *C. thuillieri*, we found that most quantitative traits measured *in situ* showed substantial spatial structure (Hardy, 2000). One population showed particularly high levels of spatial structure at quantitative traits. That population was therefore selected to test whether this structure reflected the environmental or the genetic components of the quantitative traits and, in the case of the genetic component, if it was influenced by selection. The population comprises 300 individuals on a 200-m^2 area of a grassland which is not subjected to grazing or mowing in the course of the growing season. Fifty-two individuals were randomly selected and their position was recorded on an orthogonal grid map. Seeds were collected on each of these 52 plants in August 1996 and were sown into pots. A total of 320 individuals forming 52 maternal families were thus obtained. Two months after emergence, the seedlings were transplanted in the experimental garden in a single randomised block with 45-cm spacing between individuals. Hereafter, the parental, natural population will be referred to as the 'field population', whereas the experimental population derived from seeds will be referred to as the 'garden population'.

Assessment of quantitative traits

In the field population, 17 traits were scored *in situ* in each of the 52 plants used as mothers for the garden population. In the latter, a total of 23 traits, including the 17 previous ones, were scored in each of the 320 individuals (Table 1). All but four of those traits are quantitative; the four discrete ones (DEV4, VIG14, PAP, RAY) take the form of ordinal classes and were therefore also treated as quantitative. The quantitative traits included morphological measurements of reproductive parts (size and shape of flowering heads and involucral bracts) and vegetative parts (size and shape of leaves; vegetative vigour); age at flowering was also scored. In the garden population, five characters were scored on 4-month-old plants and all others on adults (i.e. 12- to 16-month-old plants).

Where necessary, logarithmic transformation or Kleckowski's transformation (Lynch & Walsh, 1998) was applied on quantitative variables to normalize distribution and stabilize variances (Table 1). As several characters were strongly correlated, principal component analyses (PCA) were applied on data from the field and the garden populations to obtain a set of independent quantitative variables. Two nominal variables (PAP, RAY) with a low number of classes (fewer than five) were excluded from the PCA. Only those principal components with eigenvalues greater than unity were considered in subsequent analyses.

Table 1 Morphological and phenological characters. The T-leaf is the one below the last top ramification of the stem.

Character	Abbreviation
1. Capitulum diameter*	CD
2. Ratio of capitulum height to capitulum diameter*†	RCHCD
3. Bract length*	BL
4. Ratio of bract length to bract width*†	RBLBW
5. Width of central undivided part of the bract*†	WCPB
6. Number of bract teeth*‡	NBT
7. Non-serrated bract percentage*‡	NSBP
8. Development state at 4 months (six ordinal classes)	DEV4
9. Plant height at 4 months	PH4
10. Number of stems at 4 months	NST4
11. Number of flowering heads on main stem at 4 months*†	NCST4
12. Rosette diameter at 4 months	RD4
13. Rosette diameter at 12 months	RD12
14. First flowering date (second year)*	FF
15. Plant height at 14 months*	PH14
16. Plant vigour at 14 months (seven ordinal classes)	VIG14
17. T leaf length*†	LL
18. Ratio of T leaf length to T leaf width*†	RLLLW
19. Number of lobes of basal leaf*	NLO
20. Length of lobes of basal leaf*‡	LLO
21. Ratio of length to width of lobes of basal leaf*‡	RLLOWLO
22. Pappus development (four ordinal classes)*	PAP
23. Ray florets (two classes)*	RAY

* Characters assessed on both field and garden population.
†Logarithmic transformation, ‡ Kleckowski's transformation.

The experimental design does not allow us to asses the influence of maternal effects. However, maternal effects are not likely to strongly affect reproductive characters at the adult stage. Characters related to vegetative vigour are more prone to be influenced by maternal effects. The importance of maternal effects usually decreases with age as shown in another species of *Centaurea* (Weiner *et al*., 1997).

Allozyme markers

To extract proteins, 300 mg of fresh leaves were ground in a mortar with 1 mL of the following extraction buffer: 0.001 M EDTA, 0.01 M NaCl, 0.001 M $MgCl_2$, 4% PVP, and 0.1% 2-Mercaptoethanol in 0.1 M Tris-HCl at pH 7.5. The extract was centrifuged at 15 000*g* and the supernatant was stored at –75° before analysis. Electrophoreses were carried out on vertical polyacrylamide gels (8%) at 300 V for 4–6 h depending on the enzyme system. Fifteen enzymatic systems were tested using standard staining recipes (Hillis *et al*., 1996). Four systems were selected because they were polymorphic and showed interpretable banding patterns: 6-phosphogluconate dehydrogenase (PGD, EC 1.1.1.44), leucine aminopeptidase (LAP, EC 3.4.11.1), diaphorase (DIA, EC 1.8.1.4) and α and β esterases (EST). The enzymatic systems LAP, DIA and EST were monomeric and, accordingly, each band could be ascribed to one allele. PGD was dimeric so that intermediate bands corresponding to heterodimers were detected in heterozygotes. For all enzymes the relative intensity of bands was taken into account in the allelic interpretation of the tetraploid genotypes (Hardy *et al*., 2000).

Allozyme electrophoresis was carried out on the 320 individuals of the garden population, as well as on 81 individuals from the field population (for which locus DIA was not scored). Inbreeding coefficient (F_{IS}) was estimated for the field population. For autotetraploids, F_{IS} may have different expectations among loci because of the possible phenomenon of double reduction (Ronfort *et al*., 1998), and a multilocus estimate may therefore be meaningless. F_{IS} values were thus computed for each locus separately as the average correlation coefficient between all possible pairs of alleles belonging to the same individual. Multi-allelic estimates were obtained by weighting the contribution of each allele by $p(1 - p)$, where p is the allele frequency. Tests of significance of F_{IS} values were obtained by random permutations of the genes among individuals.

Spatial autocorrelation analysis

Microgeographical structure was assessed for quantitative characters and allozyme markers using spatial autocor-

relation analysis. The latter uses correlation coefficients between individuals to express similarity as a function of the spatial distance. It thus provides more information on the pattern of spatial structure than single F_{ST} estimates. Under an isolation by distance process in a two-dimensional space, pairwise correlation coefficients between individuals for a genetic trait are expected to decrease in a roughly linear fashion with the logarithm of the geographical distance (Maruyama, 1977; Hardy & Vekemans, 1999; Rousset, 1997, 2000). Hence, applying a linear regression, the negative of the corresponding slope provides a simple estimate of the magnitude of spatial structure. Such slope can also be used to estimate $4\pi D\sigma^2$, where D is the population density and σ^2 the variance of the distances of gene dispersal, which can also be interpreted as a measure of Wright's neighbourhood size (Rousset, 1997, 2000; Hardy & Vekemans, 1999).

Computations involved the construction of a matrix of geographical distances between individuals and, for each variable x and each pair of individuals i and j, a matrix of pairwise correlation coefficients. The latter were calculated as $[(x_i - \bar{x})(x_j - \bar{x})/Var(x)] + 1/(n-1)$, where $\bar{x}$ and $Var(x)$ are the estimated population average and variance, respectively, and n is the total number of individuals. The second term in this equation reduces bias of the estimator (Hardy & Vekemans, 1999). For quantitative traits, x values are the (transformed) measurements obtained for each trait, or the individual scores of each principal component provided by the PCA. For allozymes, x values are the individual genotypic values (i.e. the number of copies of a particular allele in an individual) obtained for each allele (Dewey & Heywood, 1988) such that the pairwise correlation coefficients provide estimators of Wright's coefficient of relationship, which has the advantage of being unaffected by the phenomenon of double reduction in autopolyploids (Ronfort *et al.*, 1998; Hardy & Vekemans, 1999). From the distance and correlation matrices, a linear regression of the pairwise correlation coefficients on the logarithm of the distance was performed, considering all possible pairs of individuals belonging to different families. The negative of the slope of that regression line ($-b$) provides a simple estimate of the *magnitude* of genetic structure under an isolation by distance model in two dimensions (Rousset, 1997, 2000; Hardy & Vekemans, 1999). We tested the significance of $-b$ estimates by comparing the observed values with those obtained for 1000 random permutations of families (or individuals in the case of the field population) among positions. For graphical representation, average correlation coefficients were computed over all pairs of individuals separated by given distance classes (distance classes in metres: [0,1], [1,2], [2,3], [3,4], [4,5], [5,8], [8,12], [12,20]). Those average correlation coefficients are equivalent to Moran's I statistics and will thus be referred to as $I(d)$. All calculations were performed using a computer program developed by the authors and which is available on request to O. Hardy (ohardy@ulb.ac.be).

Spatial autocorrelation analysis was performed for the field and garden populations using either individual characters, principal components or the genotypic values of each allozyme. For allozymes, in order to obtain multiallelic averages for each locus and multilocus averages, the contribution of each allele was weighted by $p(1-p)$, where p is the allele frequency. In the analysis of the garden population, the geographical distance between families is the one between the respective mother individuals in the field population.

As shown in the Appendix, codominant genetic markers and quantitative characters can be compared on the basis of the ratio of the correlation coefficients between individuals belonging to different families over the intraclass (i.e. within family) correlation coefficient, t, if we assume additive effects. Therefore, from the 52 maternal families of the garden population, intraclass correlation coefficients (t) were computed for each quantitative variable and for the genotypic values of each allele, using a one-way ANOVA (Lynch & Walsh, 1998). The significance of t values were tested by random permutations of the individuals among families. When comparing the spatial structure between quantitative characters and allozyme markers in the garden population using ratios $I(d)/t$ and $-b/t$, we considered only quantitative characters with a t value significantly greater than zero. Quantitative characters with a t value not significantly different from zero have no detectable genetic component and were therefore useless for an analysis of genetic structure.

Simulations

We attempted to model a population with similar characteristics as the one studied and conducted numerical simulations (a) to check the validity of the statistics used to compare genetic markers and quantitative traits and (b) to test if the stochastic variation of the genetic structure observed among allozyme loci and among quantitative characters is compatible with a neutral hypothesis. The simulation algorithm is similar to that described in Hardy & Vekemans (1999) but adapted to the demographic properties of the field population and the sampling scheme we used to create the garden population. Basically, we simulated a lattice model with 25×12 hermaphrodite individuals. Field observations suggest that around 20% of the individuals die every year (O. Hardy, personal observations), and hence we assigned to each individual in the population a probability of 0.2 of being replaced by a new one. As no mechanism seems to promote efficient seed dispersal in knapweeds (pappus is absent or weakly developed), seed dispersal was allowed only to five positions – the one of

the mother plant and the four adjacent, as in rook's moves – with equal probability. Pollen dispersal was assumed to follow a two-dimensional normal distribution and, as knapweeds are self-incompatible (Gardou, 1972), self-fertilization was prohibited. The variance of pollen dispersal was adjusted so that the genetic structure obtained after 100 simulation years (the time needed to reach a quasi equilibrium state) leads to similar *I(d)* values and *b* value as the ones observed for allozymes. The simulated sampling scheme considered six progenies obtained on 50 mother individuals located along two parallel transects. A genetic marker was simulated for one diallelic locus with an initial allele frequency equal to 0.4. A quantitative marker was simulated as the sum of two components: a genetic component which results from the total additive effects of 30 diallelic loci with initial allele frequencies chosen at random, and a spatially nonstructured environmental component represented by a stochastic variable following a centred normal distribution. The heritability of the quantitative marker is controlled by adjusting the variance of the environmental component. Data analysis of the 50 simulated families were identical to those described for our garden population of knapweeds. We performed 1000 replicates of this simulation in order to obtain the distribution of $-b/t$ values, and to estimate the 95% confidence intervals using the 25th and 975th ordered values.

Table 2 Statistics estimated for alleles from four loci analysed in the garden population: allele frequency (*p*), intraclass correlation coefficient (*t*), and the negative of the slope of the regression line of pairwise correlation coefficients on the logarithm of distance (*–b*).

Locus	Allele	*p*	*t*	*–b*
PGD	*a*	0.025	0.077*	0.0401*
	b	0.446	0.318*	0.0533*
	c	0.513	0.307*	0.0251*
	d	0.016	0.290*	0.0171*
	weighted mean		0.301*	0.0385*
LAP	*a*	0.025	0.005n.s.	–0.0063n.s.
	b	0.484	0.173*	–0.0013n.s.
	c	0.182	0.174*	0.0298*
	d	0.309	0.206*	–0.0015n.s.
	weighted mean		0.178*	0.0057
DIA	*a*, *b*, and weighted mean	0.436	0.159*	0.0395*
EST	*a*, *b*, and weighted mean	0.384	0.226*	0.0310*
Multilocus weighted mean			0.215*	0.0273*

* Indicates values significantly different from zero ($P < 0.05$) according to permutation test. n.s. not significant.

Results

Spatial structure of allozyme markers

The intraclass correlation coefficients (*t*) for 12 alleles ranged from 0.005 to 0.318 (weighted mean = 0.215), and all but one were significantly larger than zero (Table 2). The observed mean value was close to the one expected for true half-sib families (i.e. 0.25, assuming no inbreeding in the parental generation). Any deviation in the form of a mixture of full and half sibs, related parents or inbreeding in the parental population would increase the expected *t* value. The average inbreeding coefficients in the field population were not significantly different from zero at any locus ($F_{IS} = -0.040$ for PGD, – 0.017 for LAP, and 0.067 for EST).

In the garden population, the shape of the average correlogram *I(d)* over loci (Fig. 1) suggests that the relationship between *I(d)* values and the logarithm of the distance is approximately linear, in accordance with the prediction for an isolation by distance model (Rousset, 1997; Hardy & Vekemans, 1999). Hence, the slopes of the regression given by the *b* values are meaningful.

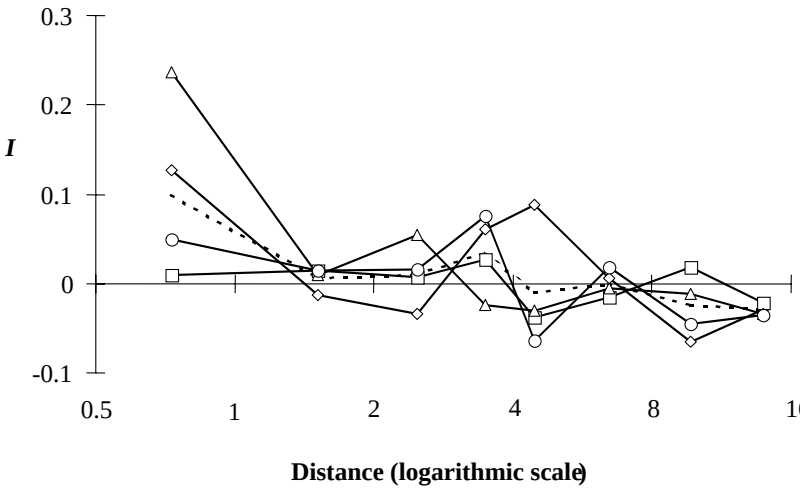

Fig. 1 Average *I(d)* values for each allozyme locus (diamonds: PGD, squares: LAP, triangles: DIA, circles: EST) and average over all loci (dotted line).

Table 3 Statistics estimated for the quantitative characters: intraclass correlation coefficient (t), and the negative of the slopes of the regression lines in the field ($-b_f$) and garden ($-b_g$) populations.

Character	t	$-b_f$	$-b_g$
CD	0.185*	−0.005n.s.	0.041*
RCHCD	0.057*	0.003n.s.	0.010n.s.
BL	0.183*	0.072n.s.	0.026*
RBLBW	0.090*	0.029n.s.	−0.008n.s.
WCPB	0.116*	0.077n.s.	−0.004n.s.
NBT	0.090*	0.231*	0.007n.s.
NSBP	0.104*	0.232*	0.007n.s.
DEV4	0.160*		0.031*
PH4	0.086*		0.014*
NST4	0.066*		0.006n.s.
NCST4	0.035n.s.	0.172*	0.059*
RD4	0.126*		0.045*
RD12	0.031n.s.		0.011n.s.
FF	0.123*	0.165*	0.041*
PH14	0.118*	0.029n.s.	0.005n.s.
VIG14	0.144*		0.018*
LL	0.055n.s.	0.019n.s.	−0.001n.s.
RLLLW	−0.041n.s.	0.166*	−0.000n.s.
NLO	0.201*	0.043n.s.	0.004n.s.
LLO	0.124*	0.106*	0.021*
RLLOWLO	0.118*	0.146*	0.011n.s.
PAP	0.073*	−0.001n.s.	−0.001n.s.
RAY	0.269*	0.059n.s.	0.036*

* Indicates values significantly different from zero ($P < 0.05$), n.s. not significant. Note: the plant height and the number of capitulum per stem assessed in the field on adult plants were compared with variables PH14 and NCST4 assessed in the garden, respectively.

Table 2 shows that nine of 12 alleles had significantly positive $-b$ values (Table 2). The average $-b$ value is 0.027. From this estimate we obtain a measure of the neighbourhood size ($4\pi D\sigma^2$) of about 80 individuals using a method described in Hardy & Vekemans (1999, eq. 15), and assuming no inbreeding.

Table 4 Statistics estimated on the seven principal components from the PCA applied on the quantitative characters from the garden population: per cent of variance represented (%), intraclass correlation coefficient (t), and negative of the slope of the regression line of pairwise correlation coefficients on the logarithm of distance ($-b$).

Principal components	%	t	$-b$
1	21	0.137*	0.0280*
2	13	0.158*	0.0046n.s.
3	9	0.197*	0.0651*
4	8	0.069*	−0.0003n.s.
5	7	0.098*	0.0005n.s.
6	6	0.053*	0.0094*
7	5	0.067*	0.0175*

* Indicates values significantly different from zero ($P < 0.05$) according to a permutation test. n.s. not significant.

Spatial structure of quantitative characters

Nineteen of 23 quantitative characters measured on the garden population had a significantly positive intraclass correlation coefficient, t (Table 3). Thus, assuming no maternal effects, variation of the corresponding characters has a genetic basis. It is worth noting that those characters that are *a priori* more prone to display maternal effects (DEV4, PH4, NST4, NCST4, RD4, RD12, PH14, VIG14) did not show higher t values (Table 3). The PCA applied on the garden population resulted in seven principal components with eigenvalues superior to one, which account for 69% of the total variance. All those components showed significantly positive t values (Table 4). In the field population, six principal components with eigenvalues superior to one explain 74% of the variance.

For brevity, correlograms of individual characters are not presented. They showed a significant spatial structure for different sets of quantitative characters in the field

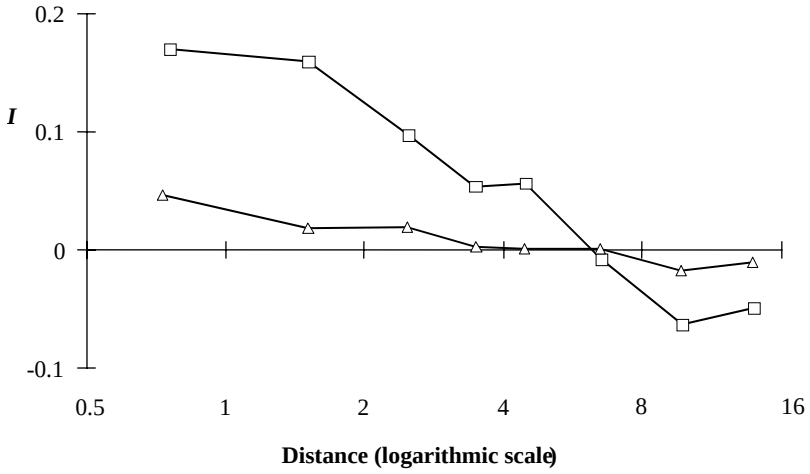

Fig. 2 Average $I(d)$ values for the principal components of the PCA in the natural population (open squares) and the experimental population (open triangles).

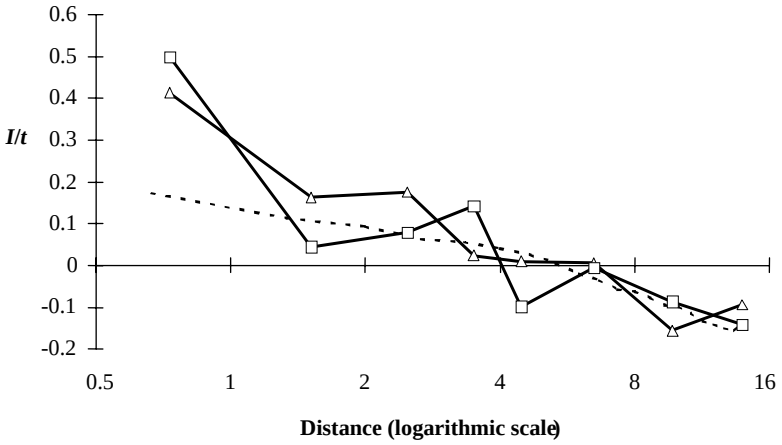

Fig. 3 Average $I(d)/t$ ratios for the allozyme markers (open squares) and the principal components of the PCA applied on the quantitative characters (open triangles). The dotted line represents the average values of the simulation model.

and garden populations. The numbers of characters showing significant positive $-b$ values are seven out of 17 in the field population, and 10 out of 23 in the garden population (Table 3). Quantitatively, the overall spatial structure for quantitative characters was much stronger in the field population than in the garden population. The difference can be visualized by comparing the average correlograms based on principal components (Fig. 2). It is interesting to note on Fig. 2 that although $I(d)$ values are markedly different between the field and garden populations, the distances of intercept ($I(d)$ equal zero) are very similar. Moreover, the relationship with the logarithm of the distance seems to be approximately linear for both correlograms. Hence the slopes contain most of the information regarding spatial structure (a linear correlogram can be defined by two parameters, a slope and an intercept, but as the average value over all pairs of individuals is constrained to zero, it actually has only one degree of freedom). The average (±SD) $-b$ values for the principal components equal 0.095 (± 0.134) in the field population (six components) and 0.018 (± 0.023) in the garden population (seven components, Table 4). Paired tests of difference in $-b$ values between the field and garden populations cannot be applied because some characters are intercorrelated. It is, however, reasonable to consider that spatial structure is globally higher in the field population given that 15 characters out of 17 had higher $-b$ values in the field population (Table 3).

Comparisons of the spatial structures between allozymes and quantitative characters were based on the ratios $I(d)/t$ and $-b/t$ estimated for the garden population. The shapes of the average correlograms $I(d)/t$ show that the patterns of spatial structure for allozyme markers and the genetic component of quantitative traits are very similar (Fig. 3). Comparisons of $-b/t$ ratios lead to the same conclusion (Fig. 4), with the following averages (±SD): 0.136 (±0.088) for allozyme markers, 0.119 (±0.113) for quantitative characters, 0.143 (±0.134) for the factors of the PCA applied on the quantitative characters.

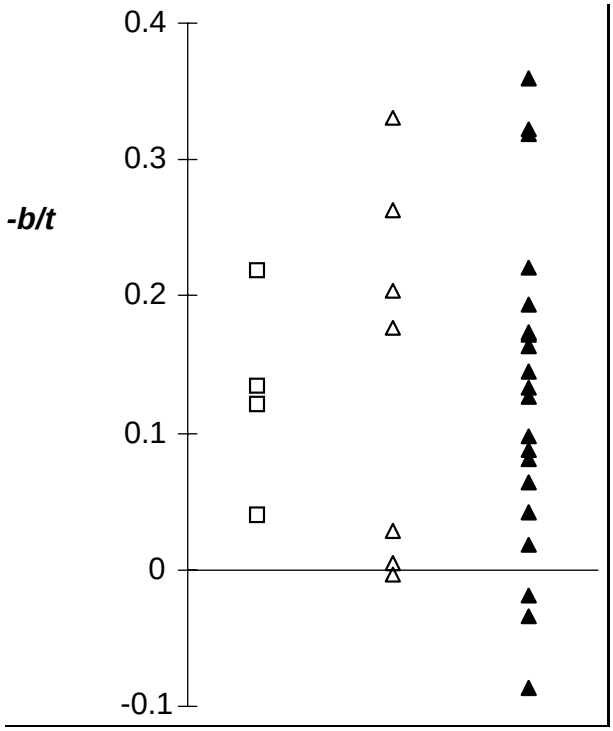

Fig. 4 Distribution of the $-b/t$ ratios for the four allozyme loci (open squares), the quantitative markers with significantly positive t values (closed triangles), and the principal components of the PCA applied on the quantitative characters (open triangles).

Simulation results

The best correspondence between simulations and allozyme data was obtained for an axial standard deviation of

pollen dispersal of 3.5 m (Fig. 3). Given our choice of seed dispersal scheme (axial standard deviation of seed dispersal distances equal to 0.8 m), the axial standard deviation for the overall gene dispersal was equal to 2.6 m. With a density of one and a half individual per square metre, the resulting measure of neighbourhood size ($4\pi D\sigma^2$) is of the order of 70 individuals, in good agreement with the estimate based on the average b value for allozymes (see above). We have checked that the structure had stabilized after 100 simulation years. The slope of the regression (b value) reached half its equilibrium value after 15 simulation years (result not shown).

The intraclass correlation coefficient (t) for the simulated genetic marker had an average value of 0.272 (95% confidence interval: 0.14–0.41). It is worth noting that all average values per locus observed for allozymes lie within this interval (Table 2). The average t value is a little higher than the expected one (0.25) for true half-sibs coming from a noninbred population. This result is expected for an isolation by distance model with non-random mating since crosses between related parents must occur.

The frequency distribution of the ratio $-b/t$ in the simulations is shown on Fig. 5 for genetic markers and quantitative traits. For the simulated quantitative trait, we selected the 790 replicates (out of 1000) with a t value at least equal to 0.05, that is approximately the minimal positive value that was considered as statistically significant ($\alpha = 0.05$) with our sample scheme. For the simulated genetic marker, t values were always larger than this threshold. From Fig. 5 it appears that the distributions are highly skewed towards positive values (skewness, i.e. third moment of the distribution, equals 1.28 and 1.24 for genetic and quantitative markers, respectively), and a significant proportion of the replicates shows values of the $-b/t$ ratio that are close to zero. The average values of the ratio $-b/t$ are not significantly different between genetic markers and quantitative traits (simulated genetic marker: 0.110; simulated quantitative trait: 0.109; Mann–Whitney U-test: $z = -1.17$, $P = 0.24$). The standard deviations were also very similar: 0.141 for the simulated genetic marker and 0.161 for the simulated quantitative trait. The 95% confidence interval for the ratio $-b/t$ is very wide: –0.065 to 0.473 for the simulated genetic marker, and –0.103 to 0.509 for the simulated quantitative trait. We can see that the observed $-b/t$ ratios for the allozyme data and the quantitative characters are all included within these confidence intervals (Fig. 4).

Discussion

This study uses a spatial autocorrelation approach to compare the genetic structure between genetic markers and quantitative characters at a microgeographical scale. Because differentiation is measured at the level of individuals, it may capture information about spatial structure at the smallest scale. If genetic markers are neutral, this approach enables us to assess the impact of natural selection on the spatial genetic structure of quantitative traits at a microgeographical scale.

Applied on a population of knapweeds, this study has shown that (a) a significant spatial genetic structure, consistent with an isolation by distance process, occurs for both the genetic markers and the quantitative characters; (b) the spatial structure of the quantitative characters observed in the field population is much more developed than the one observed in the garden population, indicating a strong spatial structure for the environmental component; (c) on average, the observed patterns of genetic structure are very similar for allozyme markers and quantitative characters, suggesting that, overall, they are subject to the same evolutionary forces; (d) the observed variances in the degree of spatial structure among allozyme loci and among quantitative characters are consistent with those inferred by simulation under a neutral hypothesis.

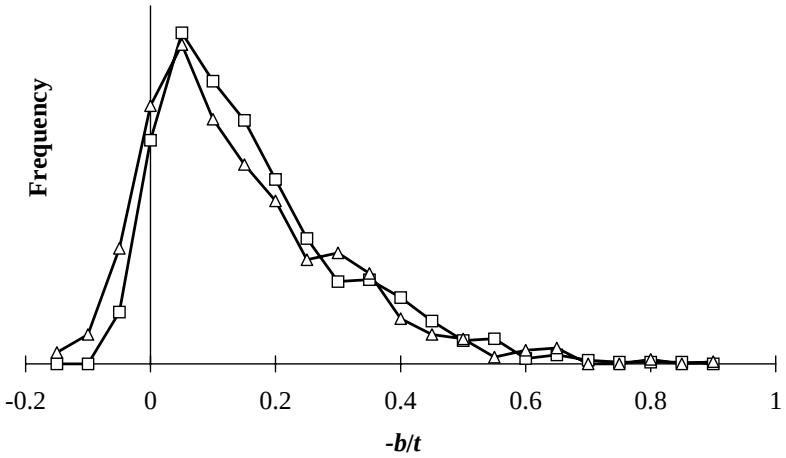

Fig. 5 Frequency distribution of the $-b/t$ ratios for a genetic marker (open squares) and a quantitative trait (open triangles) according to a simulation model (see text).

Validity of the approach

Point (c) is deduced from a model described in the Appendix and point (d) is deduced from simulations that tentatively approximate the process occurring in the real population. However, many deviations from idealized models may occur in the natural population and need to be discussed. First, the models assume genetic effects to be purely additive. Dominance and epistatis are likely to occur in nature but numerical simulations show that they do not affect strongly the ratios $I(d)/t$ and $-b/t$ (see Appendix). Second, maternal effects were also neglected. Significant maternal effects would tend to reduce $-b/t$ values. Third, no spatial clumping of individuals is possible in the simulated lattice model whereas local density is clearly not constant within natural populations. Finally, the demographic parameters used in the simulation model are not known with accuracy. However, as specified hereafter, the distribution of the estimator of spatial structure, $-b/t$, is little affected by those demographic parameters as long as the overall extent of gene dispersal is constant.

The statistics used to compare monogenic markers and quantitative characters are based on ratios of two covariances, i.e. the covariance between individuals among families and the covariance between individuals within family (see Appendix). The ratio b/t is the corresponding slope of the regression on the logarithm of the distance. In agreement with the model presented in the Appendix, our simulations have shown that the ratio $-b/t$ is not statistically different between genetic markers and quantitative characters when characters with low t values (threshold for $t = 0.05$) are rejected. However, a significant difference appeared when characters with estimated t values lower than 0.03 were included (result not shown). We suspect the reason to be that ratios of moments are biased estimators (Lynch & Walsh, 1998) and the bias probably becomes substantial when the denominator (the within-family covariance) approaches zero.

Two striking phenomena relative to the stochastic variance of the genetic structure appear from the distributions of the ratio $-b/t$ obtained from replicates of numerical simulations (Fig. 5). Firstly, the stochastic variance was hardly larger for quantitative characters than for genetic markers. This indicates that the additional variance due to the environmental component in quantitative characters was weak compared with the stochastic variance of the genetic structure. Secondly, the distributions of $-b/t$ values were strongly skewed towards positive values, and encompassed the zero and slightly negative values. Thus, the degree of spatial structure of a single genetic or quantitative marker could vary widely, from the absence of structure to a strong level of spatial structure, even in the absence of selection and with constant gene flow parameters. This observation emphasizes the importance of considering stochastic variation (Slatkin & Arter, 1991) when interpreting differences among correlograms or, more generally, differences among estimates of genetic differentiation (e.g. F_{ST}). Indeed, it has been suggested in some experimental studies that strong differences between correlograms for different allozyme loci (e.g. Epperson & Allard, 1989), or different quantitative characters (e.g. Argyres & Schmitt, 1991), resulted from differential selection. In the present study, such strong differences among individual loci or quantitative characters also occur, but all the observed values lie within the 95% confidence intervals obtained by numerical simulation. Hence, assuming that the parameters used in the simulations are representative of the field population, the observed structure for allozyme markers and quantitative traits in this particular population of knapweeds is fully compatible with a neutral hypothesis. However, this does not prove that natural selection was not acting on quantitative traits or genetic markers. The results only suggest that genetic drift was, on average, the most important mechanism generating the genetic structure.

The particular distribution described for $-b/t$ values does not result from the particular gene dispersal pattern simulated (low seed dispersal, high pollen dispersal), or the overlapping generations. Indeed, when performing simulations with nonoverlapping generations and equal seed and pollen dispersal, such that the variance of gene dispersal distances remains constant, the distribution of $-b/t$ values was barely affected (results not shown).

Genetic structure for allozymes and quantitative characters in natural populations

Natural selection may affect genetic structure of quantitative traits in different ways. Under uniform stabilizing selection, the same phenotype is selected everywhere so that genetic structure will be reduced relative to neutral expectation. Malécot (1948) showed that the impact of stabilizing selection on the genetic structure would be similar to that of an increase in the rate of mutation or immigration from a constant external source. Under diversifying selection, with the pattern of selection following a geographical gradient, genetic structure will be enhanced relative to neutral expectations. Slatkin (1973, 1978) showed that the response to changes in environmental conditions could occur only on a geographical scale larger than a 'characteristic length' equal to $\sigma/\sqrt{s}$, where σ is the standard deviation in gene dispersal distances and s the average strength of selection. An intermediate situation between stabilizing and diversifying selections may occur when environmental heterogeneity follows a fine grain pattern but without global gradient. Then, relative to neutral expectations, an increase in genetic structure may be observed on a small scale because different phenotypes are selected in adja-

cent environments, whereas a reduction in genetic structure may be observed on a large scale because the same set of phenotypes are selected all around. Hence, the potential impact of selection, a decrease or increase of differentiation, depends on whether the variance of the selected optimal phenotypes is, respectively, lower or higher than the expected variance of phenotypes under a balance between drift and gene flow alone (Latta, 1998; Hardy, 2000). Therefore, selection effects may differ substantially according to the geographical scale under study.

The principal result of this study is that the microgeographical genetic structure of quantitative characters does not differ from genetic markers. Among the previous studies that focused on large-scale differentiation among plant populations, one reported similar results to ours (Kuittinen *et al*., 1997) but most found an overall higher level of differentiation for quantitative characters, suggesting a substantial effect of diversifying selection (Podolsky & Holtsford, 1995; Bonnin *et al*., 1996; Yang *et al*., 1996; Kremer *et al*., 1997). For animal species, higher differentiation among populations has also been observed for several quantitative traits (Prout & Barker, 1993; Spitze, 1993; Long & Singh, 1995; Lynch *et al*., 1999). We may expect that the spatial scale studied affects the result. Indeed, the balance between gene flow and diversifying selection is likely to be in favour of gene flow on a small scale, and in favour of selection on a large scale, both because of the decrease of gene flow and the potential increase in environmental heterogeneity at larger scales. This is consistent with the results of Bonnin *et al*. (1996) who found similar levels of differentiation between quantitative traits and genetic markers among subpopulations 10–50 m apart, but an overall higher differentiation at quantitative traits for populations 200 km apart.

In our study we detected environmental heterogeneity in the field population (Fig. 2), and hence there is potential for diversifying selection. However, given the relatively extended gene flow that seems to occur through pollen dispersal, selection coefficients should be very high to create a significant effect on spatial structure. The effect of a stabilizing selection, which enhances genetic homogeneity, would be even more difficult to detect in our population because, as shown by simulations, a significant proportion of characters are expected to show no spatial genetic structure at all under neutrality. Hence, the absence of an effect driven by selection may not be surprising.

Given that studies carried within and among populations seem to yield contrasting results, it might be interesting to compare differentiation at genetic markers and quantitative traits over a range of geographical scales. This would enable us to assess at which scale an effect of selection becomes detectable. A spatial autocorrelation approach as the one developed in this paper would be suitable for the purpose.

Acknowledgments

This work was supported by grant no. 2.4512.97 from the Belgian National Fund for Scientific Research (F.N.R.S.) where O.J.H. is a research assistant. We are very grateful to Kent Holsinger, Henk Van Dijk and two anonymous reviewers for their helpful comments on previous drafts of the manuscript.

References

Argyres, A.Z. & Schmitt, J. 1991. Microgeographical genetic structure of morphological and life history traits in a natural population of *Impatiens capensis*. *Evolution* **45**: 178–189.

Bonnin, I., Prosperi, J.-M. & Olivieri, I. 1996. Genetic markers and quantitative genetic variation in *Medicago truncatula* (Leguminosa): a comparative analysis of population structure. *Genetics* **143**: 1795–1805.

Dewey, S.E. & Heywood, J.S. 1988. Spatial genetic structure in a population of *Psychotria nervosa*. I. Distribution of genotypes. *Evolution* **42**: 834–838.

Epperson, B.K. & Allard, R.W. 1989. Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. *Genetics* **121**: 369–377.

Felsenstein, J. 1986. Population differences in quantitative characters and gene frequencies: a comment on papers by Lewontin and Rogers. *Am. Nat.* **127**: 731–732.

Gardou, C. 1972. Recherches biosystématiques sur la Section *Jacea* Cass. et quelques sections voisines du genre *Centaurea* L. en France et dans les régions limitrophes. *Feddes Repertorium* **83**: 311–472.

Hamrick, J.L. & Godt, M.J. 1989. Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding and Genetic Resources* (A. H. D. Brown, M. T. Clegg, A. L. Kahler & B. S. Weir, eds), pp. 43–63. Sinauer, Sunderland, Mass.

Hardy, O. 2000. Structure spatiale de la variation génétique et relations entre cytotypes dans des populations d'un complexe polyplo: *Centaurea jacea* L. s.l. PhD Thesis, Université Libre de Bruxelles, Bruxelles.

Hardy, O.J., De Vanderhoeven, S., Loose, M. & Meerts, P. 2000. Ecological, morphological and allozymic differentiation between diploid and tetraploid knapweeds (*Centaurea jacea* s.l.) from a contact zone in the Belgian Ardennes. *New Phytol.* in press.

Hardy, O.J. & Vekemans, X. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**: 145–154.

Hillis, D.M., Moritz, C. & Mable, B.K. 1996. *Molecular Systematics*. Sinauer, Massachusetts.

Kremer, A., Zanetto, A. & Ducousso, A. 1997. Multilocus and multitrait measures of differentiation for gene markers and phenotypic traits. *Genetics* **145**: 1229–1241.

Kuittinen, H., Mattila, A. & Savolainen, O. 1997. Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. *Heredity* **79**: 144–152.

Lambinon, J., De Langhe, J.E., Delvosalle, L. & Duvigneaud, J. 1992. *Nouvelle flore de la Belgique, du Grand-Duché de Luxembourg, du Nord de la France et des régions voisines (Ptéridophytes et Spermatophytes)*. Meise, Belgium: Jardin botanique national de Belgique.

Lande, R. 1977. Statistical tests for natural selection on quantitative characters. *Evolution* **31**: 442–446.

Lande, R. 1991. Isolation by distance in a quantitative trait. *Genetics* **128**: 443–452.

Lande, R. 1992. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution* **46**: 381–389.

Latta, R.G. 1998. Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *Am. Nat.* **151**: 283–292.

Lewontin, R.C. 1984. Detecting population differences in quantitative characters as opposed to gene frequencies. *Am. Nat.* **123**: 115–124.

Linhart, Y.B. & Grant, M.C. 1996. Evolutionary significance of local genetic differentiation in plants. *Ann. Rev. Ecol. Syst.* **27**: 237–277.

Long, A.D. & Singh, R.S. 1995. Molecules versus morphology: the detection of selection acting on morphological characters along a cline in *Drosophila melanogaster*. *Heredity* **74**: 569–581.

Lynch, M., Pfrender, M., Spitze, K., Lehman, N., Hicks, J., Allen, D., Latta, L., Ottene, M., Bogue, F. & Colbourne, J. 1999. The quantitative and molecular genetic architecture of a subdivided species. *Evolution* **53**: 10–110.

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

Malécot, G. 1948. *Les Mathématiques de l'Hérédité*. Masson et Cie, Paris.

Maruyama, T. 1977. *Stochastic Problems in Population Genetics*. Springer-Verlag, Berlin.

Podolsky, R.H. & Holtsford, T.P. 1995. Population structure of morphological traits in *Clarkia dudleyana* I. Comparison of Fst between allozymes and morphological traits. *Genetics* **140**: 733–744.

Prout, T. & Barker, J.S.F. 1993. *F* statistics in *Drosophila buzzatii*: selection, population size and inbreeding. *Genetics* **134**: 369–375.

Rogers, A.R. 1986. Population differences in quantitative characters as opposed to gene frequencies. *Am. Nat.* **127**: 729–730.

Rogers, A.R. & Harpending, H.C. 1983. Population structure and quantitative characters. *Genetics* **105**: 985–1002.

Ronfort, J., Jenczewski, E., Bataillon, T. & Rousset, F. 1998. Analysis of population structure in autotetraploid species. *Genetics* **150**: 921–930.

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

Rousset, F. 2000. Genetic differentiation between individuals. *J. Evol. Biol.* **13**: 58–62.

Slatkin, M. 1973. Gene flow and selection in a cline. *Genetics* **75**: 733–756.

Slatkin, M. 1978. Spatial patterns in the distribution of polygenic characters. *J. Theor. Biol.* **70**: 213–228.

Slatkin, M. & Arter, H.E. 1991. Spatial autocorrelation methods in population genetics. *Am. Nat.* **138**: 499–517.

Spitze, K. 1993. Population structure in *Daphnia obtusa*: quantitative genetics and allozymic variation. *Genetics* **135**: 367–374.

Weiner, J., Martinez, S., Mûller-Schärer, H., Stoll, P. & Schmid, B. 1997. How important are environmental maternal effects in plant? A study with *Centaurea maculosa*. *J. Ecol.* **83**: 133–142.

Yang, R.-C., Yeh, F.C. & Yanchuk, A.D. 1996. A comparison of isozyme and quantitative genetic variation in *Pinus contorta* ssp. *latifolia* by *Fst*. *Genetics* **142**: 1045–1052.

Received 23 February 2000 accepted 15 March 2000

Appendix

We show that the parameter estimated by the ratio $I(d)/t$ has the same expectation for any neutral character under a set of assumptions. First we must assume that all effects are additive, i.e. no genotype–environment interactions, no dominance, no epistatic interactions. Then the phenotype of individual i is $X_i = g_i + e_i$, where g_i and e_i are the genotypic and the environmental effects, respectively. The genotypic effect can be decomposed into individual gene effects at different loci. For simplicity we consider only one locus: then $g_i = \sum_l^k a_{li}$ where a_{li} is the effect of the allele situated on the lth homologous chromosome of individual i, and k is the number of homologous chromosomes (the ploidy level assuming polysomic inheritance). The correlation between the phenotypes of two individuals i and j is $\rho(X_i,X_j) = \sigma(X_i,X_j)/[\sigma^2(X_i)\sigma^2(X_j)]^{1/2}$ where σ and σ^2 stand for the covariance and variance, respectively. Assuming that the environmental effect is a stochastic variable of null average and constant variance: $E(e_i) = 0$, $E(e_i e_j) = 0$ for any pair of individuals i and j $(i \neq j)$, $E(e^2) = \sigma_e^2 =$ constant, and $E(e_i g_j) = 0$ for any i and j, including $i = j$ (no genotype–environment covariance), where E stands for the expectation operator. Therefore, $\sigma(X_i,X_j) = \sigma(g_i,g_j)$ and $\sigma^2(X_i) = \sigma_g^2 + \sigma_e^2 = \sigma_g^2/h^2$, with σ_g^2 being the variance of genotypic effects over the population, and h^2 the heritability. Decomposing the genotypic effect into single allele effects, we obtain $\sigma(g_i,g_j) = k^2\sigma_a^2\theta_{ij}$, where θ_{ij} is the coefficient of kinship between individuals i and j (i.e. the correlation coefficient between a random allele of i and a random allele of j), and σ_a^2 is the variance of single allele effects over the population. Similarly, $\sigma_g^2 = (1 + (k - 1)f)k\sigma_a^2$, where f is the inbreeding coefficient (i.e. the correlation coefficient between two random but different alleles taken within an individual) which is assumed to have the same expectation for all individuals. Hence, $\rho(X_i,X_j) = (k\theta_{ij}h^2)/[1 + (k - 1)f)]$.

$I(d)$ statistics are estimators of the average correlation between pairs of individuals separated by given distance classes d, thus $I(d) = \hat{\rho}(X_i,X_j) = \hat{\rho}(d)$ for $|i,j| \in d$. In an isolation by distance model, assuming neutrality, the coefficient of kinship between pairs of individuals is essentially a function of the distance separating them, $\theta(d)$, hence $\rho(d) = \{kh^2/[1 + (k - 1)f)]\}\theta(d)$. Similarly, for pairs of individuals i, j belonging to the same family, the intraclass correlation coefficient $t = \hat{\rho}(X_i,X_j) = \hat{\rho}_0$, and $\rho_0 = \{kh^2/[1 + (k - 1)f)]\}\theta_0$. Hence the ratio $I(d)/t$ is an estimator of $\theta(d)/\theta_0$. The latter is therefore independent of the environmental effects, and has the same expectation for any locus submitted to the same evolutionary

forces. The extension to multiple loci necessitates the additional assumption that no linkage phase disequilibrium occurs between loci. Simulations show that the approach is relatively robust when dominance or some kind of epistatic interactions occur because the numerator and the denominator in $I(d)/t$ ratio are then reduced by similar proportions (Hardy & Vekemans, unpublished).