The spatial structure of the gene pool of a viviparous population of *Poa alpina* - environmental controls and spatial constraints

Ottar N. Bjørnstad[§], Anders Iversen[‡] and Marit Hansen[‡]

[§] Division of Zoology, Department of Biology, University of Oslo, POB 1050 Blindern, N-0316 Oslo, Norway. [‡] Division of Botany, Department of Biology, University of Oslo, POB 1066 Blindern, N-0316 Oslo, Norway.

Nordic Journal of Botany (1995) 15:347-354

Abstract

Because both the genetic make-up and the environmental conditions of a population are spatially autocorrelated, it is difficult to infer processes of selection or drift for population genetic mappings. We propose a methodology based on partial Mantel techniques and partial autocorrelation techniques to separate the action of these processes. The method is applied to data on *Poa alpina* to indicate that isolation-by-distance (drift) is the main process inducing positive autocorrelation at the scale of diaspore dispersal (<100m). The pattern for larger distances is more consistent with selection.

Keywords: Spatial autocorrelation; Selection; drift; Isolation-by-distance; gene pool; Genetics; Mantel correlation; Correlograms; Statistics

Introduction

Abundant evidence shows that plant populations typically exhibit genetic micro-differentiation (e.g., Sokal et al. 1989, Epperson 1993); that is, individuals that are located in close spatial proximity tend to be more alike than individuals at some distance apart. The populations, thus, exhibit genetic autocorrelation (e.g., Sokal & Jacquez 1991). Due to the spatial autocorrelation in most environmental and ecological factors (Sokal & Oden 1978, Legendre 1993), plants in 'ecological proximity' (e.g., measured in some species-space or environment space) will tend to be genetically similar. One possible reason for such local genetic correlation is local differences in the selective pressures, resulting in different genetic make-ups in different parts of the population (e.g. Hamrick & Holden 1979, Turkington & Aarssen 1984, Ennos 1985, Nevo et al. 1986, Sokal et al. 1989, Epperson 1993). Alternatively, the pattern may result from non-adaptive structuring due to local genetic drift through a process of isolation-by-distance (IBD), such as limited dispersal of pollen or diaspores (Hamrick & Loveless 1986, Leduc et al. 1992, Willson 1992, Epperson 1993). Spatial aggregation due to IBD and drift is, for instance, reported by Heywood & Levin (1985). In the case of structuring due to selection, the pattern emerges as a result of systematic pressures controlled by the environment ('environmental control'), whereas in the case of drift the pattern is due to spatial constraints. There are methodological problems in trying to separate the two processes because they are confounded by their inherent autocorrelation (Leduc et al. 1992); The two hypotheses share a common spatial component which need to be controlled for.

The classical way of getting around this is through experimental manipulation (e.g., transplantation; Bradshaw 1984) or finding an environment of such a checkerboard nature that strong confounding of the spatial constraints and the selective milieu is improbable (Nevo et al. 1986). Stringent manipulation is, however, not always practical, and checkerboard habitats are quite rare. In the following we propose a statistical framework which may separate the influence of the two forces from observational data on genetic data and environmental conditions (as judged by vegetational composition). The method is an adaptation of commonly used spatial autocorrelation and Mantel correlation techniques (e.g., Smouse et al. 1986, Legendre & Troussellier 1988, Sokal & Jacquez 1991, Leduc et al. 1992). We apply the method to data on a viviparous population of *Poa alpina* L. sampled along three adjacent transects across an ecological gradient from snowbed to exposed ridge. We demonstrate that the genetic pattern is largely due to spatial constraints (isolation by distance) at the scale of propagule dispersal, whereas selective forces may well generate the pattern at a larger scale.

Materials and methods

Study area and genetic analysis

This study is based on previously published ecological and genetic data (Nordal & Iversen 1993). The population analysed is situated at an altitude of 1370-1650 m, on the Leirtjønnkollen mountain in Oppdal county, Central Norway. The vegetation changes along one main gradient from snowbed to exposed ridge. Three transects running through this gradient were sampled. The transects were 170, 200, 90m long and 250m, 800m and 1200m apart, respectively.

Two-hundred-and-forty-nine individuals of the polyploid and apomictic *P. alpina* were collected for electrophoretic analyses at 15 sites along the three gradients. An average of eighteen (s.d. = 4.6) individuals from each site were analysed using enzyme electrophoresis of 12 band-phenotype

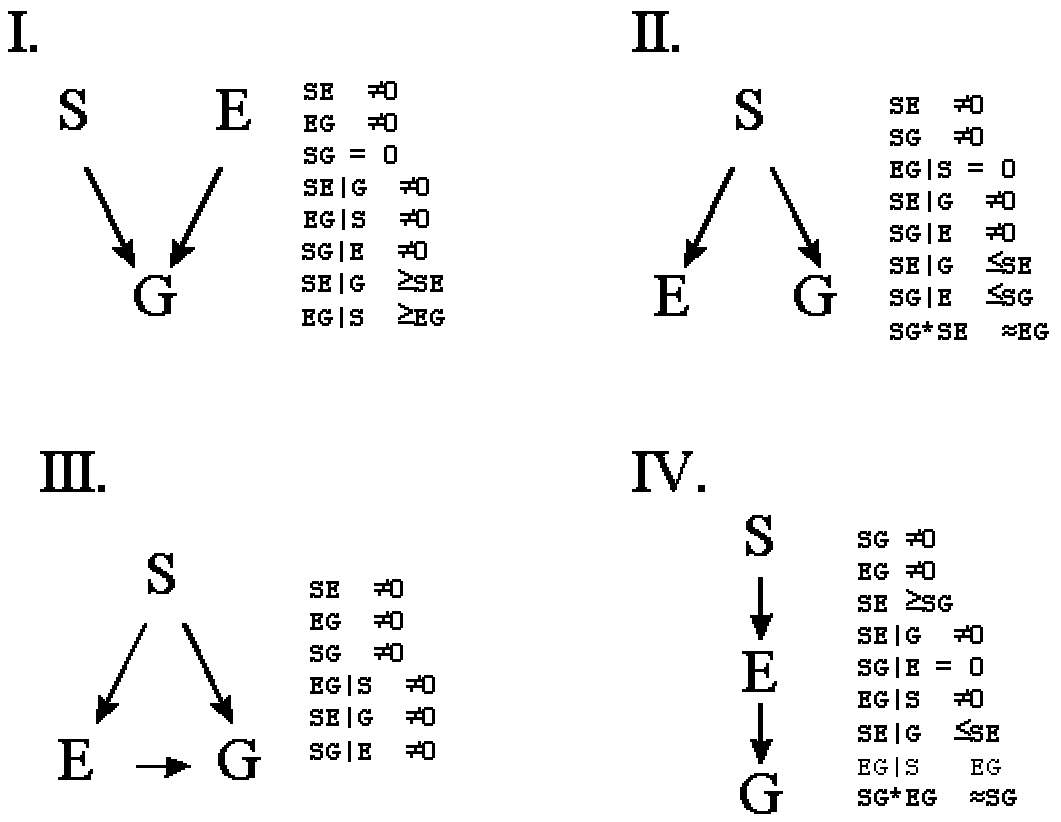

Figure 1. The four models of causal relationships involving the three distance matrices. S ('Space') represents the matrix of spatial distances between individuals, E ('Environment') represents the environmental distance matrix, and G ('Genetics') represent the matrix of genetic distances (all as defined in the text). The predictions of the models in terms of simple and partial Mantel tests (see text) are stated to the right of each model. Correlation coefficients are denoted by the symbol of the matrices, partial tests are denoted by a |. Stating that a relation is equal to zero means that the computed coefficients should not be significantly different from zero. Adapted from Legendre (1993).

complexes. Of these 3 were monomorphic and excluded from the analysis (Nordal & Iversen 1993). All sites were further analysed for floristic composition. Sixteen 0.5x0.5m quadrats were analysed at every 10 m along the transects (see Nordal & Iversen 1993 for details).

Statistical methods

Three matrices (dimension 249x249) of distances between the individuals were computed. A measure of genetic distances was calculated as the Manhattan metric distance (Sneath & Sokal 1973) between each individual in the nine-dimensional "band-phenotype space". The geographic distance between each site gave the spatial distance matrix (Matrix S). The floristic data for each sampling stations were subjected to a detrended correspondence analysis (DCA; Hill & Gauch 1980). The ecological distance between two individuals was calculated as their Euclidean distance in the space spanned by the eigenvectors accompanying the four largest eigenvalues of this ordination (matrix E). All matrices were standardised so that the off-diagonal elements had a mean of zero and a variance of unity (Manly 1991).

The two hypotheses to be tested, spatial constraint vs. environmental control, can intuitively be contrasted as follows: If the pattern is determined by limited dispersal (spatial constraint), the genetic distance between two individuals should correlate better with their spatial proximity than with their ecological proximity. If the pattern is determined by local adaptive differentiation, the genetic distance between two individuals should be a function of their ecological proximity rather than their spatial proximity. Note, however, that this may be difficult to ascertain due to spatial autocorrelation in the environment. An intermediate situation, will be that of an interaction between the two. In this way, four different causal relationships can be recognised. These are depicted graphically in figure 1. The situation in model II represents ‘spatial constraints’ alone - spatial autocorrelation in the gene pool is only spuriously correlated to the environmental variables due to a common spatial structure. Model IV represents environmental control - due to spatial

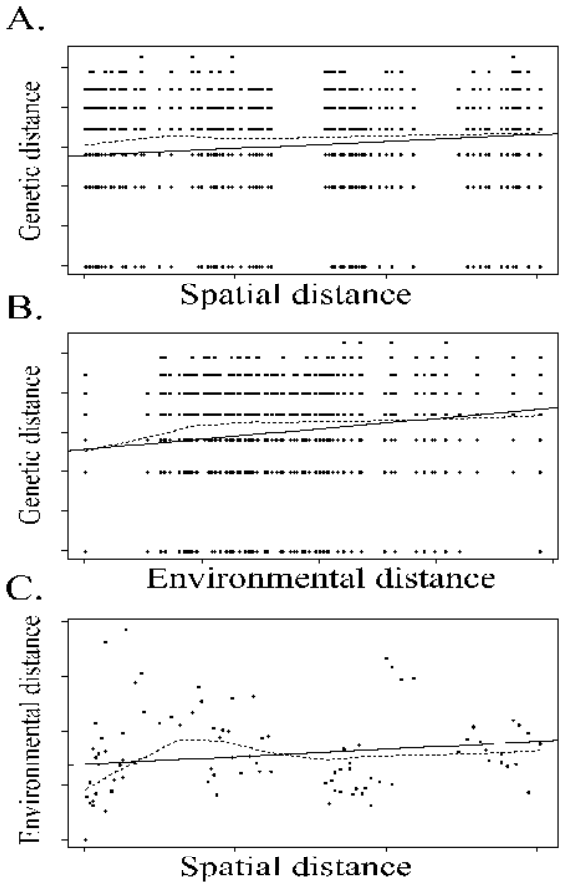

Figure 2. The three scatter plots of the genetic, the environmental and the spatial distances. The linear (solid) and the LOWESS (dotted) regression lines are included to examine any departure from linearity in the relations.

autocorrelation in the environmental variables, the gene pool will appear spatially autocorrelated. This autocorrelation is, however, not due to spatial constrains as such. The spatial autocorrelation is indirect, and due to the spatial structure of the environment. Model I and III are intermediate. In these two situations, both processes are operating and interacting. Model I, however, includes no spatial autocorrelation in the environment. This is not very likely in biological systems (e.g., Sokal & Oden 1978, Legendre 1993).

Legendre & Troussellier (1988) have derived a set of unique predictions, in terms of correlation coefficients and partial correlation coefficients between variables, to distinguish the consistency of different causal webs. Their framework can be applied for the data at hand as depicted in figure 1. The decision rule (see figure caption) is based on Mantel and partial Mantel correlations (which is the appropriate methodology for comparing distance matrices; Mantel 1967, Smouse et al. 1986, Manly 1991). The partial Mantel test is analogous to a partial correlation in that it correlates two variables whilst controlling for a third variable (denoted **AxB|C)**. This is carried out by regressing **B** on **C** and **A** on **C** and then correlating the matrices of residuals (e.g., Crawford & Duggirala 1992). Mantel statistics assume an approximately linear relationship between the distances. To check for this, we plot the distances against each other in figure 2. A linear regression line and a locally weighted regression line (LOWESS; e.g., Trexler & Travis 1993).

To investigate the consistency of the spatially and ecological distribution of the phenotype bands, Mantel correlograms (the multivariate analogue of simple correlograms; Legendre & Fortin 1989, Legendre 1993) were constructed. These correlograms depict the genetic correlation of individuals at different ecological and spatial distances apart. Correlograms are useful in identifying spatial trends or patches in the data (e.g. Sokal 1979, Barbujani 1987, Sokal & Jacquez 1991). To do this, the spatial and ecological distances must be divided in discrete classes. Ten distance classes were used. The initial distance class (the "zero'th") consists of the distance between each individual and itself. This is always zero and contain no information of interest. This class is therefore omitted from all figures and calcuclations. The remaining nine classes were constructed as follows: The extent of the first spatial distance class (10m) was chosen so as to correspond to the within-site variation. The other class limits were made to give high resolution at low distances. Due to the low number of individuals at large distances apart, these distant classes were made rather large (see e.g., Sokal & Jacquez 1991 for a discussion of generating distance classes).

An interaction between the predictor variables will lead to the conclusion that model III (or possibly I) is more consistent with the data. Such an interaction may be a result of the two processes operating at different spatial scales. One way of elucidating this is to generate partial correlograms, were the effect of the variable controlled for is removed from the dependent variable (in this case the genetic distance matrix). Note, that contrary to partial correlation (or partial Mantel test; e.g., Crawford & Duggirala 1992), the independent variable of the correlogram, which is divided in classes, cannot be corrected - the correlation between the variable controlled for, and the independent variable can not be partialled out. Doing that would alter the metric and inhibit comparison between the partial correlogram and original correlogram.

The analyses were carried out using the "R-package" (Legendre & Vaudor 1991), CANOCO (Ter Braak 1987), NTSYS-PC (Rohlf 1986), SAS version 6.08 (SAS Inst. 1990) and S-plus ver 3.2 (Statistical Sciences 1993).

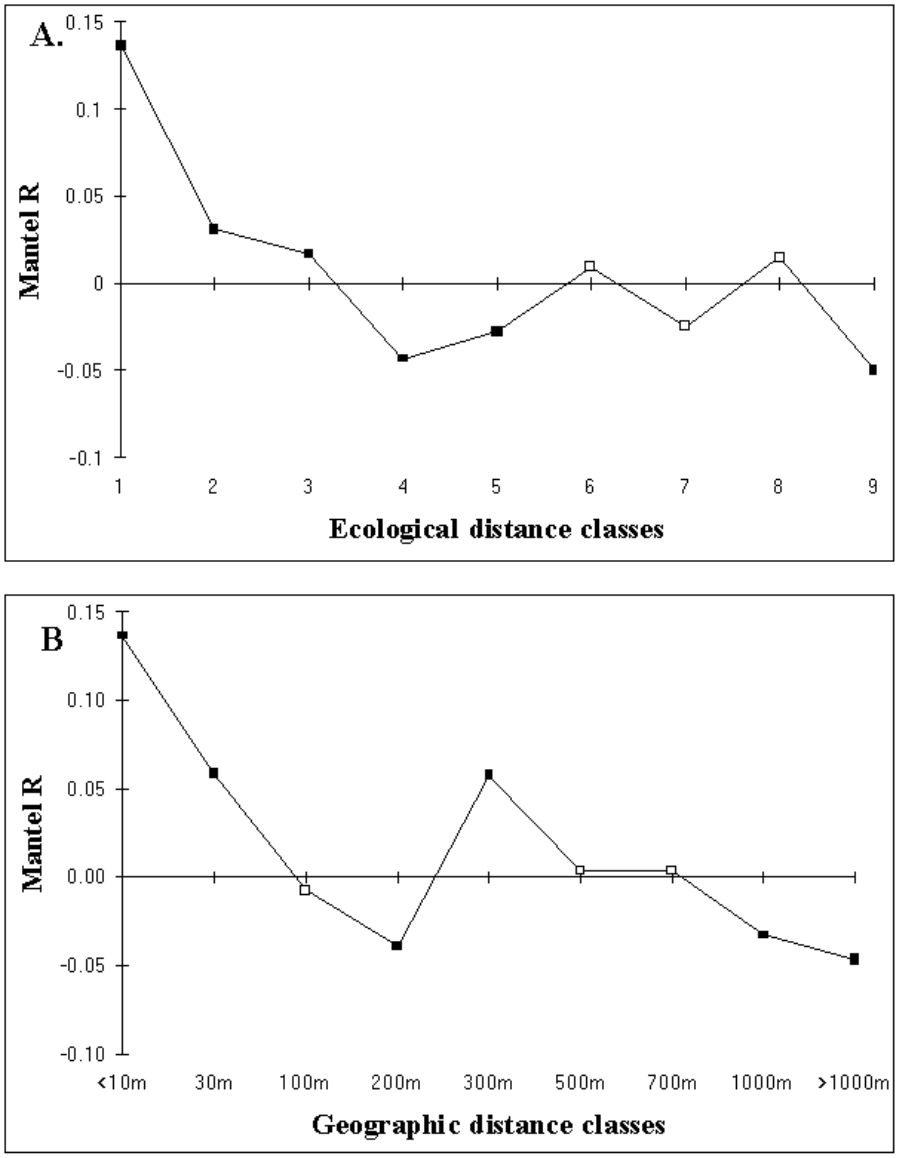

Figure 3. Mantel correlograms relating the genetic distance between individuals to (a) the ecologic distance between individuals and (b) spatial distance between the individuals (see text for definition of distances). Filled symbols represents statistically significant autocorrelations (at a Bonferroni corrected 5 % level; Rice 1989).

Results

The relationship between the distances are represented in scatter plots in figure 2. Because of the relatively few genetic traits studied, these distances appear to be discrete. This is not a problem for the statistics employed, but it hampers the visual evaluation of the linearity (of the relations). The LOWESS regression lines do not diverge strongly from the any of the linear regressions. We may, therefore, proceed with the linear statistics (but see discussion).

The simple Mantel correlograms are shown in figure 3. The gene pool is structured with respect to space as well as with respect to environmental factors. The positive autocorrelation in the smaller distance classes shows that individuals in close spatial or ecological proximity are genetically more similar than more distant individuals. This suggests that isolation by distance may be a factor. However, the negative correlation at the most distant class suggests that some systematic process, such as selection, is operating to induce clinal variation (Sokal 1979, Sokal & Jacquez 1991). The significant correlogram in environmental space may further signify a selective environment (Clearly, this may also be due to confounding; see below). Both factors are hence possible candidates as structuring forces in the gene pool of *P. alpina*.

The correlation coefficients required to discriminate between the alternate models of figure 1 are given in table 1. Allowing for multiple testing (Bonferroni correction; e.g., Rice 1989), the model where

	Mantel R	p	p_{bonf}
EG	0.17	0.001	0.006
SE	0.16	0.001	0.006
SG	0.13	0.001	0.006
SG|E	0.06	0.010	0.060
EG|S	0.05	0.017	0.102
SE|G	0.15	0.001	0.006

Table 1. The results of the Mantel and partial Mantel correlations. **E** denotes the matrix of **E**nvironmental distances, **S** the matrix of **S**patial distances, and **G** the distance of **G**enetic distances. The symbols are as explained in figure 1. Mantel R signifies the Mantel correlation, p indicates the significance level based on the permutation test (1000 permutations). Allowance should be made for multiple testing when making global conclusions (e.g., Rice 1989) such as those indicated in figure 1. pbonf indicates the Bonferroni corrected p-values.

environment and spatial constraints interact (model III) was consistent with the data. The estimated correlation coefficients indicate, however, that the direct link between the environmental and genetic distances is weaker than that between the spatial and genetic distances. The partial correlation when the spatial component is controlled for (**ExG|S**) is small and non-significant (with a Bonferroni correction).

The partial Mantel correlograms (Fig.4) highlight the interaction governing the spatial pattern of enzyme bands. In figure 4a, all of the variance that may be attributed to spatial constraints is removed from the genetic data. The residual variance is analysed with respect to the environmental matrix. Figure 4b is similar to figure 4a, but the correction is carried out with respect to the variance explained by ecology. The genetic similarity among individuals at distances up to 100 m is clearly best explained by some spatial constraint, e.g., isolation by distance due to limited powers of diaspore dispersal. This conclusion is based on the partial correlations at the most proximate distance class; the spatial autocorrelation remains significant, whereas, the environmental autocorrelation becomes insignificant when the other factor is partialled out (Fig. 4). Note that some of the spatial variation is best explained by environmental variables (distance class 3 and 4; Fig. 4a), even when spatial autocorrelation is corrected for.

Discussion

Epperson (1993) reviews the methods for estimating genetic neighbourhoods by spatial correlograms. The most common estimator is to use the lowest distance at which the correlogram crosses the x-axis. Our analyses suggest that the genetic patch size of *P. alpina* is somewhere between 30m and 100m (Fig. 4b). *P. alpina* reproduces largely by apomixis in Scandinavia (Müntzing 1940, 1966). Bulbil dispersal is the common mode of gene flow. The observed patch size compares well with empirical distributions of seed dispersal in other plants (Willson 1992). Importantly, the local spatial autocorrelation remains significant even when the environmental proximity is corrected for. The data, hence, suggest that dispersal-distribution curves determine the overall pattern at such a local scale. Both spatial correlograms (Fig. 3b and 4b) exhibit a significant positive autocorrelation at 200-300m. This is contrary to what one would expect under a pure IBD (where the correlogram would be monotonically declining). The number of distances involved in this estimate is however, very low. This result should therefore be viewed with caution. Indeed, when this class is merged with the neighbouring class (100-200m) the resultant coefficient is $r = 0.001$ ($p = 0.46$).

In the partial correlograms relating genetic make-up to the indirect measures of environmental conditions, there is evidence of positive autocorrelation across similar environmental situations even when spatial autocorrelation is corrected for (Fig. 4a). This is concordant to what we may expect as a result of selection. Furthermore, there is a significant negative correlation for the most distant environmental class in figure 4a (However note that there are rather few distances involved in the estimate). This is concordant with what might be expected from clinal selection. Note, however, that the magnitudes of the correlations are very small.

One main result emerging of this study is that indirect measures of environmental control of plant population structure (Nevo et al. 1986; Fig. 3a), should be interpreted with caution when the presence of other structuring forces, such as limited dispersal, are ignored. First, such forces may have very similar effects on the pattern, thus, inducing a problem of confounding. The present data represent a clear example; the discrepancy between figures 3 and 4 demonstrates that the environmental predictor is confounded by a strong spatial component (see also Table 1). Such confounding may not be detected when environmental transects are not replicated. Second, other structuring forces (such as limited powers of dispersal) will induce positive autocorrelation in the data set. This will invariably lead to a type I error - the hypothesis of no environmental influence will be rejected falsely (Hurlbert 1984, Legendre 1993).

We believe that the proposed statistical methodology bypasses this, as long as the sampling design is tailored loosely to such analysis. The statistical modelling is an adoption of the partial Mantel correlation (Smouse et al. 1986, Crawford & Duggirala 1992, Leduc et al. 1992). Sokal et al. (1989) proposed to use spatial autocorrelation

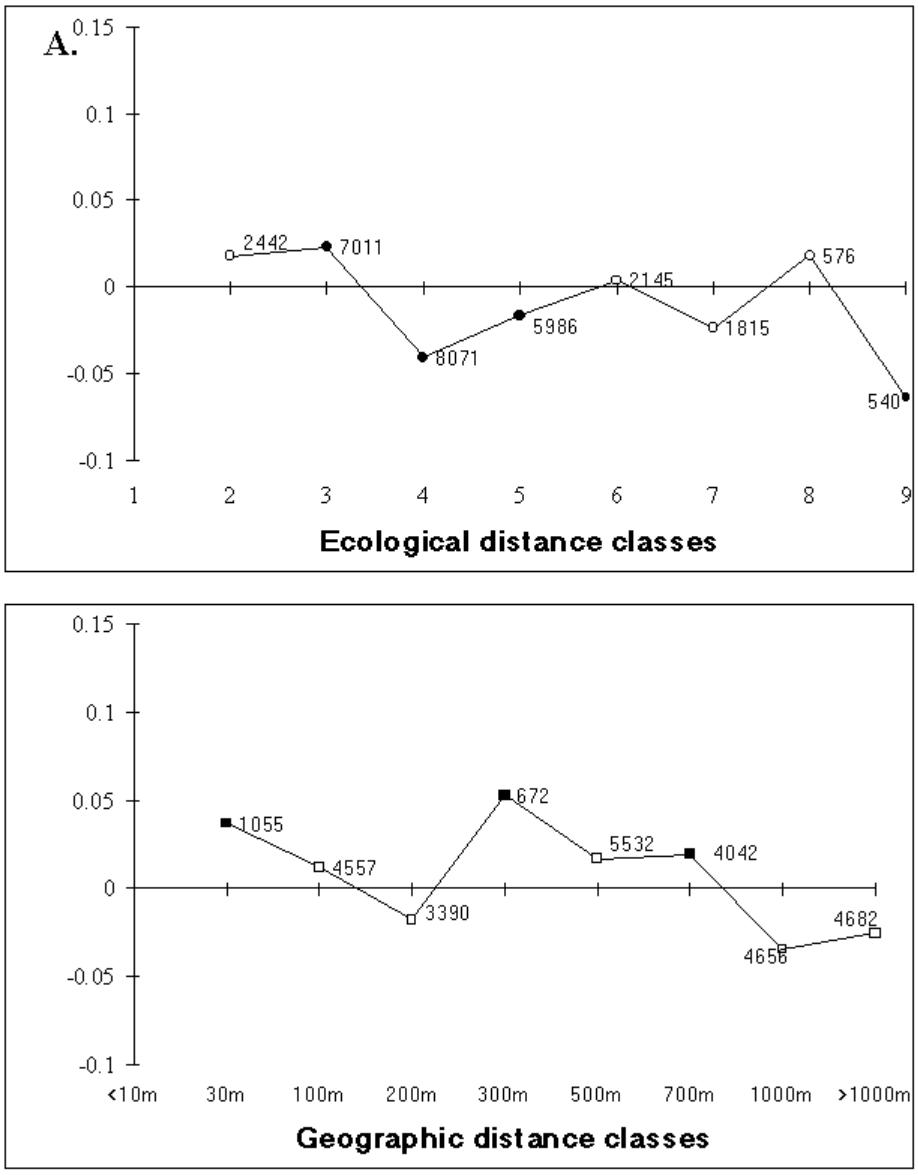

Figure 4. Partial Mantel correlograms relating the genetic distance between individuals to (a) the ecologic distance between individuals and (b) spatial distance between the individuals (see text for definition of distances). In contrast to Fig. 2, all variation in the data set that can be attributed to (a) spatial proximity, and (b) ecologic proximity has been removed prior to the analyses. Filled symbols represents statistically significant autocorrelations (at Bonferroni corrected 5 % level; Rice 1989). Note that no value is given for distance class one. This is due to exact collinearity of the two predictors (geographic distance and ecologic distance) at this distance class. The number of distances involved in each estimate is given above each estimate.

techniques to investigate selection as opposed to differentiation through isolation by distance. Our method is an extension of this by taking indirect measures of environmental conditions in to account using partial autocorrelation.

We have assumed that the relation controlled for (partialled out) is linear. There are, however, reasons for suspecting the relationship between geographical and genetic distances to be nonlinear (e.g. negative exponential; Epperson 1993, Portnoy & Willson 1993). This is reflected in the presence of some curvature in the LOWESS line of figure 2. To investigate the sensitivity of the results to the assumption of linearity we tried two other schemes for correction:

- The genetic distances where log-transformed (see Epperson 1993).

- The most flexible model was to use an ANOVA model of the genetic distances on the geographic or

environmental distance classes. The residuals of this were then used in the further Mantel correlogram. An ordinary least square method (OLS) was, in other words, used for the estimation of the residuals. OLS can be used on autocorrelated data for this sort of estimation problem, because they produce unbiased (although statistically inefficient) estimates of the parameters, and hence of the residuals (Ostrom 1978).

The results of both these exercises were virtually indistinguishable from the result shown in figure 4. They are therefore not shown.

In summary, the results presented here indicate that the genetic structure in the population can be described as an isolation by distance process for locations in close spatial proximity. However, when all variation that may be attributed to such a drift process is removed from the data, some variation remains which can be explained by local selection.

Acknowledgement: John Birks, Christian Brochmann, Rolf A. Ims, Pierre Legendre, Inger Nordal and one anonymous referee have commented on and improved the manuscript.

Literature cited:

Barbujani, G. 1987. Diversity of some gene frequencies in European and Asian populations. III. Spatial correlogram analysis. - Ann. Hum. Genet. 51:345-353.

Bradshaw, A.D. 1984. Ecological significance of genetic variation between populations. - In: Dirzo, R. & Sarukhan, J. (Eds.), Perspectives on plant population ecology. Sinauer Associates, Sunderland, pp. 213-228.

Crawford, M.H. & Duggirala, R. 1992. Digital dermatoglyphic patterns of Eskimo and Amerindian populations: relationship between geographic, dermatoglyphic, genetic, and linguistic distances. - Human Biology 64:683-704.

Ennos, R.A. 1985. Maintenance of genetic variation in plant populations. - In: Jackson, J.B.C., Buss, L.W. & Cook, R.E (eds.). Population biology and evolution of clonal organisms. Yale University Press, London, pp. 129-155.

Epperson, B.K. 1993. Recent advances in correlation studies of spatial patterns of genetic variation. - Evolutionary Biology 27:95-155

Hamrick, J.L. & Holden, L.R. 1979. Influence of microhabitat heterogeneity on gene frequency and gametic phase disequilibrium in *Avena barbarrata* - Evolution 33:521-533.

Hamrick, J.L. & Loveless, M.D. 1986. The influence on seed dispersal mechanisms on the genetic structure of plant populations. - In: Estrada, A. & Fleming, T.H. (eds.). Frugivores and seed dispersal. Dr. W. Junk Publishers, Dordrecht, pp. 211-224.

Heywood, J.S. & Levin, D.A. 1985. Associations between allozyme frequencies and soil characteristics in *Gaillardia pulchella* (Compositae). - Evolution 39:1076-1086.

Hill, M.O. & Gauch, H.G. 1980. Detrended correspondence analysis: An improved ordination technique. - Vegetatio 42:47-58.

Leduc, A., Drapeau, P., Bergeron, Y. & Legendre, P. 1992. Study of spatial components of forest cover using partial Mantel tests and path analysis. - Journal of Vegetation Science 3:67-78.

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? - Ecology 74:1659-1673.

- & Troussellier, M. 1988. Aquatic heterotrophic bacteria: Modelling in the presence of spatial autocorrelation. - Limnology and Oceography 33:1055-1067.

- & Fortin M. J. 1989. Spatial pattern and ecological analysis. - Vegetatio 80:107-138.

- & Vaudor, A. 1991. The R package: Multidimensional analysis, spatial analysis. - Département de sciences biologiques, Université de Montréal, Montréal.

Manly B.F.J. 1991. Randomization and Monte Carlo methods in biology. - Chapman & Hall, London.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. - Cancer Reserach 27:1055-1067.

Müntzing, A. 1940. Further studies in apomixis and sexuality in *Poa*. - Hereditas 40:459-516.

- . 1966. Apomixis and sexuality in new material of Poa alpina from middle Sweden. - Hereditas 54:314-337.

Nevo, E., Beiles, A., Kaplan, D., Golenberg, E.M., Olsvig-Wittaker, L. & Naveh, Z. 1986. Natural selection of allozyme polymorphisms: a microsite test revealing ecologic genetic differentiation in wild barely. - Evolution 40:13-20.

Nordal, I. & Iversen, A.P. 1993. Mictic and monomorphic versus parthenogenetic and polymorphic - a comparison of two Scandinavia mountain grasses. - Opera Botanica 21:19-27.

Ostrom, C.W. 1978. Time series analysis: Regression techniques. - Sage university paper series on quantitative applications in the social sciences no 07-009, Sage, London.

Portnoy, S. & Willson, M.F. 1993. Seed dispersal curves: behaviour of the tail of the distribution. - Evolutionary Ecology 7:25-44.

Rice, W. R. 1989. Analyzing tables of statistical tests. - Evolution 43:223-225.

Rohlf, F.J. 1986. NTSYS-pc: Numerical taxonomy system for the IBM PC microcomputer (and compatibles). - Applied Biostatistics Inc., Setauket, New York

SAS Institute. 1990. SAS/STAT user guide, Version 6, 4th ed. - SAS Institute Inc., Cary, North Carolina.

Smouse, P.E., Long, J.E. & Sokal, R.R. 1986. Multiple regression and correlation extensions of the Mantel test of matrix association. Systematic Zoology 35:627-632.

Sneath, P.H. & Sokal, R.R. 1973. Numerical taxonomy. Freeman, San Francisco.

Sokal, R.R. 1979. Ecological parameters inferred from spatial correlograms. - In: Patil, G.P. & Rosenzweig, M. (eds.). Contemporary quantitative ecology and related econometrics. ICPH, Maryland, pp. 167-196.

Sokal, R.R. & Oden, N.L. 1978. Spatial autocorrelation in biology. Biological Journal of the Linnean Society 10:199-249.

Sokal, R.R., Jacquez, G.M., Wooten, M.C. 1989. Spatial autocorrelation analysis of migration and selection. Genetics 121:845-855.

Statistical Sciences, I. 1993. S-plus for windows version 3.2 supplement. - Statistical Sciences Inc., Seattle.

Ter Braak, C.J. 1987. CANOCO - a FORTRAN program for canonical community ordination by (partial)(detrended)(canonical) correspondence analysis, principal component analysis and redundancy analysis (ver. 2.1). TNO Inst. appl. Comp. Sci., Stat. Dept. Wageningen, Wageningen.

Trexler, J.C. & Travis, J. 1993. Nontraditional regression analyses. - Ecology 74:1629-1637.

Turkington, R. & Aarssen, L.W. 1984. Local-scale differentiation as a result of competitive interactions. In: Dirzo, R. & Sarukhan, J. (eds.). Perspectives on plant population ecology. Sinauer Associates, Sunderland, pp. 107-127.

Willson, M.F. 1992. The ecology of seed dispersal. In: Fenner, M. (ed.). Seeds: The ecology of regeneration in plant communities. CAB international, pp. 61-85.