Molecular Ecology (1999) 8, 1043–1053

# Canonical correspondence analysis for estimating spatial and environmental effects on microsatellite gene diversity in brook charr (*Salvelinus fontinalis*)

BERNARD ANGERS,*† PIERRE MAGNAN,‡ MICHEL PLANTE§ and LOUIS BERNATCHEZ†
*Département des sciences biologiques, Université du Québec à Montréal, C.P.8888, Succursale Centre-Ville, Montréal, Québec, H3C 3P8, Canada, †Département de biologie, Université Laval, Sainte-Foy, Québec, G1K 7P4, Canada, ‡Département de chimie-biologie, Université du Québec à Trois-Rivières, C. P. 500, Trois-Rivières, Québec, G9 A 5H7, Canada, §Parc National de La Mauricie, 2141, Chemin Saint-Paul, Saint-Mathieu du Parc, Québec, G0X 2 V0, Canada

**Abstract**

**The understanding of the relationships between environmental factors and evolutionary forces is of importance to preserve opportunities for the continuation of dynamic evolutionary and ecological processes. This involves the identification and quantification of the relative importance of environmental factors that may influence these processes. Nevertheless, environmental factors are generally interpreted in terms of hypothetical inferences as relationships between environmental and genetic variables are often difficult to quantify. In this study, we used a statistical framework based on canonical correspondence analysis in order to determine the relative contribution of drainage pattern and environmental factors in structuring inter- and intrapopulational genetic diversity among brook charr populations as depicted by microsatellite analysis. These procedures simultaneously analyse several sets of variables and determine their relative contribution. The results revealed the influence of drainage pattern, altitude and human-induced factors on the pattern of genetic diversity and, particularly, the important role of historical events in explaining patterns of contemporary genetic diversity among populations. The statistical framework used in this study provides an efficient way to empirically relate variations of genetic diversity and descriptive variables in general.**

*Keywords*: canonical correspondence analysis, environmental factors, genetic structure, hydrography, microsatellite

*Received 23 June 1998; revision received 23 November 1998; accepted 7 February 1999*

## Introduction

The quantification of intra- and interpopulation genetic diversity is of importance in many areas of evolutionary biology and ecology. In the context of conservation biology, such information is crucial to estimate: (i) the extent of divergence among populations for recognition of evolutionary significant units (ESU), and the preservation of genetic diversity among remnant populations (Lande 1988; Waples 1991; Moritz 1994; Bernatchez 1995); and (ii) the hierarchical organization of genetic diversity within species in order to establish priority of management programmes (Ryman 1981; Frankham 1995; Moritz 1995). Preserving opportunities for the continuation of dynamic evolutionary and ecological processes is also recognized as an ultimate goal of conservation biology (Moritz 1995). Understanding the extent, causes and significance of genetic distribution in space and time is fundamental to achieve this objective. This namely implies to identify and quantify the relative importance of environmental factors that may influence the intensity of evolutionary forces. For instance, the extent of genetic differentiation among subdivided populations may differ as genetic drift will be largely influenced by their effective population size ($N_E$), which itself may be controlled by the carrying capacity of habitats. The understanding of the relationships between environmental factors and evolutionary processes also

Correspondence: P. Magnan. Fax: +1819-376-5084; E-mail: pierre_magnan@uqtr.uquebec.ca

require to distinguish the effects of dynamic natural processes from that of recent human-induced ecosystem modifications (Hindar *et al*. 1991; Carvalho 1993; Ryman *et al*. 1995). Similarly, it is of importance to measure variation related to spatial component, as the extent of gene flow may be modulated by habitat structure (Avise 1992).

Few empirical studies have been specifically designed to quantify the amount of intra- and interpopulation genetic variance imputable to the effects of individual environmental factors. Most frequently, empirical studies remain descriptive, with only speculative inferences on factors explaining organization of genetic structure. One reason why the links between environmental and genetic variables have not often been more rigorously quantified is the paucity of analytical tools that are required. For instance, Mantel's (1967) permutation test can be used to detect relationships between multivariate genetic data and other factors, such as genetic differentiation and geographical distance among populations (Roy *et al*. 1994; Estoup *et al*. 1998). Despite the usefulness of this test, it cannot be used to estimate the amount of the variation of genetic data explained by the independent variables, and how different sets of variables can be related simultaneously (Borcard *et al*. 1992). These issues, however, parallel those encountered in community ecology studies establishing the effects of environmental variables on species diversity. These studies make use of quantitative procedures that simultaneously analyse several sets of environmental variables and determine their relative contribution in structuring species communities (e.g. Legendre & Troussellier 1988; Borcard *et al*. 1992; Magnan *et al*. 1995; Rodriguez & Magnan 1995). One of the most efficient statistical tools to relate variation in species composition to different sets of predictor variables is the canonical correspondence analysis (CCA) (ter Braak 1988a). CCA is an eigenvalue ordination technique designed for direct analysis of the relationships between multivariate ecological data tables (ter Braak 1986; Legendre & Legendre 1998). It integrates regression and ordination techniques, and avoids the assumption of linearity in the response of species to environmental variables. Interestingly, ter Braak (1988a) suggested this analysis to relate variation of genetic components to ecological factors. To our knowledge, the efficiency of this approach has not been evaluated empirically in this field.

The objective of this study is to investigate the usefulness of CCA to estimate the individual effects of variables describing the habitat of populations on intra- and interpopulation genetic diversity. In a recent study, we documented the fine-scale genetic population structure of brook charr, *Salvelinus fontinalis*, populations in La Mauricie National Park (LMNP, Québec, Canada) using microsatellite loci (Angers & Bernatchez 1998). Populations were organized in seven well differentiated groups, which were not fully congruent with the hydrographic system; based on hierarchical analysis of population structure, more important genetic differentiation was observed among these groups than among drainage systems. The amount of intrapopulation diversity estimated from number of alleles and heterozygosity at individual loci was also highly variable among populations belonging to either the same or different groups. Altogether, these results indicated that the drainage pattern was a major, although not the unique, factor responsible for the extent of genetic partitioning among populations. Detailed ecological data and information on human activities are available for the majority of lakes in LMNP, over the time period since Canada Park Services has managed this territory (1970). The large diversity of lake habitat conditions found in LMNP (e.g. lake sizes, low altitude vs. high altitude), as well as the potential differential effects of human activities (e.g. sport fishing allowed for some lakes but prohibited for others, past stocking or not) may have influenced the genetic structure of brook charr populations. LMNP therefore provides an appropriate system to assess the usefulness of CCA in the statistical framework proposed by Magnan *et al*. (1995) in order to investigate: (i) what variables, either drainage pattern or environmental (altitude, lake characteristics, human activities) can best account for the extent of genetic diversity among and within populations; and (ii) the relative contribution and statistical significance of these variables in explaining the various components of genetic variance.

## Materials and methods

### *Study site and genetic data*

The La Mauricie national park (LMNP) is located in the middle south of Québec, Canada (46°46′ N 73°00′ W). Brook charr from 26 lakes (mean sample size = 30 individuals), and representative of seven different minor drainage systems (Fig. 1) were analysed with five microsatellites; SFO-8, SFO-12, SFO-18 and SFO-23 designed for brook charr (Angers *et al*. 1995) and MST-85 from brown trout, *Salmo trutta* (Presa & Guyomard 1996). Details of the genetic analyses and results have been previously presented (Angers & Bernatchez 1998).

### *The approach*

CCA integrates ordination and multiple regression techniques. In conventional ordination techniques such as principal component and correspondence analyses, the variation in ecological data is reduced in a few dimensions, in an ordination diagram mapping dependent and independent variables. The patterns obtained are usually interpreted in relation to external data such as correlation

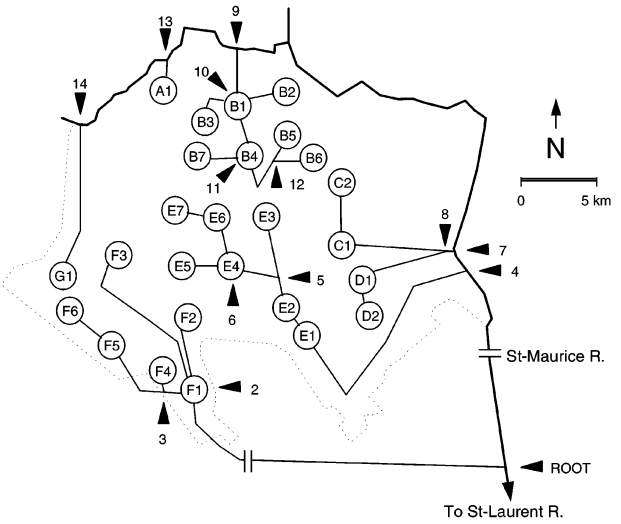


**Fig. 1** Schematic representation of geographical locations and interconnections among the study lakes in the La Mauricie national park. Codes correspond to names of lakes as published previously (Angers & Bernatchez 1998). Nodes are indicated by black arrows.

coefficients between independent variables and ordination axes or with multiple regression of the ordination axes on independent variables (ter Braak 1988a; Legendre & Legendre 1998). One difficulty with these techniques is that the ordination axes are just particular orthogonal directions in the ordination diagram and that other directions may well be better related to the environmental variables (ter Braak 1988a). Contrasting with conventional techniques, in CCA a regression model is inserted in the ordination model with the result that the ordination axes appear in order of variance explained by linear combinations of independent variables (ter Braak 1988a). The multiple regression thus constrains the ordination scores (ter Braak 1988a). As the tests implemented in CCA are based on permutation techniques, there are no specific assumptions regarding data distribution. It is, however, preferable to log-transform variables that contain extreme values to reduce their effect on central tendency. Finally, CCA assumes a unimodal relationship between the dependent and independent variables, in contrast to ordination techniques which assume a linear or monotonic relationship between them (ter Braak 1988a).

In the present analysis, allele frequencies, number of alleles and the expected heterozygosity were related separately to two sets of descriptive variables: drainage pattern and environmental variables. In contrast to the Mantel's test that uses two or more similarity or distance matrices, CCA relates real data organized in a 'variable-by-population' matrix. Of particular interest in population genetics is the fact that CCA is suited to cope with matrices containing many empty cells (see ter Braak 1988a; section 6.3), a situation commonly encountered with allele frequency data. Basically, CCA maps a dependent variables matrix to a descriptive variables matrix, maximizes the mutual fit between them and then assesses the quality of that fit. CCA analysis allows the independent variables to be selected that contributed significantly to the explanation of the variations in dependent variables, using a forward selection procedure, estimates the contribution of independent variables, and assesses the statistical significance of these relations. Finally, CCA and partial CCA was used to extract the amount of variation shared by both sets of descriptive variables (Borcard *et al*. 1992; Borcard & Legendre 1993).

### *Dependent variables*

Genetic data represent the dependent variables in this analysis. In order to take into account intra- and inter-population components of genetic diversity, data entries were structured into seven separate matrices. Variation of the genetic diversity among populations was inferred from the variation of the relative abundance of each allele at a given loci (allele frequencies). A different matrix was constructed for each of the five loci and took the form of an 'alleles of a given locus by lake' matrix. On the other hand, variation (among populations) of the intrapopulation diversity was inferred from the variation of the number of alleles ($A$) or the expected heterozygosity ($H_E$) by loci. Thus, genetic diversity within populations took the form of a 'level of variation of loci by lake' matrix (Appendix I). While both intrapopulational diversity estimators are calculated from allele frequencies, this information is different in that two populations may carry the same intrapopulational diversity without sharing any common allele. Furthermore, the number and frequency of alleles may vary substantially for the same expected heterozygosity values in situations where populations are not at mutation-drift equilibrium.

### *Independent variables*

Two matrices, one for drainage pattern and the other for environmental variables, were constructed to discriminate the influence of the hydrographic network from other factors on genetic structuring. The drainage pattern matrix represents the organization of populations by the pattern of lake interconnections, as suggested by Magnan *et al*. (1995), and was constructed as follows. A hydrographic network mapping the connections among lakes was built (Fig. 1). Each branching point represents a node and a number was assigned to each of them. The root is defined as the common branching point for all of the populations. This is analogous to methods used in phylogenetic reconstruction (e.g. Kluge & Farris 1969). Each

lake is then characterized by the sequence of nodes from the root to its position along the hydrographic tree (for instance, lake G1 had the following sequence from the root: nodes 4, 7, 9, 13 and 14; Fig. 1). A 'lake-by-nodes' matrix was then constructed by assigning, for each lake, a value of 1 to all nodes connecting it to the root and 0 to all other nodes (for instance, lake G1 was described by the sequence 10010010100011 for nodes 1–14, respectively). Such a matrix translated the numerical coding of the drainage basin pattern by the hydrographic tree. According to Magnan *et al*. (1995), lakes forming a node (e.g. F1, E1, E2 in Fig. 1) were coded 0 for that node, because they were not located downstream from that node, while lakes positioned on branches between two bifurcations (e.g. F5 and F6) were assigned the same node values as they belong to the same subdrainage.

Other variables expected to influence genetic diversity were represented by three different sets of estimators pooled into a single environmental matrix (Appendix II). Altitude of the lakes, which is highly variable (201–368 m) among and within drainages, may impose different environmental conditions and act as a barrier to gene flow. In the absence of detailed census data for all lakes studied, we a priori used lake morphological parameters (surface area, mean depth, perimeter, and volume) as a surrogate of population size (Johnson *et al*. 1992). Activities expected to have potentially altered genetic diversity (human and beaver dams, sport fishing since 1970, past occurrence of logging, stocking of domestic brook charr) were coded either as 1 or 0 for the occurrence (actual or historical) or absence, respectively, of a given activity. In addition, fish species composition was considered in this matrix (Magnan 1989), and coded by the total number of species reported in a given lake (data on their abundance not being available). Frequency distribution of the predictive variables were examined graphically and transformations were applied when necessary to reduce the effect of extreme values on central tendency.

### *Statistical analyses*

The statistical approach used in the present study is detailed in Magnan *et al*. (1995). Briefly, in order to determine the variation of dependent variables related to independent variables, each of the seven matrices encoding for genetic variation among and within populations was related to the spatial and environmental variables separately, by CCA, using version 3.1 of the program CANOCO (ter Braak 1988b; program available from C. J. F. ter Braak, Agricultural Mathematics Group, TNO Institute for Applied Computer Science, Box 100, NL-6700 AC Wageningen, The Netherlands). For both drainage pattern and environmental variables, the variables that contributed most to the explanation of the variations were selected using a forward selection procedure available in CANOCO, with a cut-off point of $P = 0.10$ (i.e. alpha-to-enter, as in multiple regression analysis), based on 1000 Monte Carlo permutations (see ter Braak 1988b). Contribution of each set of variables was estimated independently using the sum of canonical eigenvalues and the statistical significance was assessed by Monte Carlo permutation tests of the sum of all eigenvalues, using 1000 permutations as implemented in CANOCO (ter Braak 1990).

As environmental variables may exhibit spatial heterogeneity, the variation explained by both sets of variables may be partially redundant. In order to determine whether interpretation of the role of environmental variables in shaping genetic diversity could be biased by their dependence with the drainage pattern, variation in genetic diversity was analysed against the drainage pattern and environmental variables together, using the method of variation partitioning proposed by Borcard *et al*. (1992) and Borcard & Legendre (1993). The total amount of genetic variance explained by independent variables was consequently partitioned into four independent components using CCA and partial CCA (ter Braak 1986; 1988c). (i) Following the estimation of variation accounted for by the environmental variables using CCA, the pure component explained by environmental variables is computed by removing the effect of drainage pattern using partial CCA. (ii) Similarly, the pure component explained by drainage pattern is computed by removing the effect of the environmental variables. (iii) The shared component between the drainage pattern and environmental variables corresponds to the fraction removed by partial CCA from the total variation explained by a set of predictors. (iv) The unexplained component of genetic variance is obtained by subtracting the pure environmental component, the pure drainage pattern component, and the shared component from the total genetic variance. The statistical significance of the environmental and drainage pattern components of genetic variance was evaluated by Monte Carlo permutation tests of the sum of all eigenvalues, using 1000 permutations as implemented in CANOCO (ter Braak 1990).

To correct for multiple uses of the same set of observations (drainage pattern and environmental variables) we applied the sequential Bonferroni correction (Rice 1989) starting at $a/k$ where $a = 0.05$, $k = 7$, with the number of different matrices simultaneously tested against either the environmental or drainage pattern matrix.

## Results

### *Diversity among populations*

The CCA revealed the strong influence of drainage pattern on the interpopulation component of genetic diversity in

**Table 1** Summary statistics for canonical correspondence analysis of genetic diversity and drainage pattern variables (nodes in the drainage pattern; see text and Fig. 1). For each locus, eigenvalues are given in parentheses. For some loci, the selection of a single variable provides eigenvalues on a single axis

| Locus<br>Predictor | Canonical coefficients<br>Axis 1 | <br>Axis 2 | *P* model | Explained variation (%) |
|---|---|---|---|---|
| Interpopulation genetic diversity | | | | |
| SFO-8 | (0.744) | (0.666) | 0.001 | 52.4 |
| N 3 | 0.52 | −0.26 | | |
| N 4 | −0.05 | −0.06 | | |
| N 8 | 0.65 | −0.32 | | |
| N 13 | 0.21 | −0.28 | | |
| N 14 | 0.35 | 0.87* | | |
| SFO-12 | (0.801) | (0.528) | 0.001 | 65.7 |
| N 3 | 0.58 | −0.53 | | |
| N 4 | −0.31 | −0.64 | | |
| N 8 | 0.25 | 0.60 | | |
| N 14 | 0.50 | −0.06 | | |
| SFO-18 | (0.494) | (0.252) | 0.001 | 50.4 |
| N 4 | −0.67* | 0.22* | | |
| N 8 | 0.46* | 0.36* | | |
| N 14 | 0.24* | 0.20* | | |
| SFO-23 | (0.837) | (0.559) | 0.001 | 56.4 |
| N 3 | −0.05* | 0.07 | | |
| N 4 | −0.06* | 0.08 | | |
| N 8 | 0.01 | −0.02 | | |
| N 13 | 0.11* | 0.05 | | |
| N 14 | −0.08* | −0.03 | | |
| MST-85 | (0.783) | (0.682) | 0.001 | 69.5 |
| N 3 | −0.09 | −0.02 | | |
| N 4 | −0.11 | −0.02 | | |
| N 7 | −0.01 | 0.01 | | |
| N 12 | −0.01 | −0.01 | | |
| N 13 | −0.02 | 0.11 | | |
| N 14 | 0.01 | −0.08 | | |
| Intrapopulation genetic diversity | | | | |
| Number of alleles | (0.009) | — | 0.009 | 12.68 |
| N10 | 0.097* | — | | |
| Heterozygosity | (0.036) | 0.002 | 0.001 | 28.79 |
| N5 | 0.101* | 0.042 | | |
| N9 | 0.196* | −0.011 | | |

*$P < 0.05$.

brook charr (Table 1). Seven nodes (representing the drainage pattern variables) out of 14 considered in the CCA were retained by the forward selection procedure as significant predictors of the variation in allele frequencies. These accounted for 50.4% (SFO-18) to 69.5% (MST-85) of the total variation in allele frequencies, depending on loci. It is noteworthy that nodes 4 and 14 were selected at all loci, and that nodes 3, 8 and 13 were, respectively, selected for four, four, and three loci out of five (Table 1). The relationships between allele frequencies and drainage pattern variables was statistically significant at all loci ($P < 0.001$).

A significant relationship between environmental variables and allele frequencies variation was also detected, although to a much lower extent than drainage pattern variables. The environmental variables selected accounted for 7.8% (SFO-8) to 31.3% (MST-85) of the total variation in allele frequencies, depending on loci (Table 2). Only three out of 11 environmental variables considered in the CCA were retained as significant predictive variables. Lake altitude was selected for all except the SFO-18 locus, past stocking with domestic brook charr was retained for two loci (SFO-18 and MST-85), and number of fish species in lakes was retained for MST-85 (Table 2). The relationships between allele frequencies and environmental variables was statistically significant at all loci ($0.001 < P < 0.043$, Table 2).

Ordination scores were used to create a scattergram that simultaneously represents alleles of a given locus, with populations and effects of variables retained. The

| Locus Predictor | Canonical coefficients Axis 1 | Axis 2 | *P* model | Explained variation (%) |
|---|---|---|---|---|
| Interpopulation genetic diversity | | | | |
| SFO-8 | (0.436) | — | 0.001 | 7.8 |
| Altitude | −0.66* | — | | |
| SFO-12 | (0.506) | — | 0.003 | 22.9 |
| Altitude | −0.71* | — | | |
| SFO-18 | (0.152) | — | 0.043 | 9.3 |
| Stocking | 0.39 | — | | |
| SFO-23 | (0.392) | — | 0.001 | 10.3 |
| Altitude | −0.69* | — | | |
| MST-85 | (0.459) | (0.260) | 0.001 | 31.3 |
| Altitude | −0.44* | 0.50* | | |
| Stocking | −0.42* | 0.14 | | |
| No. of species | 0.42* | 0.57* | | |
| Intrapopulation genetic diversity | | | | |
| No. of alleles | (0.009) | — | 0.010 | 11.27 |
| altitude | −0.096* | | | |
| Heterozygosity | (0.021) | — | 0.009 | 15.91 |
| altitude | −0.146* | | | |

**Table 2** Summary statistics for canonical correspondence analysis of genetic diversity and environmental variables. For each locus, eigenvalues are given in parentheses. For some loci, the selection of a single variable provides eigenvalues on a single axis

*$P < 0.05$.

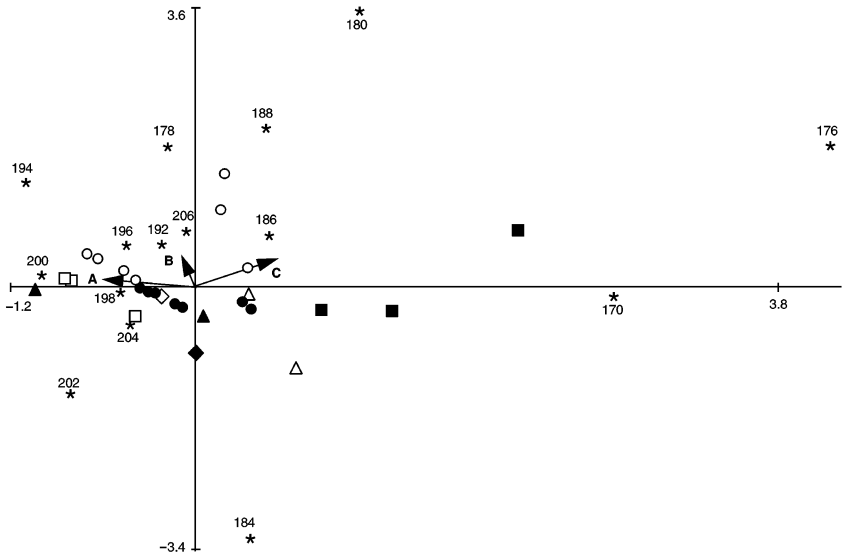


**Fig. 2** Canonical correspondence ordination of populations (A1 = filled diamond; B1–B7 = filled circles; C1–C2 = filled triangles; D1–D2 = open triangles; E1–E7 = open circles; F1–F3 = filled squares; F4–F6 = open squares; G1 = open diamond), alleles (asterisk) and environmental variables (arrows) for the locus MST-85. The environmental variable arrows are drawn from the centroid of the population dispersion; A = altitude; B = stocking; C = number of species.

ordination diagram of MST-85 (Fig. 2) showed that populations are mainly discriminated along the first axis (altitude and number of species) and to a lesser extent by the second axis (past stocking). The usefulness of such a plot is that it shows relationships between dependent and descriptive variables. As an example, alleles 194 and 200 predominated in lakes of higher altitude and with a lower number of species than alleles 170 and 176. Similarly, alleles 178, 180 and 188 showed their highest frequencies in lakes that were stocked in the past. These interpretations, based on the ordination plot, could be confirmed by direct observation of allele frequencies in the lakes concerned. The effect of given environmental variables cannot always be separated out because of their multicollinearity (ter Braak 1988a). For instance, in the present locus, there is an inverse relationship between altitude and number of fish species, clearly illustrated by the opposite direction of arrows.

Results of the partial CCA partitioning of the variation (Table 3) indicated that the pure drainage pattern

**Table 3** Partition of the total variation into four independent components: pure environmental, pure drainage pattern, shared (spatial component of environmental variation) and unexplained

| | Pure environmental | Pure drainage pattern | Shared | Unexplained |
|---|---|---|---|---|
| Interpopulation genetic diversity | | | | |
| SFO-8 | 5.3 (0.006) | 49.9 (0.001) | 2.5 | 42.3 |
| SFO-12 | 4.3 (0.04 NS) | 47.0 (0.001) | 18.7 | 30.1 |
| SFO-18 | 6.1 (0.02 NS) | 47.3 (0.001) | 3.2 | 43.5 |
| SFO-23 | 3.3 (0.07 NS) | 49.4 (0.001) | 6.9 | 40.4 |
| MST-85 | 9.0 (0.003) | 47.1 (0.001) | 22.3 | 21.6 |
| Intrapopulation genetic diversity | | | | |
| No. of alleles | 9.9 (0.021 NS) | 11.3 (0.020) | 1.4 | 77.4 |
| Heterozygosity | 5.3 (0.143 NS) | 18.2 (0.004) | 10.6 | 65.9 |

Numbers in parentheses refer to *P*-values after 1000 permutations with a Monte Carlo permutation test.
NS indicates nonsignificant results after sequential Bonferroni correction.

component significantly ($P < 0.001$) contributed in almost the same proportion (between 47.0% and 49.9%) to the total variation in allele frequencies at all loci. The pure environmental component to the total variation in allele frequencies was much less important (between 3.3% and 9.0%) and was significant for only two loci (SFO-8 and MST-85), following the sequential Bonferroni correction. The component of variation shared by both drainage pattern and environmental variables varied between 2.5% and 22.3% depending on locus. Altogether, spatial and environmental variables accounted for between 56.6% and 78.4% of the total variation in allele frequencies depending on loci (Table 3).

### *Diversity within populations*

The CCA also indicated a significant effect of the drainage pattern on the variation of the intrapopulation genetic diversity, but to a lesser extent than on interpopulation diversity (Table 1). These variables accounted for 12.7% of the variance in numbers of alleles ($A$) and 28.8% of the variance in heterozygosity ($H_E$) within populations. Three nodes were retained by the forward selection procedure: nodes 5 and 9 for heterozygosity and node 10 for the number of alleles. Both $A$ and $H_E$ were significantly related to drainage pattern ($0.001 < P < 0.009$, Table 1). A weak, but significant effect of environmental variables on intrapopulation diversity was also detected (Table 2). A single predictor (altitude) was retained by the forward selection procedure and accounted for 11.3% of the variance in numbers of alleles and 15.9% of the variance in heterozygosity within populations. Both $A$ and $H_E$ and environmental variables were significantly related ($0.009 < P < 0.010$, Table 2).

Results of the partial CCA of variance partitioning (Table 3) indicated that the pure drainage pattern component contributed significantly to the total variation of both $A$ (11.3%) and $H_E$ (18.2%). The contribution of the pure environmental component, however, was not significant for either $A$ or $H_E$ following the Bonferroni correction. An important proportion of the explained variance in heterozygosity was shared by the drainage pattern and environmental variables (10.4%). Altogether, both sets of descriptive variables accounted for a relatively low portion of the total variance of intrapopulation genetic ($A = 22.6\%$ and $H_E = 34.1\%$, Table 3).

## Discussion

### *Diversity among populations*

The objective of this study was to investigate the usefulness of the CCA to estimate the contribution of drainage pattern and environmental factors on genetic variation among salmonid populations. Results of the CCA are congruent with previous observations (Angers & Bernatchez 1998) and clearly indicated the strong influence of drainage pattern on the variation of genetic diversity among populations.

It is noteworthy that a similar trend is observed for all loci when considered separately, as five out of the seven nodes selected to explain diversity among populations were significant for nearly all the five loci. Significant nodes retained as predictors could be interpreted as the discontinuity of the genetic diversity between populations located upstream and downstream to these nodes. In most cases, sections of the hydrographic system defined by the selected nodes corresponded to distinct drainages. Nodes 8, 13 and 14 referred to the drainages harbouring populations C1-C2-D1-D2, A1 and G1, respectively, while node 3 referred to the subdrainage including F4–F6 populations. Node 4 referred to the E drainage. As no nodes were selected between node 4 and B drainage (except for the MST-85 locus), B populations shared the same

node sequence as E populations. Population assemblages included within these drainage sections are consistent with those defined using genetic distances and phylogenetic inferences (Angers & Bernatchez 1998). Congruence between both approaches therefore strongly supports the efficiency of the CCA-based method to relate heterogeneity of allele frequencies among populations to drainage pattern.

CCA also revealed the influence of environmental factors as a significant environmental trend (pure environmental component) in the allelic composition remained at two loci after accounting for the main effect measured by the drainage pattern. Among environmental factors that contribute to explain the variation of allele frequencies, altitude accounted for a significant portion of the genetic diversity among populations. Such a pattern could be expected along hybrid zones or multiple phase invasion related to the historical availability of habitats. For instance, strong altitudinal distribution was observed in brown trout in northern UK (Hamilton *et al*. 1989) and this pattern was attributed to a two-phase colonization event, where a first group invaded regions of high altitude, while the following group did not have access to these zones. This analysis has also allowed us to detect a small but significant effect of human activities on the variation of genetic diversity. The impact of stocking has been evidenced by the higher frequencies of characteristic alleles in stocked vs. nonstocked lakes. Stocking may rapidly disturb the genetic integrity of populations, by introduction of new alleles and homogenization of well structured populations (Hindar *et al*. 1991; Kriegler *et al*. 1995; Largiadèr & Scholl 1995). The statistical framework used here therefore offers a powerful alternative to more traditional approaches, because as well as stocking effects (environmental effects) being detected, its importance in explaining a proportion of overall genetic variance among populations was quantified.

A major point highlighted by CCA is that a large proportion of the allele frequency variation could be related to both drainage pattern and environmental variables, the shared component. For instance, altitude between subdrainage sections harbouring populations highly differentiated could be quite different (e.g. F1–F3 = 217–249 m; F4–F6 = 315–343 m). Altitude could be considered as another dimension characterizing the drainage pattern, and change in altitude may represent an important step between lakes. Such results highlight the potential contribution of altitude in the distribution of the genetic diversity among populations as well as the hydrographic pattern itself. It is therefore important in such circumstances to consider interactions between selected variables in order to not overemphasize or misinterpret the effects of one or more factors (Magnan *et al*. 1995). Using mutational information of microsatellite and genetic distance linear with time (Goldstein *et al*. 1995), the estimated time of divergence of population assemblages observed in the LMNP largely predated their contemporary hydrographic distribution (Angers & Bernatchez 1998). Hence, both hydrographic pattern and altitude may have strongly modulated the colonization processes following Pleistocene glacier retreat.

### *Diversity within populations*

The important variation in the extent of the number of alleles and heterozygosity reported by Angers & Bernatchez (1998) suggested a priori that environmental and/or human-induced factors affected the genetic diversity of founding populations in a lake-specific manner. No such trend, however, was detected, and none of the variables included in the analysis was significantly correlated to genetic diversity. Hence, a large part of intrapopulation genetic variance remained unexplained, in contrast to the large proportion of genetic diversity explained among populations. In the absence of detailed demographic information for lakes studied, we used lake dimensions as a surrogate of population census, with the expectation that this could partly explain variation in genetic diversity within lakes. Our results, however, revealed that lake size was clearly not a good indicator of brook charr population capacity to maintain a given level of genetic diversity. Ranges in both lake size (2.63–512 ha) and genetic diversity (total number of alleles per population: 10–43, $H_E$: 0.17–0.79) were expected to be wide enough to detect an existing relationship between these variables. Hence, the lack of resolution was probably not involved and rather suggests a lack of correspondence between lake size and genetic diversity. Lack of correspondence between the $N_E$ estimation based on the temporal method and lake size has been reported in brown trout (*Salmo trutta*) populations (Jorde & Ryman 1996). Populations from lakes that are similar in size and other apparent characteristics showed similar levels of genetic variation based on heterozygosity, but carried estimated $N_E$ varying from 52 to 480 individuals (Jorde & Ryman 1996).

Significant relationships between drainage pattern and genetic diversity within population suggests that the observed values of genetic diversity were partially representative of founding populations. Nodes selected for $H_E$ referred to a single phylogenetic lineage (populations B1–B7, E3–E7, Angers & Bernatchez 1998). Hence, the pattern of intrapopulation genetic diversity is partly representative of historical diversity rather than the resulting effect of contemporary demographic factors. This suggests that mutation/drift equilibrium have not been reached for all populations and loci. Microsatellite results must also be interpreted cautiously, as several aspects of their modes of evolution remain partially understood

(Jarne & Lagoda 1996; Feldman *et al*. 1997; Goldstein & Pollock 1997). This means that one cannot rule out the possibility that a given locus may have evolved under different constraints in different populations. As an example, size homoplasy (identical alleles from distinct mutational events) may frequently occur at microsatellite loci and underestimate the real diversity (Angers & Bernatchez 1997; Viard *et al*. 1998). It has also been shown that for a given locus, propensity to generate mutations could be different among alleles (Jin *et al*. 1996; Primmer *et al*. 1996). As these processes are purely stochastic, a larger number of loci have to be considered to analyse variation of the genetic diversity within populations (Cornuet & Luikart 1996). Finally, according to the simplification of the system analysed here, a large proportion of unexplained variance could also be the result of parameters not being properly quantified. For instance, fine descriptors of recent population fluctuations could be analysed (Cornuet & Luikart 1996). This means that such analyses possibly require the inclusion of additional descriptive ecological variables, and stresses the need for further fusion of ecology with population genetics.

## Conclusions

The statistical framework used in this study allowed us to detect the influence of drainage pattern, altitude and human-induced factors on the pattern of genetic diversity among and within populations (based on allele frequencies). The main interest of such an analysis was to empirically estimate the relative importance of these variables, which had not been attempted so far in genetic studies of population differentiation. This approach can not, however, use the genetic information based on microsatellite length variation within and between samples. In this context, alternative methods such as AMOVA and partial correlation analysis performed on distance matrix could be used to complement CCA, to make a full use of information contained in molecular data. We conclude that despite its potential limitations in explaining the pattern of intrapopulation genetic variance in La Mauricie National Park, the statistical framework used in this study provides an efficient way to empirically relate variations of genetic diversity and descriptive variables. The usefulness of this analysis in population genetics could therefore be valuable for conservation biology to investigate factors responsible for population divergence, as well as in fundamental research to orientate and design further experiments.

## Acknowledgements

We thank Drs Marco Rodríguez and Laurent Excoffier, and three anonymous referees who gave helpful comments on earlier versions of the manuscript. This work was financially supported by Parks Canada, and from grants from the Natural Sciences and Engineering Research Council of Canada to L. Bernatchez and P. Magnan.

## References

Angers B, Bernatchez L (1997) Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Molecular Biology and Evolution*, **14**, 230–238.

Angers B, Bernatchez L (1998) Combined use of SMM and non-SMM methods to infer fine structure and evolutionary history of closely related brook charr (*Salvelinus fontinalis*, Salmonidae) populations from microsatellites. *Molecular Biology and Evolution*, **15**, 143–159.

Angers B, Bernatchez L, Angers A, Desgroseillers L (1995) Specific microsatellite loci for brook charr (*Salvelinus fontinalis* Mitchill) reveal strong population subdivision on a microgeographic scale. *Journal of Fish Biology*, **47** (Suppl. A), 177–185.

Avise JC (1992) Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *OIKOS*, **63**, 62–76.

Bernatchez L (1995) A role for molecular systematics in defining evolutionarily significant units in fishes. In: *Evolution and Aquatic Ecosystem: Defining Units in Population Conservation*. (ed. Nielsen JL), pp. 114–132. American Fisheries Society Symposium, 17, Bethesda, Maryland.

Borcard D, Legendre P (1993) Environmental control and spatial structure in ecological communities, with an example on Oribatid mites (*Acari*, *Oribatei*). *Journal of Environment Statistics*, **1**, 55–76.

Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.

ter Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis in ecology. *Ecology*, **67**, 1167–1179.

ter Braak CJF (1988a) CANOCO—a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal component analysis and redundancy analysis (Version 2.1). Technical report LWA-88-02, Agricultural Mathematics Group, AC Wageningen, The Netherlands.

ter Braak CJF (1988b) CANOCO—an extension of DECORANA to analyze species-environment relationships. *Vegetatio*, **75**, 159–160.

ter Braak CJF (1988c) Partial canonical correspondence analysis. In: *Classification and Related Methods of Data Analysis*. (ed. Block HH), pp. 551–558. North Holland Press, Amsterdam.

ter Braak CJF (1990) Update notes: CANOCO, Version 3.10. Agricultural Mathematics Group, Wageningen, The Netherlands (Mimeo.).

Carvalho GR (1993) Evolutionary aspects of fish distribution: genetic variability and adaptation. *Journal of Fish Biology*, **43** (Suppl. A), 53–73.

Cornuet J-M, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Estoup A, Rousset F, Michalakis Y, Cornuet J-M, Adriamanga M, Guyomard R (1998) Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). *Molecular Ecology*, **7**, 339–353.

Feldman MW, Bergman A, Pollock DD, Goldstein DB (1997) Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics*, **145**, 207–216.

Frankham R (1995) Conservation genetics. *Annual Review in Genetics*, **29**, 305–327.

Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity*, **88**, 335–342.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the USA*, **92**, 6723–6727.

Hamilton KE, Ferguson A, Taggart JB, Tómasson T (1989) Post-glacial colonization of brown trout, *Salmo trutta* L. Ldh-5 as a phylogeographic marker locus. *Journal of Fish Biology*, **35**, 651–664.

Hindar K, Ryman N, Utter F (1991) Genetic effects of cultured fish on natural fish populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **48**, 945–957.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Evolution and Ecology*, **11**, 424–429.

Jin L, Macaubas C, Hallmayer J, Kimura A, Mignot E (1996) Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. *Proceedings of the National Academy of Sciences of the USA*, **93**, 15285–15288.

Johnson SL, Rahel FJ, Hubert WA (1992) Factors influencing the size structure of brook trout populations in beaver ponds in Wyoming. *North American Journal of Fisheries Management*, **13**, 118–124.

Jorde PE, Ryman N (1996) Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change of allele frequencies. *Genetics*, **143**, 1369–1381.

Kluge AG, Farris JS (1969) Quantitative phyletics and the evolution of anurans. *Systematic Zoology*, **18**, 1–32.

Kriegler FJ, McCracken GF, Habera JW, Strange RJ (1995) Genetic characterization of Tenessee brook trout populations and associated management implications. *North American Journal of Fisheries Management*, **15**, 804–813.

Lande R (1988) Genetics and demography in biological conservation. *Science*, **241**, 1455–1460.

Largiadèr CR, Scholl A (1995) Effects of stocking on the genetic diversity of brown trout populations of the Adriatic and Danubian drainages in Switzerland. *Journal of Fish Biology*, **47** (Suppl. A), 220–225.

Legendre P, Legendre L (1998) *Numerical Ecology. Development in Environmental Modeling, 20*. (2nd edn). Elsevier, Amsterdam.

Legendre P, Troussellier M (1988) Aquatic heterotrophic bacteria: Modeling in the presence of spatial autocorrelation. *Limnology and Oceanography*, **33**, 1055–1067.

Magnan P (1989) The impact of cyprinid and catastomid introductions on brook charr, *Salvelinus fontinalis*, populations: a review. *Physiol. Ecol. Japan*, Special Volume **1**, 337–356.

Magnan P, Rodríguez MA, Legendre P, Lacasse S (1995). Dietary variation in a freshwater fish species: relative contribution of biotic interactions, abiotic factors, and spatial structure. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 2856–2865.

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

Moritz C (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Evolution and Ecology*, **9**, 373–375.

Moritz C (1995) Uses of molecular phylogenies for conservation. *Philosophical Transactions of the Royal Society of London B*, **349**, 113–118.

Presa P, Guyomard R (1996) Conservation of microsatellites in three species of salmonids. *Journal of Fish Biology*, **49**, 1326–1329.

Primmer CR, Ellegren H, Saino N, Moller AP (1996) Directional evolution in germline microsatellite mutations. *Nature Genetics*, **13**, 391–393.

Rice WR (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

Rodriguez MA, Magnan P (1995) Application of multivariate analyses in studies of the organization and structure of fish and invertebrate communities. *Aquatic Sciences*, **57**, 199–216.

Roy MS, Greffen E, Smith D, Ostrander EA, Wayne RK (1994) Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution*, **11**, 553–570.

Ryman N (1981) Conservation of genetics resources: experience from the brown trout (*Salmo trutta*). *Ecological Bulletin (Stockolm)*, **99**, 147–151.

Ryman N, Utter F, Laikre L (1995) Protection of intraspecific biodiversity of exploited fishes. *Reviews in Fish Biology and Fisheries*, **5**, 417–446.

Viard F, Franck P, Dubois M-P, Estoup A, Jarne P (1998) Variation of microsatellite size homoplasy across electromorphs, loci and populations in three invertebrate species. *Journal of Molecular Evolution*, **47**, 42–51.

Waples RS (1991) Pacific salmon, *Oncorhynchus* spp. & the definition of a 'species' under the Endangered Species Act. *Marine Fisheries Review*, **53**, 11–22.

---

B.A. is interested in the evolutionary biology of populations and molecular evolution and is doing a postdoctoral fellowship in France on the evolution of microsatellite sequences in Salmonid species. P.M. is interested in the structure and function of freshwater aquatic communities and trophic polymorphism in brook charr. M.P. is the aquatic habitat manager in La Mauricie national park and responsible for the survey of aquatic ecosystems. L.B.'s research interests are in the understanding of patterns and processes of organismal and molecular evolution, as well as their significance to conservation issues.

---

**Appendix I** Number of alleles and expected heterozygosity of the five microsatellites loci used to estimate the intrapopulational diversity. Refer to Fig. 1 for lake localizations

| | No. of alleles | | | | | Heterozygosity | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Lake | Sfo-8 | Sfo-12 | Sfo-18 | Sfo-23 | MST-85 | Sfo-8 | Sfo-12 | Sfo-18 | Sfo-23 | MST-85 |
| A1 | 3 | 3 | 1 | 2 | 1 | 0.42 | 0.10 | 0.00 | 0.39 | 0.00 |
| B1 | 5 | 1 | 3 | 12 | 3 | 0.67 | 0.00 | 0.65 | 0.80 | 0.35 |
| B2 | 2 | 1 | 3 | 3 | 1 | 0.21 | 0.00 | 0.55 | 0.39 | 0.00 |
| B3 | 3 | 1 | 4 | 8 | 2 | 0.35 | 0.00 | 0.62 | 0.85 | 0.19 |
| B4 | 8 | 1 | 4 | 7 | 4 | 0.56 | 0.00 | 0.52 | 0.73 | 0.34 |
| B5 | 5 | 1 | 4 | 5 | 4 | 0.54 | 0.00 | 0.59 | 0.63 | 0.50 |
| B6 | 5 | 1 | 3 | 5 | 4 | 0.57 | 0.00 | 0.55 | 0.66 | 0.48 |
| B7 | 3 | 1 | 3 | 6 | 3 | 0.48 | 0.00 | 0.54 | 0.75 | 0.49 |
| C1 | 4 | 3 | 3 | 6 | 3 | 0.43 | 0.39 | 0.38 | 0.66 | 0.27 |
| C2 | 6 | 2 | 3 | 7 | 6 | 0.54 | 0.13 | 0.34 | 0.68 | 0.45 |
| D1 | 9 | 2 | 3 | 9 | 5 | 0.81 | 0.29 | 0.60 | 0.72 | 0.62 |
| D2 | 3 | 2 | 2 | 2 | 3 | 0.52 | 0.49 | 0.10 | 0.36 | 0.55 |
| E1 | 15 | 4 | 5 | 11 | 8 | 0.86 | 0.56 | 0.76 | 0.86 | 0.61 |
| E2 | 15 | 5 | 6 | 9 | 8 | 0.86 | 0.65 | 0.80 | 0.82 | 0.74 |
| E3 | 9 | 3 | 4 | 9 | 4 | 0.76 | 0.16 | 0.65 | 0.81 | 0.48 |
| E4 | 5 | 2 | 3 | 4 | 6 | 0.51 | 0.12 | 0.38 | 0.73 | 0.32 |
| E5 | 3 | 3 | 4 | 7 | 5 | 0.16 | 0.51 | 0.53 | 0.80 | 0.37 |
| E6 | 6 | 1 | 4 | 6 | 5 | 0.73 | 0.00 | 0.65 | 0.79 | 0.66 |
| E7 | 6 | 1 | 3 | 8 | 5 | 0.72 | 0.00 | 0.59 | 0.73 | 0.72 |
| F1 | 12 | 5 | 3 | 10 | 6 | 0.85 | 0.50 | 0.51 | 0.76 | 0.58 |
| F2 | 14 | 4 | 5 | 8 | 6 | 0.87 | 0.52 | 0.46 | 0.57 | 0.37 |
| F3 | 7 | 2 | 3 | 6 | 1 | 0.62 | 0.10 | 0.32 | 0.62 | 0.00 |
| F4 | 6 | 2 | 2 | 4 | 3 | 0.66 | 0.12 | 0.46 | 0.70 | 0.45 |
| F5 | 7 | 4 | 5 | 9 | 4 | 0.42 | 0.13 | 0.57 | 0.66 | 0.21 |
| F6 | 10 | 4 | 6 | 10 | 5 | 0.72 | 0.29 | 0.67 | 0.74 | 0.46 |
| G1 | 7 | 4 | 3 | 8 | 5 | 0.75 | 0.25 | 0.58 | 0.70 | 0.38 |

**Appendix II** Variables included in the environmental matrix for CCA. Refer to Fig. 1 for lake localizations

| Lake | Altitude (m) | Perimeter (m) | Mean depth (m) | Surface area (ha) | Volume ($10^5$ $m^3$) | Human dam | Beaver dam | Sport fishing | Past logging | Past stocking | No. of species |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A1 | 272 | 2605 | 5.0 | 16.2 | 8.2 | 0 | 1 | 0 | 0 | 0 | 1 |
| B1 | 277 | 18 105 | 15.2 | 301.4 | 457.6 | 1 | 0 | 1 | 1 | 0 | 3 |
| B2 | 305 | 972 | 3.3 | 4.4 | 1.4 | 0 | 0 | 0 | 0 | 0 | 2 |
| B3 | 325 | 1981 | 9.1 | 21.2 | 19.4 | 0 | 1 | 0 | 0 | 0 | 2 |
| B4 | 282 | 15 902 | 8.8 | 264.2 | 235.0 | 1 | 0 | 1 | 0 | 0 | 3 |
| B5 | 321 | 1822 | 19.4 | 15.7 | 30.4 | 0 | 1 | 1 | 0 | 0 | 3 |
| B6 | 308 | 845 | 4.2 | 3.7 | 1.6 | 0 | 1 | 1 | 0 | 0 | 2 |
| B7 | 332 | 699 | 4.7 | 2.6 | 1.2 | 0 | 0 | 0 | 0 | 0 | 2 |
| C1 | 282 | 17 747 | 5.7 | 174.2 | 99.7 | 1 | 1 | 1 | 0 | 1 | 2 |
| C2 | 289 | 1817 | 1.6 | 3.7 | 0.6 | 0 | 1 | 0 | 1 | 0 | 2 |
| D1 | 201 | 1351 | 2.0 | 6.3 | 1.3 | 1 | 1 | 1 | 0 | 1 | 5 |
| D2 | 224 | 1862 | 7.4 | 10.8 | 8.0 | 0 | 1 | 1 | 0 | 0 | 5 |
| E1 | 259 | 2770 | 3.6 | 10.3 | 3.7 | 1 | 0 | 1 | 1 | 1 | 8 |
| E2 | 277 | 7500 | 11.4 | 111.4 | 130.1 | 1 | 0 | 1 | 0 | 1 | 8 |
| E3 | 299 | 4000 | 8.6 | 28.8 | 24.7 | 0 | 1 | 1 | 0 | 0 | 5 |
| E4 | 335 | 9557 | 8.5 | 139.5 | 118.9 | 1 | 0 | 1 | 1 | 0 | 2 |
| E5 | 368 | 1999 | 6.1 | 19.8 | 12.1 | 0 | 1 | 1 | 0 | 0 | 2 |
| E6 | 344 | 2885 | 4.9 | 23.7 | 11.6 | 0 | 1 | 1 | 0 | 0 | 2 |
| E7 | 361 | 710 | 3.6 | 3.2 | 1.2 | 0 | 0 | 1 | 0 | 0 | 2 |
| F1 | 217 | 40 536 | 15.6 | 512.8 | 801.2 | 0 | 0 | 1 | 1 | 0 | 9 |
| F2 | 221 | 987 | 3.8 | 3.4 | 1.3 | 0 | 1 | 0 | 0 | 0 | 7 |
| F3 | 249 | 2850 | 11.5 | 65.1 | 75.0 | 0 | 0 | 1 | 0 | 0 | 3 |
| F4 | 315 | 1366 | 3.1 | 8.2 | 2.5 | 0 | 1 | 1 | 0 | 0 | 1 |
| F5 | 361 | 14 620 | 13.3 | 391.5 | 521.8 | 1 | 0 | 1 | 1 | 0 | 1 |
| F6 | 363 | 6730 | 11.7 | 88.8 | 99.2 | 0 | 1 | 1 | 1 | 0 | 1 |
| G1 | 319 | 8050 | 11.9 | 59.0 | 70.7 | 1 | 0 | 1 | 0 | 0 | 2 |