Molecular Ecology (1999) **8**, 419–431

Hierarchical analyses of genetic differentiation in a hybrid zone of *Sorex araneus* (Insectivora: Soricidae)

N. LUGON-MOULIN, H. BRÜNNER, A. WYTTENBACH, J. HAUSSER and J. GOUDET
Institut de Zoologie et d'Ecologie Animale, Bâtiment de Biologie, Université de Lausanne, CH-1015 Lausanne-Dorigny, Switzerland

Abstract

Microsatellites are used to unravel the fine-scale genetic structure of a hybrid zone between chromosome races Valais and Cordon of the common shrew (*Sorex araneus*) located in the French Alps. A total of 269 individuals collected between 1992 and 1995 was typed for seven microsatellite loci. A modified version of the classical multiple correspondence analysis is carried out. This analysis clearly shows the dichotomy between the two races. Several approaches are used to study genetic structuring. Gene flow is clearly reduced between these chromosome races and is estimated at one migrant every two generations using *R*-statistics and one migrant per generation using *F*-statistics. Hierarchical *F*- and *R*-statistics are compared and their efficiency to detect inter- and intraracial patterns of divergence is discussed. Within-race genetic structuring is significant, but remains weak. F_{ST} displays similar values on both sides of the hybrid zone, although no environmental barriers are found on the Cordon side, whereas the Valais side is divided by several mountain rivers. We introduce the exact *G*-test to microsatellite data which proved to be a powerful test to detect genetic differentiation within as well as among races. The genetic background of karyotypic hybrids was compared with the genetic background of pure parental forms using a CRT–MCA. Our results indicate that, without knowledge of the karyotypes, we would not have been able to distinguish these hybrids from karyotypically pure samples.

Keywords: common shrew, exact test, genetic differentiation, hybrid zone, microsatellite, multivariate analysis

Received 11 March 1998; revision received 22 September 1998; accepted 13 October 1998

Introduction

The use of microsatellites in population genetics has grown rapidly in the last few years mainly because they have the interesting properties of being highly polymorphic and codominant (Jarne & Lagoda 1996), two features that are essential for an accurate assessment of population structure. However, this high polymorphism also implies a high mutation rate, which could deflate traditional measures of differentiation such as Wright's F_{ST} (Wright 1965). To overcome this problem, Slatkin (1995) recently proposed a statistic that better accounts for high mutation rates and specific mode of mutation. It has been suggested that new microsatellite alleles are generated by addition or deletion of one or a few repeat units (Jarne & Lagoda 1996). Under this scenario, alleles of similar sizes probably share a more recent common ancestor than alleles of very different sizes, and Slatkin (1995) devised the statistic R_{ST}, based on the variance of allele size, to account for this fact. Under which circumstances should R_{ST} be preferred to F_{ST}? One way to tackle this problem is to investigate the effect of restricted gene flow at different scales. Hybrid zones are ideal for this purpose, as their two-level hierarchical structure offers an interesting empirical opportunity to compare these alternative statistics. We present here hierarchical analyses of microsatellite-derived genotypic frequencies particularly suited to hybrid zones.

The common shrew (*Sorex araneus*, L.) shows impressive chromosomal variation, both within (polymorphism) and between populations (polytypy). Variation in the number of chromosomes and racial differentiation is principally due to fusions of acrocentric chromosomes to form metacentric chromosomes (Robertsonian fusions; Wojcik & Searle 1988). At present, 50 separate chromosome races

Correspondence: J. Goudet. Fax: +41 21 692 41 05; E-mail: jerome.goudet@izea.unil.ch

(the term race will be used for chromosome race throughout this study) are recognized (Zima *et al*. 1996) and several chromosomal hybrid zones have been discovered in this species. The majority of these hybrid zones have been studied using chromosomes only, except for a few cases where protein markers were also used (e.g. Brünner & Hausser 1996). As protein polymorphism in this species is very low (for references, see Wojcik & Wojcik 1994), no sufficiently polymorphic markers to study within-race structuring have been found so far in most of the zones studied (for references, see e.g. Brünner & Hausser 1996).

A hybrid zone involving the Valais and the Cordon chromosome races is being studied in the Alps, at Les Houches (Mont Blanc, Upper Savoy, France; Hausser *et al*. 1991; Brünner & Hausser 1996; Lugon-Moulin *et al*. 1996). There, the races Valais and Cordon meet and hybridize at a mountain river, the torrent de la Griaz. These two races are well differentiated on the basis of chromosomes (Wojcik 1993) and allozymes (Hausser *et al*. 1991; Neet & Hausser 1991). Using albumins and urinary pepsins as genetic markers, Brünner & Hausser (1996) found a significant F_{ST} at Les Houches hybrid zone. However, the F_{ST} value reflected the genetic differentiation among these races. When analysing each side of the zone separately, the authors did not find any significant intraracial structuring. In a preliminary study, Lugon-Moulin *et al*. (1996) used microsatellites in this hybrid zone. They found a significant F_{ST}, as did Brünner & Hausser (1996) with allozymes. However, intraracial genetic structuring was not studied with these markers, and therefore intraracial patterns of structuring are not clearly understood.

This aim of this study is to assess the levels of gene flow within as well as between the chromosome races. We determine the genetic constitution of karyotypic hybrids to compare it with the genetic constitution of karyotypically pure parental forms. To address these questions, several statistical approaches are used. We use the constant row total–multiple correspondence analysis (CRT–MCA), a new ordination technique with a direct link to the population genetics parameters F_{ST} (Guinand 1996). Hierarchical *F*-statistics are used in conjunction with hierarchical *R*-statistics to detect different levels of structuring. *F*- and *R*-statistics are compared and their respective efficiencies in detecting structuring at different levels is discussed. Finally, we use the exact *G*-test of differentiation for diploid populations (Goudet *et al*. 1996).

Materials and methods

Characterization of the hybrid zone

The study area is a 13-km long transect in the Arve Valley between Les Houches and Chamonix at the foot of the northern slopes of the Mont Blanc massif in the western Alps (Upper Savoy, France; Fig. 1). The transect runs through a hybrid zone between the chromosome races Valais and Cordon of *Sorex araneus*.

The karyotype of the common shrew is characterized by a sex trivalent in the male (XX/XY_1Y_2) and metacentric chromosomes *af*, *bc* and *tu*, the remaining chromosome arms *g-r* being subject to polymorphism and polytypy (arm designation following Searle *et al*. 1991).

The Cordon race has one of the most primitive karyotypes yet found in *S. araneus* (Wojcik & Searle 1988), consisting principally of acrocentric chromosomes. Only the metacentric *jl* is polymorphic, that is, *j* and *l* may be found either as acrocentrics or fused together, forming a metacentric chromosome (notation: *j/l*, the slash indicating that the Robertsonian fusion is polymorphic). This race is found westward of the torrent de la Griaz (Fig. 1) and, after 4 km, it is replaced by a sibling species, the Millet's shrew (*S. coronatus*).

The Valais race is characterized by the four metacentrics, *gi*, *hj*, *kn*, and *l/o*, the latter being polymorphic. The Valais race is found eastward of the torrent de la Griaz (Fig. 1). Its distribution eastward spreads through the passes of the Arve valley leading to the Valais region (Switzerland). On the Valais side of the hybrid zone, the topography is characterized by the presence of several mountain rivers originating from the Mont Blanc glaciers. These rivers are frequently flooded during a thaw and some of them represent wide channels with up to 20 m wide stony beds.

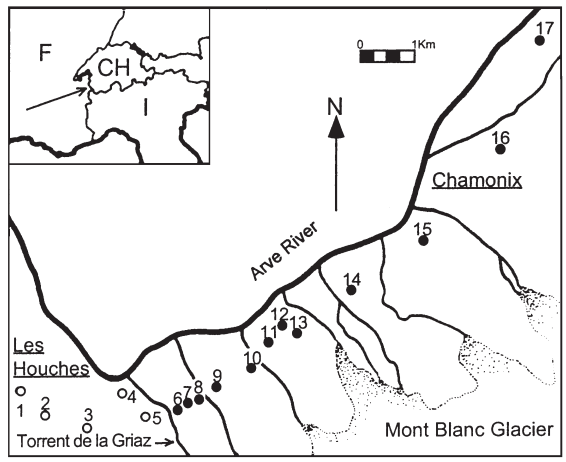


Fig. 1 Map of the study area representing the 17 sampling sites (open and closed circles, 1–17). The two chromosome races meet at the torrent de la Griaz, at Les Houches. The Cordon race is located westward of the torrent de la Griaz (open circles 1–5) while the Valais race is found eastward (closed circles 6–17). The bold line represents the Arve River and the fine lines represent mountain rivers originating from the Mont Blanc Glaciers.

Collection of individuals

A total of 273 individuals was collected with Longworth traps on a 13-km transect from September to December during the years 1992–95 (Brünner & Hausser 1996). This period corresponds to the time when juveniles disperse (Churchfield *et al*. 1995). DNA sampling was performed by toe-clipping. Phalanxes were stored in 70% ethanol and the shrews were released at their collecting site. We retained only samples consisting of a strict minimum of four individuals to carry out statistical analyses. Therefore 269 individuals from 17 sampling sites were retained for the present study (Table 1; Fig. 1). Most sites were not sampled each year. Eighty-one individuals were collected on the Cordon side and 188 on the Valais side (Table 1).

DNA isolation, amplification and electrophoresis, and karyotype analysis

Total genomic DNA was isolated from ethanol-preserved phalanxes by proteinase K digestion. This was followed by phenol and chloroform purification steps and precipitation by NH_4 acetate/ethanol (Kocher *et al*. 1989). The DNA was then washed twice with 70% ethanol, resuspended in 100 μL of double-distilled, sterile water and stored at – 20 °C.

Seven microsatellite loci, designated L9, L16, L45, L57, L62, L67 and L69, following Wyttenbach *et al*. (1997), were used. They were amplified in a Perkin-Elmer Cetus 480 or a Biometra thermal cycler. The 10 μL reaction mixture contained 50–100 ng of template DNA, 0.5 μM of each oligodeoxinucleotide primer, 80 μM of dNTP (dATP concentration 1:10), 1.5 mM $MgCl_2$ (2 mM for L57), 0.02 μL of [^{33}P]-dATP at 1000 Ci/mmol, 1× Extra*Pol*II reaction buffer, 0.1 μL of bovine serum albumin and 0.5 units of Extra*Pol*II *Taq* polymerase (Chemie Brunschwig AG). For L69, we used 200 μM of dNTP (dATP concentration 1:10).

After amplification, aliquots of the solutions were mixed to 0.5 volumes of formamide loading buffer. The samples were heated (80–90 °C, 2 min) before being run on a denaturing polyacrylamide gel (6%, 8 M urea). Fixation, drying and autoradiography were realized following standard procedures (Sambrook *et al*. 1989). A sequencing reaction was used as a size marker. We deliberately did not know to which site or race individuals belonged when reading banding patterns.

Karyotype preparation and analysis methods have been decribed previously (Brünner & Hausser 1996). Karyotypic hybrids were discovered in this hybrid zone. All were backcrosses; no F_1 hybrids were discovered and only a single possible F_2 was found (Brünner & Hausser 1996). Individuals possessing extra-racial chromosomes are referred to as hybrids. For example, if an individual presents a normal Cordon karyotype but for a *kn* metacentric chromosome, then this individual is considered as a hybrid. Indeed, arms *k* and *n* are always found as acrocentric chromosomes in the Cordon race, whereas the metacentric chromosome *kn* is typical for the Valais race.

CRT–MCA

A multivariate ordination technique called CRT–MCA that deals explicitly with allele frequency data has been

No.	Trapping locality	1992	1993	1994	1995	Total
1.	Rau des Chavants	–	–	–	8	8
2.	Nant Jorland	–	–	–	6	6
3.	Les Lavouets	–	–	5	–	5
4.	Vallée Patinoire	–	–	7	–	7
5.	Les Gens	15	22	8	10	55
6.	Granges des Faux I	13	–	8	6	27
7.	Granges des Faux II	8	–	18	–	26
8.	Granges des Faux III	6	17	9	4	36
9.	Le Trembley	6	9	–	6	21
10.	Taconnaz	–	–	–	9	9
11.	Vers le Nant	–	–	–	12	12
12.	Les Granges	–	4	–	–	4
13.	Les Tissières	–	6	–	9	15
14.	Chalet J. Balmas	–	7	–	7	14
15.	Chamonix Téléphérique	–	4	–	–	4
16.	Les Coverays	–	–	10	–	10
17.	Les Tines	–	–	10	–	10
Total		48	69	75	77	269

Table 1 Number of individuals of *Sorex araneus* collected, per year and per trapping locality (1–17), according to Fig. 1. Individuals from sites 1–5 belong to the Cordon race, and individuals from sites 6–17 are of the Valais race

–, indicates that the site was not visited that year.

described recently (Guinand & Easteal 1996; Guinand 1996). The main interest of this method when compared to other multivariate analyses is that the correlation ratio of variables (loci) with the factorial score of each axis is a variance ratio similar to F_{ST} (Guinand 1996), and the axis eigenvalues correspond to the overall F_{ST}. These two measures of genetic differentiation are obtained for each axis of the analysis, therefore allowing distinction between main and secondary differentiation patterns. This method was shown to be a coherent and polyvalent tool for use in population genetics and was successfully applied to express the successive phases of colonization in *Bufo marinus* populations (Guinand & Easteal 1996). The software package ADE-4 (Thioulouse *et al*. 1997) was used to perform the analysis. As sample sizes are unbalanced, samples were weighted by their relative sizes.

Genetic polymorphism, genetic structuring, and linkage disequilibrium

The software FSTAT version 2.0 (updated from Goudet 1995; available upon request) was used to calculate allele frequencies, observed heterozygosities (H_O), and expected heterozygosities within (H_S) and between (H_T) samples following Nei (1987). *F*-statistics according to Weir & Cockerham (1984) were also obtained with this program. *R*-statistics were estimated using a hierachical analysis of variance carried out on the number of repeats of alleles (Michalakis & Excoffier 1996; Rousset 1996) using the statistical package S-PLUS (Statistical Sciences, 1995). Genotypic linkage disequilibrium was investigated for each pair of loci using the program GENEPOP 2, updated from Raymond & Rousset (1995).

Temporal changes in allele frequencies were analysed using FSTAT 2.0. We calculated F_{ST} to measure the differentiation of the same sampling sites from different dates. Indeed, a test of temporal F_{ST} (namely the homogeneity test of allele frequency) was shown to be a powerful method to detect the differentiation of populations from different years (Viard *et al*. 1997). Here, we used the permutation test of FSTAT 2.0, which was shown to be similar in power (Goudet *et al*. 1996). Only sites sampled during at least two different years were used for analysis. These are sites 5, 6, 7, 8, 9, 13 and 14 (see Table 1). Using FSTAT 2.0, these sites were analysed in turn to detect allele frequency changes over years (e.g. F_{ST} for site 14, from 1993 and 1995). We tested the values of F_{ST} for significant departure from zero using 5000 permutations of genotypes among samples. Fisher's procedure (Sokal & Rohlf 1981) was used to obtain overall significance for allele frequency changes of the seven sampling sites.

To test whether populations are differentiated, we used the exact *G*-test on allelic frequencies advocated by Goudet *et al*. (1996). Contingency tables of alleles in columns against samples in rows are first obtained, and the log likelihood ratio statistic *G* (Sokal & Rohlf 1981; Goudet *et al*. 1996) is calculated. Genotypes are then permuted among samples, and a *G*-statistic is obtained from the permuted data set. The procedure is repeated 5000 times. An unbiased estimator of the probability that the samples correspond to arbitrary groups from the same population is then obtained as the number of *G*-statistics larger or equal to the observed (Goudet *et al*. 1996); to test whether there is differentiation among samples within races and among the two races, we modified the sampling procedure of Goudet *et al*. (1996). To test for differentiation within races, genotypes of individuals are permuted among samples but kept within their chromosome race. To test for differentiation among races, the exact *G*-test is calculated on contingency tables of alleles vs. races. The units to be permuted this time are the samples rather than the genotypes. Note that this latest procedure has the inconvenience of altering the number of genotypes entering in each race, because our samples are not all of equal sizes (on the other hand, the number of samples per race is kept constant, Excoffier *et al*. 1992). The overall probability of no genetic differentiation was estimated by summing the *G*-statistics of individual loci.

Hierarchical estimates of *F*-statistics (Weir 1996) were estimated using the computer package S-PLUS (Statistical Sciences, 1995) in order to obtain indirect estimates of gene flow between the different sampling units. F_{IS} estimates the heterozygote deficit within samples, F_{SR} between samples within races, and F_{RT} between the two races. The effective number of migrants, Nm (under the assumption of an island model at equilibrium between migration and drift) can be estimated both among samples within races and between races using the formula $F = 1/[4Nmd/(d-1)+1]$ [Slatkin & Voelm (1991); N represents the effective population size, and m, the migration proportion; the correction for the number of samples (d) is not squared here, because our *F* estimates are based on the variance component method (Rousset 1996; eqn 13)]. For F_{SR}, Nm represents the number of migrants among samples within races while, for F_{RT}, it is the number of individuals exchanged between the two races. Recently, Slatkin (1995) showed that F_{ST} values overestimate the number of migrants if the loci used are microsatellites evolving under the stepwise model of mutation (SMM). He suggested a novel estimator of genetic differentiation based on the variance in the number of repeats of alleles, R_{ST}. Here, we obtained hierarchical *R* estimates (R_{IS}, R_{SR} and R_{RT}) from an analysis of variance carried out on the number of repeats of alleles (Michalakis & Excoffier 1996; Rousset 1996) using the statistical package S-PLUS (Statistical Sciences, 1995). The number of migrants was estimated with the same relation as above.

In order to compare the estimates of *F*- and *R*-statistics, we generated the bootstrapped distribution (bootstrapping

5000 times over loci) for each. Under the assumption that the mutation model underlying microsatellite evolution is a SMM, we expect *R*-statistics to be larger than *F*-statistics (Slatkin 1995). Therefore, we took the difference $R^* - F^*$ (where the asterisk denotes an element of the bootstrap distribution) for all 5000 bootstrapped values. The proportion of these differences which is less than or equal to zero estimates the probability that *R* is not larger than *F*.

Results

Polymorphism, genetic variability of microsatellites, linkage disequilibrium and genetic variation over time

The number of alleles per locus ranged from three to 26 (Table 2). Expected heterozygosities within samples (H_S) ranged from 0.29 to 0.87, with an average of 0.73, whereas expected heterozygosities between samples (H_T) averaged 0.79 (range: 0.36–0.93; Table 2). Observed heterozygosity (H_O) values ranged from 0.24 to 0.90. The average value was 0.72 (Table 2). Our analyses did not evidence any significant linkage disequilibrium for any pair of loci (data not shown). All the loci were therefore considered as statistically independent.

Sampling sites for which data were available over several years did not show significant gene frequency changes over time, except for a single site out of the seven tested (site 8; $F_{ST} = 0.021$, $P = 0.022$). When probabilities were combined over all sites (including site 8), no significant departure from Hardy–Weinberg equilibrium was observed for any of the seven loci and overall. Therefore, as allele frequency variation over time did not show significant changes, samples from the same sampling site were pooled over years to increase sample size. Therefore, each sampling site is now represented by a single sample.

CRT–MCA

The first axis of the CRT–MCA captures 29% of the total inertia contained in the data set and the second axis about 13%. The eigenvalues associated with the first two factorial axes are 0.408 and 0.181, respectively. These eigenvalues are correlation ratios, and could therefore represent an estimation of the mean F_{ST} related to a given factorial axis (Guinand 1996).

Figure 2a gives the position of the 17 samples on the main factorial plane described by the two main axes of the analysis. Valais and Cordon samples form two distinct clusters relative to the first axis of the multivariate analysis. The geographical information is expressed in Fig. 2b. Coordinates resulting from the first axis of the CRT–MCA are projected on a background map (open circles for negative values; full squares for positive values). The factorial coordinates of samples on the first axis when projecting on a factor map clearly illustrate the two main geographical areas separated by the torrent de la Griaz (Fig. 2b).

Table 2 Number of alleles (*Na*), observed heterozygosity (H_O) and expected heterozygosities within (H_S) and between (H_T) samples, per locus and over all loci

Locus	*Na*	H_O	H_S	H_T
L9	25	0.90	0.87	0.91
L16	3	0.24	0.29	0.36
L45	8	0.60	0.59	0.64
L57	26	0.88	0.87	0.93
L62	19	0.83	0.80	0.89
L67	13	0.84	0.82	0.89
L69	21	0.79	0.85	0.90
Overall	115	0.72	0.73	0.79

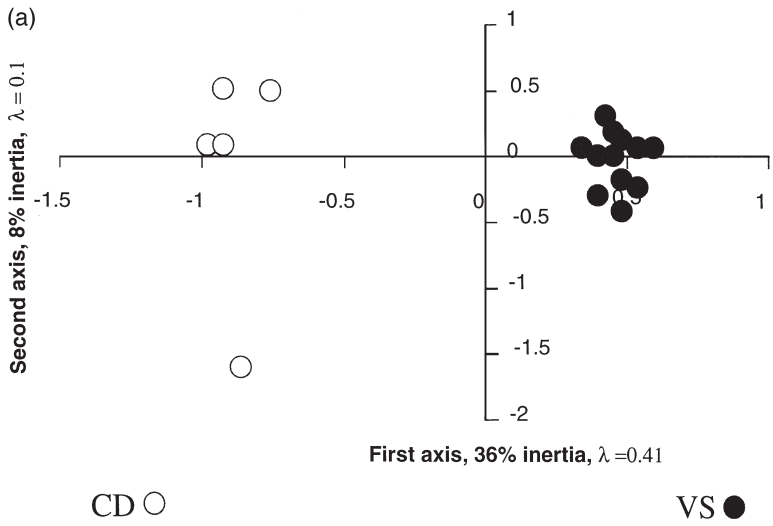
Genetic structuring

We first estimated *F*-statistics according to Weir & Cockerham (1984). The overall within-population heterozygote deficit F_{IS} ($F_{IS} = 0.012$) is not significant. The relatively high and significant overall F_{ST} ($F_{ST} = 0.076$) suggests a moderate genetic structuring. Exact *G*-tests for each locus and overall loci are highly significant. As two chromosome races are involved, *F*-statistics as well as *R*-statistics were estimated for Cordon and Valais samples independently (Table 3).

Considering only Cordon samples, a significant overall loci $F_{SR(CD)}$ is found ($F_{SR(CD)} = 0.024$, exact *G*-test $P < 0.01$; Table 3). However, when loci are considered independently, only locus L16 is significant ($F_{SR(CD)} = 0.197$, $P < 0.0002$), so that the significance of the mean $F_{SR(CD)}$ can certainly be attributed to the presence of this highly significant locus (L16). Looking more closely at this locus across samples, we notice that sample 4 is monomorphic (a single allele present), while all other Cordon samples contain the three alleles present at locus L16. The same analysis carried out on Cordon samples either without sample 4 or without locus L16 is no longer significant (overall *G*-test: $P = 0.09$; $P = 0.06$, respectively).

Interestingly, the overall loci $R_{SR(CD)}$ value is much lower than the corresponding $F_{SR(CD)}$ ($R_{SR(CD)} = 0.009$; Table (3). All $R_{SR(CD)}$ values, except for a single locus (L57), are lower than their corresponding $F_{SR(CD)}$, although the difference is not significant (Bootstrap test, $P = 0.76$).

When considering Valais samples only, the overall loci $F_{SR(VS)}$ is highly significant ($F_{SR(VS)} = 0.022$, $P < 0.0002$). All loci are highly significant when considered independently (Table 3). This indicates that there is a structuring on the Valais side. $R_{SR(VS)}$ values are not significantly larger than their corresponding $F_{SR(VS)}$ (bootstrap test, $P = 0.99$).



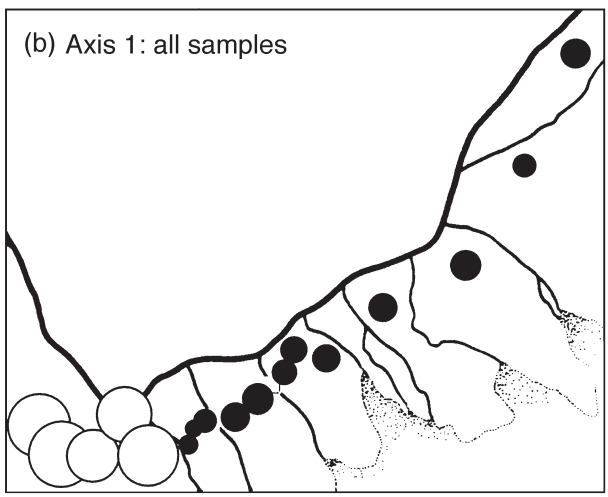


Fig. 2 (a) Factor map of the two main factorial axes of the CRT–MCA. Cordon and Valais samples are represented by open and closed circles, respectively. (b) Geographical interpretation of the main patterns of genetic differentiation by projection of the factorial scores of each sample on the physical map. Open and closed circles indicate negative and positive factorial coordinates, respectively. Their size is proportional to their value. According to the first factorial axis of the CRT–MCA, samples of both races form distinct clusters separated by the torrent de la Griaz.

In a second step, hierarchical *F*- and *R*-statistics were estimated. Per locus and overall loci values are presented in Table 4. Inter-racial (F_{RT} and R_{RT}) and intraracial (F_{SR} and R_{SR}) statistics were evaluated. Exact *G*-tests per locus are all significant and F_{RT} values range from 0.039 for locus L69 to 0.277 for locus L16 (Table 4). Overall *G*-test is highly significant ($P < 0.0002$) with an overall F_{RT} of 0.103. R_{RT} values range from – 0.005 to 0.432 for locus L57 (Table 4). The high latter value can be explained by the very different allelic distribution in the Cordon and Valais race at this locus (Fig. 3). Indeed, at locus L57 in the Valais race, we have the presence of many longer alleles which are virtually absent in the Cordon race (Fig. 3). It is noteworthy that the R_{RT} values for most loci are quite higher than their F_{RT} counterparts. Indeed, the bootstrap test result is marginally significant at the 5% level ($P = 0.1$) (Table 4). F_{SR} and R_{SR} values are rather low, all being inferior to 0.030 except for locus L16 presenting a quite high F_{SR} of 0.191. All *G*-tests within race are significant, with six of the seven loci being highly significant ($P < 0.0002$; Table 4). However, the intraracial *F*- and *R*-values remain low, and reflect a weak intraracial structuring. It should be noted that R_{SR} values are not significantly larger than their F_{SR} counterparts (bootstrap test, $P = 1.00$).

The number of migrants between populations was estimated both from *F*- and *R*-statistics. The number of migrants exchanged between the two chromosome races is 0.5 and 1.1 individuals per generation when calculated from R_{RT} and F_{RT}, respectively. The number of migrants between populations of the same race is 8.4 (calculated from F_{SR}), but is of little meaning because many assumptions behind the approximation are probably violated (island structure, large population, small migration). *Nm* cannot be inferred from R_{SR}, because its value is negative.

Discussion

The multivariate analysis clearly demonstrates the dichotomy between Cordon and Valais populations. It confirms the already well-documented distinction between these two chromosome races (Hausser *et al*. 1991; Wojcik 1993; Taberlet *et al*. 1994). The Valais race would have diverged in Italy during the last glaciations before colonizing the Valais (Switzerland), probably from the Simplon pass (Hausser *et al*. 1991). It then progressed westwards and reached the Arve valley where it met the Cordon race.

Table 3 Values of *F*- and *R*-statistics, per locus over all populations, calculated independently for each side of the torrent de la Griaz to detect within-race genetic structuring. Subscripts I, S, R, CD and VS stand for individuals, samples, races, Cordon and Valais, respectively. Asterisks indicate significant genetic structuring among Cordon and among Valais samples, respectively, obtained with the exact *G*-test

		Cordon samples					
	Exact *G*-test	*F*-statistics			*R*-statistics		
Locus	Among Cordon samples	$F_{IS(CD)}$	$F_{SR(CD)}$	$F_{IR(CD)}$	$R_{IS(CD)}$	$R_{SR(CD)}$	$R_{IR(CD)}$
L9	NS	−0.049	0.008	−0.041	−0.201	−0.002	−0.203
L16	***	0.087	0.197	0.267	0.058	0.063	0.117
L45	NS	−0.080	0.022	−0.057	−0.165	0.009	−0.154
L57	NS	0.010	−0.009	0.001	0.091	0.057	0.142
L62	NS	0.050	0.004	0.053	0.048	−0.015	0.033
L67	NS	0.075	−0.002	0.073	0.133	−0.033	0.104
L69	NS	0.198	−0.006	0.194	0.134	−0.016	0.120
All loci	**	0.048	0.024	0.071	0.014	0.009	0.023
		Valais samples					
	Exact *G*-test	*F*-statistics			*R*-statistics		
Locus	Among Valais samples	$F_{IS(VS)}$	$F_{SR(VS)}$	$F_{IR(VS)}$	$R_{IS(VS)}$	$R_{SR(VS)}$	$R_{IR(VS)}$
L9	***	0.000	0.014	0.014	−0.055	−0.021	−0.077
L16	**	0.190	0.019	0.205	0.077	−0.005	0.073
L45	**	−0.055	0.041	−0.012	−0.034	0.010	−0.024
L57	**	0.008	0.007	0.016	0.056	−0.001	0.055
L62	***	−0.048	0.031	−0.015	0.031	0.034	0.064
L67	***	−0.007	0.029	0.022	−0.020	0.011	−0.009
L69	***	0.046	0.015	0.060	0.096	−0.004	0.092
All loci	***	0.003	0.022	0.024	0.022	0.003	0.025

NS, non significant; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table 4 Hierarchical *F*- and *R*-statistics per locus and over all loci. Subscripts I, S, R, T stand for individuals, samples, races and total, respectively. Asterisks indicate significant genetic differentiation, within race and between race, respectively, obtained with the exact *G*-test

	Exact *G*-test		Hierarchical *F*-statistics				Hierarchical *R*-statistics			
Locus	Within race	Among race	F_{IS}	F_{SR}	F_{RT}	F_{IT}	R_{IS}	R_{SR}	R_{RT}	R_{IT}
L9	***	***	−0.020	0.019	0.061	0.061	−0.092	−0.014	0.001	−0.107
L16	***	**	0.047	0.191	0.277	0.443	0.109	−0.038	0.260	0.315
L45	***	***	−0.041	0.014	0.093	0.069	−0.047	0.010	−0.005	−0.042
L57	*	***	0.005	0.009	0.102	0.115	0.096	−0.048	0.432	0.462
L62	***	***	−0.013	0.020	0.099	0.106	0.052	−0.004	0.118	0.161
L67	***	***	0.017	0.024	0.128	0.163	0.021	−0.015	0.343	0.347
L69	***	***	0.092	0.007	0.039	0.134	0.093	0.021	0.243	0.328
All loci	***	***	0.012	0.027	0.103	0.137	0.033	−0.013	0.199	0.209

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

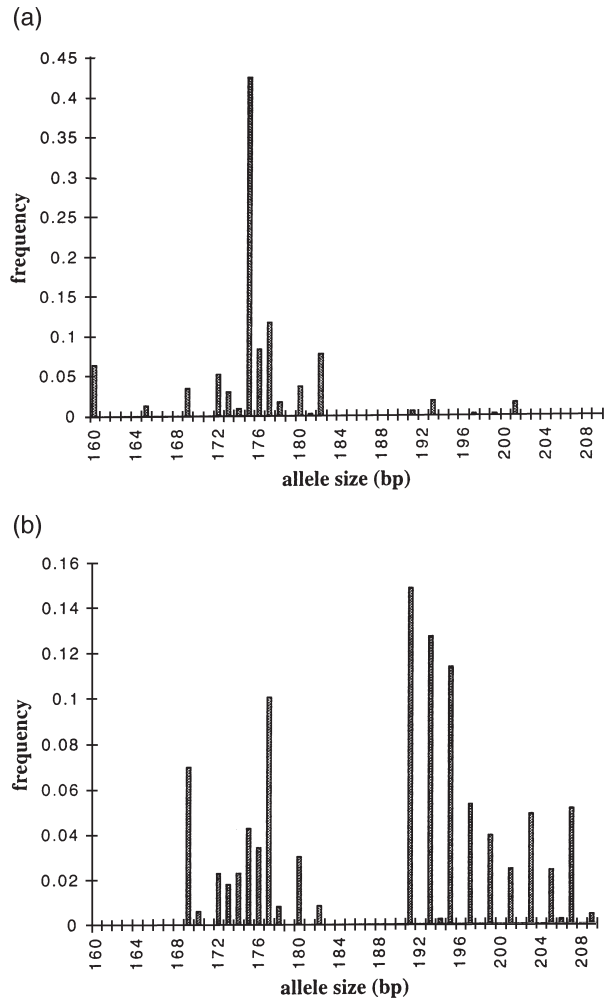


Fig. 3 Allele distribution at locus L57 in the Cordon (a) and Valais (b) races. Allele size is indicated in base pairs. Longer-sized alleles are far more common in the Valais race.

Intraracial genetic structuring

Using allozymes, Brünner & Hausser (1996) did not detect significant genetic structuring on neither side of this hybrid zone (N = 273; F_{ST} = [-0.095; 0.067]). Here, a significant intraracial structuring is found (F_{SR} = 0.027). Cordon and Valais G-tests are highly significant and F_{ST} values ($F_{SR(CD)}$ = 0.024 and $F_{SR(VS)}$ = 0.022, respectively) are similar to the ones reported for the Bretolet race in another alpine valley (F_{ST} = [0.015; 0.026]; Wyttenbach & Hausser 1996). It is intriguing that the Cordon side displays an F_{ST} value similar to the Valais F_{ST}. Indeed, the geographical scale on the Cordon side is small and no rivers are found on this side of the zone, whereas the Valais side is divided by several mountain rivers (see Fig. 1). However, it should be remembered that only locus L16 was significant on the Cordon side, sample 4 being monomorphic at this locus. As the analyses carried out without locus L16 or without sample 4 are no longer significant, it may be suggested that there is no genetic structuring on the Cordon side. However, it should be stressed here that sample sizes on the Cordon side are generally small, so that a significant structuring could be undetected. Rivers were not found to have a significant effect on reducing gene flow in this hybrid zone (Lugon-Moulin *et al*. 1996). However, their sample size was smaller than the present one (N = 77 vs. N = 269 for this study). In the present study, the significant genetic structuring found on the Valais side suggests that these rivers, or some of them, may have a significant effect on gene flow.

Gene flow between the two chromosome races

Gene flow is clearly reduced between these chromosome races. It has been estimated at one migrant every two generations using R-statistics and one migrant per generation using F-statistics. If we remove the two loci giving the largest and the smallest R_{RT}, respectively, we obtain an overall R_{RT} of 0.18 and an Nm of 0.56, which is very similar to the original value. The difference between R_{RT} and F_{RT} estimates is marginally significant at the 5% level and the P-value of 0.1 suggests that a test based on a larger number of loci might be significant. While no systematic ecological data exist on animal movement between these two chromosome races, a pure Valais karyotype was found only once on the Cordon side of the hybrid zone, and no pure Cordon has been found on the Valais side during the four sampling years (Brünner & Hausser 1996). This observation, even if it does not confirm the indirect estimate, is compatible with it. Moreover, the finding that no F_1 karyotypic hybrids, but only backcrossed hybrids, were discovered in this hybrid zone (Brünner & Hausser 1996) indicates that successful, fertile inter-racial matings are infrequent. This is also to some extent in accordance with our low inter-race Nm estimates.

It should be noted that F_1 hybrids between the Cordon and Valais race have been produced in the laboratory (Castagné *et al*. 1994). However, data on hybrid fertility and interracial sexual compatibility such as assortative mating are lacking. Therefore, it is not yet possible to disentangle the various possible causes of this reduced interracial gene flow.

Genetic constitution of karyotypic hybrids

We showed that the two races are well differentiated in terms of microsatellites in this hybrid zone. Therefore, if karyotypic hybrids are of intermediate genetic composition, they should fall in between the samples composed of pure Valais and pure Cordon individuals, respectively.

To verify this prediction, the following analysis was realized. All shrews presenting a hybrid karyotype were identified (23 individuals; Brünner & Hausser 1996; H. Brünner, unpublished results) and two virtual hybrid populations (VHP) were created. The first VHP consisted of all hybrids collected on the Cordon side (10 hybrids; CD–VHP), and the second, of all those collected on the Valais side (13 hybrids; VS–VHP). Individuals of unknown karyotype were discarded from these analyses. Cordon site 3 will therefore disappear, because all individuals from this sampling site are either hybrids or of unknown karyotype. A CRT–MCA was performed with these two virtual hybrid populations.

The multivariate analysis clearly shows that hybrids found on a given side (i.e. Cordon or Valais side of the hybrid zone) are indistinguishable on the basis of microsatellites from nonhybrid karyotypes found on the same side (Fig. 4). This implies that without knowledge of the karyotypes, we would not have been able to distinguish these hybrids from karyotypically pure individuals. Therefore, hybrids must be backcrossed individuals representing the trace of a rare inter-racial hybridization event which took place several generations ago. This is confirmed by the karyotypes, because no F_1 hybrids and only a possible F_2 were found (Brünner & Hausser 1996).

Comparison of different statistics to estimate population structuring

Recently, Valsecchi *et al*. (1997) compared various methods such as Slatkin's (1995) R_{ST}, unbiased R_{ST} (UR_{ST}; Goodman 1997) and Wright's (1965) F_{ST} to calculate genetic distances from microsatellite data. They concluded 'unbiased R_{ST} (Goodman 1997) is the most reliable statistic because, unlike the other methods, it allows for unequal sample sizes'. The reason for Goodman (1997) to develop the statistic UR_{ST} was not the unequal sample sizes, but rather it was to account for loci with vast differences in the variance of allele sizes. However, it has not been demonstrated that this statistic is any better than the one suggested by Rousset (1996). This issue needs further investigation.

Moreover, it must be emphasized here that it is not a general property of F_{ST} estimators to be sample size dependent. Weir & Cockerham (1984) devised an estimator of F_{ST} that is essentially unbiased. Similarly, Michalakis & Excoffier (1996), and Rousset (1996) have shown that under an analysis of molecular variance (AMOVA) or a hierarchical analysis of variance framework, one can obtain unbiased estimates of R_{ST}. In the present study, both F_{ST} and R_{ST} are unbiased methods and allow for unequal sample sizes. The comparison between the two families of estimators can therefore be carried out.

Under a SMM or a two-phase model (TPM), Slatkin (1995) showed that F_{ST} overestimates the number of migrants; that is, the expected value of F_{ST} is smaller than that of R_{ST}. Here, we obtained opposite results at the two scales investigated: on the one hand, among races differentiation measured with R_{ST} is higher in magnitude than that measured with F_{ST}. This is congruent with what is expected from Slatkin's results. On the other hand, when comparing among samples, within race levels of differentiation, the allele size-based statistic does not detect any structuring, while F_{ST} is positive. What could be the reasons for this discrepancy? First, the mutation patterns of our markers might not follow either a SMM or a TPM. For example, a strict SMM seemed inappropriate in a study on the geographical structure of green turtle populations in Australia (FitzSimmons *et al*. 1997). Angers & Bernatchez (1998) found relative similarity of F_{ST} and R_{ST} averaged

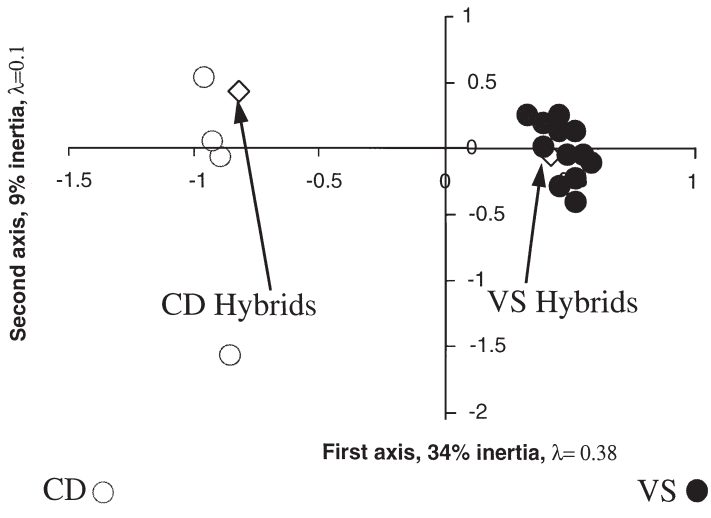


Fig. 4 Factor map of the two main factorial axes of CRT–MCA carried out with two virtual karyotypic hybrid populations (VHP; see text for details). Cordon and Valais samples are represented by open and closed circles, respectively. The 'Cordon' VHP (CD hybrids), represented by an open diamond, lies at the edge of the cluster formed by karyotypically pure Cordon samples while the 'Valais' VHP (VS hybrids), also represented by an open diamond, lies among karyotypically pure Valais samples.

across loci, but major discrepancies were observed at individual loci, one of which was known to strongly deviate from SMM expectations. Following Cornuet & Luikart (1997), a study by A. Wyttenbach (unpublished results) on a different sample of the same species and loci than the ones used in the present study was unable to conclude whether the mutation pattern was better approximated by a SMM or an infinite allele model (IAM). But if the discrepancy is only due to the mutation process, we should have observed similar differences within and among races, which is not the case. Second, as Slatkin (1995) noticed, while R_{ST} is less biased than F_{ST}, it has a larger variance. Indeed, omitting locus L16 (an outlier for F_{SR}), the range of values covered by R_{SR} [– 0.048; 0.021] is larger than that covered by F_{SR} [0.007; 0.027]. Third, and not exclusive, the mutation process might not matter at the intraracial scale while it might matter between races.

Table 5 presents a cursory comparison of several recent studies in which both *F*- and *R*-statistics were used. In their comparison of both type of estimators, Ross *et al*. (1997) concluded that 'use of the equidistant metric [F_{ST}] appeared to detect structure more efficiently than use of the euclidian metric [R_{ST}] for the microsatellites'. Indeed, when levels of differentiation are low, such as in the Ross *et al*. (1997) study, more structuring seems to be detected when allele size is not accounted for: in Table 5, when R_{ST} is < 0.06, only a single R_{ST} value is higher in magnitude than its corresponding $F_{ST.}$ However, this R_{ST} equals zero when a locus with a complex series of allele sizes is omitted (for more details, see Lehmann *et al*. 1997). This is corroborated by the simulation study of Slatkin (1995). When the number of migrants is high, *R*-statistics have large variances and therefore confidence intervals will overlap zero. Moreover, as stated by Slatkin (1995), 'if a typical mutation rate at a microsatellite locus is 10^{-3}, then F_{ST} can be used if it is known from other information that the time scales of interest are tens or hundreds of generations'. On the other hand, when levels of differentiation are higher, allele-based statistics point to a higher differentiation: in Table 5, when R_{ST} is ≥ 0.06, all R_{ST} values are higher than or equal in magnitude to their corresponding F_{ST}, except for a single value ($R_{ST} \approx 0.100$) we estimated from fig. 4 in Forbes *et al*. (1995) (Table 5). In this case, the information content of allele size will allow us to detect more accurately any differentiation.

In this study, neither hierarchical *F*- nor *R*-statistics were tested as such. Goudet *et al*. (1996) showed that *F*-statistics are not the most powerful statistics to use for testing population differentiation, particularly when sampling is not balanced. This result is probably even more pronounced for *R*-statistics, as they are well known to have large variances (e.g. Slatkin 1995; Goodman 1997). As pointed out by Rousset & Raymond (1997) 'a test statistic is generally chosen to maximize the power of the test when some specified alternative hypothesis is true'. This is the case for the *G*-test here. Not only would it be bad statistical practice to go on testing *F*-statistics and *R*-statistics from the same data set, because this would probably inflate the risk of type I error, but it could also lead to awkward situations: in our case, as R_{SR} is negative, one would conclude from it that there is no differentiation among samples within races, whereas the exact *G*-test applied among samples within races clearly shows for all loci a significant differentiation.

Conclusions

What can be said about the potential usage of microsatellites in hybrid zones? While they are adequate to detect fine-grained spatial structure, they showed their limits for the detection of hybrid karyotypes. This is certainly because the few hybrids are the results of multiple backcrosses with one or the other parental race. From the statistical analyses perspective, while the CRT–MCA gave results corroborated by the other analyses, one could just draw attention to the interpretation of the eigenvalues, considered by Guinand (1996) as F_{ST}. It is puzzling to find that the CRT–MCA-based estimate of the first axis is much larger than F_{RT} (first axis eigenvalue = 0.408, whereas the largest F_{RT} value (for locus L16) is 0.277). This matter will need further investigation before CRT–MCA eigenvalues can be taken at face value. While this multivariate analysis lacks testing procedures (Guinand 1996), important patterns of population genetic structure can be revealed when it is used in conjunction with other approaches for which testing procedures exist. Hierarchical statistics enabled us to estimate levels of migration across the hybrid zone as well as among the populations making contact. Although females are strictly territorial during most of their lifespan and males are nomadic during the period of reproduction (e.g. Shillito 1963; Croin Michielsen 1966; Cantoni 1990), both sexes undergo juvenile dispersal (Hanski *et al*. 1991). Simultaneous use of mitochondrial DNA and Y-chromosome loci would be of interest to estimate dispersal in both sexes, which could bring a better understanding to the processes acting in this contact zone. The simultaneous use of hierarchical *F*- and *R*-statistics in studies of hybrid zones is of great interest, because each statistic seems to be appropriate to detect structuring at different levels. Although F_{RT} and R_{RT} estimates are not statistically different in this hybrid zone, the *P*-value of 0.1 suggests that a test based on a larger number of loci might be significant. It would therefore be meaningful to add several loci to our data set to verify this hypothesis. Finally, the exact *G*-test proved to be powerful in detecting structuring at the two hierarchical levels.

Table 5 Cursory comparison of some recent studies having simultaneously used *F*- and *R*-statistics. F_{ST} and R_{ST} represent overall loci values. The higher values of the two statistics are represented in bold characters. We arbitrarily made two sections according to the level of differentiation based on R_{ST} values: low to moderate ($R_{ST} < 0.060$), and moderate to high ($R_{ST} \geq 0.060$). Therefore, references in this table are classified according to R_{ST} values, from the lowest to the highest. See text for discussion

Species investigated	No. of populations sampled and geographical scale	No. of loci	F_{ST}	R_{ST}*	References
A. Low to moderate levels of differentiation ($R_{ST} < 0.060$)					
Common shrew (*Sorex araneus*)	Among 17 populations kept within their chromosomal race (≈ 13 km)	7	**0.027**	−0.013	This study
Anopheles gambiae	Among six houses (< 10 km)	5	0.000	0.000 (**0.001**†)	Lehmann *et al*. (1997)
Green turtle (*Chelonia mydas*)	Among four regions 1000–4700 km apart	4	**0.014**	0.007	FitzSimmons *et al*. (1997)
Anopheles gambiae	Among two regions (≈ 50 km)	5	0.000	0.013 (0.001†)	Lehmann *et al*. (1997)
Brook charr (*Salvelinus fontinalis*)	Among 26 populations (3–42 km apart) within groups within lineages	5	**0.068**	0.036	Angers & Bernatchez (1998)
Anopheles gambiae	Among two regions 700 km apart (sampled the same year)	9	**0.072**	0.037	Lehmann *et al*. (1998)
Anopheles gambiae	Among two regions 700 km apart (regions not sampled the same year)	9	**0.100**	0.038	Lehmann *et al*. (1998)
Domestic sheep (*Ovis aries*)	Among three herds (Israel, Spain and the UK)	8	**0.085**	≈0.050‡	Forbes *et al*. (1995)
B. Moderate to high levels of differentiation ($R_{ST} \geq 0.060$)					
Freshwater snail (*Bulinus truncatus*)	Among samples in the same pond	4	0.010	**0.060**	Viard *et al*. (1996)
Bighorn sheep (*Ovis canadensis*)	Among five herds sampled from Alberta to Colorado (USA)	8	**0.224**	≈0.100‡	Forbes *et al*. (1995)
Freshwater snail (*Bulinus truncatus*)	Among samples in the same pond	4	0.070	**0.100**	Viard *et al*. (1996)
African buffalo (*Syncerus caffer*)	Among 11 localities in eastern and southern Africa	6	0.085§	**0.110**	Simonsen *et al*. (1998)
European harbour seal (*Phoca vitulina vitulina*)	Among 12 localities, (≈ 250 km to > 1000 km)	7	0.172	**0.181**	Goodman (1998)
Common shrew (*Sorex araneus*)	Among two chromosomal races (17 populations; ≈ 13 km)	7	0.103	**0.199**	This study
Brook charr (*Salvelinus fontinalis*)	Among populations within groups in drainage	5	0.159	**0.201**	Angers & Bernatchez (1998)
Brook charr (*Salvelinus fontinalis*)	Among groups in drainage	5	0.238	**0.287**	Angers & Bernatchez (1998)
Brown trout (*Salmo trutta*)	Among 11 samples in two drainages; up to ≈ 850 km	5	0.279	**0.287**	Estoup *et al*. (1998)
Brown trout (*Salmo trutta*)	Among nine populations in one drainage; waterway distances: 2.5–328 km	6	0.263	**0.292**	Estoup *et al*. (1998)
Freshwater snail (*Bulinus truncatus*)	Among samples in the same pond	4	0.300	0.300	Viard *et al*. (1996)
Brook charr (*Salvelinus fontinalis*)	Among 26 populations (3–42 km apart)	5	0.370	**0.460**	Angers & Bernatchez (1998)
Brook charr (*Salvelinus fontinalis*)	Among groups in lineages	5	0.352	**0.482**	Angers & Bernatchez (1998)
Freshwater snail (*Bulinus truncatus*)	Among three ponds several hundred km apart	4	0.420	**0.640**	Viard *et al*. (1996)
Domestic & Bighorn sheep (*Ovis aries & O.canadensis*)	Between the two species	8	≈0.200‡	≈**0.650**‡	Forbes *et al*. (1995)

*Different derivations of R_{ST} could occur among the studies cited.
†With four loci (a locus with complex series of allele sizes omitted-see Lehmann *et al*. (1997) for more details)
‡The overall value is not explicitly given in Forbes *et al*. (1995), we evaluated it approximately from Fig. 4 (Forbes *et al*. 1995)
§This study uses essentially R_{ST}; this overall F_{ST} value of 0.085 is the only one given.

Acknowledgements

This work was supported by the Swiss National Science Foundation (grants no. 31–34057.92 and 31–43161.95 to J.H. and 31–43443.95 to J.G.) and by a grant of the French Embassy in Switzerland. We thank John Aeschimann, Hedy Brack, Cédric Cretegny, Christian Koenig, Patrick Patthey, Andrea Persico and Hendrik Turni for their invaluable help in the field; and Anne-Marie Mehmeti, Frédéric Busslinger and Hedy Brack for help in cell culture and karyotype preparation, Nelly Di Marco for technical assistance, William Brown for improving our English, and Laurent Keller, François Balloux, Michel Solignac and an anonymous referee for helpful comments on the manuscript.

References

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.

Brünner H, Hausser J (1996) Genetic and karyotypic structure of a hybrid zone between the chromosomal races Cordon and Valais in the common shrew, *Sorex araneus*. *Hereditas*, **125**, 147–158.

Cantoni D (1990) Etude en milieu naturel de l'organisation sociale de trois espèces de musaraignes, *Crocidura russula, Sorex coronatus*, et *Neomys fodiens* (Mammalia, Insectivora, Soricidae). PhD Thesis, University of Lausanne.

Castagné C, Mehmeti AM, Hausser J (1994) Interbreedings between karyotypic Alpine races of the common shrew *Sorex araneus* (Insectivora, Mammalia). *Caryologia*, **47**, 11–18.

Churchfield S, Hollier J, Brown K (1995) Population dynamics and survivorship in the common shrew *Sorex araneus* in southern England. *Acta Theriologica*, **40**, 53–68.

Cornuet JM, Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Croin Michielsen N (1966) Intraspecific and interspecific competition in the shrews *Sorex araneus* L. & *S. minutus* L. *Archives Néerlandaises de Zoologie*, **17**, 73–174.

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

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

FitzSimmons NN, Moritz C, Limpus CJ, Pope L, Prince R (1997) Geographic structure of mitochondrial and nuclear gene polymorphisms in Australian green turtle populations and male-biased gene flow. *Genetics*, **147**, 1843–1854.

Forbes SH, Hogg JT, Buchanan FC, Crawford AM, Allendorf FW (1995) Microsatellite evolution in congeneric mammals: domestic and Bighorn sheep. *Molecular Biology and Evolution*, **12**, 1106–1113.

Goodman SJ (1997) R_{ST} Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology*, **6**, 881–885.

Goodman SJ (1998) Patterns of extensive genetic differentiation and variation among European harbor seals (*Phoca vitulina vitulina*) revealed using microsatellite DNA polymorphisms. *Molecular Biology and Evolution*, **15**, 194–118.

Goudet J (1995) FSTAT, Version 1.2: a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.

Goudet J, Raymond M, DeMeeüs T, Rousset F (1996) Testing genetic differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Guinand B (1996) Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations. *Biological Journal of the Linnean Society*, **58**, 173–195.

Guinand B, Easteal S (1996) Multivariate patterns of genetic differentiation support complex colonization schemes in *Bufo marinus* populations. *Evolution*, **50**, 944–951.

Hanski I, Peltonen A, Kaski L (1991) Natal dispersal and social dominance in the common shrew *Sorex araneus*. *Oïkos*, **62**, 48–58.

Hausser J, Bosshard E, Taberlet P, Wojcik J (1991) Relationship between chromosome races and species of the *Sorex* of the *araneus* group in the western Alpes. *Mémoires de la Société Vaudoise de Sciences Naturelles*, **19**, 79–95.

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

Kocher TD, Thomas WK, Meyer A, *et al*. (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the USA*, **86**, 6196–6200.

Lehmann T, Besansky NJ, Hawley WA, *et al*. (1997) Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. *Molecular Ecology*, **6**, 243–253.

Lehmann T, Hawley WA, Grebert H, Collins FH (1998) The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. *Molecular Biology and Evolution*, **15**, 264–276.

Lugon-Moulin N, Wyttenbach A, Brünner H, Goudet J, Hausser J (1996) Study of gene flow through a hybrid zone in the common shrew (*Sorex araneus*) using microsatellites. *Hereditas*, **125**, 159–168.

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellites loci. *Genetics*, **142**, 1061–1064.

Neet CR, Hausser J (1991) Biochemical analysis and determination of living individuals of the Alpine karyotypic races and species of the *Sorex araneus* group. *Mémoires de la Société Vaudoise de Sciences Naturelles*, **19**, 97–106.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, NY, USA.

Raymond M, Rousset F (1995) GENEPOP, Version 1.2: population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Ross KG, Krieger MJB, Shoemaker DD, Vargo EL, Keller L (1997) Hierarchical analysis of genetic structure in native fire ant populations: results from three classes of molecular markers. *Genetics*, **147**, 643–655.

Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.

Rousset F, Raymond M (1997) Statistical analyses of population genetic data: new tools, old concepts. *Trends in Ecology and Evolution*, **12**, 313–317.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Searle JB, Fedyk S, Fredga K, Hausser J, Volobouev VT (1991) Nomenclatue for the chromosomes of the common shrew *Sorex araneus*. *Mémoires de la Société Vaudoise de Sciences Naturelles*, **19**, 13–22.

Shillito JF (1963) Observations on the range and movements of a woodland population of the common shrew, *Sorex araneus* L. *Proceedings of the Zoological Society of London*, **140**, 533–546.

Simonsen BT, Siegismund HR, Arctander P (1998) Population structure of African buffalo inferred from mtDNA sequences and microsatellite loci: high variation but low differentiation. *Molecular Ecology*, **7**, 225–237.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Slatkin M, Voelm L (1991) F_{ST} in a hierarchical island model. *Genetics*, **127**, 627–629.

Sokal RP, Rohlf FJ (1981) *Biometry*. Freeman, San Francisco.

Statistical Sciences (1995) S-PLUS guide to statistical and mathematical analysis, Version 3.3. StatSci, a division of MathSoft, Inc., Seattle.

Taberlet P, Fumagalli L, Hausser J (1994) Chromosomal versus mitochondrial DNA evolution: tracking the evolutionary history of the southwestern European populations of the *Sorex araneus* group Mammalia, Insectivora. *Evolution*, **48**, 623–636.

Thioulouse J, Chessel D, Doledec S, Olivier JM (1997) ADE-4: a multivariate analysis and graphical display software. *Statistical Computing*, **7**, 75–83.

Valsecchi E, Palsboll P, Hale P, *et al*. (1997) Microsatellite genetic distances between oceanic populations of the humpback whale (*Megaptera novaeangliae*). *Molecular Biology and Evolution*, **14**, 355–362.

Viard F, Bremond P, Labbo R, *et al*. (1996) Microsatellites and the genetics of highly selfing populations in the freshwater snail *Bulinus truncatus*. *Genetics*, **142**, 1237–1247.

Viard F, Justy F, Jarne P (1997) Population dynamics inferred from temporal variation at microsatellite loci in the selfing snail *Bulinus truncatus*. *Genetics*, **146**, 973–982.

Weir BS (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA, USA.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wojcik JM (1993) Chromosome races of the common shrew *Sorex araneus* in Poland: a model of karyotype evolution. *Acta Theriologica*, **38**, 315–338.

Wojcik JM, Searle JB (1988) The chromosome complement of *Sorex granarius* – the ancestral karyotype of the common shrew *Sorex araneus*? *Heredity*, **61**, 225–229.

Wojcik JM, Wojcik AM (1994) Protein variation in the common shrew (*Sorex araneus* L.) in Poland, in relation to karyotype. *Folia Zoologica*, **43**, 53–61.

Wright S (1965) The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.

Wyttenbach A, Favre L, Hausser J (1997) Characterization of simple sequence repeats in the genome of the common shrew. *Molecular Ecology*, **6**, 797–800.

Wyttenbach A, Hausser J (1996) The fixation of metacentric chromosomes during the chromosomal evolution in the common shrew (*Sorex araneus*, Insectivora). *Hereditas*, **125**, 209–217.

Zima J, Fedyk S, Fredga K, *et al*. (1996) The list of the chromosome races of the common shrew (*Sorex araneus*). *Hereditas*, **125**, 97–107.

NLM's PhD project focuses on population genetics, phylogeography and the study of hybrid zones in the common shrew using molecular markers. HB is undertaking a comparative study of gene flow through different hybrid zones between chromosome races of the common shrew. AW works on molecular evolution of microsatellites and its application in population genetics. His focus is now on triplet repeat expansions in human neurodegenerative diseases. JH is Professor at the University of Lausanne (Switzerland), and its prime research interests are population genetics and the evolutionary history of soricine shrews. JG enjoys playing with FSTAT and obscure statistical packages.
