 doi: 10.1111/j.1365-294X.2006.02891.x

Genetic and karyotypic structure in the shrews of the *Sorex araneus* group: are they independent?

P. BASSET, G. YANNIC and J. HAUSSER
Department of Ecology and Evolution, Biology Building, University of Lausanne, CH-1015 Lausanne, Switzerland

Abstract

The species of the common shrew (*Sorex araneus*) group are morphologically very similar but exhibit high levels of karyotypic variation. Here we used genetic variation at 10 microsatellite markers in a data set of 212 individuals mostly sampled in the western Alps and composed of five karyotypic taxa (*Sorex coronatus*, *Sorex antinorii* and the *S. araneus* chromosome races Cordon, Bretolet and Vaud) to investigate the concordance between genetic and karyotypic structure. Bayesian analysis confirmed the taxonomic status of the three sampled species since individuals consistently grouped according to their taxonomical status. However, introgression can still be detected between *S. antinorii* and the race Cordon of *S. araneus*. This observation is consistent with the expected low karyotypic complexity of hybrids between these two taxa. Geographically based cryptic substructure was discovered within *S. antinorii*, a pattern consistent with the different postglaciation recolonization routes of this species. Additionally, we detected two genetic groups within *S. araneus* notwithstanding the presence of three chromosome races. This pattern can be explained by the probable hybrid status of the Bretolet race but also suggests a relatively low impact of chromosomal differences on genetic structure compared to historical factors. Finally, we propose that the current data set (available at www.unil.ch/dee/page7010_en.html#1) could be used as a reference by those wanting to identify *Sorex* individuals sampled in the western Alps.

Keywords: common shrew, genetic structure, introgression, karyotypic structure, *Sorex araneus* group

Received 2 June 2005; revision accepted 23 September 2005

Introduction

Closely related species and even populations of the same species can exhibit a large amount of karyotype variation. In such situations, the real impact of karyotypic differences on the genetic relationships among taxa is often particularly difficult to evaluate (King 1993; Rieseberg 2001). The closely related shrews of the *Sorex araneus* group constitute an illustrative example. This group of morphologically very similar species is characterized by an XY1Y2 sex chromosome complex in males (Sharman 1956) and is well known for its spectacular chromosomal evolution. Considerable autosomal variation (mainly Robertsonian changes) can be observed not only among the species of this group but also within its type species, *S. araneus*. At least 60 chromosomal races distributed all over Europe and Siberia make this species one of the most chromosomally polymorphic among mammals (Wójcik *et al*. 2002). Different models of chromosomal evolution have been developed to account for the large chromosomal variation found in these species (for recent review see Searle & Wójcik 1998) but the comparative analyses of karyotypic, biochemical or mitochondrial DNA data often show contradictory results (Taberlet *et al*. 1994; Fumagalli *et al*. 1996; Ratkiewicz *et al*. 2002).

During the last Pleistocene glaciations, the Alpine barrier played a major role in separating the Italian peninsula from the rest of the continent. For numerous taxa (Taberlet *et al*. 1998), this geographical isolation led to genetic divergence and thereafter influenced postglacial recolonization of Europe. The shrews of the *S. araneus* group did not escape this common pattern and at least five chromosomal races and/or species of this group meet in this region.

Correspondence: Patrick Basset, Fax: +41 21 692 41 65;
E-mail: patrick.basset@unil.ch

© 2006 Blackwell Publishing Ltd

Actually, *Sorex coronatus* would have diverged in glacial refugia situated in southwestern France or Spain (Hausser 1978) and then would have colonized the pre-Alpine lowlands and large Alpine valley from the west. The refugia of *Sorex antinorii* were certainly situated in the Apennine peninsula (Brünner *et al*. 2002b). This species (formerly considered as a chromosome race of *S. araneus*, Brünner *et al*. 2002a) crossed several lower Alpine passes in the Swiss and French Alps but mostly remained restricted to Italy (Lugon-Moulin & Hausser 2002). Finally, *S. araneus* certainly presents the most complex evolutionary history. This species may have been restricted to several refugia during the past glacial periods (Taberlet *et al*. 1994) but probably colonized the Alps mostly from southeastern refugia. Three genetically closely related chromosome races of this species (Cordon, Bretolet and Vaud) are presently distributed in the western Alps.

Species and chromosome races of the *S. araneus* group are morphologically very similar and impossible to tell apart in the field. Identification of individuals can be performed from karyotypes and chromosome counts, allozymes (Hausser & Zuber 1983; Neet & Hausser 1991), morphometric measures (Hausser *et al*. 1991) or analysis of diagnostic markers (Basset & Hausser 2003) but all these methods require destructive sampling or have drawbacks in field studies (reviewed in Basset & Hausser 2003). Recently, Pfunder *et al*. (2004) proposed an attractive microarray-based diagnostic test for shrew species; however, it did not allow discrimination between all *Sorex* species present in the Alps (e.g. between *S. antinorii* and *S. araneus*). Additionally, this method could be somewhat expensive and time consuming when used at small scale.

Recently, numerous methods based on genetic assignation of individuals to a given group have been developed and seem especially effective (Paetkau *et al*. 1995; Rannala & Mountain 1997; Cornuet *et al*. 1999; Pritchard *et al*. 2000; Vázquez-Dominguez *et al*. 2001; Wilson & Rannala 2003). However, before using these techniques to allocate unknown samples to taxa, it is necessary to check the agreement between the genetic and the taxonomic structure. This last point should not be underestimated in the *S. araneus* group, as the genetic relationships among taxa are not straightforward. European species and chromosome races of this group show parapatric distributions and typically form various contact or hybrid zones. Several of these zones have been studied all over Europe (Searle & Wójcik 1998; Fredga & Narain 2000; Ratkiewicz *et al*. 2003; Andersson *et al*. 2004) including the Alps (Neet & Hausser 1990; Neet 1992; Lugon-Moulin *et al*. 1996; Brünner *et al*. 2002b). These zones generally showed a large variety of levels of gene exchange between adjacent populations, from total isolation to almost free gene flow. For example, species identity was clearly the greatest factor structuring the hybrid zone between *S. antinorii* and *S. araneus* (Brünner *et al*. 2002b) whereas Andersson *et al*. (2004) found a similar level of genetic structure within and between two chromosome races belonging to two different karyotypical groups (group of chromosome races characterized by some shared metacentrics (Searle & Wójcik 1998)). Actually, the effect of chromosomal differences on the gene flow between two populations is much debated (Rieseberg 2001; Navarro & Barton 2003; Panithanarak *et al*. 2004) and still needs to be assessed in the case of the *S. araneus* group. Almost no real trial has been carried out to link genetic and karyotypic structure at a larger scale than the hybrid zone. Generally, to estimate population structure and/or assign individuals to a population it is necessary to a priori define discrete populations following subjective criteria. However, it seems important to check whether these a priori assumptions match genetic data in natural populations, particularly when population limits are not totally clear (for example in hybridizing taxa). Some of the recent assignment methods (Pritchard *et al*. 2000; Vázquez-Dominguez *et al*. 2001; Wilson & Rannala 2003) allow description of population structure without requiring predefined groups, providing new opportunities for checking the relationships between expected and real population structure. The methods of Pritchard *et al*. (2000) and Wilson & Rannala (2003) are particularly informative when studying a possibly hybridizing group of species as they consider that an individual could originate from more than one population.

In the present study we genotyped 212 individuals at 10 microsatellite loci in order to (i) check the concordance between genetic structure and karyotypic structure in the *S. araneus* group; (ii) identify potential cryptic substructure; (iii) estimate the utility of genetic markers in the identification of different species or chromosomal races of the *S. araneus* group in the Alps; and (iv) develop a genetic reference to allocate individuals of unknown origin to species and/or population with Bayesian assignment techniques.

Materials and methods

Sample collection

A total of 33 Jersey shrews (*Sorex coronatus*), 83 Valais shrews (*Sorex antinorii*) and 96 common shrews (*Sorex araneus*), subdivided into three chromosome races: 30 *S. araneus* Cordon, 25 *S. araneus* Bretolet and 41 *S. araneus* Vaud were analysed during this study. All these individuals were sampled by various collectors (DEE collection, Lausanne University) from 1985 to 2003. Distribution of sampling localities is shown in Fig. 1 and covers a large part of the European distribution of the studied taxa. Species and chromosome race identification of most individuals followed karyotype analysis but in a few unambiguous

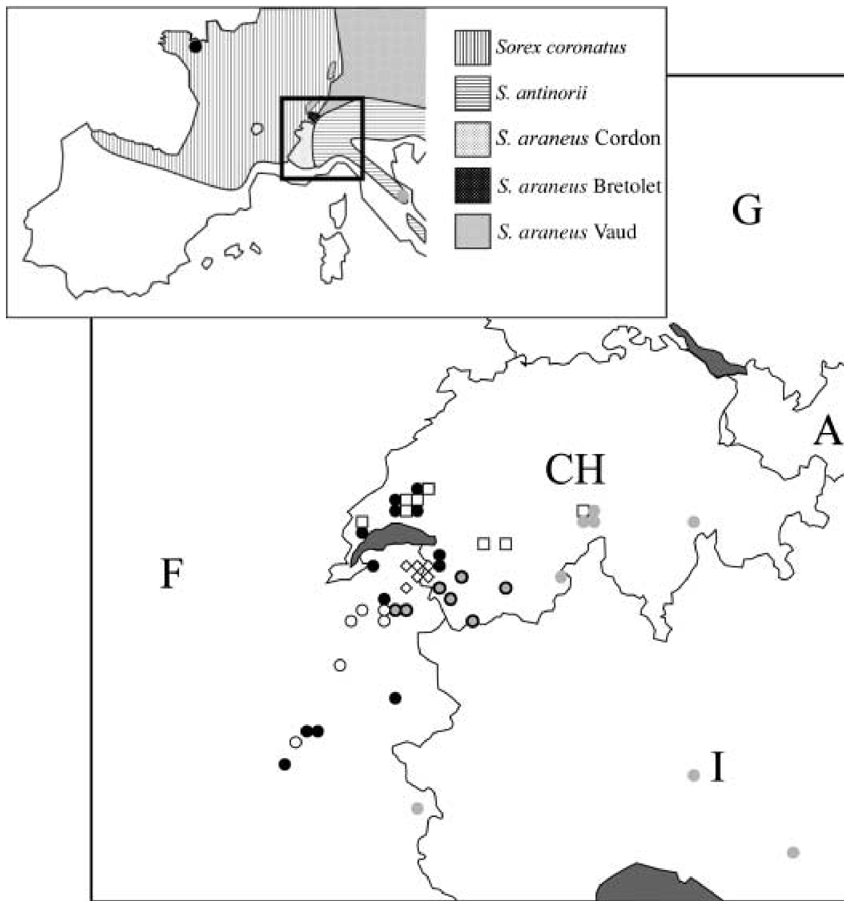

Fig. 1 Insert map: approximate distribution of the studied taxa in southwestern Europe, including remote sampling localities. Below, sampling localities of the five taxa studied; black circles, *Sorex coronatus*; grey circles, *Sorex antinorii* (framed grey circles: St-Bernard group as defined by our analysis); open circles, *Sorex araneus* Cordon; open squares, *Sorex araneus* Vaud; and open diamonds, *Sorex araneus* Bretolet.

cases, it was deduced from sampling localities and morphological analysis.

DNA extraction and microsatellite analysis

Tissue samples (liver, heart or spleen) were stored at –70 °C and total genomic DNA was extracted using the DNeasy Tissue Kit (QIAGEN).

Ten microsatellite loci were used in this study and included loci L9, L13, L67, L99 (polymerase chain reaction (PCR) conditions given in Balloux *et al*. 1998 and Lugon-Moulin *et al*. 2000) and B3, B5, B10, B15, C5, C19 (Table 1). These last six loci were extracted from two enriched *S. araneus* microsatellite libraries developed by a commercial company (Genetic Identification Services). PCR conditions for these loci were as follows: 0.2 mM dNTPs, 0.325 μM of each primer, 1× PCR buffer (QIAGEN) and 0.4 U *Taq* polymerase (QIAGEN). $MgCl_2$ concentration as well as annealing temperature varied (Table 1). For all primers, PCR amplifications were performed in a 20-μL total volume and cycling was carried out in a PE 9700 (Applied Biosytems) using the following profile: 95 °C for 5 min, 35 cycles of 30 s at 94 °C, 30 s at the annealing temperature (Table 1), 30 s at 72 °C; and a final extension at 72 °C for 4 min. One primer of each pair was labelled with a fluorescent dye on the 5′–end, which allowed analyses on an ABI 377XL sequencer (Applied Biosystems). Data collection, sizing of the bands and analyses were carried out using GENESCAN software (Applied Biosystems).

Hardy–Weinberg and linkage equilibrium

The software package GENETIX 4.02 (Belkir *et al*. 2001; www.univ-montp2.fr/~genetix/genetix.htm) was used to calculate the allele frequencies, allele number, observed (H_O) and expected (H_E) heterozygosities for each species or chromosome races. Genotypic linkage disequilibria were tested using GENEPOP version 3.4 (updated from Raymond & Rousset 1995; http://wbiomed.curtin.edu.au/genepop). Deviations from Hardy–Weinberg equilibrium and the significance of Weir & Cockerham (1984) *F*-statistics were evaluated using FSTAT 2.9.4 (Goudet 1995; www.2.unil.ch/popgen/sofwares/fstat.html).

Bayesian analyses

To check the concordance between karyotypic and genetic structure, all the genotypes were screened using a Bayesian

© 2006 Blackwell Publishing Ltd, *Molecular Ecology*, 10.1111/j.1365-294X.2006.02891.x

Table 1 Primer sequence and PCR conditions of six microsatellite loci isolated from two *Sorex araneus* microsatellite libraries developed by Genetic Identification Services

Locus	Primer sequence (5′–3′)	Annealing temp. (°C)	$MgCl_2$ (mM)	Size (bp)	Nb repeats	Accession nos
B3	F: CTTGCCACATTCCCACATC R: AGCCCCACAGCTTTCTCC	57	1	208	30	DQ074646
B5	F: ATGTCTTGCTGGCTGAAGG R: CTGCTGTTCACAAACTCCAAG	55	1.5	196	19	DQ074647
B10	F: CTCCAAACCCTAACACTCTGTC R: TTCACGTGTTCTTTGCTTCC	55	1.5	434	18	DQ074648
B15	F: GTAGAGTTGCTGGCTCAAAGG R: ATGGGAAGACATTGGATTGG	55	1.5	299	18	DQ074650
C5	F: TAGATGACTCTGTGTTCAGGC R: GTTGGGAAGGTAAGATCAGG	55	1.5	236	16	DQ074649
C19	F: TGCCATAAACACCACTTACC R: GTGATCAATACCCTGTGGAG	60	1.5	211	12	DQ074651

admixture procedure implemented in STRUCTURE 2.1 (Pritchard *et al*. 2000; http://pritch.bsd.uchicago.edu). This model was designed to identify the K (unknown) genetic clusters (or populations) of origin of individuals, and simultaneously to probabilistically assign individuals to one cluster or more than one cluster if they are genetically admixed as a result of hybridization. STRUCTURE was run with the 'admixture model', and five repetitions of 100 000 iterations following a burn-in period of 20 000 iterations.

We first assessed population structure using the total data set ($n = 212$), assuming that sampled individuals belong to an unknown number of K genetically distinct clusters. Posterior probability values for K ['Log probability of data'; $L(K)$] were estimated assigning a prior from 1 to 10. Using only this parameter as described by Pritchard *et al*. (2000) it was not obvious which number of clusters (K) best fits our data set (Fig. 2A). Therefore, we followed the recommendation of Evanno *et al*. (2005) and calculated the ΔK statistic, which is based on the rate of change in the 'Log probability of data' between successive K values. We chose the value of $K = 3$, which showed the highest ΔK and then evaluated the individual membership coefficient (q_{ind}) to the three inferred clusters. Individuals with a proportion of membership to each cluster $q_{ind} < 0.90$ (admixed individual) were assigned to more than one cluster whereas individuals with $q_{ind} \geq 0.90$ were assigned to only one cluster. The threshold value of 0.90 was arbitrarily defined to be sure that at least 90% of an individual's genome is assigned to one cluster (Manel *et al*. 2002; Cegelski *et al*. 2003). Then, we assessed the average membership coefficient (q_{group}) of each taxon (species or chromosomal race) to each cluster. Similarly, each sampled taxon (species or chromosome race) was assigned to one cluster if its q_{group} was ≥ 0.90, or jointly to more than one cluster, if its q_{group} to each cluster was < 0.90 (admixed taxon).

In cases of hierarchical population structure, STRUCTURE is known to preferentially detect the uppermost structure level (Evanno *et al*. 2005). First analysis of our data set and the presence of three chromosome races within *S. araneus* suggested that such a situation was present in our case. To explore whether substructure could be detected within each species, the data set of each species was then analysed independently.

Additionally, we investigated the power of our data set to act as a reference to identify individuals of unknown origin. Therefore, we used STRUCTURE with $K = 3$, using the available prior population information (species classification), and option USEPOPINFO = 1. In this way, each shrew of the data set was forced to have its genotype assigned to one of the three species, or, if admixed to more than one species. The same analysis was then performed on the common shrew data set only, to discriminate among the three chromosome races. Finally, we tested the real efficiency of our data set in the identification of unknown individuals using a 'leave one out' procedure. We chose a random subset of individuals (representing about 10% of each taxon) as test individuals. We ran STRUCTURE using the available prior species information ($K = 3$) for all individuals (USEPOPINFO = 1) except for the test individuals (4 *S. coronatus*, 9 *S. antinorii*, 3 *S. araneus* Cordon, 3 *S. araneus* Bretolet and 4 *S. araneus* Vaud) treated as having unknown origin (USEPOPINFO = 0). The same analyses were then repeated 10 times, each time randomly selecting the same number of test individuals. Assignation results of the test individuals were then used to estimate the percentage of correct assignations (individuals correctly assigned with $q_{ind} \geq 0.90$), unassigned individuals (showing admixture $0.10 < q_{ind} < 0.90$) and assignment mistakes (shrews assigned to an incorrect cluster with high probability $q_{ind} \geq 0.90$). The same procedure was then repeated on the *S. araneus* data set only.

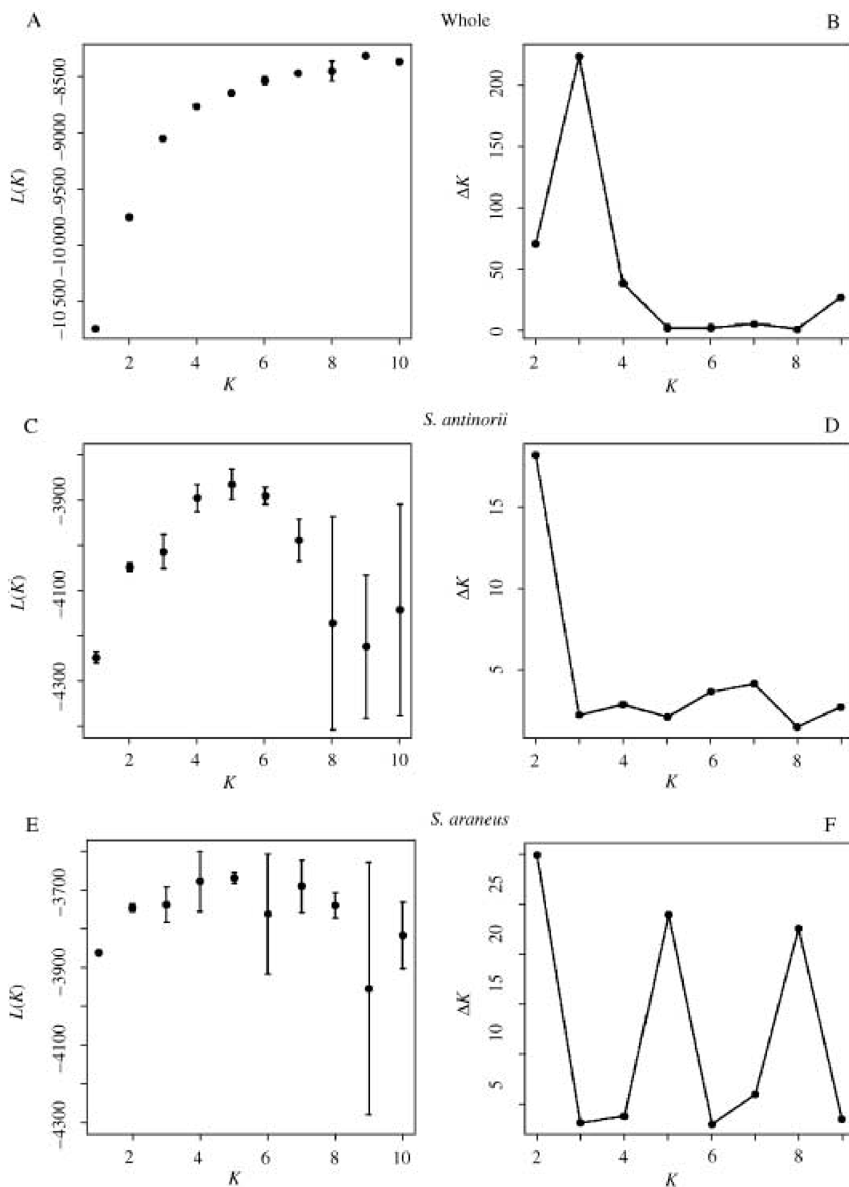

Fig. 2 Detection of the number of groups in the whole data set (A and B), *Sorex antinorii* data set (C and D) and *Sorex araneus* data set (E and F). (A) (C) and (E): Mean $L(K)$ ($\pm$ SD) over five runs as a function of K. (B) (D) and (F): ΔK following Evanno *et al*. (2005) as a function of K.

Results

Population genetic diversity

We determined the individual genotypes at 10 microsatellite loci in 212 shrews. All microsatellite markers were polymorphic and the number of alleles per locus ranged from 5 to 43 (average 28.5 ± 13.4). In the following analyses, the data set was first subdivided into the three species included in this study (*Sorex coronatus*, *Sorex antinorii* and *Sorex araneus*). Then *S. araneus* was subdivided into the three chromosome races, *S. araneus* Cordon, *S. araneus* Bretolet and *S. araneus* Vaud. Number of alleles per taxon ranged from 70 (*S. coronatus*) to 221 (*S. antinorii*). The number of private alleles ranged from 3 (*S. araneus* Bretolet) to 75 (*S. antinorii*) with an average of 43.7 ± 28.0 when pooling the three chromosome races of *S. araneus* or 24.2 ± 29.2 when subdividing *S. araneus* in the three races (Table 2).

Tests of fit to Hardy–Weinberg and linkage equilibria and divergence among shrew taxa

All five species and chromosome races showed H_O values lower than expected, with average F_{IS} values that were positive (from 0.232 to 0.454), highly significantly different from 0 ($P \leq 0.002$) indicating deviation from Hardy–Weinberg equilibrium (Table 2). This suggests a strong Wahlund effect, probably resulting from the pooling of individuals of geographically and genetically different origins. At least a part of this homozygote excess could be

© 2006 Blackwell Publishing Ltd,

Table 2 Genetic diversity in the three species of shrews and in the three chromosome races of *Sorex araneus* over the 10 microsatellite loci

	No. of alleles	No. of private alleles	H_E	H_O	Hardy–Weinberg equilibrium (F_{IS})
1 *S. coronatus* (n = 33)	70	21	0.558 (0.273)	0.315 (0.239)	0.454***
2 *S. antinorii* (n = 83)	221	75	0.779 (0.256)	0.561 (0.284)	0.286***
3 *S. araneus* all (n = 96)	168	35	0.738 (0.273)	0.527 (0.295)	0.291***
3.1 Cordon (n = 30)	126	14	0.727 (0.307)	0.573 (0.300)	0.232***
3.2 Bretolet (n = 25)	104	3	0.717 (0.241)	0.542 (0.289)	0.263***
3.3 Vaud (n = 41)	111	8	0.688 (0.283)	0.486 (0.310)	0.305***
Overall	285		0.829 (0.192)	0.501 (0.271)	0.296***

H_E, expected heterozygosity without bias (Nei 1978); H_O, observed heterozygosity; standard deviation in parentheses.
F_{IS}, Deviation from Hardy–Weinberg equilibrium following Weir & Cockerham (1984), ***$P \leq 0.002$.

Table 3 Estimates of pairwise genetic differentiation (F_{ST}) among shrew taxa. ***$P \leq 0.01$

Taxa	*Sorex coronatus*	*Sorex antinorii*	*Sorex araneus* Cordon	Bretolet	Vaud
S. coronatus	—	0.172***	0.300***	0.312***	0.326***
S. antinorii		—	0.127***	0.153***	0.174***
S. araneus					
Cordon			—	0.046***	0.064***
Bretolet				—	0.028***
Vaud					—

explained by the presence of null alleles (Pemberton *et al*. 1995). To test for this effect, the correlation between the number of nonamplifying samples for each locus and F_{IS} was examined. In each taxa, individuals were grouped into geographic sampling localities. Localities with less than three individuals were left out of the analysis. A significant correlation was found only for locus L99 ($R^2 = 0.942$, $P \leq 0.001$). This result can easily be explained by the low allele number of this locus, which is confirmed by the fact that this locus did not show homozygote excess in any of the populations tested. Thus, it seems more probable that the general homozygote excess is due to the pooling of genetically differentiated individuals. After Bonferroni correction for multiple comparisons, the loci B10 and L13 were not in linkage equilibrium in *S. antinorii*. However, these loci map to different chromosomes (P. Basset, unpublished).

Genetic diversity was significantly partitioned among the three species ($F_{ST} = 0.199$, $P \leq 0.001$) and chromosome races ($F_{ST} = 0.047$, $P \leq 0.001$). All pairwise F_{ST} values (Table 3) were significantly different from 0 ($P \leq 0.01$).

Genetic admixture and assignment analysis

We used Bayesian analyses (STRUCTURE, POPINFO = 0) to detect admixture and possible cryptic substructure in our data set (n = 212). Using only the 'Log probability of data' (as described in Pritchard *et al*. 2000) it was not clear which number of clusters best fits our data (Fig. 2a). However, the statistic ΔK described by Evanno *et al*. (2005) clearly indicates that the sample included at least three distinct groups (the highest ΔK was obtained with $K = 3$, Fig. 2B).

The average proportions of membership (q_{group}) of each sampled taxa in the three clusters (Table 4) showed that all

Table 4 Bayesian clustering analyses for the *Sorex* reference data set (212 individuals; 10 loci) performed using STRUCTURE (Pritchard *et al*. 2000)

Taxa	Cluster I	II	III
S. coronatus (n = 33)	**0.991**	**0.004**	**0.005**
S. antinorii (n = 83)	**0.007**	**0.969**	**0.024**
Ind n°2883	0.003	0.899	0.098
Ind n°3070	0.003	0.845	0.152
Ind n°3071	0.002	0.885	0.113
Ind n°3134	0.004	0.560	0.436
Ind n°5319	0.003	0.893	0.104
S. araneus Cordon (n = 30)	**0.006**	**0.105**	**0.890**
Ind n°3336	0.002	0.249	0.749
Ind n°3337	0.002	0.322	0.676
Ind n°3341	0.002	0.423	0.575
Ind n°3342	0.003	0.672	0.325
Ind n°3348	0.009	0.444	0.547
Ind n°3373	0.003	0.167	0.831
Ind n°3379	0.009	0.313	0.677
Ind n° LC2	0.044	0.185	0.771
S. araneus Bretolet (n = 25)	**0.019**	**0.024**	**0.956**
Ind n°2345	0.209	0.003	0.787
Ind n°3274	0.185	0.005	0.810
Ind n°3275	0.006	0.454	0.540
S. araneus Vaud (n = 41)	**0.005**	**0.007**	**0.988**

NB: In bold, average proportion of membership (q_{group}) of each predefined population in each of $K = 3$ inferred clusters. Admixed individuals ($q_{ind} < 0.90$) are indicated under each population lines.

the *S. coronatus* individuals grouped in cluster I ($q_{coronatus}$ I = 0.99) while *S. antinorii* was significantly assigned to cluster II ($q_{antinorii}$ II = 0.97). However, five individuals (out of 83) of this species showed signs of admixture (q_{ind} II < 0.90) with *S. araneus*. Cluster III represented the '*S. araneus* cluster' since this species grouped in this cluster with $q_{araneus}$ III = 0.95. Cordon was the only chromosome race to show admixture with *S. antinorii* (q_{Cordon} III = 0.89 and 8/30 individuals had q_{ind} III between 0.33 and 0.83). In contrast, every Vaud and all but three Bretolet individuals grouped in cluster III (q_{Vaud} III = 0.99 and $q_{Bretolet}$ III = 0.96, respectively). To summarize, of the 212 individuals tested, no individual grouped with a cluster different from its putative origin and only 16 (7.5%) showed signs of admixture. Half of these individuals were karyotypically identified as *S. araneus* Cordon and showed signs of admixture with *S. antinorii*.

Preliminary analysis and the presence of three different chromosome races within *S. araneus* prompted us to explore whether substructure could be detected within each of the three species. No substructure was detected within *S. coronatus* as the most probable *K* for this species is 1. For *S. antinorii* and *S. araneus* two distinct groups were detected within each of these species. Again, it was necessary to estimate the ΔK statistic to decide which *K* best fits the data (Fig. 2C–F). For *S. araneus*, a careful comparison of this statistic with the $L(K)$ was necessary as more than one ΔK peaks were detected.

The *S. antinorii* data set was split into two well geographically differentiated clusters (Fig. 3A): individuals sampled in Italy, eastern Switzerland and the southern French Alps (Hautes-Alpes) grouped in one cluster while individuals sampled in western Switzerland and the northern part of the French Alps grouped in a second cluster (framed grey circles in Fig. 1). Two clusters were also detected within the *S. araneus* data set ($q_{araneus}$ I = 0.536, Fig. 3B). Each of the three chromosome races showed signs of admixture between the two clusters with q_{Cordon} I = 0.865, $q_{Bretolet}$ I = 0.743 and q_{Vaud} I = 0.168.

Shrew ancestry was then estimated using prior information (POPINFO = 1) about species ($K = 3$). Each species grouped in their respective cluster with high probability ($q_{species} \geq 0.99$). Of the 212 individuals tested, none was assigned to a species different from its putative origin but five (2.3%) showed signs of admixture (ind. n°3071, 3134, 3341, 3379 and 2345). Finally, ancestry was estimated on the *S. araneus* data set using prior information (POPINFO = 1) about chromosomal race ($K = 3$). Individuals clustered into three groups corresponding to the three chromosome races studied (Cordon, Bretolet and Vaud) with high probability ($q_{race} \geq 0.96$). Again, of the 96 individuals tested, only five (5.2%) showed signs of admixture (ind. n°3349, LC6, 2596, 3021 and 3261).

We tested the efficiency of our data set to act as reference to identify individuals with unknown species identity and to identify *S. araneus* individuals with unknown chromosome

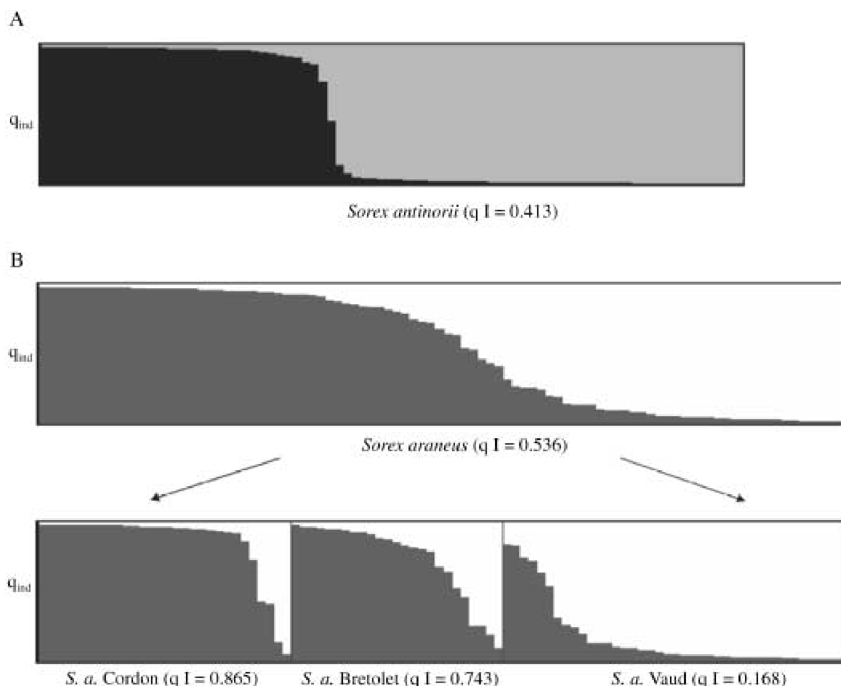

Fig. 3 Distruct plots (Rosenberg 2004) for *Sorex antinorii* (A) and *Sorex araneus* (B). Each individual is represented by a line partitioned into two colours ($K = 2$) representing its genotypic assignment to one cluster or the other (q_{ind}). In both species, individuals are sorted according to decreasing q_{ind} values. The *Sorex araneus* data set is then sorted according to the three chromosome races Cordon, Bretolet and Vaud. Mean q I-values are given in parentheses.

Table 5 Percentage of correct assignation ($q_{ind} \geq 0.90$), assignation with admixture ($q_{ind} < 0.90$) and wrong assignation ($q_{ind\ wrong} \geq 0.90$ for wrong taxa) of species (A) and *Sorex araneus* chromosome race (B)

A

	S. coronatus	*S. antinorii*	*S. araneus*
Correct ($q_{ind} \geq 0.90$)	100.0%	92.3%	86.4%
Admix ($q_{ind} < 0.90$)	0.0%	7.7%	13.6%
Wrong ($q_{ind\ wrong} \geq 0.90$)	0.0%	0.0%	0.0%

B

	S. araneus		
	Cordon	Bretolet	Vaud
Correct ($q_{ind} \geq 0.90$)	48.3%	22.5%	41.5%
Admix ($q_{ind} < 0.90$)	48.3%	75.0%	56.1%
Wrong ($q_{ind\ wrong} \geq 0.90$)	3.4%	2.5%	2.4%

race identity. All individuals showing admixture signs in the previous analyses were left out from this reference data set. Then assignment tests were performed on our global sample (including admixed individuals) using this reduced data set ($n = 207$ for the species data set and $n = 91$ for the *S. araneus* data set). Species identification was correct in 100% of *S. coronatus*, 92.2% of *S. antinorii* and 86.4% of *S. araneus* using the criterion $q_{ind} \geq 0.90$ (Table 5A). For these last two species, 8.8% and 13.4%, respectively, of the individuals were identified as admixed ($q_{ind} < 0.90$ level). None of the individuals with q_{ind} values ≥ 0.90 were assigned to a cluster different from its correct origin (assignment mistakes).

Chromosome race identification within *S. araneus* was much lower (Table 5B). Only 48.3%, 22.5% and 41.5% of individuals belonging to the Cordon, Bretolet and Vaud race, respectively, were correctly assigned using the criterion $q_{ind} \geq 0.90$. A large percentage of the individuals (respectively 48.3%, 75.0% and 56.1%) could not be assigned to any cluster and showed clear admixture. Finally, in 3.4%, 2.5% and 2.4% of the cases, respectively, individuals were assigned to a wrong cluster with a high q_{ind} value ($q_{ind} \geq 0.90$).

Discussion

Species introgression

Species and chromosome races of the *Sorex araneus* group form numerous contact or hybrid zones (Searle & Wójcik 1998). This observation is particularly true in the Alpine region, which is known to be a suture zone for numerous taxa (Taberlet *et al*. 1998; Hewitt 2001). Recent and/or past introgression among these taxa explains why relationships among taxa of the *S. araneus* group in the Alps are challenging issues that are much debated (Taberlet *et al*. 1994; Brünner *et al*. 2002b; Lugon-Moulin & Hausser 2002). However, no trial has been carried out to study genetic structure without using prior population information and to detect hidden or unexpected structure. The first goal of our work was to check if the genetic structure matched the structure defined by the different karyotypes involved in this group (defined as the taxonomical structure) without using prior information about population. Results of this study confirmed the taxonomic status of the three species *Sorex coronatus*, *S. antinorii* and *S. araneus* since three distinct clusters were detected by STRUCTURE. Each cluster consistently grouped individuals according to their taxonomic status but it is interesting to note that weak introgression can still be detected between *S. antinorii* and *S. araneus* (Table 4). It should be noticed that a specific status was given to *S. antinorii* (Brünner *et al*. 2002a) notwithstanding a very limited gene flow with adjacent populations of *S. araneus* (see Discussion Brünner *et al*. 2002a).

Five individuals (out of 83) of *S. antinorii* showed signs of admixture with *S. araneus*. About the same proportion of individuals of *S. araneus* (11 out of 96) showed admixture with *S. antinorii*. However, eight of these individuals belonged to the chromosome race Cordon and admixed individuals of this race showed a much higher level of introgression ($0.33 \leq q_{ind}$ III ≤ 0.83) than the other taxa. Such introgression could be explained by the presence of at least one hybrid zone with *S. antinorii* (Brünner & Hausser 1996). This hypothesis is confirmed by the sampling localities of five admixed individuals situated close to the contact zone. However, we did not detect any introgression of *S. antinorii* into the *S. araneus* Vaud chromosome race notwithstanding the occurrence of hybridization between these two taxa (Brünner *et al*. 2002b). Differences in the karyotype complexity of these two chromosome races could explain the disparity in introgression level. The Cordon race is known for its acrocentric karyotype; hybrids with *S. antinorii* should form mostly trivalents and encounter only mild problems at meiosis (Brünner *et al*. 2002b). In contrast, the Vaud race is defined by a much more metacentric karyotype showing up to eight monobrachial homologies with *S. antinorii*. Hybrids should then meet more chromosomal incompatibilities (Brünner *et al*. 2002b) and gene flow between these two taxa should be greatly reduced.

Only a few (5 out of 83) individuals of *S. antinorii* showed signs of admixture with *S. araneus*, therefore introgression seems to be unidirectional (*S. antinorii* into *S. araneus*). This pattern could be explained by differences in the sampling distribution of the two taxa, i.e. most Cordon individuals were sampled close to some *S. antinorii* localities, whereas

sample distribution of *S. antinorii* is much larger and only a few of these individuals were sampled close to Cordon localities. However, more comprehensive studies of contact zones between these two taxa are necessary to clarify this pattern.

Structure within species

Evanno *et al.* (2005) showed that when confronted with complex migration schemes, STRUCTURE detects the uppermost hierarchical level of population structure. Our study illustrates this phenomenon as we detected substructure into two out of three species, a result not evidenced by the first analyses of the data set.

Bayesian clustering of the *S. antinorii* data set revealed two different, geographically based clusters. One group contained all individuals sampled in the northern part of the French Alps and western Switzerland (St-Bernard region, Fig. 1). The second group contained the remaining individuals sampled in Italy, eastern Switzerland and the southern French Alps. This geographical subdivision confirmed the possible presence of at least two different recolonization routes (i.e. Simplon and St-Bernard pass) of *S. antinorii* from the Italian peninsula, as postulated by Lugon-Moulin & Hausser (2002) and Fivaz *et al.* (2003).

In addition, the present work clarified some of the hypotheses developed by the same authors to account for the near absence of gene flow between the two regions. These authors postulated that two distinct male lineages could have already been differentiated in Italy. In our study, all individuals sampled in Italy (Apennine region) clustered with the eastern Switzerland and southern French Alps group. Further Italian samples are needed to properly address the recolonization of Switzerland by *S. antinorii*, and in particular to verify if the secluded Aosta valley, leading to the St-Bernard pass, actually hosts populations differentiated from the other Italian ones.

Despite the presence of three chromosome races, analyses of the *S. araneus* data set revealed only two different clusters. However, partition of individuals in these clusters generally followed karyotypic identification, i.e. most individuals of the race Cordon grouped in one cluster (q_{Cordon} I = 0.865) while most individuals of the Vaud race grouped in the second cluster (mean q_{Vaud} I = 0.168). The chromosome race Bretolet showed an admixed status between these two clusters although it shows closer links to the Cordon cluster ($q_{Bretolet}$ I = 0.743), suggesting the hybrid status of this chromosome race. Actually, this karyotypic group was postulated to be formed by introgression of 'Vaud' metacentric chromosomes into Cordon populations (Hausser *et al.* 1991).

We did not find any population structure within *S. coronatus* but it should be noted that the sampling effort was weaker for this species ($n = 33$). This species has a larger allelic diversity compared to *S. araneus* suggesting that substructure could still be hidden.

Effect of karyotype on genetic structure

Differences in karyotypes are thought to reduce gene flow among different karyotypic groups (Rieseberg 2001; Navarro & Barton 2003; Panithanarak *et al.* 2004). Data concerning the *S. araneus* group are scarce but in our case, this should be translated into reduced gene flow among the three chromosome races. We were therefore expecting three distinct clusters within *S. araneus* (corresponding to the three chromosome races sampled) and only one cluster within *S. antinorii* (this species has a homogenous karyotype). Surprisingly, our study did not confirm this conclusion as two clusters were detected in both species. In addition, levels of introgression between clusters seemed larger within *S. araneus* (Fig. 2B) compared to *S. antinorii* (Fig. 2A). This is also suggested by the rather low F_{ST} values calculated among chromosome races within *S. araneus* ($F_{ST\ araneus}$ from 0.028 to 0.064) compared to the larger F_{ST} between the two geographical clusters within *S. antinorii* ($F_{ST\ antinorii} = 0.082$). These results suggest that in the taxa studied, karyotypic differences played only a minor role in structuring the populations relative to historical and/or geographical factors. Andersson *et al.* (2004) obtained similar results in a hybrid zone between the North (Abisko race) and West (Sidensjö race) European karyotypic group of *S. araneus*. However, the real impact of karyotypic changes on the genetic structure among taxa of the *S. araneus* group can only be addressed through detailed individual analyses of the chromosomes involved in these changes.

Species identification

Sibling species can be defined as those sister species that are very difficult or virtually impossible to distinguish by morphological characters (Futuyma 1986). This applies to species and chromosome races of the *S. araneus* group, and identification methods have already been the topic of several papers (Hausser *et al.* 1991; Wójcik *et al.* 1996; Basset & Hausser 2003; Pfunder *et al.* 2004). These authors used morphometric measures, diagnostic microsatellite markers or microarray-based analysis to distinguish the different species of this group. However, drawbacks were found with each method (reviewed in Basset & Hausser 2003) and only morphometric analyses gave information about the possible hybrid status of an individual; a crucial point when studying potentially hybridizing groups of species. Our method seemed to perform well at the species level since we identified with a $q_{ind} \geq 0.90$ criterion, 100% of the *S. coronatus* individuals, 92.2% of *S. antinorii* and

© 2006 Blackwell Publishing Ltd,

86.4% of *S. araneus* (Table 5A). These values are close to those obtained by Hausser & Jammot 1984) and Hausser *et al*. (1991) using morphological measures combined with discriminant analyses. These authors found correct classification in more than 95% of the cases but this technique is particularly time-consuming, and requires that the same person analyses the reference and the individuals to assign.

Identification efficiency strongly decreased when we tried to discriminate among the different chromosomal races within *S. araneus*. Using the $q_{ind} \geq 0.90$ criterion, identification within this species did not exceed 48.3% (Cordon race). Such low assignment power could be explained by the low number of individuals sampled for each of these chromosome races (Evanno *et al*. 2005). But F_{ST} values among these taxa were also the lowest observed in this study (Table 3). Assignment power of STRUCTURE is known to be well correlated with measures such as F_{ST} (Berry *et al*. 2004). *S. coronatus* also illustrates this relationship, as this species showed the highest F_{ST} values with all other taxa ($F_{ST} \geq 0.17$) (Table 3) and was always correctly identified or separated from the other taxa. Berry *et al*. (2004) have shown that increasing the number of genetic markers could increase the assignment power particularly when F_{ST} values are rather low ($F_{ST} \leq 0.08$). Genetic discrimination between other chromosome races of *S. araneus* has already proven to be difficult (Andersson *et al*. 2004).

In conclusion, Bayesian assignment analyses revealed large-scale introgression of *S. antinorii* into one chromosome race (Cordon) of *S. araneus*. In addition, geographically based cryptic substructure was discovered within *S. antinorii*, a pattern consistent with the different putative postglacial recolonization routes of this species. This study furthermore illustrates the relatively low impact of chromosomal differences on the genetic structure compared to historical factors. Finally, the data set described in this article is available to researchers at www.unil.ch/dee/page7010_en.html#1, so that it can be used as a reference dataset to help identify unknown *Sorex* individuals sampled in the western Alps or clarify the relationships among individuals sampled in hybrid zones between species of the *S. araneus* group. This last point is indeed particularly important as it is generally very difficult to obtain reference datasets when working on hybrid zones.

Acknowledgements

We are particularly grateful to Laura Brändli, Guillaume Evanno, Julie Jaquiéry, Lori Handley and Valérie Vogel for helpful comments on the manuscript. This work was supported by the Swiss NSF (grants no: 3100-05943), a Société Académique Vaudoise (Switzerland) grant to P.B. and a Fondation Agassiz (Switzerland) grant to P.B.

References

Andersson AC, Narain Y, Tegelström H, Fredga K (2004) No apparent reduction of gene flow in a hybrid zone between the West and North European karyotypic groups of the common shrew, *Sorex araneus*. *Molecular Ecology*, **13**, 1205–1215.

Balloux F, Ecoffey E, Fumagalli L, Goudet J, Wyttenbach A, Hausser J (1998) Microsatellite conservation, polymorphism, and GC content in shrews of the genus *Sorex* (Insectivora, Mammalia). *Molecular Biology and Evolution*, **15**, 473–475.

Basset P, Hausser J (2003) Identification of three *Sorex* species with microsatellite markers. *Mammalia*, **68**, 245–252.

Belkir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2001) *Genetix 4.03, logiciel sous Windows™ pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Berry O, Tocher MD, Sarre SD (2004) Can assignment tests measure dispersal? *Molecular Ecology*, **13**, 551–561.

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.

Brünner H, Lugon-Moulin N, Balloux F, Fumagalli L, Hausser J (2002a) A taxonomical re-evaluation of the Valais chromosome race of the common shrew *Sorex araneus* (Insectivora: Soricidae). *Acta Theriologica*, **47**, 245–275.

Brünner H, Lugon-Moulin N, Hausser J (2002b) Alps, genes, and chromosomes: their role in the formation of species in the *Sorex araneus* group (Mammalia, Insectivora), as inferred from two hybrid zones. *Cytogentic and Genome Research*, **96**, 85–96.

Cegelski CC, Waits LP, Anderson NJ (2003) Assessing population structure and gene flow in Montana wolverines (*Gulo gulo*) using assignment-based approaches. *Molecular Ecology*, **12**, 2907–2918.

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics*, **153**, 1989–2000.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Fivaz F, Basset P, Lugon-Moulin N, Hausser J (2003) Postglacial recolonization of the Valais (Switzerland) by the shrew *Sorex antinorii*: is dispersal sex-biased? A preliminary study. *Mammalia*, **68**, 253–262.

Fredga K, Narain Y (2000) The complex hybrid zone between the Abisko and Sidenjö chromosome races of the *Sorex araneus* in Sweden. *Biological Journal of the Linnean Society*, **70**, 285–307.

Fumagalli L, Hausser J, Taberlet P, Gielly L, Stewart DT (1996) Phylogenetic structure of the Holarctic *Sorex araneus* group and its relationships with *S. samniticus*, as inferred from mtDNA sequences. *Hereditas*, **125**, 191–199.

Futuyma DJ (1986) *Evolutionary Biology*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

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

Hausser J (1978) Répartition en Suisse et en France de *Sorex araneus* L., 1758 et de *Sorex coronatus* Millet, 1828 (Mammalia, Insectivora). *Mammalia*, **42**, 329–341.

Hausser J, Jammot D (1984) Etude biométrique des mâchoires chez les *Sorex* du groupe *araneus* en Europe continentale (Mammalia, Insectivora). *Mammalia*, **38**, 324–343.

Hausser J, Zuber N (1983) Détérmination spécifique d'individus vivants des deux espèces jumelles *Sorex araneus* et *S. coronatus*, par deux techniques biochimiques (Insectivora, Soricidae). *Revue Suisse de Zoologie*, **90**, 857–862.

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

Hewitt GM (2001) Speciation, hybrid zones and phylogeography — or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

King M (1993) *Species Evolution. The Role of Chromosomal Changes*. Cambridge University Press, Cambridge.

Lugon-Moulin N, Hausser J (2002) Phylogeographical structure, postglacial recolonization and barriers to gene flow in the distinctive Valais chromosome race of the common shrew (*Sorex araneus*). *Molecular Ecology*, **11**, 785–794.

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.

Lugon-Moulin N, Balloux F, Hausser J (2000) Genetic differentiation of common shrew (*Sorex araneus*) population among different alpine valleys revealed by microsatellites. *Acta Theriologica*, **45**, 103–117.

Manel S, Berthier P, Luikart G (2002) Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Conservation Biology*, **16**, 650–659.

Navarro A, Barton NH (2003) Chromosomal speciation and molecular divergence — accelerated evolution in rearranged chromosomes. *Science*, **300**, 321–324.

Neet CR (1992) The use of fur colour characters to distinguish the sibling species *Sorex araneus* and *Sorex coronatus* (Insectivora, Soricidae) — a field test in a zone of parapatric contact. *International Journal of Mammalian Biology*, **56**, 176–178.

Neet CR, Hausser J (1990) Habitat selection in zones of parapatric contact between the common shrew *Sorex araneus* and Millet's shrew *S. coronatus*. *Journal of Animal Ecology*, **59**, 235–250.

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

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, **4**, 347–354.

Panithanarak T, Hauffe HC, Dallas JF, Glover A, Ward RG, Searle JB (2004) Linkage-dependent gene flow in a house mouse chromosomal hybrid zone. *Evolution*, **58**, 184–192.

Pemberton JM, Slate J, Bancroft DR, Barrett JA (1995) Nonamplifying alleles at microsatellite loci-a caution for parentage and population studies. *Molecular Ecology*, **4**, 249–252.

Pfunder M, Holzgang O, Frey E (2004) Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I vs. cytochrome *b* as genetic markers. *Molecular Ecology*, **13**, 1277–1286.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences, USA*, **4**, 9197–9201.

Ratkiewicz M, Fedyk S, Banaszek A *et al.* (2002) The evolutionary history of the two karyotypic groups of the common shrew, *Sorex araneus*, in Poland. *Heredity*, **88**, 235–242.

Ratkiewicz M, Banaszek A, Jadwiszczak K, Chetnicki W, Fedyk S (2003) Genetic diversity, stability of population structure and barriers to gene flow in a hybrid zone between two *Sorex araneus* chromosome races. *Mammalia*, **67**, 275–283.

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

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, **16**, 351–358.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology*, **4**, 137–138.

Searle JB, Wójcik JM (1998) Chromosomal evolution: the case of *Sorex araneus*. In: *Evolution of Shrews* (eds Wójcik JM, Wolsan M), pp. 219–262. Mammal Research Institute, Polish Academy of Sciences, Bialowieza, Poland.

Sharman GB (1956) Chromosome of the common shrew. *Nature*, **177**, 941–942.

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.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Vázquez-Dominguez E, Paetkau D, Tucker N, Hintens G, Moritz C (2001) Resolution of natural groups using iterative assignment tests: an example from two species of Australian native rats (*Rattus*). *Molecular Ecology*, **10**, 2069–2078.

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

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Wójcik JM, Wójcik AM, Zalewska H, Rychlik L (1996) Allozyme differentiation in four sympatric species of European shrews (Soricidae: Mammalia). *Biochemical Systematics and Ecology*, **24**, 291–298.

Wójcik JM, Ratkiewicz M, Searle JB (2002) Evolution of the common shrew *Sorex araneus*: chromosomal and molecular aspects. *Acta Theriologica*, **47**, 139–167.

This research is part of Patrick Basset's PhD thesis which focuses on the role of chromosomal rearrangements in the speciation process of shrews. Glenn Yannic (PhD student) is studying population genetics, phylogeography and hybrid zones in the common shrew focusing especially on sex-specific markers. Jacques Hausser is Patrick and Glenn's supervisor and his interests include chromosome polymorphism, speciation process and phylogeography in shrews and snails.