

Estimating gene flow across hybrid zones: how reliable are microsatellites?

François BALLOUX, Nicolas LUGON-MOULIN and Jacques HAUSSER

Balloux F., Lugon-Moulin N. and Hausser J. 2000. Estimating gene flow across hybrid zones: how reliable are microsatellites? [In: Evolution in the *Sorex araneus* group: Cytogenetic and molecular aspects. J. B. Searle and J. M. Wójcik, eds]. Acta Theriologica 45, Suppl. 1: 93–101.

Due to their high polymorphism, microsatellites have become one of the most valued genetic markers in population biology. We review the first two published studies on hybrid zones of the common shrew based on microsatellites. Both reveal surprisingly high interracial gene flow. It can be shown that these are overestimates. Indeed, in classical population genetics models as *F*-statistics, mutation is neglected. This constitutes an acceptable assumption as long as migration is higher than mutation. However, in hybrid zones where genetic exchanges can be rare, neglecting mutation will lead to strong overestimates of migration when working with microsatellites which display mutation rates up to 10^{-3} . As there seems to be no straightforward way to correct for this bias, interracial gene flow estimates based on microsatellites should be taken with caution. This problem should however not conceal the enormous potential of microsatellites to unravel the genetics of hybrid zones.

Institut d'Ecologie, Laboratoire de Zoologie et d'Ecologie Animale, Université de Lausanne, 1015 Lausanne-Dorigny, Switzerland, e-mail: Francois.Balloux@ie-zea.unil.ch

Key words: *Sorex araneus*, mutation rate, population structure, gene exchange, contact zones

Analysing hybrid zones with genetic markers

One of the central questions in biology is how new species evolve (Coyne and Orr 1998). If it is generally accepted that only reproductively isolated populations can give rise to distinct species, barriers to gene flow as diverse as geography, morphology, phenology, behaviour, chromosomes or genetic incompatibilities have been invoked (Barton and Hewitt 1989). However, the relative roles of these barriers to gene exchanges are still under debate. Actually, once populations are differentiated, it is extremely difficult to find out which initial causes promoted speciation (Coyne and Orr 1998). One approach is to study situations where gene flow is impeded but not completely suppressed as in hybrid zones (Hewitt 1988, Harrison 1990).

The common shrew *Sorex araneus* Linnaeus, 1758 is an interesting model species to study barriers to gene flow and in particular the possible role of

chromosomes, since its karyotype is characterised by dramatic chromosomal variation. Indeed, more than 50 chromosomal races have been characterised so far (Zima *et al.* 1996) and several interracial hybrid zones have been described (review in Searle and Wójcik 1998). However, shrews as a model system are also characterised by some severe limitations. Being inconspicuous as most small mammals, they are very difficult to study in the field. In addition, shrews of the genus *Sorex* are difficult to maintain in captivity and have long generation times. Therefore, one of the most efficient ways to study hybrid zones in the common shrew is through indirect estimates drawn from genetic markers.

Genetic studies should ideally be undertaken with selectively neutral and polymorphic loci. There are several classes of genetic markers. The most obvious class for the study of chromosome races are karyotypes. Polymorphic Robertsonian fusions can be analysed as di-allelic loci. Karyotypes provide precious information over the geographical localisation of the hybrid zones and insights into the frequency of interracial gene exchanges. The problem is that different chromosomal fusions are themselves potential barriers to gene flow, and cannot therefore be considered as selectively neutral. In most cases, allozymes are not very helpful since their diversity is extremely low both within and among karyotypic races (review in Ruedi 1998). Another class of genetic markers are microsatellites; they seem to have all the required qualities to study gene flow in hybrid zones of the common shrew. To date, seventeen microsatellite markers for *Sorex araneus* have been deposited in Genbank. Here we will review the advantages and also the limitations of microsatellites for the study of hybrid zones of the common shrew.

Microsatellite markers

In recent years, microsatellites have become the marker of choice in population genetics (Jarne and Lagoda 1996). Indeed, these sequences of repeated DNA combine some of the most desirable properties for genetical studies. They generally display an impressive polymorphism. It is not rare to observe loci with heterozygosities over 0.9 and more than 20 alleles in reasonably sized datasets. Although some microsatellites may be in linkage disequilibrium with genes under selection, most microsatellites seem to behave neutrally. Another quality of microsatellite markers is their codominance. In other words, it is possible to determine at each given locus whether an individual is homozygous or heterozygous. Codominance allows full advantage to be taken of the most complete population genetics models. This is a major advantage over other molecular markers such as RAPDs, multilocus fingerprints or AFLP, which involve scoring the presence or absence of anonymous bands. In addition, microsatellite loci are rather easy to use, once developed. Visualisation of an individual's alleles can be obtained through a PCR reaction with the specific primers followed by electrophoresis. As the PCR reaction amplifies the target sequence to millions of copies, minute amounts of tissues are needed, thus permitting non-invasive sampling. These properties have allowed population

genetics studies at a much more detailed scale than was possible with the less polymorphic allozymes (eg Amos *et al.* 1993, Garza *et al.* 1997, Balloux *et al.* 1998).

Published studies

Two hybrid zones between chromosome races of the common shrew have been analysed with microsatellites so far. The first is located at Les Houches in the French Alps at the foot of the Mont Blanc glacier. There, the Cordon and Valais races meet at a mountain torrent. The intra- and interracial genetic structure of this zone has been analysed with microsatellite markers in a series of papers by Lugon-Moulin *et al.* (1996, 1999a, b).

Genetic structuring among the two races was found to be highly significant. Interracial gene exchange was found to be reduced and estimated at one migrant every two generations using R -statistics and one every generation using F -statistics. Furthermore, chromosomal hybrids were genetically indistinguishable from karyotypically pure individuals, suggesting that hybrids are backcrossed individuals of deep hybrid ancestry, and not F1 or F2 progenies. This latter analysis suggests that interracial gene exchanges are uncommon at the present time, although it does not provide a quantitative estimate.

The second published study of a *S. araneus* hybrid zone analysed with microsatellites is that of Wyttenbach *et al.* (1999a). These authors studied intra- and interracial gene flow in the Hällefors and Uppsala races that meet at the river Kolbäck in central Sweden. They found very low interracial structuring. Actually, their interracial estimates of gene flow were as small as intraracial estimates. This is a surprising result since the two races are well differentiated on the basis of their chromosomes. Furthermore, few karyotypic hybrids could be detected outside the close vicinity of the very narrow contact zone.

At a first glance, the microsatellite analyses in both studies seem to give rather different results. In the Les Houches hybrid zone, gene flow between races is reduced, whereas in Sweden no reduction can be detected between the hybridizing races. The biological significance of this difference needs additional investigation. However, in the present paper we will focus on the evidence that gene flow between races is strongly overestimated in both studies. The first indication of this stems from the comparison between estimates based on several classes of genetic markers. In the Les Houches hybrid zone, the number of migrants inferred from microsatellites are at least ten times higher than that estimated from karyotypes, allozymes (Brünner and Hausser 1996) and mtDNA (Balloux *et al.* in press). In Sweden, the absence of any interracial barrier to gene flow is puzzling since the karyotypes of the same individuals form very steep clines, indicating very rare interracial gene exchange. Indeed, the frequencies of the individual race-specific chromosomes vary from nearly complete fixation to virtually zero over less than one km around the centre of the hybrid zone (Narain and Fredga 1996). Of course, it is known that neutral alleles will introgress more than counterselected ones. The

differences between microsatellites and other markers are however far too strong to be explained by differential selection alone. This puzzling discrepancy urgently asks for an additional explanation. In the following sections we will explore reasons why interracial migration rates inferred from microsatellites are overestimates. Actually, the key to this problem lies in the violation of some simplifying assumption central to population genetics.

Gene flow estimates

Underlying theory

The genetic structuring of populations is traditionally estimated by means of Wright's F -statistics (Wright 1951). The F -statistic most relevant to the study of gene flow is F_{ST} , the differentiation between populations. There are different ways to estimate this quantity. In the present paper we will limit ourselves to the classical approach of hierarchical heterozygosities. F_{ST} can be defined as:

$$F_{ST} = \frac{H_t - \bar{H}_s}{H_t} \quad (1)$$

where H_t and H_s are respectively the expected proportion of heterozygous individuals in the total population and within populations. Expected heterozygosities (H) can be obtained for any hierarchical level (eg population, geographical area, race) and are computed as:

$$H_x = 1 - \sum p_{ix}^2 \quad (2)$$

where p_{ix} stands for the frequency of the i th allele at the hierarchical level x .

The popularity of F_{ST} stems from its simple relationship with the number of migrants a population receives per generation. However, these estimates rely on a series of simplifying assumptions incorporated in the so-called island model (Wright 1931). In this model there are an infinite number of equally sized populations. Geography is completely neglected since all population exchange at random the same proportion of gametes with all other populations. Furthermore, the island model assumes absence of mutation and selection. As all real populations violate at least some of these assumptions, the value of migration estimates based on fixation indices has been questioned (Whitlock and McCauley 1999). Despite this lack of realism, estimates of effective migrants inferred from F_{ST} are very popular. Indeed, they allow formulating genetic differentiation into an intuitive equivalence. Following Wright (1943), the relation between F_{ST} at equilibrium and migration under the island model is defined as:

$$\hat{F}_{ST} = \frac{(1 - m)^2}{2N - (2N - 1)(1 - m)^2} \quad (3)$$

where m stands for the proportion of migrants and N for the effective population size. If terms involving m , m^2 and Nm^2 are neglected, equation 3 reduces to Wright's (1951) famous approximation in population genetics:

$$\hat{F}_{ST} \approx \frac{1}{4Nm + 1} \quad (4)$$

As mentioned above, the island model assumes an infinite number of populations. If this number is fairly large, the latter relation is only slightly biased. However, when estimating interracial gene flow across hybrid zones, the number of populations equals two and equation 4 must be corrected following Slatkin and Voelm (1991):

$$\hat{F}_{ST} = \frac{1}{[4Nmd / (d - 1)] + 1} \quad (5)$$

where d stands for the number of populations.

Effect of high mutation rate when migration is reduced

If we take into account mutation, F -statistics give estimates not of Nm but of $N(m + \mu)$, the product of local population size times the sum of migration and mutation. For most population genetics studies, neglecting mutation is reasonable, since migration is generally several orders of magnitude higher than mutation. However, in some situations as hybrid zones, where migration can be very low, the mutation term μ for highly polymorphic loci can be higher than migration. In this case F_{ST} is more influenced by the effective number of mutants $N\mu$, rather than migrants Nm . This problem of marker variability on migration estimates has already been noticed by Wright (1978: 83). However, it is only with the new generation of molecular markers with extremely high mutation rates such as microsatellites, that this bias is so important (Charlesworth 1998, Nagylaki 1998, Hedrick 1999). To give an idea of the importance of this bias on F_{ST} in hybrid zones, we can create a hypothetical example. Let us imagine two populations, which exchange no migrants at all. If these populations have been separated for enough time, we can imagine that they are characterised by strictly non-overlapping allele distributions. When estimating F_{ST} with a biallelic locus, we will get a value of 1, which correctly reflects the complete absence of gene flow. Now imagine that we estimate F_{ST} with a locus which has ten alleles at the same frequencies in each population, none of them being shared between the two populations. In this case, equation 2 leads to $H_t = 0.95$ and $H_s = 0.90$, which will translate into a F_{ST} estimate of about 0.05, although no allele is present in both populations. This hypothetical example exemplifies the bias on F_{ST} inferred from very variable loci, when migration is low.

Homoplasy

In the previous example, we stated that, since the two populations did not exchange any migrants, they had no allele in common. Indeed, the mutation model underlying F -statistics, which is called the infinite allele model (IAM), assumes that every mutation event creates a novel allele. The high mutation rate of microsatellites combined with a finite number of possible allelic states will change identity by descent to identity by state (Rousset 1996). Indistinguishable alleles will arise from different mutational events, a phenomenon called homoplasy. For microsatellites, homoplasy can be very high (Ortí *et al.* 1997), and this will further lower F_{ST} estimates.

Alternative approaches

Under mutation models generating homoplasy, such as the single step mutation model (SSM) which has been proposed for microsatellites (eg Di Rienzo *et al.* 1998, Weber and Wong 1993), the relation between F_{ST} and $N(m+\mu)$ no longer holds (Rousset 1996). An alternative approach called R -statistics, has been specifically defined for loci undergoing stepwise mutations (Slatkin 1995). Under the latter mutation model, R -statistics estimates are independent of the mutation rate. There are however two problems with these estimates. First, they are characterised by a large variance (Slatkin 1995, Lugon-Moulin *et al.* 2000), and reliable estimates would therefore require a very large number of loci. A second more serious limitation stems from their sensitivity to random deviations from the strict stepwise mutation model. Balloux *et al.* (in press) showed by simulations that even if only 5% of mutations are random rather than stepwise, this will strongly bias R_{ST} estimates downwards.

Empirical work on microsatellite mutations has shown a complex mutation scheme, with frequent additions or deletions of several repeat units (Fitzsimmons 1998, Primmer *et al.* 1998, Schlötterer *et al.* 1998). Furthermore, several population genetics analyses have reached the conclusion that microsatellite loci do not fit a stepwise process (eg Estoup *et al.* 1995, Ross *et al.* 1997). This result is also confirmed by Wyttenbach *et al.* (1999b), who tested whether the variation shown by six microsatellites of the common shrew was better explained by a stepwise or an infinite allele model, and found that neither model could explain their allele distributions.

Another approach is provided by absolute genetic distances such as Nei's (1972). These are not subject to biases due to high mutation rates. However, as there is no link to population parameters such as population size or migration, they do not permit the estimation of gene flow. A completely different approach based on assignment tests developed to determine the population membership of specific individuals (review in Waser and Strobeck 1998) could also reveal valuable information in the context of hybrid zones. These tests are not biased by the polymorphism of genetic markers. However, they require a rather large number of loci.

Conclusions

Fixation indices give estimates of both mutation and migration rate. It is therefore important to keep in mind the possible effects of marker polymorphism when interpreting F_{ST} values. In hybrid zones, genetic structuring estimates based solely on microsatellites can lead to strong overestimates of gene flow. This problem should not conceal the enormous potential of microsatellite markers for the study of hybrid zones in the common shrew. Indeed, for the first time, a reasonable number of polymorphic genetic markers are available for *S. araneus*. The high statistical power associated to very polymorphic loci enabled for instance Lugon-Moulin *et al.* (1999a) to test for several potential causes responsible for the structuring within the Valais side of the Les Houches hybrid zone. With use of partial Mantel tests, these authors could show that introgression from the Cordon populations was the main factor. There is also a great potential for comparative work among different hybrid zones. Comparing absolute genetic distances from several hybrid zones in the light of the karyotypic background of the hybridising races could give insights into the role of chromosomes as barriers to gene flow.

As we plan to develop at least 50 microsatellite markers and to map them on the chromosomes, even more ambitious projects will be possible in the future. For instance, with a panel of many mapped markers, it would be possible to estimate, with means of absolute distances, the number of chromosomal segments involved in reproductive isolation by estimating interracial differentiation on different regions of the karyotype across hybrid zones (Rieseberg *et al.* 1999). The localization of these segments will indicate the relative role of chromosomal rearrangements as barriers to gene flow. Indeed, if these segments were preferentially located on the differentiated chromosomes, this would indicate that Robertsonian rearrangements play a major role. On the other hand, random distribution of the segments would indicate that chromosomes have a minor effect.

Acknowledgements: We would like to thank M. Reuter for critical suggestions, A. Kohler and J. Kohler for moral support, and the University of Lausanne for financial support.

References

- Amos B., Schlötterer D. and Tautz D. 1993. Social structure of pilot whales revealed by analytical DNA profiling. *Science* 260: 670–672.
- Balloux F., Brünnner H., Lugon-Moulin N., Hausser J. and Goudet J. 2000. Microsatellites can be misleading: an empirical and simulation study. *Evolution* (in press)
- Balloux F., Goudet J. and Perrin N. 1998. Breeding system and genetic variance in the monogamous, semi-social shrew, *Crocidura russula*. *Evolution* 52: 1230–1235.
- Barton N. H. and Hewitt G. M. 1989. Adaptation, speciation and hybrid zones. *Nature* 341: 497–502.
- Brünnner H. and 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.
- Charlesworth B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* 15: 538–543.

- Coyne J. A. and Orr H. A., 1998. The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society London B* 353: 287–305.
- Di Rienzo A., Donnely P., Toomajian A., Sisk B., Hill A., Petzl-Erler P. L., Haines G. K. and Barch D. H. 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics* 148: 1269–1284.
- Estoup A., Garnery L., Solignac M. and Cornuet J. M. 1995. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: Hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* 140: 679–695.
- Fitzsimmons N. N. 1998. Single paternity of clutches and sperm storage in the promiscuous green turtle (*Chelonia mydas*). *Molecular Ecology* 7: 575–584.
- Garza J. C., Dallas J., Duryadi D., Gerasimov S., Croset H. and Boursot P. 1997. Social structure of the mound-building mouse *Mus spicilegus* revealed by genetic analysis with microsatellites. *Molecular Ecology* 6: 1009–1017.
- Harrison R. G. 1990. Hybrid zones: windows on evolutionary process. [In: *Oxford Surveys in Evolutionary Biology*. D. Futuyama and J. Antonovics, eds]. Oxford University Press, London: 69–128.
- Hedrick P. W. 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53: 313–318.
- Hewitt G. M. 1988 Hybrid zones-Natural laboratories for evolutionary studies. *Trends in Ecology and Evolution* 3: 158–167.
- Jarne P. and Lagoda P. J. L. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* 11: 424–429.
- Lugon-Moulin N., Balloux F. and Hausser J. 2000. Genetic differentiation of common shrew *Sorex araneus* populations among different alpine valleys revealed by microsatellites. [In: *Evolution in the Sorex araneus group*. Cytogenetic and molecular aspects. J. B. Searle and J. M. Wójcik, eds]. *Acta Theriologica* 45, Suppl. 1: 103–117.
- Lugon-Moulin N., Brünner H., Balloux F., Hausser J. and Goudet J. 1999a. Do riverine barriers, history or introgression shape the genetic structuring of a common shrew (*Sorex araneus*) population? *Heredity* 83: 155–161.
- Lugon-Moulin N., Brünner H., Wyttenbach A., Hausser J. and Goudet J. 1999b. Hierarchical analyses of genetic differentiation in a hybrid zone of *Sorex araneus* (Insectivora, Soricidae). *Molecular Ecology* 8: 419–431.
- Lugon-Moulin N., Wyttenbach A., Brünner H., Goudet J. and Hausser J. 1996. Study of gene flow through a hybrid zone in the common shrew (*Sorex araneus*) using microsatellites. *Hereditas* 125: 159–168.
- Nagylaki T. 1998. Fixation indices in subdivided populations. *Genetics* 148: 1325–1332.
- Narain Y. and Fredga K. 1996. A hybrid zone between the Hällefors and Uppsala chromosome races of *Sorex araneus* in central Sweden. *Hereditas* 125: 137–145.
- Nei M. 1972. Genetic distances between populations. *The American Naturalist* 105: 385–398.
- Ortí G., Pearse D. E. and Avise J. C. 1997. Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Sciences of the USA* 94: 10745–10749.
- Primmer C., Saino R. N., Møller A. P. and Ellegren H. 1998. Unravelling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. *Molecular Biology and Evolution* 15: 1047–1058.
- Rieseberg L. H., Whitton J. and Gardner K. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152: 713–727.
- Ross K. G., Krieger M. J. B., Shoemaker D. D. and 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.
- Ruedi M. 1998. Protein evolution in shrews. [In: *Evolution of shrews*. J. M. Wójcik and M. Wolsan, eds]. *Mammal Research Institute, Polish Academy of Science, Białowieża*: 269–294.

- Schlötterer C., Ritter R., Harr B. and Brem G. 1998. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution* 15: 1269–1274.
- Searle J. and Wójcik J. M. 1998. Chromosomal evolution: The case of *Sorex araneus*. [In: Evolution of shrews. J. M. Wójcik and M. Wolsan, eds]. Mammal Research Institute Polish Academy of Science, Białowieża: 219–262.
- Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457–462.
- Slatkin M. and Voelm L. 1991. F_{st} in a hierarchical island model. *Genetics* 127: 627–629.
- Waser P. M. and Strobeck C. 1998. Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution* 13: 43–44.
- Weber J. L. and Wong C. 1993. Mutation of human short tandem repeats. *Human Molecular Genetics* 2: 1123–1128.
- Whitlock M. C. and McCauley D. E. 1999. Indirect measures of gene flow and migration: $F_{st} \approx 1/(4Nm+1)$. *Heredity* 82: 117–125.
- Wright S. 1931. Evolution in mendelian populations. *Genetics* 16: 97–159.
- Wright S. 1943. Isolation by distance. *Genetics* 28: 114–138.
- Wright S. 1951. The genetical structure of populations. *Annals of Eugenics* 15: 323–354.
- Wright S. 1978. Evolution and the genetics of populations. Vol. 4. Variability within and among natural populations. The University of Chicago Press, Chicago and London: 1–580.
- Wytenbach A., Goudet J., Cornuet J.-M. and Hausser J. 1999a. Microsatellite variation reveals low genetic subdivision in a chromosome race of *Sorex araneus* (Mammalia, Insectivora). *Journal of Heredity* 90: 323–327.
- Wytenbach A., Narain Y. and Fredga K. 1999b. Genetic structuring and gene flow in a hybrid zone between two chromosome races of the common shrew (*Sorex araneus*, Insectivora) revealed by microsatellites. *Heredity* 82: 79–88.
- Zima J., Fedyk S., Fredga K., Hausser J., Mishta A., Searle J. B., Volobouev V. T. and Wójcik J. M. 1996. The list of the chromosome races of the common shrew (*Sorex araneus*). *Hereditas* 125: 97–107.

Received 1 September 1999, accepted 28 January 2000.