Letter to the Editor

Microsatellite Conservation, Polymorphism, and GC Content in Shrews of the Genus Sorex (Insectivora, Mammalia)

François Balloux, Eric Ecoffey,* Luca Fumagalli,† Jérôme Goudet,* Andreas Wyttenbach,‡ and Jacques Hausser**

*Institut de Zoologie et d'Ecologie Animale, University of Lausanne, Switzerland; †Department of Zoology, University of Queensland, Brisbane, Australia; and ‡Department of Genetics, Uppsala University, Uppsala, Sweden

Several studies have shown that microsatellites developed from various species can amplify a homologous and polymorphic end-product in nonsource species. Conservation of microsatellite loci has been documented in many organisms, such as mammals (e.g., Schlötterer, Amos, and Tautz 1991; Kondo et al. 1993; Deka et al. 1994; Pepin et al. 1995), birds (Primmer, Møller, and Ellegren 1996), insects (Chapuisat 1996), and plants (Lagercrantz, Ellegren, and Anderson 1993; Whitton, Riesenberg, and Ungerer 1997). The conservation of the priming sites is very variable and can sometimes be spectacular. Some loci have been shown to be conserved over time spans of 300 Myr among sea turtles (Fitzsimmons, Moritz, and Moore 1995) and 470 Myr among fish species (Rico, Rico, and Hewitt 1996).

The possibility of amplifying a microsatellite locus in another species depends on the conservation of its flanking regions, which is a function of the phylogenetic distance between taxa. While the distance to the source species should be minimal to perform successful cross-species amplification, it would be useful to have other clues for selecting loci. Recently, Glenn et al. (1996) reported some unexpected positive correlations between the mutation rates of alligator microsatellites and their flanking regions. The more the loci were polymorphic in alligators, the less they amplified in other crocodilian genera. These authors also reported a negative correlation between the GC content of the flanking regions of a locus and its polymorphism in the source species. These results are potentially very important both theoretically and practically. Assuming that the polymorphism of a locus reflects a mutation rate, as does the conservation of priming sites, then the correlation between microsatellite polymorphism and flanking sequence conservation suggests that some regions in the genome experience higher mutation rates than others. Furthermore, the GC content of a particular region in the genome could be used as a relative measure of mutation rate. Practically, these correlations might be valuable in choosing a particular set of microsatellites; for instance, primers for GC-poor clones should be developed preferentially to obtain highly polymorphic markers in the source species. On the other hand, less polymorphic loci should be chosen for cross-species amplification (at the possible cost of low variability in nonsource species).

The potential usefulness of these results prompted us to test whether these correlations hold for other taxa. We chose to work on the soricine shrew *Sorex araneus* (Insectivora). *Sorex araneus* seems to be a good model with which to explore microsatellite conservation at different phylogenetic levels. Indeed, the species is subdivided into numerous chromosomal races (Zima et al. 1996), and the genus *Sorex* is itself divided into different subgenera (George 1988). Microsatellites were developed on the chromosomal race Valais of *S. araneus* (Wyttenbach, Favre, and Hausser 1997; this study). We tested these loci on two other *S. araneus* chromosomal races (Vaud and Cordon), two sister species (*S. granarius* and *S. coronatus*), three species belonging to the same subgenus *Sorex* (*S. samniticus*, *S. raddei*, and *S. alpinus*), one species of the subgenus *Otisorex* (*S. cinereus*), and, finally, a species of another genus in the same subfamily (*Neomys anomalus*).

In addition to amplification success, we checked whether it was possible to predict the level of polymorphism in nonsource species. To this end, nine individuals (three individuals from three different populations about 50 km apart) were sampled in each taxon. The 12 microsatellites used in the present study are given in table 1. Since most loci used by Glenn et al. (1996) were AC repeats, we restricted our choice to this motif. In order to avoid as much as possible a bias toward long and polymorphic loci (Ellegren et al. 1997), we used very relaxed criteria of selection: all loci with more than five perfect repeats and a clear amplification pattern in the source species were employed irrespective of their polymorphism.

PCR amplifications were carried out in 10 μl reaction volume overlayed with mineral oil. For all loci, the following conditions were used: ≈5 ng of genomic DNA, 1 × *Taq* polymerase buffer, 0.7 mM dNTPs, 1.0 mM $MgCl_2$, 0.2 μg/μl BSA, 1 μM primers, 0.15 μCi [^{33}P] ATP, and 0.25 U *Taq* polymerase (Eurobio). Thermal profiles were as follows: an initial denaturation at 95°C for 2 min, followed by 30 cycles consisting of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and polymerization at 72°C for 1 min; the last polymerization step was extended to 7 min. End-products were run on polyacrylamide sequencing gels (6%, 8 M urea), together with the amplified clone as a size marker.

For 11 loci, the results were unambiguous, all individuals of the same taxon either failed to amplify or amplified a scorable end-product similar in size to the clone. Locus 69 did amplify a very faint end-product in *S. alpinus* which appeared to be impossible to read and

Key words: Soricinae, *Sorex*, microsatellites, interspecific priming, polymorphism, GC content.

Address for correspondence and reprints: François Balloux, Institut de Zoologie et d'Ecologie Animale, Université de Lausanne, CH-1015 Lausanne, Switzerland. E-mail: francois.balloux@izea.unil.ch.

Mol. Biol. Evol. 15(4):473–475. 1998

Table 1
Number of Alleles of the 12 Microsatellite Loci Used in this Study

			Number of Alleles									
Name	Core Repeat[a]	GC%	*Sorex araneus* Valais (0)[b]	*Sorex araneus* Vaud (0.0182)[b]	*Sorex araneus* Cordon (0.0196)[b]	*Sorex granarius* (0.0196)[b]	*Sorex coronatus* (0.0413)[b]	*Sorex samniticus* (0.0911)[b]	*Sorex raddei* (0.1222)[b]	*Sorex alpinus* (0.1438)[b]	*Sorex cinereus* (0.1517)[b]	*Neomys anomalus* (0.2066)[b]
L69[c].	$(AC)_{17}$	25.5	14	11	12	11	5	10	10	0	0	0
L45[c].	$(AC)_{10}$	29.6	4	4	4	5	3	7	8	1	0	0
L16[c].	$(AC)_{13}$[e]	31.7	1	2	2	1	1	1	1	0	0	0
L14[d]	$(AC)_{14}$	34.5	11	8	14	9	10	9	5	0	9	0
L67[c].	$(AC)_{17}$	35.7	7	10	8	8	1	6	12	6	1	0
L9[c].	$(AC)_{30}$[e]	37.7	13	11	12	8	3	5	0	0	0	0
L62[c].	$(AC)_{16}$	40.0	13	10	7	9	8	4	7	2	8	0
L57[c].	$(AC)_{11}$[f]	41.0	10	13	9	12	7	0	7	0	0	0
L68[d]	$(AC)_{11}$	42.0	10	6	4	4	3	7	9	0	0	0
L92[d]	$(AC)_{6}$	43.8	6	8	6	5	3	1	3	1	1	1
L97[d]	$(AC)_{46}$	44.8	9	9	6	3	1	1	1	4	6	0
L33[d]	$(AC)_{19}$	46.5	12	9	9	8	4	5	5	1	0	1

[a] Sequenced allele.

[b] Kimura two-parameter distance to the source taxon *S. araneus* Valais.

[c] Primers published in Wyttenbach, Favre, and Hausser (1997). GenBank accession numbers: U82711–U82717.

[d] GenBank accession numbers: AF032911–AF032912. Primers may be downloaded from the following Internet server: http://www.unil.ch/izea/pages/recherche/shrewgenpop.html.

[e] Microsatellite repeat motif contains some imperfections.

[f] Longest perfect repeat of a more complex microsatellite: $(ACC)_6(AAC)_3A(AC)_{11}$.

was therefore considered nonamplifying. The same locus did not amplify two *S. coronatus* individuals despite three successive attempts, suggesting the presence of a null allele at high frequency. Table 1 summarizes the number of alleles found for each locus in the different samples.

We first tested the prediction of a negative correlation between GC content and polymorphism in the source species by means of a Spearman rank correlation. GC content was measured on the entire sequence of the cloned insert after exclusion of the repeated motif of the microsatellite. GC content was unrelated to polymorphism (rho = 0.028, $P = 0.935$). We applied the same test to the 10 loci developed by Favre and Balloux (1997) on another shrew species (*Crocidura russula*) and did not obtain a significant correlation for those loci either. The success of cross-species amplification of a given locus and its polymorphism was tested against the phylogenetic distance between taxa, the polymorphism of the locus in the source species, and the GC content of its flanking regions. Phylogenetic distances to the source taxon (*S. araneus* Valais) were computed as Kimura two-parameter distances (Kimura 1980) based on 731 bp of the mitochondrial DNA cytochrome *b* gene sequenced on homologous taxa (Taberlet, Fumagalli, and Hausser 1994; unpublished data). The polymorphism of the loci was quantified by the number of alleles in the source species.

A permutation-based multiple-regression analysis (Manly 1991, pp. 108–111) using phylogenetic distance, GC content, and polymorphism in the source taxon as descriptors was performed to predict the success of amplification. The model explained 44% of the variance and gave phylogenetic distance as the only significant negative effect ($P < 0.0001$; 10,000 permutations). Neither GC content nor polymorphism showed any correlation with success of amplification. These results hold independently of the order in which the factors were entered in the model. A second multiple regression was carried out using the same independent variables to predict the polymorphism in cross-species amplification. The polymorphism was computed as the number of alleles of a locus in a given species. An unsuccessful amplification was coded as zero alleles, and a monomorphic end-product was coded as one allele. The model explained 53% of the variance. Phylogenetic distance to the source taxon was a negative significant variable ($P < 0.0001$; 10,000 permutations), whereas the number of alleles in the source taxon was a positive significant variable ($P < 0.0002$; 10,000 permutations). Loci with more alleles in the source species had a tendency to be more polymorphic in cross-species amplification. Again, GC content did not affect polymorphism, irrespective of the order in which the variables were entered.

This result has to be taken with some caution, since we included into this analysis both loci that did not amplify and those that were monomorphic in some taxa. Absence of amplification is informative about variation in the flanking regions and not about the polymorphism of the repeated sequence itself, whereas monomorphism, if it confirms the conservation of the priming sites, can be obtained by different means: fixation of an allele or deletion or important alteration of the repeated sequence (Angers and Bernatchez 1997). We therefore also conducted an analysis excluding the ''species × locus'' observations with zero or one allele detected. In this instance, we obtained a model explaining 30% of the variance with the number of alleles in the source taxon as main effect ($P < 0.0001$) and phylogenetic distance ($P < 0.05$) as side-effect.

These results indicate that the correlations obtained by Glenn et al. (1996) cannot be generalized to all taxa.

At the intraspecific level, we did not observe a negative correlation between GC content and polymorphism. At the interspecific level, we did not confirm the negative correlation between polymorphism in the source species and the success of cross-species amplification. The only factor in our study which is correlated to the success of amplification is phylogenetic distance. This suggests that the mutation rate in the flanking regions of microsatellites is a function of the time of divergence between taxa irrespective of GC content of the flanking regions or polymorphism in the source species. However, this latter variable turned out to be a good predictor of the number of alleles in cross-species amplification. This suggests that some microsatellite loci could have a tendency to be more polymorphic than others throughout the range of species in which they amplify. As we detected no negative correlation between polymorphism and success of amplification but a positive one between polymorphism in source and nonsource taxa, we would recommend choosing polymorphic loci from the closest species for cross-species amplification.

Acknowledgments

We are indebted to P. Vogel for providing samples. We also wish to thank N. Lugon-Moulin, L. Keller, and an anonymous referee for valuable comments.

LITERATURE CITED

ANGERS, B., and L. BERNATCHEZ. 1997. Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. Mol. Biol. Evol. **14**:230–238.

CHAPUISAT, M. 1996. Characterization of microsatellite loci in Formica-Lugubris-B and their variability in other ant species. Mol. Ecol. **5**:599–601.

DEKA, R., M. D. SHRIVER, L. M. YU, L. JIN, C. E. ASTON, R. CHAKRABORTY, and R. E. FERRELL. 1994. Conservation of human-chromosome-13 polymorphic microsatellite (Ca)(N) repeats in chimpanzees. Genomics **22**:226–230.

ELLEGREN, H., S. MOORE, N. ROBINSON, K. BYRNE, W. WARD, and B. C. SHELDON. 1997. Microsatellite evolution—a reciprocal study of repeat lengths at homologous loci in cattle and sheep. Mol. Biol. Evol. **14**:854–860.

FAVRE, L., and F. BALLOUX. 1997. Characterization of microsatellite loci in the greater white-toothed shrew: *Crocidura russula*. Mol. Ecol. **6**:595–596.

FITZSIMMONS, N. N., C. MORITZ, and S. S. MOORE. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Mol. Biol. Evol. **12**: 432–440.

GEORGE, S. B. 1988. Systematics, historical biogeography, and evolution of the genus *Sorex*. J. Mammal. **69**:443–461.

GLENN, T. C., W. STEPHAN, H. C. DESSAUER, and M. J. BRAUN. 1996. Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. Mol. Biol. Evol. **13**:1151–1154.

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

KONDO, Y., M. MORI, T. KURAMOTO, J. YAMADA, J. S. BECKMANN, D. SIMON-CHAZOTTES, X. MONTAGUTELLI, J. L. GUENET, and T. SERIKAWA. 1993. DNA segments mapped by reciprocal use of microsatellite primers between mouse and rat. Mamm. Genome **4**:571–576.

LAGERCRANTZ, U., H. ELLEGREN, and L. ANDERSON. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. **21**: 1111–1115.

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

PEPIN, L., Y. AMIGUES, A. LEPINGLE, J. L. BERTHIER, A. BENSAID, and D. VAIMAN. 1995. Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. Heredity **74**:53–61.

PRIMMER, C. R., A. P. MØLLER, and H. ELLEGREN. 1996. A wide-range survey of cross-species microsatellite amplification in birds. Mol. Ecol. **5**:365–378.

RICO, C., I. RICO, and G. HEWITT. 1996. 470 million years of conservation of microsatellite loci among fish species. Proc. R. Soc. Lond. B Biol. Sci. **263**:549–557.

SCHLÖTTERER, C., B. AMOS, and D. TAUTZ. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature **354**:63–65.

TABERLET, P., L. FUMAGALLI, and J. HAUSSER. 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.

WHITTON, J., L. H. RIESENBERG, and M. C. UNGERER. 1997. Microsatellites loci are not conserved across the Asteraceae. Mol. Biol. Evol. **14**:204–209.

WYTTENBACH, A., L. FAVRE, and J. HAUSSER. 1997. Characterization of simple sequence repeats in the genome of the common shrew. Mol. Ecol. **6**:797–800.

ZIMA, J., S. FEDYK, K. FREDGA, J. HAUSSER, A. MISHTA, J. B. SEARLE, V. T. VOLOBOUEV, and J. M. WÓJCIK. 1996. The list of the chromosome races of the common shrew. Hereditas **125**:97–107.

MASAMI HASEGAWA, reviewing editor

Accepted December 12, 1997