Microsatellite Evolution—A Reciprocal Study of Repeat Lengths at Homologous Loci in Cattle and Sheep

Hans Ellegren,* Steve Moore,† Nick Robinson,‡ Keren Byrne,† Wayne Ward,‡ and Ben C. Sheldon§

*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala; †CSIRO, Tropical Agriculture, Molecular Animal Genetic Centre, University of Queensland, Australia; ‡Victorian Institute of Animal Science, Victoria, Australia; and §Institute of Cell, Animal and Population Biology, University of Edinburgh, United Kingdom

The application of microsatellites in evolutionary studies requires an understanding of the patterns governing their evolution in different species. The finding that homologous microsatellite loci are longer, i.e., containing more repeat units, in human than in other primates has been taken as evidence for directional microsatellite evolution and for a difference in the rate of evolution between species. However, it has been argued that this finding is an inevitable consequence of biased selection of longer-than-average microsatellites in human, because cloning procedures are adopted to generate polymorphic and, hence, long markers. As a test of this hypothesis, we conducted a reciprocal comparison of the lengths of microsatellite loci in cattle and sheep using markers derived from the bovine genome as well as the ovine genome. In both cases, amplification products were longer in the focal species, and loci were also more polymorphic in the species from which they were originally cloned. The crossing pattern that we found suggests that interspecific length differences detected at homologous microsatellite loci are the result of biased selection of loci associated with cloning procedures. Hence, comparisons of microsatellite evolution between species are flawed unless they are based on reciprocal analyses or on genuinely random selection of loci with respect to repeat length.

Introduction

Microsatellites or short tandem repeats (STRs) are well-known DNA markers used in a variety of applications that require highly polymorphic and locus-specific genetic systems, e.g., in linkage analysis, paternity testing, and population and evolutionary genetics. They consist of iterated simple sequences, usually occurring as monotonous tandem repeats of 1–5 bp. The number of iterations may vary between loci but most polymorphic markers display alleles in the range of 10–30 repeat units. Rarely, some loci may harbor alleles with a substantial number of repeat units. Allelic variation seems to be almost exclusively due to varying numbers of tandem repeats, a pattern also found in other types of repetitive sequences, like minisatellite DNA.

Although some authors have sought to explain the ubiquitous occurrence of microsatellites in terms of a functional significance (Haniford and Pulleyblank 1983; Hamada et al. 1984; Stallings et al. 1991), most recent models of simple repeat evolution, in the absence of close linkage to a selected locus ("genetic hitchhiking"; Slatkin 1995), assume selective neutrality (Tachida and Iizuka 1992; Shriver et al. 1993; Valdes, Slatkin, and Freimer 1993; Di Rienzo et al. 1994). Several lines of evidence indicate that the formation of long tandemly repetitive and polymorphic tracts is due to replication slippage, i.e., strand displacement followed by an out-of-frame pairing of repeated sequences either resulting in a net gain or loss of one or more repeat units (Levinson and Gutman 1987a, 1987b; Slottner and Tautz 1992). There has been considerable interest in understanding microsatellite evolution, particularly the patterns of underlying mutational events. In accordance with a slippage mechanism of mutational change, current ideas of microsatellite evolution favor a stepwise model, possibly extending to a two-phase model which allows mutations by more than one repeat unit (Di Rienzo et al. 1994). Data on direct observations of microsatellite mutations in pedigrees are now accumulating, and a majority of mutational events apparently involve increase or decrease by a single repeat unit, although rare cases of changes by two or more repeat units have been documented (Weber and Wong 1993; Amos et al. 1996; Primmer et al. 1996).

Recently, Rubinsztein et al. (1995) have suggested that microsatellite evolution may be directional and proceed with different paces in different lineages, in this case illustrated by the proposition that human microsatellites are evolving faster than homologous loci in other primates. This conclusion was reached from the observation that highly polymorphic microsatellites isolated from the human genome amplified significantly shorter products in chimpanzees (Pan troglodytes) and other primates, a pattern suggesting that the primate homologs generally harbor alleles with a smaller number of repeat units than humans. The conclusion is a matter of controversy (Ellegren, Primmer, and Sheldon 1995; Amos and Rubinsztein 1996; Jarne and Lagoda 1996). An alternative explanation is that a simple selection bias, with a nonrandom choice of loci with particularly long repeats in the focal species (i.e., the species from which genome markers have originally been cloned and isolated), will inevitably give shorter repeats at homologous loci in related species. We, and others, proposed that comparison of the rate and direction of microsatellite evolution in two related species requires reciprocal analysis of markers developed from both species (Elle-
glen, Primmer, and Sheldon 1995; Forbes et al. 1995). In this paper, we present the first study along these lines, involving a reciprocal comparison of microsatellite allele lengths at homologous loci in cattle (Bos taurus) and sheep (Ovis aries), a pair of species that diverged some 16–28 MYA (Allard et al. 1992). These new data support our previous suggestion that a selection bias will generally result in inappropriate conclusions concerning relative rates and directions of microsatellite evolution if markers are not applied reciprocally.

Materials and Methods

Markers and Animals

Allele frequencies at 13 microsatellite loci isolated from the bovine genome and 14 from the ovine genome were analyzed. Despite being developed in different laboratories, all markers had originally been isolated to obtain polymorphic genetic markers. Hence, library screenings emphasized the selection of long stretches of repetitive arrays (as probably is the case for virtually all microsatellites isolated from library screenings). All markers consisted of (CA)n repeats. For nine bovine markers (RM012, 024, 088, 103, 113, 179, E11, and E23; markers originally cloned from cow DNA), data on allele frequencies were extracted from the literature as were data on allele frequencies for some bovine markers in sheep (RM012, 024, 041, 088, 103, 113, 153, 179, and E11). Similarly, all allele frequencies for ovine markers (originally cloned in sheep) were from the literature. For the remaining marker/species combinations, genotype data were gathered by us for the purpose of this study. Detailed marker information including primer sequences can be found in the references listed in table 1.

All animals included in the study were unrelated individuals. The particular animals used for genotyping differed between markers. Breed affiliations among animals used for different markers are indicated in table 1.

Genotyping

Microsatellites were analyzed by radioactive or fluorescent detection. For the former, one primer in each pair was end-labeled with γ[3P]-dATP (Amersham) using T4 polynucleotide kinase (Promega). PCR was carried out in 10-μl volumes containing 100 ng genomic DNA, 25 nM of both forward and reverse primer, 1 unit Taq polymerase (Toyobo), 200 μM of each dNTP, 100 mM Tris-Cl (pH 9.0), 50 mM KCl, 1.5 mM MgCl₂, 0.1% Triton X-100, and 0.01% gelatin. Amplification was achieved on an Omnigene thermal cycler (Hybaid) with two 1-min steps per cycle, extension occurring during the transition between the annealing and denaturation (+94°C) steps. Annealing of primer to template DNA was done at temperatures indicated in original marker descriptions. PCR products were electrophoresed through 6% denaturing polyacrylamide sequencing gels and visualized by autoradiography. Microsatellite alleles were sized and scored with reference to lanes of M13/pUC19 sequence and DNA standards.

For fluorescent typing, one primer in each pair was synthesized with either 6FAM, TET, or HEX dyes (Perkin-Elmer/ABI) and was used in amplifications as above. Genotypes were determined using an ABI 373A model DNA sequencer with the aid of an internal lane standard (GS-TAMRA350; Perkin-Elmer/ABI). Scoring was accomplished using Genescan and Genotyper software (Perkin-Elmer/ABI).

Experimental Design

We randomly chose previously isolated bovine and ovine microsatellite markers available in our laboratories and surveyed the primer pairs for their ability to amplify a specific PCR product in the other species. It was imperative for the purpose of the study that we had no prior information on cross-species performance for any of these primer pairs. For markers that could be used across species, unrelated individuals of both species were genotyped to estimate the respective allele fre-

Table 1

<table>
<thead>
<tr>
<th>MARKER</th>
<th>No. of CHROMOSOMES</th>
<th>Cattle</th>
<th>Sheep</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSSM31</td>
<td>6</td>
<td>22a</td>
<td>22a</td>
<td>Moore et al. (1994)</td>
</tr>
<tr>
<td>RM011</td>
<td>53a</td>
<td>40b</td>
<td>40b</td>
<td>Kossarek et al. (1994)</td>
</tr>
<tr>
<td>RM012</td>
<td>46a</td>
<td>22d</td>
<td>22d</td>
<td>Kossarek et al. (1994)</td>
</tr>
<tr>
<td>RM024</td>
<td>42a</td>
<td>22d</td>
<td>22d</td>
<td>Kossarek et al. (1994)</td>
</tr>
<tr>
<td>RM041</td>
<td>50a</td>
<td>24a</td>
<td>24a</td>
<td>Kossarek et al. (1995a)</td>
</tr>
<tr>
<td>RM044</td>
<td>50a</td>
<td>40d</td>
<td>40d</td>
<td>Kossarek et al. (1995b)</td>
</tr>
<tr>
<td>RM088</td>
<td>46a</td>
<td>24d</td>
<td>24d</td>
<td>Kossarek et al. (1995a)</td>
</tr>
<tr>
<td>RM103</td>
<td>46a</td>
<td>24d</td>
<td>24d</td>
<td>Kossarek et al. (1995a)</td>
</tr>
<tr>
<td>RM113</td>
<td>46a</td>
<td>24d</td>
<td>24d</td>
<td>Kossarek et al. (1995a)</td>
</tr>
<tr>
<td>RM153</td>
<td>46a</td>
<td>24d</td>
<td>24d</td>
<td>Kossarek et al. (1995b)</td>
</tr>
<tr>
<td>RM179</td>
<td>46a</td>
<td>24d</td>
<td>24d</td>
<td>Kossarek et al. (1995b)</td>
</tr>
<tr>
<td>RME11</td>
<td>46a</td>
<td>22d</td>
<td>22d</td>
<td>Grosse et al. (1995)</td>
</tr>
<tr>
<td>RME23</td>
<td>46a</td>
<td>38d</td>
<td>38d</td>
<td>Grosse et al. (1995)</td>
</tr>
<tr>
<td>C.SRD24</td>
<td>10c</td>
<td>100b</td>
<td>99b</td>
<td>Davies et al. (1994)</td>
</tr>
<tr>
<td>C.SRD2105</td>
<td>40c</td>
<td>92c</td>
<td>92c</td>
<td>Davies et al. (1996)</td>
</tr>
<tr>
<td>C.SRD2171</td>
<td>40c</td>
<td>94d</td>
<td>94d</td>
<td>Davies et al. (1996)</td>
</tr>
<tr>
<td>Mc070</td>
<td>18d</td>
<td>162f</td>
<td>162f</td>
<td>Buchanan and Crawford (1992b)</td>
</tr>
<tr>
<td>Mc099</td>
<td>12e</td>
<td>812f</td>
<td>812f</td>
<td>Buchanan and Crawford (1992a)</td>
</tr>
<tr>
<td>Mc058</td>
<td>40f</td>
<td>94b</td>
<td>94b</td>
<td>Hulme et al. (1994)</td>
</tr>
<tr>
<td>Mc064</td>
<td>35h</td>
<td>90d</td>
<td>90d</td>
<td>Hulme et al. (1994)</td>
</tr>
<tr>
<td>Mc0136</td>
<td>8f</td>
<td>102d</td>
<td>102d</td>
<td>Hulme et al. (1995)</td>
</tr>
<tr>
<td>Mc0147</td>
<td>40h</td>
<td>80h</td>
<td>80h</td>
<td>Smith et al. (1995)</td>
</tr>
<tr>
<td>Mc0373</td>
<td>40h</td>
<td>98h</td>
<td>98h</td>
<td>Hulme et al. (1995)</td>
</tr>
<tr>
<td>OARCP34</td>
<td>20h</td>
<td>48f</td>
<td>48f</td>
<td>Ede, Pierson, and Crawford (1995)</td>
</tr>
<tr>
<td>OARFC85</td>
<td>14d</td>
<td>24f</td>
<td>24f</td>
<td>Buchanan, Galloway, and Crawford (1994)</td>
</tr>
<tr>
<td>OARFBC128</td>
<td>8e</td>
<td>70f</td>
<td>70f</td>
<td>Buchanan and Crawford (1993)</td>
</tr>
<tr>
<td>OARJMP8</td>
<td>18e</td>
<td>64b</td>
<td>64b</td>
<td>Penty et al. (personal communication)</td>
</tr>
</tbody>
</table>

* Africander, Brahman, and Hereford.
* b Merino, Border Leister, Suffolk, Romney, Karakul, Finnish Landrace, Poll Dorset, and Carpet Master.
* In Bos taurus X Bos indicus cross. Precise breeds making up this cross were not known.
* d Merino.
* e Parents of the international reference families for bovine genome mapping.
* f Bos taurus × Bos indicus, N'Dama, Boran, and Drangus.
* g Merino and Border Leicester.
* h Murray-Grey and Holstein-Fresian.
* i Merino and Romney Marsh.
quences. We also searched the literature for bovine or ovine microsatellites that had been found to amplify in the other species and for which allele frequencies had been reported. The search was restricted to original descriptions of new markers in the focal species and hence was not biased toward markers for which data had been reported solely because the marker had been found to be polymorphic in the other species. By these two approaches, allele frequency data for 13 bovine and 14 ovine (CA)₆ microsatellites were obtained (table 1).

In the following analysis, we use length of PCR products (bp) as an indication of repeat length harbored within amplified fragments. It is clear that interspecific size variation at individual loci might not be due solely to varying lengths of the repetitive tract, but could also be a consequence of indels in regions flanking the repeat. However, since there is no reason why size variation caused by such processes should be biased in either direction in a sample of loci or systematically between species (cf. Rubinsztein et al. 1995), we feel confident that this approach will accurately reflect general trends of microsatellite length variation between the two species.

Results

Comparisons of Repeat Lengths and Variability in Reciprocal Amplification

Repeat lengths were clearly longer in the focal species than in the other species (table 2 and figs. 1 and 2). Eleven of 13 bovine markers revealed longer amplification products in cattle than in sheep (z = 2.27, P = 0.023; Wilcoxon’s signed-ranks test, two-tailed, on mean allele length at each locus). Conversely, 12 of 14 ovine markers revealed longer products in sheep than in cattle (z = 2.98, P = 0.003). Size differences between the two species varied considerably between loci, with a mean of 13.3 bp (±6.2 SE) longer alleles in cattle than in sheep for bovine microsatellites and a mean of 10.6 bp (±9.2 SE) longer alleles in cattle than in sheep for ovine microsatellites, averages corresponding to differences of 5–7 repeat units. There was no indication that the magnitude of difference in repeat length between the focal and congeneric species differed between bovine and ovine markers (z = 0.87, P = 0.38; Mann-Whitney U-test). This could have been expected if the rate of microsatellite evolution differed significantly between the two species and if evolution was directional (for instance, if bovine microsatellites tended to become longer at a higher rate than homologous loci in sheep, we would have expected the average size difference (focal species – congeneric species) for bovine markers to be greater than that for ovine markers).

There was a pronounced difference in the degree of variability displayed in homologous and heterologous amplification, respectively (table 2 and figs. 1 and 2). While all markers were polymorphic in the focal species, six bovine microsatellites were monomorphic in sheep and seven ovine microsatellites were monomorphic in cattle. Accordingly, heterozygosities were significantly higher in the focal species (bovine markers: z = 2.29, P = 0.022; ovine markers: z = 2.86, P = 0.005). We found no evidence for a correlation between the degree of variability in the focal species or in the other species and the size difference between the two species (r = 0.03, P = 0.88 and r = 0.11, P = 0.60, respectively; data from both species pooled).

Discussion

Following observations of longer microsatellite repeats in the focal species than in related species among such divergent taxa as primates (Deka et al. 1995; Rubinsztein et al. 1995; Watanabe et al. 1996), sheep (Forbes et al. 1995), birds (Ellegren, Primmer, and Sheldon 1995), and turtles (FitzSimmons, Moritz, and Moore 1995), we show here that this trend is retained even in a reciprocal comparison. Microsatellites developed from the bovine genome are longer in cattle than in sheep and, conversely, microsatellites from the ovine genome are longer in sheep than in cattle. Similar results from a unidirectional analysis might, at the first glance, be interpreted as evidence for a difference in evolutionary
rate between species (Rubinsztein et al. 1995; Rubinsztein, Leggo, and Amos 1995). Our findings demonstrate that this interpretation is not reliable. Because the trend of one species having longer repeats than the other goes in opposite directions in a reciprocal comparison, the observed patterns must be attributed to a systematic bias associated with the loci selected for analysis.

The potential source of this selection bias has been related to the cloning procedure (Ellegren, Primmer, and Sheldon 1995; Forbes et al. 1995). Microsatellites are generally developed to be used as polymorphic markers, and library screening procedures are therefore optimized so that longer-than-average repeats are selected, repeat length usually being positively correlated with degree of variability (e.g., Weber 1990). This nonrandom choice of loci introduces an ascertainment bias in the sense that homologs in congeners should not be expected to have evolved into equally long repeats or, alternatively, should not have retained equally long and/or uninterrupted repeats, depending on the ancestral repeat state. Sequence analyses from several species confirm that homologous loci in congeners generally contain shorter repeats (Schlötterer, Amos, and Tautz 1991; Pépin et al. 1995; Watanabe et al. 1996).

Similar to the observed patterns of repeat lengths in cattle and sheep, the degree of genetic variability was considerably higher in the focal species than in the congener. Although this relationship is as expected from an
ascertainment bias, our data must be treated cautiously, since polymorphism may vary between breeds. Moreover, polymorphism is likely to be higher if animals from divergent lines are analyzed for a particular marker. In this case, for instance, the majority of the bovine microsatellites were genotyped in the parents of international cattle reference families, including N'Dama, Boran, Brangus, and Bos indicus lines, hence representing a wide gene pool. In contrast, most bovine markers were genotyped in a single sheep breed. We believe that this situation alone cannot explain the differing degrees of genetic variability in the two species, since about half of the bovine microsatellites were not only less polymorphic but in fact monomorphic in sheep, and since a similar pattern was evident for sheep microsatellites despite most of these being genotyped in just a few cattle as well as sheep breeds.

In this context, it should be emphasized that our collection of markers itself might be biased to some extent in terms of level of polymorphism. Although we randomly selected markers from the literature, it is possible that the literature contains an unproportionally high share of polymorphic microsatellites, since markers failing to detect polymorphism may not always be reported. Potentially, such bias could partly explain the observed higher levels of genetic variability in the focal species than in the congener. It is hard to evaluate the importance of this potential bias, but we believe that it has not greatly affected our results, since the majority of microsatellites with some 12 or more repeat units (a
H. R. U. 9. Similarly, Amos et al. (1996), by compiling data from human (CA)n microsatellites, found significantly more gains than losses. It remains to be determined whether these data are representative for microsatellites in general.

It might be expected that loci found to be monomorphic in the congener would be more likely to amplify shorter products relative to the focal species than would loci that did show polymorphism. The reasoning behind this is that monomorphic loci are those that are most likely to have lost or not gained significant repeat lengths. The fact that we did not observe such a relationship may simply be related to sample size coupled with the “noise” caused by length variation in repeat flanking regions. Moreover, mutations introducing interruptions in repeat structures in the congener may have decreased genetic variability while still yielding comparatively long amplification products.

It seems plausible that longer repeats and greater polymorphism in the focal species can be explained by an ascertainment bias. Our reciprocal tests lend additional support to this theory, inasmuch as they negate one of the arguments used to invoke directional evolution as a cause of longer repeats and greater polymorphism (Rubinsztein et al. 1995). According to the latter idea, a larger effective population size in the focal species should promote genetic variability and, coupled with a mutational bias toward an increase in repeat units, favor the generation of long alleles. Clearly, this argument cannot apply to a reciprocal test.

Interestingly, the reciprocal design used in this study does provide a means to test whether the rate of microsatellite expansion differs between two species, in this case cattle and sheep. If the difference in mean repeat lengths in the two sets of comparisons in itself differed significantly, then the rate of microsatellite evolution ought to be higher in the species in which focal markers revealed the greatest size difference (focal species – congeneric species). We did not find any evidence for such a difference in the comparison of bovine and ovine microsatellites, suggesting that the rates of microsatellite evolution in these two species are not dramatically different.

It must be emphasized that we do not adduce the present data to be evidence against directional microsatellite evolution; they serve to demonstrate that directional evolution cannot be inferred from unidirectional comparison of homologous amplification products. In fact, some recent studies based on recorded events of in vivo mutations suggest that at least some microsatellite loci may indeed evolve directionally. Primmer et al. (1996) found a significant excess of gains over losses of repeat units at the hypervariable and hypermutable swallow Hirundo rustica tetranucleotide repeat locus HrU9. Similarly, Amos et al. (1996), by compiling data from human (CA)n microsatellites, found significantly more gains than losses. It remains to be determined whether these data are representative for microsatellites in general.

Acknowledgments

H. E. is supported by grants from the Swedish Research Councils for Natural Sciences and for Forestry and Agriculture, N.R. is supported by the International wool secretariat and the Department of Natural Resources and Environment, and B.C.S. is supported by a NERC (U.K.) postdoctoral research fellowship. The manuscript benefited from comments by Rodney L. Honeycutt.

LITERATURE CITED


RODNEY L. HONEYCUTT, reviewing editor

Accepted May 6, 1997