Isolation of Y Chromosome-specific Microsatellites in the Horse and Cross-species Amplification in the Genus Equus

B. Wallner, F. Piumi, G. Brem, M. Müller, and R. Achmann

From the Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien Veterinärplatz 1, A-1210 Wien, Austria (Wallner, Brem, and Müller); INRA, Centre de Recherches de Jouy, Département de Genetique Animale, Jouy-en-Josas, France (Piumi); and Ludwig Boltzmann-Institut für immuno-, zyto- und molekulargenetische Forschung, Veterinärplatz 1, A-1210 Wien, Austria (Brem and Achmann). We thank W. Zimmermann, C. Waltzer, H. Korb, E. Dvorak, H. Siegismund, the Zoological Society of San Diego, and the Zoological Garden Munich for samples. We are grateful to M. Wigler for a detailed RDA protocol. Thanks to G. Muir for comments on the manuscript. This study was financially supported by the Hochschuljubiläumsstiftung der Stadt Wien (H-234/99) and Agrobiogen GmbH.

Address correspondence to Roland Achmann, genteQ GmbH, Falkenried 88, D-20251 Hamburg, Germany, or e-mail: achmann@genteQ.de.

Y chromosome polymorphisms such as microsatellites or single nucleotide polymorphisms represent a paternal counterpart to mitochondrial DNA (mtDNA) for evolutionary and phylogeographic studies. The use of Y chromosome haplotyping in natural populations of species other than humans is still hindered by the lack of sequence information necessary for polymorphism screening. Here we used representational difference analysis (RDA) followed by a screen of a bacterial artificial chromosome (BAC) library for repetitive sequences to obtain polymorphic Y-chromosomal markers. The procedure was performed for the domestic horse (Equus caballus) and we report the first six Y-chromosomal microsatellite markers for this species. Three markers were also useful for haplotyping taxa of the zebra/ass lineage. Y-chromosomal microsatellite markers show a single haplotype in the domestic horse, whereas notable variation has been observed in the other members of the genus Equus.

Recent advances in genome analysis and population genetics in humans suggest that Y chromosome haplotyping may develop into an important tool for studying natural populations (Hurles and Jobling 2001). For equids, where mtDNA and autosomal microsatellites have been used to investigate the phylogenetic relationship within the genus (Oakenfull et al. 2000) and horse domestication (e.g., Jansen et al. 2002; Vila et al. 2001), polymorphic Y-chromosomal markers are not available at present. The construction of a haploid chromosome-specific library requires the preselection of a particular chromosome by flow cytometry (Bergstrom et al. 1998) or microdissection (Ponce de Leon et al. 1996). Both methods are technically demanding and suffer from limitations such as separation efficiency and minute amounts of DNA as starting material (Claussen et al. 1997). Alternatively, haploid markers can be isolated by the use of sequence tags (STTs) as probes for selecting chromosome-specific clones from genomic libraries containing large inserts. The inserts of identified clones can be further screened for microsatellite markers.

Here we used representational difference analysis (RDA) (Lisitsyn et al. 1993) for the isolation of Y chromosome-specific STTs. RDA is a subtractive approach that permits isolating
the cloning of differences between similar genomes. The principle behind RDA is that unique target sequences present in one DNA population (tester, heterogametic XY) can be purified through repetitive steps of difference enrichment by hybridization with an excess of a second DNA population (driver, homogametic XX). RDA-enriched Y-specific STSs were used as probes for screening a bacterial artificial chromosome (BAC) library to subsequently isolate microsatellites by subcloning. We demonstrate the practicability of this strategy to isolate Y-specific STSs and microsatellite markers for the domestic horse (Equus caballus).

Moreover, we present data on the variability of the first six equine Y-chromosomal microsatellite markers in the domestic horse and other equine species.

Materials and Methods

Representational Difference Analysis

The RDA procedure was applied essentially as previously described (Lisitsyn et al. 1993), using the restriction enzyme BglII and corresponding amplification adapter oligonucleotides. Genomic DNA from a Lipizzan stallion and Lipizzan mares was used as tester and driver, respectively. In order to avoid a significant reduction in the kinetic enrichment of the target due to the “Cot effect” (Mathieu-Daudé 1996), the number of polymerase chain reaction (PCR) cycles used to reamplify the product after each subtractive round was kept to a minimum (first round: 12 cycles; third round: 22 cycles).

Fragments of sizes between 500 and 1000 bp of the ampiclon pools after two and three rounds of subtraction were cloned into pUC19 (MBI Fermentas). Inserts from 30 clones, 7 from the library after two subtractive rounds and 23 after three subtractive rounds, were sequenced with BigDye terminator cycle sequencing chemistry (Applied Biosystems) using standard M13 primers. Sequences were classified by their similarity into different sequence families using Autoassembler software (Applied Biosystems). For each family a randomly selected sequence was designated as the reference sequence. Sequences were compared for homologies to known sequences by using BLASTN. Repetitive elements were identified by the program Repeat Masker (Smit and Green 2003; program available at http://ftp.genome.washington.edu/RM/RepeatMasker.html).

Primers were designed for each reference sequence (Table 1) and Y chromosome specificity was tested by PCR amplification of genomic DNA from males and females. Amplification was performed in 15 μl reactions containing 20 ng DNA, 0.5 μM of each primer, 0.25 mM of each dNTP, 2 mM MgCl2, 1X AmpliTaq buffer, and 0.5 U AmpliTaq Gold (Applied Biosystems). PCR conditions were 10 min at 95°C, followed by 35 cycles of 30 s at 95°C, 40 s at annealing temperature (see Table 1), 90 s at 72°C, and a final extension of 10 min at 72°C. Ethidium bromide-stained PCR products were visualized on 2% agarose gels. DNA samples (one per breed) from 14 different breeds (Akhal-Teke, Andalusian, Arabian, Austrian Warmblood, Icelandic Horse, Kladruber, Lipizzan, Mongolian Domestic Horse, Noric, Quarter Horse, Shetland Pony, Tarpan, Thoroughbred, Trakehner) were directly sequenced for male-specific STSs.

Isolation of Microsatellites from Selected BAC Clones

A horse BAC library was screened for Y-chromosomal clones as described (Godard et al. 1998) with primers specific for eight Y-chromosomal STSs isolated by RDA (we excluded locus SH3-B-14 because of difficulties in PCR amplification) and a horse SRY-specific primer pair (Hasegawa et al. 1999). We isolated BAC DNA as described at http://dga.jouy.inra.fr/grafra/ (last visited December 2003). BAC DNA was prescreened for the presence of (CA/GT)n microsatellites by dot-blot hybridization. Microsatellite enriched libraries were established as described by Armour et al. (1994). The enrichment of (CA/GT)n-containing elements was performed as described in Traxler et al. (2000), with the modification that adapters RBgl24 and RBgl12 (Lisitsyn et al. 1993) were used for reamplification. Twenty randomly chosen microsatellite-enriched subclones originating from each BAC clone were subjected to colony PCR using standard M13 primers. Amplicons were controlled for the presence of (CA/GT)n sequences by dot-blot hybridization (see above). For positive clones, plasmid DNA was extracted and sequenced.

Microsatellite Genotyping

We designed specific PCR primers for identified microsatellite sequences and checked their Y-chromosomal location as described above. One primer of each pair was labeled with a fluorescent dye to allow simultaneous analysis of fragments on an ABI310 instrument (Applied Biosystems). PCR was performed in a 15 μl volume containing primers (for concentration see Table 2) and the PCR reaction mix as described above. The DNA was initially denatured at 95°C for 10 min followed by 35 cycles of 30 s at 95°C, 40 s at annealing temperature (see Table 2), 90 s at 72°C, and a final extension of 30 min at 72°C. Alleles were sized relative to the internal size standard Tamra GS 500 using Genescan version 2.1 (Applied Biosystems).

Microsatellite variability was evaluated with DNA from 49 male horses of 32 different breeds representing the domestic horse (Akhal-Teke, Andalusian, Appaloosa, Arabian, Austrian Warmblood, Barb, Connemara, Icelandic Horse, Irish Tinker, Kladruber, Lipizzan, Mangolarga Marchador, Minishetty, Missouri Foxtrotter, Mongolian Domestic Horse, New Forest, Noric, Norwegian Fjord, Oldenburger, Paint, Pinto, Quarter Horse, Shagya Arabian, Shetland Pony, Saddlebred, Shire Horse, Tarpan, Thoroughbred, Trakehner, Trotter, Welsh Pony, Wurttemberger). In addition DNA samples from other equids were analyzed: 10 Przewalski’s Horses (Equus przewalskii), 3 Donkeys (Equus asinus), 1 Onager (Equus hemionus onager), 3 Kiangs (Equus kiang), 3 Grevy’s Zebras (Equus grevyi), 14 Hartmann’s Zebras (Equus zebra hartmannae), 5 Damara Zebras (Equus burchelli antiquorum), and 2 Grant’s Zebras (Equus burchelli boehmi).
# Table 1. Details for RDA selected horse STSs

<table>
<thead>
<tr>
<th>Reference sequence</th>
<th>Number of isolates (two/three rounds)</th>
<th>Insert size (bp)</th>
<th>Primer sequence (5’ to 3’)</th>
<th>Product size (bp)</th>
<th>$T_a$ (°C)</th>
<th>Accession number</th>
<th>Y specificity</th>
<th>Sequence similarities</th>
<th>Number of positive BAC clones/address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eca SH-2-A1</td>
<td>1/4</td>
<td>842</td>
<td>CGGTGTCAAGGTTTTGGGACTTT</td>
<td>747</td>
<td>56</td>
<td>BV005744</td>
<td>Yes</td>
<td>none</td>
<td>4 / 71E7, 144G4, 233F1, 614D12</td>
</tr>
<tr>
<td>Eca SH2-B-5</td>
<td>4/1</td>
<td>549</td>
<td>AAGGATTCGCTGTCGCCCATCAT</td>
<td>394</td>
<td>60</td>
<td>BV005721</td>
<td>No</td>
<td>LINE 1</td>
<td>Not applied</td>
</tr>
<tr>
<td>Eca SH-2-B7</td>
<td>2/1</td>
<td>558</td>
<td>CTCCACCCATACGGAGTTTCTGAGTCCTGAC</td>
<td>502</td>
<td>60</td>
<td>BV005722</td>
<td>No</td>
<td>LINE 1</td>
<td>Not applied</td>
</tr>
<tr>
<td>Eca SH2-B-17</td>
<td>1/1</td>
<td>592</td>
<td>TTTGGCTCTCTTGGCTCTCA</td>
<td>528</td>
<td>60</td>
<td>G72335</td>
<td>Yes</td>
<td>None</td>
<td>1 / 1056F3</td>
</tr>
<tr>
<td>Eca SH3-B-1</td>
<td>0/2</td>
<td>655</td>
<td>TGGTTTATGGGAGTTGGTG</td>
<td>501</td>
<td>60</td>
<td>G72336</td>
<td>Yes</td>
<td>LINE 1</td>
<td>2 / 255A1, 255B2</td>
</tr>
<tr>
<td>Eca SH3-B-2</td>
<td>0/1</td>
<td>496</td>
<td>AGACAAAGCAGAGAGCCCA</td>
<td>368</td>
<td>60</td>
<td>BV005746</td>
<td>No</td>
<td>none</td>
<td>Not applied</td>
</tr>
<tr>
<td>Eca SH3-B-5</td>
<td>0/1</td>
<td>493</td>
<td>ACCCTGAGATTTTTGAGTCTGG</td>
<td>369</td>
<td>60</td>
<td>BV005723</td>
<td>No</td>
<td>SINE (MIRs)</td>
<td>Not applied</td>
</tr>
<tr>
<td>Eca SH3-B-6*</td>
<td>0/2</td>
<td>580</td>
<td>AGAGTTGGAATTTTGAGTGG</td>
<td>492</td>
<td>60</td>
<td>BV005718</td>
<td>Yes</td>
<td>ERV class I</td>
<td>2 / 68B8, 804B12</td>
</tr>
<tr>
<td>Eca SH3-B-7</td>
<td>0/1</td>
<td>534</td>
<td>TGAAAGGGCTAAGGAGCTG</td>
<td>436</td>
<td>56</td>
<td>BV005719</td>
<td>Yes</td>
<td>Line 1</td>
<td>1 / 18SA12</td>
</tr>
<tr>
<td>Eca SH3-B-8</td>
<td>0/2</td>
<td>668</td>
<td>CCCCAGTTCCGCTGAGAT</td>
<td>472</td>
<td>60</td>
<td>G72337</td>
<td>Yes</td>
<td>None</td>
<td>2 / 447F12, 601D9</td>
</tr>
<tr>
<td>Eca SH3-B-11</td>
<td>0/1</td>
<td>554</td>
<td>AGGCAAGTTCAAGTGAAGCTATG</td>
<td>353</td>
<td>58</td>
<td>BV005724</td>
<td>No</td>
<td>LINE 1, MER1_type</td>
<td>Not applied</td>
</tr>
<tr>
<td>Eca SH3-B-12</td>
<td>0/1</td>
<td>499</td>
<td>GGAGGGACTGAGAGAAGCTCAA</td>
<td>400</td>
<td>64</td>
<td>G72338</td>
<td>Yes</td>
<td>None</td>
<td>2 / 180F9, 927H5</td>
</tr>
<tr>
<td>Eca SH3-B-14*</td>
<td>0/3</td>
<td>518</td>
<td>GTGACCCGAGGACGTGTG</td>
<td>486</td>
<td>66</td>
<td>BV005745</td>
<td>Yes</td>
<td>ERV class I</td>
<td>Not applied</td>
</tr>
<tr>
<td>Eca SH3-B-19</td>
<td>0/1</td>
<td>580</td>
<td>TCCGACAGCTTCGTCCCTCTC</td>
<td>307</td>
<td>60</td>
<td>BV005720</td>
<td>Yes</td>
<td>Human Y, AC006371</td>
<td>2 / 180F9, 927H5</td>
</tr>
<tr>
<td>EcaSRY</td>
<td></td>
<td></td>
<td>GGATGCGAGTGAAAAAGTCGTT</td>
<td>364</td>
<td>66</td>
<td>AB004572</td>
<td>Yes</td>
<td>None</td>
<td>1 / 616B11</td>
</tr>
</tbody>
</table>

$T_a$, annealing temperature.

Y chromosome specificity was tested by amplifying male and female DNA.

Reference sequences marked with * occur in multiple copies on the horse Y chromosome (see text).

BAC clone addresses refer to the library constructed by Godard et al. (1998). Note that Eca SH3-B-12 and Eca SH3-B-19 were found in the same BAC clones. The reference sequence for EcaSRY is from Hasegawa et al. (1999).
Results and Discussion

Representational Difference Analysis

In order to isolate specific markers from the equine Y chromosome we performed RDA with male equine genomic DNA as tester and female genomic DNA as driver. Figure 1 shows amplicons after the first, second, and third round of RDA. The second and third round products show an enrichment of discrete bands and a reduction in the intensity of smearing background DNA. Second and third round amplicons were size fractionated and subcloned. Thirty randomly selected clones were sequenced and sequences clustered into 14 families. Eight of the 14 sequence families showed similarity to common genome-wide repeats and one clone showed homology to a human Y-chromosomal sequence (Table 1).

For 9 of the 14 reference sequences, PCR on male samples amplified a single band, whereas female samples could not be amplified. The exclusive amplification of male DNA strongly supports a Y-chromosomal origin of the isolated sequences (Table 1).

Because four of the nine Y-specific sequences show a high homology to known repetitive elements (LINEs or ERVs), we tested for multiple copies on the Y chromosome by direct sequencing of PCR products from 14 male horses from different breeds. Two STSs with partial homologies to endogenous retroviral sequences showed heterogeneous sequencing signals at some nucleotide positions (Eca SH3-B-6 at 4 positions, Eca SH3-B-14 at 10 positions), indicating multiple, divergent copies on the horse Y chromosome. Because of the repetitive nature of the Y chromosome (Foote et al. 1992) and the preferential isolation of repeats due to kinetic enrichment in the amplification steps (e.g., Panaud et al. 2002), the isolation of Y-specific repeats is not surprising. Sequencing of the remaining seven Y-chromosomal sequences showed no variable positions over a total length of 3033 nucleotides among the 14 individuals from different breeds. This observation is not unexpected, because the diversity on the Y chromosome is known to be extremely low (Shen et al. 2000).

Screening of a Horse BAC Library

We screened a male horse BAC library (Godard et al. 1998) with primers specific for nine Y-chromosomal STSs and selected 15 BAC clones in total (Table 1). Three markers exclusively identified a single clone and four markers simultaneously amplified in two to four clones. STSs SH3-B-12 and SH3-B-19 gave positive signals for the same two clones, indicating that they are located very closely on the Y chromosome. Hybridization of a (CA/GT)n probe to BAC DNA dot blots identified 11 BAC clones containing simple tandem repeat sequences. The application of an enrichment method (Armour et al. 1994, Traxler et al. 2000) permitted a quick isolation of dinucleotide microsatellites from BACs. The screening of microsatellite-enriched BAC DNA (20 random subclones per original BAC clone) by dot blot hybridization revealed that out of 220 subclones tested for...
microsatellite sequences, 56 showed a positive signal. We sequenced the insert of all positive subclones and identified 11 different microsatellite loci. Six of eight loci were successfully amplified from horse DNA (Table 2) and proved to be male specific. Electropherograms of Y-specific markers displayed the typical peak pattern (main peak preceded by one or two stutter peaks) of dinucleotide microsatellites (Figure 2a). However, marker Eca.YH12 produced a peak pattern with three main peaks (Figure 2a). Eca.YH12 could not be amplified from female DNA and amplification from the corresponding BAC clone 68B8 revealed a single main peak (Figure 2b). We assume that this microsatellite has multiple copies on the horse Y chromosome. The duplication or triplication of a larger Y-chromosomal region, followed by a change in the number of repeats within the microsatellite is a frequently observed phenomenon in humans (Kayser et al. 2000).

We assessed the variability of the six equine Y-specific microsatellites in a sample of 49 male domestic horses. Although males from 32 distinct breeds from different continents were investigated, we detected only a single haplotype (see Table 3).

The Przewalski's Horse (n = 10) haplotypes differed from that of the domestic horse at two loci (Eca.YH12 and Eca.YA16; Table 3). In contrast to the domestic horse, we identified two haplotypes, which were distinguished from the domestic horse haplotype by the same two loci. Eca.YH12 displayed either one or two detectable alleles (Figure 2c,d), which can be explained by the fact that Eca.YH12 is not triplicated in the Przewalski's Horse. Alternatively, the locus could be triplicated, but shows identical alleles at two or three loci. Finally, mutations in the primer binding sites may affect the amplification of this locus, so that one or two copies are not amplified at all (nonamplifiable alleles; Achmann et al. 2001). Our data show that although E. przewalskii has undergone a severe bottleneck (Boyd and Houpt 1994), at least two stallions must have contributed to the present-day population.

Figure 1. Agarose gel electrophoresis of the whole genomic fraction of Tester amplicon (T) and the difference products, obtained after the first (1), second (2), and third (3) cycles of RDA. The positions of the 1000 bp (upper arrow) and the 500 bp (lower arrow) bands of a DNA ladder (L) are indicated in the margin.

Figure 2. (A) Electropherogram showing amplification products of six Y-chromosomal microsatellite markers for an E. caballus DNA sample. Note the multiallelic peak pattern of Eca.YH12. (B) Eca.YH12 amplified from BAC clone. (C,D) Eca.YH12 from E. przewalskii with one or two alleles detected.
We further extended the analysis to several other equine species (Table 3). Three markers (Eca.YP9, Eca.YM2, and Eca.YE1) could be amplified under less stringent PCR conditions (Table 2). For E. ainsu, E. bennius onager, E. kiang, E. zebra hartmannae, E. burchelli antiquorum, E. burchelli boehmi, and E. greyi unique haplotypes were observed, which allow the discrimination between each species/subspecies (Table 3). We discovered interspecific haplotypic variation in E. ainsu, E. greyi, and E. zebra hartmannae (Table 3), although we analyzed only a small number of animals (n = 3–14) in these species.

The six newly characterized Y-specific microsatellites displayed no variability in the domestic horse. This is in contrast to the genetic variation normally observed for equine autosomal markers such as microsatellites (e.g., Vila et al. 2001) and polymorphic genes (Shubitowski et al. 2001) as well as for mtDNA (Jansen et al. 2001; Vila et al. 2001). Although some of the isolated Y-chromosomal microsatellite loci for the horse are imperfect and have a low total number of repeats, they show variability in other equine species. Thus a low mutation rate alone cannot account for the observed Y chromosome uniformity in the domestic horse.

We suggest that the observation of a single Y chromosome haplotype in the horse compared to the high levels of mtDNA variation (Jansen et al. 2002; Vila et al. 2001) may reflect a strong bias toward females in breeding and trade. The effective population size of the Y chromosome corresponds to the number of breeding males in a population. Intensive breeding strategies in which selected breeding stallions are used for many mares (Levine 1999), together with upgrading many breeds by crossing in Arabian and thoroughbred stallions, have probably resulted in the fixation of a single Y chromosome haplotype. Further investigations are necessary to confirm this hypothesis, because a global selective sweep, due to advantageous mutations in a Y-specific gene, could also have led to the fixation of a single haplotype. Whatever the explanation for the apparently monomorphic Y chromosome in domestic horses is, it seems, that today’s horse Y chromosomes can be traced back to a recent common ancestor.

In summary, we show that RDA in combination with the screen of a BAC library is a quick and effective strategy to isolate Y-chromosomal microsatellite markers. Since BAC libraries are available for several domestic animals, our approach for the isolation of Y-specific markers can be easily performed in other species. An increasing number of Y-specific markers could help to extend our knowledge about sex-specific differences in ecology, behavior, or migration not only in humans, but also in many other mammalian species.

Table 3. Observed allele sizes for Y-specific microsatellites and Y-chromosomal haplotype frequencies in equids

<table>
<thead>
<tr>
<th>Species (sample size)</th>
<th>Eca.YH12</th>
<th>Eca.YA16</th>
<th>Eca.YP9</th>
<th>Eca.YM2</th>
<th>Eca.YE1</th>
<th>Eca.YJ10</th>
<th>Haplotype frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equus caballus (49)</td>
<td>96, 100, 102</td>
<td>157</td>
<td>214</td>
<td>116</td>
<td>196</td>
<td>213</td>
<td>1.0</td>
</tr>
<tr>
<td>Equus przewalskii (10)</td>
<td>100</td>
<td>161</td>
<td>214</td>
<td>116</td>
<td>196</td>
<td>213</td>
<td>0.5</td>
</tr>
<tr>
<td>Equus asinus (3)</td>
<td>—</td>
<td>—</td>
<td>197</td>
<td>110</td>
<td>191</td>
<td>—</td>
<td>0.66</td>
</tr>
<tr>
<td>Equus burchelli antiquorum (15)</td>
<td>—</td>
<td>—</td>
<td>197</td>
<td>120</td>
<td>191</td>
<td>—</td>
<td>1.0</td>
</tr>
<tr>
<td>Equus greyi (3)</td>
<td>—</td>
<td>—</td>
<td>204</td>
<td>110</td>
<td>189</td>
<td>—</td>
<td>0.66</td>
</tr>
<tr>
<td>Equus zebra hartmannae (14)</td>
<td>—</td>
<td>—</td>
<td>197</td>
<td>122</td>
<td>189</td>
<td>—</td>
<td>0.71</td>
</tr>
<tr>
<td>Equus burchelli boehmi (2)</td>
<td>—</td>
<td>—</td>
<td>197</td>
<td>124</td>
<td>189</td>
<td>—</td>
<td>0.14</td>
</tr>
<tr>
<td>Equus przewalskii (3)</td>
<td>—</td>
<td>—</td>
<td>195</td>
<td>112</td>
<td>189</td>
<td>—</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Allele designations correspond to allele sizes measured by capillary electrophoresis (ABI310; Applied Biosystems). Microsatellite marker Eca.YH12 shows a multiple peak pattern in E. caballus and E. przewalskii (see Figure 2 and text).

References


Received May 29, 2003
Accepted January 15, 2004

Corresponding Editor: Oliver A. Ryder