Original article

A genetic analysis of variable number of tandem repeats (VNTR) polymorphism in the horse

G Guérin [1], M Bertaud [1], B Billoud [1], JC Mériaux [2]

[1] INRA, Laboratoire de Génétique Biochimique ;
[2] INRA, Laboratoire des Groupes Sanguins, 78352 Jouy-en-Josas Cedex, France

Summary – Restriction fragment length polymorphism detected with the 33.6 minisatellite probe (Jeffreys *et al*, 1985) was analysed in 3 horse families of paternal half-sibs and in a sample of stallions from 4 breeds. Among the bands detected on *Hae* III genomic DNA digests, it was found that strongly hybridizing fragments behaved as alleles at 2 different loci. These 2 loci, showing 7 and 3 detectable alleles, were not closely linked to each other nor to informative blood groups and protein markers. No neo-mutation in allele length was observed at these 2 loci in the 3 families. In the stallion sample, 8 alleles were detected at the first locus and no differences were found between their frequencies in the 4 breeds. Heterozygosity and polymorphism information content (PIC) values estimated in the 4 breeds show that the Thoroughbred is the least variable breed and the Arab the most. In the whole population the PIC values were 0.73 and 0.70 respectively.

horse / variable number of tandem repeats / linkage / genetic marker / polymorphism

Résumé – **Analyse génétique du polymorphisme du nombre variable de répétitions en tandem (VNTR) chez le cheval.** *Le polymorphisme de longueur de fragment de restriction détecté par la sonde minisatellite 33.6 (Jeffreys* et al, *1985) a été analysé chez le cheval dans trois familles de demi-germains paternels et dans un échantillon d'étalons de quatre races. Parmi toutes les bandes révélées à partir d'ADN génomique digéré par* HaeIII, *les fragments donnant un signal intense se comportent comme des allèles ségrégeant à deux locus différents. Ceux-ci, qui possèdent respectivement sept et trois allèles détectables, ne sont pas étroitement liés entre eux ni avec les locus informatifs des groupes sanguins et des protéines du sang. Aucune néo-mutation de longueur des allèles n'a été détectée dans les trois familles. Dans l'échantillon des étalons, huit allèles ont été observés au premier locus et aucune différence de fréquence n'apparaît entre les quatre races. Les valeurs d'hétérozygotie et les valeurs informatives des polymorphismes (PIC) estimées sur cet échantillon montrent que la race de Pur-sang est la moins variable à l'opposé de*

la race Arabe. Les valeurs estimées sur la population totale sont respectivement de 0,73 et de 0,70.

cheval / nombre variable de répétitions en tandem / liaison / marqueur génétique / polymorphisme

INTRODUCTION

It is now well established that the genome of a number of animal and plant species contains regions that are highly variable, among which are the minisatellites. These regions are detected by probes from a different origin containing sequences several base-pairs long, tandemly repeated. Initially, highly polymorphic patterns displaying a large number of fragments on Southern blots resulting from variation in the number of tandem repeats and called fingerprints were obtained in man by Jeffreys *et al* (1985). These authors used a repeat found in the second intron of the myoglobin gene to design a probe that revealed related human sequences, some of which were themselves later used as probes. This was the case for the 33.6 and 33.15 regions that revealed extensive polymorphism in a number of organisms. Other probes containing repeated sequences such as a sequence from the M13 bacteriophage (Vassart *et al*, 1987), a sequence homologous to the *Drosophila Per* gene (Shin *et al*, 1985), in the 3′ region of the α globin gene (Jarman *et al*, 1986) and a number of others, cloned from the human genome (see for example Nakamura *et al*, 1987), were found to detect very high levels of polymorphism.

In pets and domestic farm animals, polymorphisms were found in dogs and cats (Jeffreys and Morton 1987), in cattle (Georges *et al*, 1988), in poultry (Hillel *et al*, 1989), in sheep (Drinkwater *et al*, 1990) and in pigs (Coppieters *et al*, 1990) using either human, viral or homologous probes.

In horses, restriction fragment length polymorphisms (RFLP) have been detected using different probes such as M13, Jeffreys' minisatellites or core sequence, *Per*, human VNTR, (TG)n, a homologous horse probe and synthetic tandem repeats (Georges *et al*, 1988; Massina *et al*, 1989; Troyer *et al*, 1989; Bernoco and Byrns, 1991; Broad *et al*, 1991; Hopkins *et al*, 1991; Ellegren *et al*, 1992; Giulotto *et al*, 1992; Mariat *et al*, 1992). However, very little or no information about segregation of the fragments and linkage analysis has been reported as limited families were typed.

In this report, we show the segregation of DNA fragments hybridized with a human minisatellite probe in 3 large horse pedigrees and give some genetic characteristics of the regions detected in 4 horse breeds.

MATERIALS AND METHODS

Animals

Three horse families consisting of a stallion and half or full paternal sibs, along with their mothers, were collected. The number of mother (74) – offspring (87)

combinations were 30, 27 and 30 for the 3 stallions, Iris Landais, Artichaut and Matador du Bois, respectively. All animals were of Selle Français origin. A sample of stallions from 4 breeds, Arabs (AR, $N = 47$), Thoroughbred (TB, $N = 43$), Selle Français (SF, $N = 40$) and Trotteur Français (TF, $N = 43$) were also obtained from different state stud-farms. All animals were blood typed by the Laboratoire des Groupes Sanguins des Chevaux, INRA-CRJ, Jouy-en-Josas, for the blood group systems (A, C, D, K, P, Q, U) and the protein systems (Alb, Tf, A1B, Es, Pi, Gpi, Pgd, Pgm) and all offspring qualified.

Purification of DNA

Blood samples were drawn in tubes containing tripotassium EDTA (Vacutainer) and DNA was extracted according to the method of Jeanpierre (1987). Blood was treated twice in a solution of 10 mM NaCl, 10 mM EDTA, pH 8, for 30 min at +4°C. Pellets containing leucocytes were homogenized in a solution of 5 M guanidine HCl, 0.5 M ammonium acetate, 1.25% N-lauroylsarcosine containing 0.1 mg/ml proteinase K for 1 h at 60°C. DNA was then precipitated with 2 vol ethanol and washed twice with 70% ethanol.

Restriction endonuclease digestion and electrophoresis

Seven μg samples of horse DNA were digested for 20 h at 37°C by restriction endonucleases *Hae*III or *Hinf*I (BRL or Boehringer) according to the manufacturer's recommendations. Complete digestion was achieved with a total addition of 5 units per μg DNA. Restriction endonuclease DNA fragments were subjected to horizontal electrophoresis in 1% agarose gels 25 cm in length in Tris-acetate buffer (40 mM Tris, 20 mM AcOH, 2 mM EDTA; pH 8.3). Electrophoresis was performed with circulating buffer at 55 V (*Hae*III) or 60 V (*Hinf*I) for 18 h. Calibration was achieved by adding a sample of *Bst*EII-*Kpn*I–*Sma*I digested lambda phage DNA.

Transfer

Transfer was performed on Biodyne B membrane (Pall Industry, SA). The DNA digested fragments in the gels were depurinated in 0.15 N HCl for 10 min, denatured in a 0.4 N NaOH solution for 30 min and transferred using the method of Southern (1975) at room temperature for 18 h with 0.4 N NaOH.

Hybridization probes

The 2 human VNTR probes, 33.6 and 33.15 were kindly provided by AJ Jeffreys. They were labelled by the primer extension method according to Jeffreys (1985), except that the radiolabelled insert was not extracted from the plasmid.

Prehybridization and hybridization

Prehybridization was carried out for at least 4 h at 65°C in a mixture containing 6 × SSC (0.09 M trisodium citrate, 0.9 M NaCl), 0.25% skimmed milk, 0.5% Denhardt's solution, 0.5% SDS). Hybridization was performed in the same solution

except that 0.05% sodium pyrophosphate was added and sufficient radiolabelled probe to provide $\approx 1 \times 10^6$ cpm/ml. Membranes were then washed twice for 15 min at room temperature in 2 × SSC, 0.1% SDS then in the same solution for 30 min at 65°C (in more stringent solution at the same temperature if necessary) and finally rinsed in 2 × SSC at room temperature. The restriction endonuclease fragments were vizualized by autoradiography in a cassette with a Kodak XAR film and Cronex LI-PLUS intensifying screens at −70°C.

Dehybridization

This was performed by washing the blot for 1 h at 37°C with 0.4 N NaOH, followed by washing for 1 h in a 0.1 × SSPE, 0.5% SDS, 0.2 M Tris–HCl pH 7.5 solution (1 × SSPE = 0.18 M NaCl, 0.01 M NaH_2PO_4, 1 mM EDTA).

Linkage analysis

Calculations were performed according to the Lod score method of Morton (1955).

RESULTS

Probes and enzymes

DNA extracted from our group of horses was digested with *Hae*III and with *Hin*fI. Whatever the enzyme used, the 33.15 probe gave a large number of fragments with rather homogeneous intensities and no clear polymorphism (results not shown). Conversely, the 33.6 probe detected fewer bands with different hybridization signal intensities. With *Hin*fI, all strongly hybridizing bands were detected in the range of small fragment lengths (less than 3.5 kb) (fig 1a). *Hae*III gave more evenly distributed bands spread over the autoradiograms (fig 1b). Since, with DNA digested with *Hin*fI, the different levels of band migration appeared to be more heterogeneous than with *Hae*III, it was decided to take into account only results obtained with the 33.6-*Hae*III probe–enzyme combination.

Fragment identification and polymorphism

Blots usually consisted of DNA digests from 1 sire and its offspring along with DNA digests from their mothers. The 3 sires were also run on the same gel to compare the respective mobilities of their bands (fig 2). Two major groups of bands with a strong radioactive signal appeared in the range of medium (2.5 to 5 kb) and small fragment length inferior to 2 kb (fig 2) on *Hae*III blots probed with 33.6. A total of 15 fragments were given a letter as an identification code (A to O) according to their mobility and intensity while the rest of the bands giving low signals were not included in our interpretation. All strongly hybridizing fragments were found to be polymorphic.
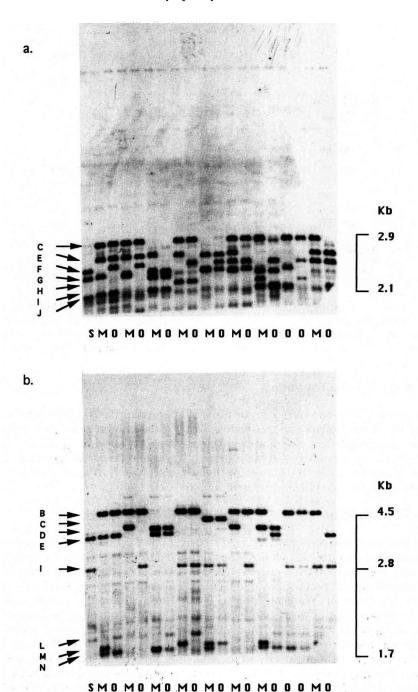
Fig 1. Patterns obtained on a horse pedigree whose DNA was probed with 33.6. a. *Hinf*I digest; b. *Hae*III digest. Fragment designation is indicated on the left and fragment size in kilobases (kb) on the right. Pedigree is represented as follows: S = sire; M = mare; O = offspring.
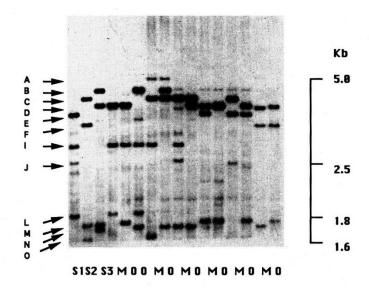
Fig 2. Patterns obtained with *Hae*III horse DNA digest probed with 33.6. The 3 sires and a partial sire pedigree are represented. All offspring are paternal half or full sibs from stallion No 3 and are preceeded by their mare. S = sire; M = mare; O = offspring. Fragment designation is indicated on the left and fragment size in kilobases (kb) on the right.

Segregation of the fragments

Transmission of single paternal fragments was followed in the 3 sire progenies. Only informative dam(s)–offspring combinations, that is dam negative for each fragment, were taken into account. In this way, the presence or absence of a band in the offspring could be unambiguously interpreted. It was observed that a given fragment present in the sire was transmitted to half of his progeny (table I) suggesting a Mendelian inheritance of the corresponding allele.

Locus identification

Two approaches were consecutively used to determine the number of loci controlling the fragment polymorphism. One is based on the analysis of the relative segregation of paternal fragments in their progeny, the other considers that, in the same animal, any fragment should not triplicate with others already assigned to a locus. In the medium size range of fragments (2.5 to 5 kb), sires exhibited 2 bands that were found to behave as alleles at one locus. As no dam or offspring possesses more than 2 of these bands, it was concluded that these fragments were alleles at a single locus. Six alleles present in the sires were thus identified (B, C, D, E, F, I). In addition, one fragment (H) which was only present in some mares and part of their offspring, but did not triplicate with the 6 previously defined alleles, was also assigned to this first locus. The same arguments were employed for the smaller fragments which were shown to segregate as 3 alleles M, N and O at a second locus

Table I. Segregation of paternal fragments in 3 families.

Fragment identification	Fragment size (kb)	Transmission of fragments to progeny		No of informative offspring	χ^2	Significance (P < 0.05)
		+	−	Total		
B	4.47	12	9	21	0.43	NS
C	4.12	8	17	25	3.24	NS
D	3.87	7	5	12	0.33	NS
E	3.62	9	11	20	0.20	NS
F	3.36	11	9	20	0.20	NS
I	2.84	9	16	25	1.96	NS
M	1.75	17	19	36	0.11	NS
N	1.70	9	10	19	0.05	NS
O	1.63	8	11	19	0.47	NS

but the existence of negative animals in this band category suggests that there are other fragments generated by this locus not detected in our conditions. Some other fragments were not allocated to any specific other loci mainly because insufficient genetic information could be obtained (bands A, G, J, K and L).

Mutation rate in allele length

Each fragment belonging to 1 of 2 loci present in an offspring was also present in one of its parents which strongly suggests that no mutation or mutation-like event had occurred in these 3 families. The upper limit of the 5% confidence interval of the mutation rate based on the Poisson distribution is thus 1.8% at the first locus and 4.4% at the second locus according to the number of observed fragments transmitted to the progeny (168 and 66 respectively).

Linkage analysis

No close linkage was found between these 2 loci and between them and any of the other markers that were heterozygous in at least 1 of the 3 sires (A, D, P, Q and U blood groups and Al, Tf, Es, A1B, Pi, Pgd and Pgm for the proteins). No Lod scores exceeded 1.0 for recombination values ranging between 0.01 and 0.5.

Polymorphism in the 4 breeds

As for the families, all strongly hybridizing fragments were polymorphic. The F fragment, detected in families, was split into F1 and F2 fragments that were considered to be alleles. Conversely, H and I were difficult to distinguish as these 2 fragments are not frequent and migrate very close to each other and thus were pooled into a single fragment (H). No interpretation is given for the second locus because band intensities were sometimes variable and could become a source of error. The 4 populations were found to be in Hardy–Weinberg equilibrium based

on the albumin locus (mean value $\chi^2 = 1.28, df = 3$). Allelic frequencies for the first minisatellite locus are given in table II and no differences were detected between the populations ($\chi^2 = 0.94$, $df = 21$). Heterozygosities and polymorphism information content (PIC) values (Botstein *et al*, 1980) are given in table III.

Table II. Allelic frequencies at the first 33.6 locus in the 4 stallion populations.

Alleles	AR	TB	SF	TF	All
B	0.00	0.00	0.04	0.06	0.02
C	0.20	0.09	0.05	0.05	0.11
D	0.24	0.59	0.40	0.49	0.43
E	0.07	0.02	0.11	0.06	0.07
F1	0.11	0.00	0.00	0.05	0.04
F2	0.29	0.28	0.19	0.21	0.24
G	0.01	0.00	0.04	0.05	0.02
H	0.07	0.01	0.15	0.03	0.07

AR: Arab; TB: Thoroughbred; SF: Selle Français; TF: Trotteur Français.

Table III. Heterozygosities and polymorphism information content (PIC) values in the 4 stallion populations.

Breed	Heterozygosity	PIC value
AR	0.79	0.77
TB	0.63	0.50
SF	0.75	0.73
TF	0.75	0.67
All	0.73	0.70

Abbreviations: see table II.

DISCUSSION

In an initial screening, horse genomic DNA was digested with *Hae*III or *Hin*fI and Southern blots were subjected to hybridization with either the 33.6 or 33.15 probe. It appeared that the 33.15 probe gave about 25 hybridizing bands per individual with a low degree of polymorphism and some background signal. On the contrary, 33.6, in the same experimental conditions allowed the detection of a smaller number of more distinct bands per individual. Comparison of horse DNA digested with *Hae*III or *Hin*fI and probed with 33.6 revealed that the 2 enzymes detected roughly similar banding patterns with a greater number of fragment migration levels with *Hae*III than with *Hin*fI. For example in figure 1, the *Hae*III-B fragment corresponds exactly to the *Hin*fI-C fragment while the *Hae*III-E is included in the *Hin*fI-J, which means that the J fragment is heterogeneous (the second locus was not

observable on *Hin*fI DNA digests). This observation confirms the prediction that the polymorphism detected with VNTR originates from the number of tandem repeats and not from mutation in the DNA restriction site of the enzyme.

The banding pattern of the 33.6-*Hae*III probe–enzyme combination resembles more a single-locus VNTR situation than a fingerprint and shows that at least 2 loci are predominantly recognized due to a great similarity between the horse minisatellite core sequence and the human probe sequence. On the contrary, the 33.15 probe detects a family of DNA fragments, the sequences of which are more similar to each other and are most probably generated by a greater number of loci.

The 33.6-*Hae*III combination allowed an analysis of the segregation of strongly hybridizing fragments in sire progenies, an approach which has proved to be very tentative or even impossible with true fingerprints. Two loci were identified with 7 and 3 detectable alleles in these 3 families. No close linkage was detected between any of the red cell antigen, protein and DNA loci detected in this study and no mutation event could be observed which indicates that these loci are rather stable.

In the stallion populations, 8 alleles were detected at the first locus whose frequencies did not significantly differ between the 4 breeds. However, this observation has to be confirmed on a larger number of animals. On the other hand, some bands at this locus migrated very close to each other and we have certainly underestimated the real level of polymorphism of this system. It can be observed that the lowest PIC value is found in the Thoroughbred, which confirms the well-known homogeneity of this breed (Bowling and Clark 1985; Guérin and Mériaux, 1986; Ellegren *et al*, 1992). Conversely, the Arab sample is the most heterogeneous of the 4 breeds. At the second locus in the population sample, identification of alleles was rendered somewhat hazardous due to the low power of resolution in this part of the gel. Bands were found to be less distinct with variable intensities and levels of migration heterogeneous. These characteristics were much less important in the families probably because animals were interrelated, thus reducing the variability observed in unrelated animals of the same breed.

In spite of some limits imposed by the resolution of the technique, the high polymorphism detected with the 33.6 probe makes it a very useful marker for horse identification and linkage analysis in this species.

ACKNOWLEDGMENTS

REFERENCES

Bowling AT, Clark RS (1985) Blood group and protein polymorphism gene frequencies for seven breeds of horses in the United States. *Anim Blood Groups Biochem Genet* 16, 93-108

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using RFLP. *Am J Hum Genet* 32, 314-331

Broad TE, Lewis PE, Phua SH, Crellen LJ, Forrest JW, Pugh PA, Anderson IL, Farndale B (1991) DNA fingerprinting of horses. *Anim Genet* 22, supp 1, 83-84

Bernoco D, Byrns GR (1991) DNA fingerprint variation in horses. *Anim Biotechnol* 2, 145-160

Coppieters W, Van de Weghe A, Depicker A, Bouquet Y, Van Zeveren A (1990) A hypervariable pig DNA fragment. *Anim Genet* 21, 29-38

Drinkwater RD, Hediger R, Hetzel DJS (1990) Polymorphism and location of minisatellite sequences in sheep. *In: 4th World Congr Animal Genetics*, vol 13, 129-132

Ellegren H, Andersson L, Johansson M, Sandberg K (1992) DNA fingerprinting in horses using a simple (TG)n probe and its application to population comparisons. *Anim Genet* 23, 1-9

Georges M, Lequarre AS, Castelli M, Hanset R, Vassart G (1988) DNA fingerprinting in domestic animals using four different minisatellite probes. *Cytogenet Cell Genet* 47, 127-131

Giulotto E, Bisson N, Semino O, Vergnaud G, Brega A, Di Sansebastiano G, Santachiara Benerecetti S (1992) DNA polymorphism in horse identification. *In: 43rd Annu Meet EAAP*. Madrid, 13-17 September, vol 2, 558

Guérin G, Mériaux JC (1986) La distribution des marqueurs sanguins dans les races équines. Analyse sur un échantillon de Pur-sang, Trotteur Français et Selle Français. *In: 12ème Journée d'Etude du CEREOPA*, 2-13

Hillel J, Plotzy Y, Haberfeld A, Lavi U, Cahaner A, Jeffreys AJ (1989) DNA fingerprints of poultry. *Anim Genet* 20, 145-155

Hopkins B, O'Connel, Hopkins J (1991) Use of DNA fingerprinting in paternity analysis of closely related Exmoor ponies. *Equine Vet J* 23, 277-279

Jarman AP, Nicholls RD, Weatherall DJ, Clegg JB, Higgs DR (1986) Molecular characterisation of a hypervariable region downstream of the human α-globin gene cluster. *EMBO J* 5, 1857-1863

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. *Nature (Lond)* 314, 67-73

Jeanpierre M (1987) A rapid method for the purification of DNA from blood. *Nucleic Acids Res* 15, 9611

Jeffreys AJ, Morton DB (1987) DNA fingerprints of dogs and cats. *Anim Genet* 18, 1-15

Mariat D, Guérin G, Bertaud M, Vergnaud G (1992) Modulation of polymorphic loci detection with synthetic tandem repeats variants. *Mammalian Genome* 3, 546-549

Masina P, Di Gregorio P, Rando A, Blasi M (1989) L'impronta del DNA per individuare paternità errate nel cavallo. *Atti Soc Ital Sci Vet* XLIII, 1789-1793

Morton NE (1955) Sequential tests for the detection of linkage. *Am J Hum Genet* 7, 177-318

Nakamura Y, Leppert M, O'Connel P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235, 1616-1622

Shin HS, Bargiello TA, Clar BT, Jackson FR, Young MW (1985) An unusual coding sequence from a *Drosophila* clock gene is conserved in vertebrates. *Nature (Lond)* 317, 445-448

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98, 503-517

Troyer D, Howard D, Leipold HW, Smith JE (1989) A human minisatellite sequence reveals DNA polymorphism in the equine species. *J Vet Med* 36, 81-83

Vassart G, Georges M, Monsieur R, Brocas H, Lequarre AS, Christophe D (1987) A sequence M13 phage detects hypervariable minisatellites in human and animal DNA. *Science* 235, 683-684