Breed demarcation and potential for breed allocation of horses assessed by microsatellite markers

G. Bjørnstad and K. H. Røed

Department of Morphology, Genetics and Aquatic Biology, The Norwegian School of Veterinary Science, Oslo, Norway

Summary

Population demarcation of eight horse breeds was investigated using genotype information of 306 horses from 26 microsatellite loci. The breeds include the indigenous Norwegian breeds Fjord Horse, Nordland/Lyngen Horse, Døle Horse and Coldblooded Trotter together with Icelandic Horse, Shetland Pony, Standardbred and Thoroughbred. Both phylogenetic analysis and a maximum likelihood method were applied to examine the potential for breed allocation of individual animals. The phylogenetic analysis utilizing simple allele sharing statistics revealed clear demarcation among the breeds; 95% of the individuals clustered together with animals of the same breed in the phylogenetic tree. Even breeds with a short history of divergence like Døle Horse and Coldblooded Trotter formed distinct clusters. Implementing the maximum likelihood method allocated 96% of the individuals to their source population, applying an assignment stringency of a log of the odds ratio larger than 2. Lower allocation stringency assigned nearly all the horses. Only three individuals were wrongly allocated a breed by both methods. In conclusion, the study demonstrates clear distinction among horse breeds, and by combining the two assignment methods breed allocation could be determined for more than 99% of the individuals.

Keywords genetic structure, horse breed, microsatellite, population assignment, simple allele sharing, WHICHRUN.

Introduction

The development of molecular DNA markers has brought about great advances during recent years because of their highly polymorphic nature. Application of the DNA markers reveal extensive capability to distinguish among individuals, and this ability has been utilized in analyses of reproductive success, kinship and parentage (e.g. Queller *et al.* 1993; Marklund *et al.* 1994; Coltman *et al.* 1999). They have also proven highly advantageous for detection of genetic differentiation among populations (e.g. MacHugh *et al.* 1998; Cañon *et al.* 2000) and establishing phylogenetic relationship among populations or higher taxonomic groupings (e.g. Estoup *et al.* 1995; Hanslik *et al.* 2000).

Up till now, relatively little attention has been given to the possibility of using individual genotype information to determine the population membership of a single individual. However, as the development of DNA markers has expanded, the potential for recognizing population specific profiles is a topic receiving current interest (Blott *et al.* 1999; Roques *et al.* 1999). Potential for discrimination among individuals is essential for effective and accurate management of both natural populations and livestock breeds. Furthermore, a test for breed identity would be valuable for the validation of quality and origin of livestock products.

The degree of reproductive isolation and time since domestication of animals will influence the development of distinct breed characteristics. Cattle and sheep have a domestic history dating back 10 000–12 000 years, and clear differentiation is demonstrated among breeds of both sheep (Buchanan *et al.* 1994) and cattle (MacHugh *et al.*

Address for correspondence

Gro Bjørnstad, Department of Morphology, Genetics and Aquatic Biology, The Norwegian School of Veterinary Science, PO Box 8146 Dep., N-0033 Oslo, Norway.
E-mail: gro.bjornstad@veths.no

Accepted for publication 5 October 2000

1998; Blott *et al.* 1999). The horse was domesticated several thousand years later (cf. Bökönyi 1996) and thus there might be less demarcation among horse breeds as compared to species like cattle and sheep. Analysis of four horse breeds, each represented by five animals, indicated that DNA-fingerprinting has the potential to distinguish among horse breeds (Ellegren *et al.* 1992a). These results are supported by Cañon *et al.* (2000), who suggested that microsatellites could be useful for population assignment of horses, although the assignment success differed between breeds.

For investigation of the breed of origin, it is essential to quantify breed differentiation. Therefore, the first objective of the present study was to investigate the genetic demarcation of eight horse breeds, of which four are native to Norway, by microsatellite analysis. Secondly, the potential for breed allocation of individual animals was evaluated using both phylogenetic analysis and a method based on maximum likelihood estimates. The relative strength of these two methods was evaluated.

Materials and methods

Selection of animals and breeds

A total of 306 horses, representing eight of the most prevalent horse breeds of Norway, were included in the study. The panel of breeds covered the four native breeds (sample sizes are shown in parenthesis): Fjord Horse (40), Nordland/Lyngen Horse (30), Døle Horse (40) and Coldblooded Trotter (44). These indigenous breeds originated from different parts of the country; Fjord Horse in western counties, Nordland/Lyngen Horse in northern counties, Døle Horse in eastern counties and the Coldblooded Trotter was developed as a subline of Døle Horse during the last 100 years (http://www.fao.org/dad-is/). The four other breeds included in the study were Icelandic Horse (37), Shetland Pony (34), Standardbred (41) and Thoroughbred (40). Care was taken to sample individuals that were not closely related. Genotyping data for six of the breeds included in the present study have also been included in an earlier study (Bjørnstad *et al.* 2000a).

Microsatellites

Genotype information from a total of 26 microsatellites was used, including 10 loci incorporated in the PE Applied Biosystems StockMark system (Foster City, CA, USA) (Bozzini *et al.* 1996): *AHT4*, *AHT5* (Binns *et al.* 1995), *ASB2* (Breen *et al.* 1997), *HMS2*, *HMS6*, *HMS7* (Guérin *et al.* 1994), *HTG4*, *HTG6* (Ellegren *et al.* 1992b), *HTG7* (Marklund *et al.* 1994) and *VHL20* (van Haeringen *et al.* 1994). Two StockMark loci (*HMS3* and *HTG10*) were not included because of large deviations from Hardy–Weinberg equilibrium in four of the eight breeds. We also used the following 16 microsatellites: *ASB17* (Breen *et al.* 1997), *HTG14* (Marklund *et al.* 1994), *LEX20* (Coogle *et al.* 1996), *NVHEQ5*, *NVHEQ11*, *NVHEQ18* (Røed *et al.* 1997), *NVHEQ29*, *NVHEQ40*, *NVHEQ43*, *NVHEQ100* (Røed *et al.* 1998), *NVHEQ21*, *NVHEQ54*, *NVHEQ70*, *NVHEQ79*, *NVHEQ82* (Bjørnstad *et al.* 2000b) and *UCDEQ425* (Eggleston-Stott *et al.* 1997). A mean of 24·3 loci was genotyped for each horse (range 10–26). The genotypic data are available on request. Details of laboratory protocols, including DNA extraction and microsatellite genotyping, were described in Bjørnstad *et al.* (2000a,b).

Data analyses

Simple allele sharing statistics (Bowcock *et al.* 1994) were applied to investigate the genetic structure of the breeds. The approach was also applied to examine the potential for breed allocation of individual animals. An interindividual genetic distance matrix based on allele sharing statistics was generated using Microsat 1·5d (Minch *et al.* 1995). A neighbour-joining tree (Saitou & Nei 1987) was constructed from the genetic distance matrix using PHYLIP 3·57c (Felsenstein 1995) with individuals as operational taxonomic units. The phylogenetic tree was visualized using TreeView (Page 1996). The number of animals allocated to the correct breed cluster was used to assess the assignment success.

The phylogenetic procedure for resolving the breed of origin outlined above does not predict the certainty of a particular assignment. By using the allele frequency distributions of the multilocus genotypic data and implementing a maximum likelihood approach, breed allocation of individuals and the certainty of these allocations were estimated using WHICHRUN 3·2 (Banks & Eichert 2000). Incorporating jack-knife iterations, this procedure samples individuals one at a time and recalculates the allele frequency in the absence of each genotype before determining the most likely source population of the particular individual. To resolve the stringency of an allocation, WHICHRUN utilizes the log of the odds (LOD) ratio for the two most likely source populations. By restricting breed allocation to apply only for assignments that have a LOD ratio of at least two, a particular assignment will have a 1/100 chance of error or less.

The strength of the breed demarcation was investigated by increasing the number of implemented loci randomly one by one and examining the proportion of misallocated animals. Animals with missing genotype data for more than one locus were excluded from this analysis.

Results

Phylogenetic analysis

The neighbour-joining phylogenetic tree shows clear differentiation among the horse breeds (Fig. 1). Of the 306 individuals, only 30 (10%) associated with a breed other than the major cluster of source breed individuals. Icelandic Horse divided into two clades containing 22 and 15 individuals, respectively. Accepting that the Icelandic Horse constituted two clades implies that only 15 animals (5%) associated with another breed, with between zero and four animals in each breed incorrectly allocated. Fjord Horse had two misallocations, Nordland/Lyngen Horse one, Coldblooded Trotter two, Shetland Pony one, Standardbred four and Thoroughbred none. In addition, five misclassifications included assignment of Døle Horse to Coldblooded Trotter or vice versa.

Maximum likelihood estimates

The breed assignment pattern obtained using WHICHRUN is shown in Table 1. Of the 306 individuals, 295 (96%) were correctly allocated to their source population with a LOD ratio of at least two. Another seven individuals were allocated to the correct breed, but with lower stringency (LOD < 2). Only one horse was allocated to the wrong breed with high stringency, while three individuals were allocated to the wrong breed but with lower stringency than LOD ratio 2.

All breeds had high level of correct assignments, but the lowest number of correct allocations was seen in the Døle Horse. The close genetic relationship between Døle Horse and Coldblooded Trotter marginally complicates the separation between these breeds. When Coldblooded Trotter was removed from the analysis, all Døle horses were allocated correctly.

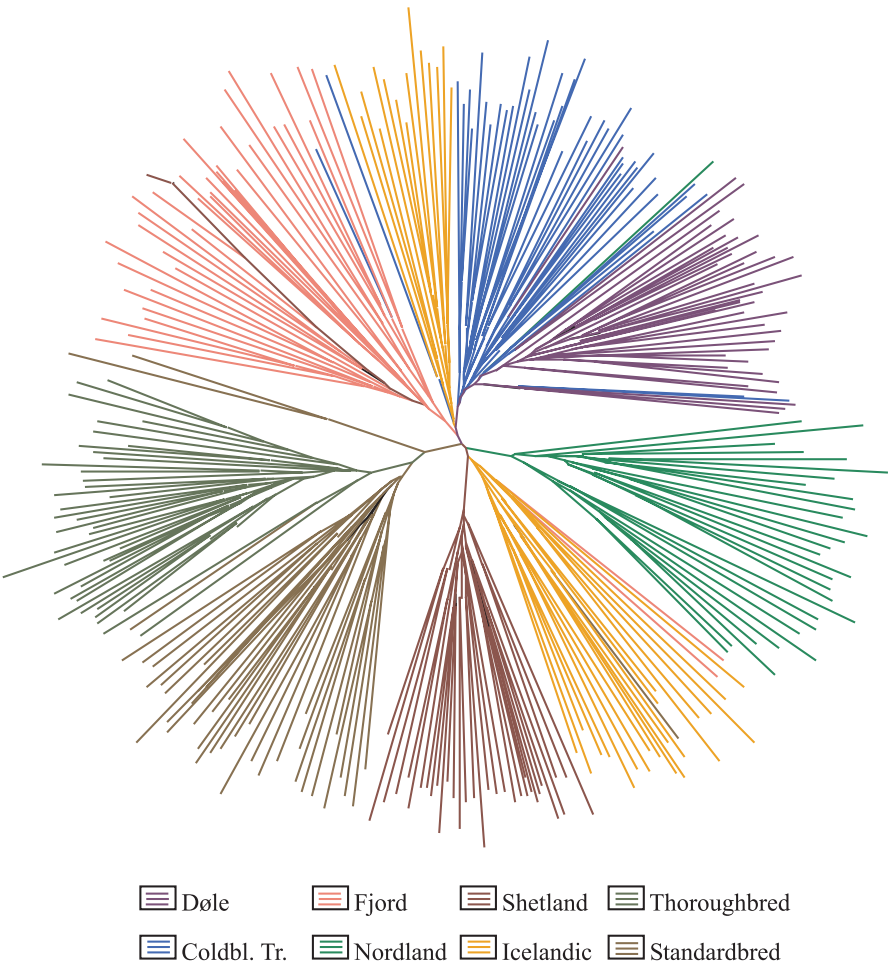

Figure 1 Radial presentation of the neighbour-joining dendrogram constructed from simple allele sharing statistics among 306 animals representing eight horse breeds.

Table 1 Breed assignments of individuals from eight horse breeds using genotype information of 26 microsatellite loci. For each breed, numbers in the first row represent individuals assigned a breed with a LOD ratio exceeding 2, while any numbers present in the second row represent animals assigned a breed with a LOD ratio less than 2.

Breed	*n*	Icelandic	Shetland	Nordland/ Lyngen	Fjord	Døle	Coldblooded Trotter	Standard-bred	Thorough-bred
Icelandic	37	35	–	–	–	–	–	–	–
		2							
Shetland	34	–	34	–	–	–	–	–	–
Nordland/Lyngen	30	–	–	29	–	–	–	–	–
						1			
Fjord	40	–	–	–	40	–	–	–	–
Døle	40	–	–	–	–	34	1	–	–
						3	2		
Coldblooded Trotter	44	–	–	–	–	–	42	–	–
							2		
Standardbred	41	–	–	–	–	–	–	41	–
Thoroughbred	40	–	–	–	–	–	–	–	40

Figure 2 illustrates the relationship between the number of implemented markers and the proportion of misallocated animals. Three allocation stringencies were applied. The highest stringency (LOD > 2) gave an error rate of about 25% when 10 random microsatellites were implemented, while 20 microsatellites were required to achieve an error rate close to 5%. The moderate stringency (LOD > 1) allocated 90% of the horses when applying 10 loci, while 20 loci allocated 96% of the horses. When the stringency was loosened further (LOD > 0) 10 loci allocated 95% of the horses, and 20 loci allocated 98% of the horses.

Discussion

Both phylogenetic analysis, utilizing simple allele sharing statistics, and analysis based on likelihood estimates have

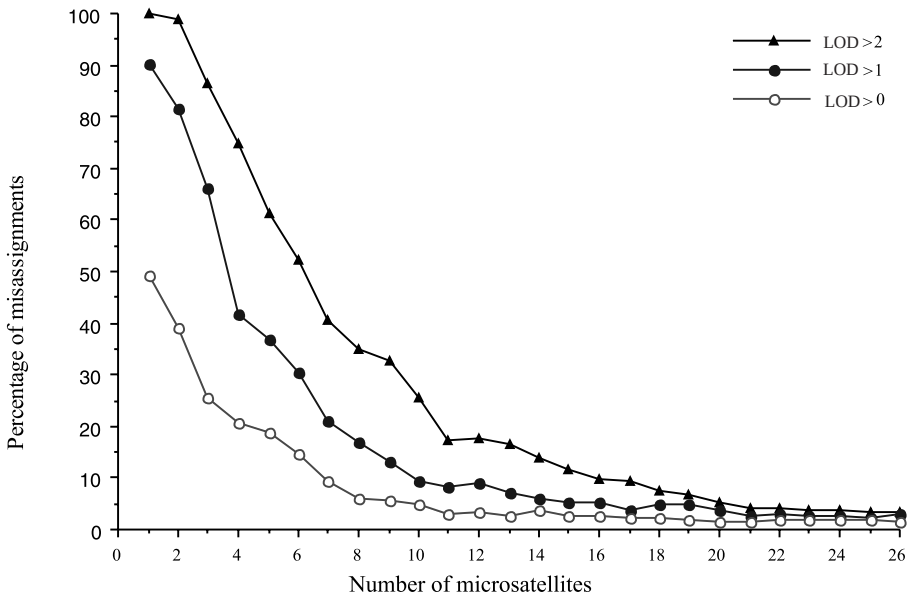

Figure 2 Percentage of misallocated animals using a varying number of microsatellite loci implemented in maximum likelihood analysis based on allele frequency distributions. Three LOD ratios were used to assess the strength of the allocation pattern. Number of implemented loci was randomly increased one by one in the following order *ASB2*, *NVHEQ18, HTG4*, *NVHEQ21*, *ASB17*, *NVHEQ79*, *NVHEQ29*, *AHT5*, *NVHEQ82*, *HMS7*, *HTG7*, *VHL20*, *NVHEQ54*, *HMS6*, *NVHEQ70*, *HMS2*, *NVHEQ40*, *AHT4*, *NVHEQ100*, *HTG6*, *NVHEQ11*, *LEX20*, *NVHEQ5*, *HTG14*, *UCDEQ425*, *NVHEQ43*.

the power to identify the source of human populations (Bowcock *et al.* 1994), fish stocks (Roques *et al.* 1999) and livestock breeds (Buchanan *et al.* 1994; MacHugh *et al.* 1998; Blott *et al.* 1999). The two approaches allocated a large proportion of the horses in the present study to their correct breed. The phylogenetic approach allocated 95% correctly and the analysis based on likelihood estimates correctly allocated 96% of the horses using the allocation stringency that the LOD ratio had to exceed 2. This high stringency means that the suggested source breed is more than 100 times more likely than the second most likely breed. Furthermore, investigating individuals with better than even odds of belonging to a particular breed allocated 302 of the 306 horses. Only one animal was misallocated with high stringency. A combination of the two statistical approaches could contribute to elucidate the population of origin of individual horses. Three individuals were either misallocated or allocated with low stringency (LOD > 2) by both methods. This gives an overall error rate of <1%, and the results suggest high probabilities for correct breed identification.

Breed protection in the form of herd book systems was established 100–200 years ago for Døle Horse, Fjord Horse, Standardbred and Thoroughbred, while herd books for the other breeds appeared at a later date (http://www.fao.org/dad-is/). Genetic exchange among breeds was abundant prior to herd book establishment. As an example, traditional trade routes between eastern and western counties introduced Døle Horse stallions into the Fjord Horse breeding pool and vice versa. Despite the formerly frequent cross-breeding, the present study shows very distinct demarcation among these horse breeds.

All Icelandic Horses clustered together with animals of their own breed, but the breed emerged in two distinct clusters. One cluster is situated between Shetland Pony and Nordland/Lyngen Horse, while the other cluster is located between Fjord Horse and Coldblooded Trotter. A mixed origin of the Icelandic Horse could explain the division of the breed. Horses originating from both the British Isles and Norway were introduced to Iceland during the Viking period a millennium ago (Adalsteinsson 1981), and similarity to both Shetland Pony and the Norwegian breeds is thus expected.

The Coldblooded Trotter was developed during the last 100 years or so, by breeding a subline representing light individuals of the Døle Horse. *F*-statistics suggest a differentiation between Døle Horse and Coldblooded Trotter of about 8% (Bjørnstad *et al.* 2000a). The demarcation observed between these two breeds was surprisingly clear, with only five animals of the one breed classified as the other in the phylogenetic tree. The herd books for these two breeds are still mutually open. However, the clear breed demarcation described here suggests that these breeds should be preserved as discrete genetic entities.

The remaining six breeds had between zero and four animals allocated to a breed cluster other than their source breed. All Thoroughbreds were consigned in one cluster. This homogeneity could reflect the limited origin of Thoroughbred consisting of Arabian horses and that the breed is relatively old, founded in the eighteenth century (cf. Bökönyi 1996). Three Standardbred animals associated with the cluster were formed by Thoroughbreds. A genetic similarity among these breeds is anticipated, as the formation of the Standardbred consisted of several breeds, including the Thoroughbred (cf. Bökönyi 1996).

The clustering pattern of the horse breeds in the present study is comparable with (or moderately stronger than) the pattern observed for cattle breeds (MacHugh *et al.* 1998), despite a longer domestic history for cattle. There has obviously been reasonable reproductive isolation between these horse populations during the relatively short period of breed development and/or prior to domestication, resulting in the discrete genetic structure for these breeds.

The number of implemented loci will influence the clustering pattern (Bowcock *et al.* 1994) and allocation success (e.g. Buchanan *et al.* 1994; MacHugh *et al.* 1998; Blott *et al.* 1999). A positive relationship between the number of analysed loci and allocation success was also found in the present study. We used three stringencies to assess the allocation pattern. Implementing the highest stringency (LOD > 2) required analysis of 20 loci to obtain an error rate of 5%. On the other hand, when implementing a better than even odds stringency for individuals to belong to a particular breed (LOD > 0), a large proportion of the horses were correctly allocated when applying considerably fewer markers. The curves showing the relationship between the proportion of misassignments and the number of implemented loci become flatter after reaching an error rate of about 5%. Extrapolating these curves indicate that a very high number of markers must be analysed to obtain an error rate of <1%. Thus, there seems to be relatively little to gain by analysing more than 10–20 loci, depending on the preferred allocation stringency. However, the selection of loci was random, and other markers could be better in detecting differences among breeds (Blott *et al.* 1999).

In a study of the genetic structure of Spanish Celtic horse breeds, about 90% of the horses were assigned correct breed using 13 microsatellites (Cañon *et al.* 2000). The weaker aggregation observed for the Spanish breeds was, however, not a function of the number of analysed loci as analysis of the similar number of markers still gave a stronger allocation pattern for the north-European breeds as shown in

Fig. 2. Lower breed differentiation, measured as *F*-statistics, was described among the Spanish breeds (Cañon *et al.* 2000) compared with the north-European breeds (Bjørnstad *et al.* 2000a; Bjørnstad *et al.* unpublished data). Thus, the genetic differentiation among breeds will influence the breed allocation potential.

In summary, a strong demarcation is observed among the north-European horse breeds. Evaluation of the assignment methods revealed that both phylogenetic analysis incorporating simple allele sharing statistics and maximum likelihood estimates based on the allele frequency distributions achieved high power in the breed allocation. Application of a combination of the two methods will assign more than 99% of the horses to the correct breed, and the procedure could significantly contribute to clarifying population background of individuals with unknown or suspect origin.

Acknowledgements

Ellen Gunby is acknowledged for laboratory assistance. We also thank Astri and Birger Torsteds Legacy for covering the printing costs of the colour presentation. This work was supported by the Norwegian Research Council by grant number 110113/112 and the Norwegian Trotter Association.

References

Adalsteinsson S. (1981) Origin and conservation of farm animal populations in Iceland. *Journal of Animal Breeding and Genetics* **98,** 258–64.

Banks M.A. & Eichert W. (2000) WHICHRUN (Version 3.2): a computer program for population assignment of individuals based on multilocus genotype data. *Journal of Heredity* **91,** 87–9.

Binns M.M., Holmes N.G., Holliman A. & Scott A.M. (1995) The identification of polymorphic microsatellite loci in the horse and their use in Thoroughbred parentage testing. *British Veterinary Journal* **151,** 9–15.

Bjørnstad G., Gunby E. & Røed K.H. (2000a) Genetic structure of Norwegian horse breeds. *Journal of Animal Breeding and Genetics* **117,** 307–17.

Bjørnstad G., Midthjell L. & Røed K.H. (2000b) Characterization of ten equine dinucleotide microsatellite loci: *NVHEQ21*, *NVHEQ54*, *NVHEQ67*, *NVHEQ70*, *NVHEQ75*, *NVHEQ77*, *NVHEQ79*, *NVHEQ81*, *NVHEQ82* and *NVHEQ83*. *Animal Genetics* **31,** 78–9.

Blott S.C., Williams J.L. & Haley C.S. (1999) Discriminating among cattle breeds using genetic markers. *Heredity* **82,** 613–9.

Bökönyi S. (1996) Horse. In: *A World Dictionary of Livestock Breeds, Types and Varieties* (Ed. by I.L. Mason) 4th edn, pp. 162–73. CAB International, Wallingford.

Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R. & Cavalli-Sforza L.L. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368,** 455–7.

Bozzini M., Fantin D., Ziegle J., van Haeringen H., Jacobs W., Ketchum M., Spencer M. & Bates S. (1996) Automated equine paternity testing. *Animal Genetics* **27**(Suppl. 2), 32.

Breen M., Lindgren G., Binns M.M., Norman J., Irvin Z., Bell K., Sandberg K. & Ellegren H. (1997) Genetical and physical assignments of equine microsatellites – first integration of anchored markers in horse genome mapping. *Mammalian Genome* **8,** 267–73.

Buchanan F.C., Adams L.J., Littlejohn R.P., Maddox J.F. & Crawford A.M. (1994) Determination of evolutionary relationships among sheep breeds using microsatellites. *Genomics* **22,** 397–403.

Cañon J., Checa M.L., Carleos C., Vega-Pla J.L., Vallejo M. & Dunner S. (2000) The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. *Animal Genetics* **31,** 39–48.

Coltman D.W., Bancroft D.R., Robertson A., Smith J.A., Clutton-Brock T.H. & Pemberton J.M. (1999) Male reproductive success in a promiscuous mammal: behavioural estimates compared with genetic paternity. *Molecular Ecology* **8,** 1199–209.

Coogle L., Reid R. & Bailey E. (1996) Equine dinucleotide repeat loci *LEX015-LEX024*. *Animal Genetics* **27,** 217–8.

Eggleston-Stott M.L., DelValle A., Bautista M., Dileanis S., Wictum E. & Bowling A.T. (1997) Nine equine dinucleotide repeats at microsatellite loci *UCDEQ136*, *UCDEQ405*, *UCDEQ412*, *UCDEQ425*, *UCDEQ437*, *UCDEQ467*, *UCDEQ487*, *UCDEQ502* and *UCDEQ505*. *Animal Genetics* **28,** 370–1.

Ellegren H., Andersson L., Johansson M. & Sandberg K. (1992a) DNA fingerprinting in horses using a simple $(TG)_n$ probe and its application to population comparisons. *Animal Genetics* **23,** 1–9.

Ellegren H., Johansson M., Sandberg K. & Andersson L. (1992b) Cloning of highly polymorphic microsatellites in the horse. *Animal Genetics* **23,** 133–42.

Estoup A., Garnery L., Solignac M. & 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,** 679–95.

Felsenstein J. (1995) *PHYLIP (Phylogeny Inference Package), Version 3.57c*. Department of Genetics, University of Washington, Seattle, USA.

Guérin G., Bertaud M. & Amigues Y. (1994) Characterization of seven new horse microsatellites: *HMS1*, *HMS2*, *HMS3*, *HMS5*, *HMS6*, *HMS7* and *HMS8*. *Animal Genetics* **25,** 62.

van Haeringen H., Bowling A.T., Stott M.L., Lenstra J.A. & Zwaagstra K.A. (1994) A highly polymorphic horse microsatellite locus: *VHL20*. *Animal Genetics* **25,** 207.

Hanslik S., Harr B., Brem G. & Schlötterer C. (2000) Microsatellite analysis reveal substantial genetic differentiation between contemporary New World and Old World Holstein Friesian populations. *Animal Genetics* **31,** 31–8.

MacHugh D.E., Loftus R.T., Cunningham P. & Bradley D.G. (1998) Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Animal Genetics* **29,** 333–40.

Marklund S., Ellegren H., Eriksson S., Sandberg K. & Andersson L. (1994) Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Animal Genetics* **25,** 19–23.

Minch E., Ruiz-Linares A., Goldstein D.B., Feldman M.W. & Cavalli-Sforza L.L. (1995) *Microsat (Version 1.4d): a Computer Program for Calculating Various Statistics on Microsatellite Allele Data.* http://human.stanford.edu/microsat/microsat.html.

Page R.D.M. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12,** 357–8.

Queller D.C., Strassmann J.E. & Hughes C.R. (1993) Microsatellites and kinship. *Trends in Ecology and Evolution* **8,** 285–8.

Røed K.H., Midthjell L., Bjørnstad G. & Olsaker I. (1997) Equine dinucleotide repeat microsatellites at the *NVHEQ5*, *NVHEQ7*, *NVHEQ11*, *NVHEQ18* and *NVHEQ24* loci. *Animal Genetics* **28,** 381–2.

Røed K.H., Midthjell L. & Bjørnstad G. (1998) Eight new equine dinucleotide repeat microsatellites at the *NVHEQ26*, *NVHEQ29*, *NVHEQ31*, *NVHEQ40*, *NVHEQ43*, *NVHEQ90*, *NVHEQ98* and *NVHEQ100* loci. *Animal Genetics* **29,** 470.

Roques S., Duchesne P. & Bernatchez L. (1999) Potential of microsatellites for individual assignment: the North Atlantic redfish (genus *Sebastes*) species complex as a case study. *Molecular Ecology* **8,** 1703–17.

Saitou N. & Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4,** 406–25.