

# JOURNAL OF ANIMAL SCIENCE

*The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science*

## **Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis**

C. Maudet, G. Luikart and P. Taberlet

*J Anim Sci* 2002. 80:942-950.

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://jas.fass.org>



**American Society of Animal Science**

[www.asas.org](http://www.asas.org)

# Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis<sup>1</sup>

C. Maudet<sup>2</sup>, G. Luikart, and P. Taberlet

Laboratoire de Biologie des Populations d'Altitude, F-38041 Grenoble, France

**ABSTRACT:** Genetic variability and relationships among six native French cattle breeds (Abondance, Tarentaise, Villard de Lans, Montbéliarde, Limousin, and Charolais) and one foreign breed (Holstein) were investigated using 23 microsatellite markers. These breeds were also compared with four Swiss breeds genotyped in a previously published study. Interestingly, the French alpine breeds have smaller population sizes but showed higher genetic variability than the larger Holstein breed. Neighbor-joining trees and PCA (principal components analysis) showed that alpine breeds tend to cluster together. Abondance and Tarentaise breeds were closely related, whereas the Holstein was highly differentiated from all breeds analyzed. Two different

assignment tests for determining the breed of origin of individuals were compared: "direct" and "exclusion-simulation" approaches. The exclusion-simulation significance test correctly assigns fewer individuals than the direct approach but provides a confidence level (e.g.,  $P < 0.01$ ) for each individual being assigned. Accurate assignment with high statistical confidence is required for animal traceability. Unfortunately, the accuracy of assignment greatly decreases as the threshold level of confidence of assignment increases (e.g., from  $P < 0.05$  to  $P < 0.001$ ). Assignment accuracy also greatly declines as the level of population differentiation decreases below the level often found between related breeds (e.g.,  $F_{ST} < 0.1$ ).

Key Words: Genetic Variation

©2002 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2002. 80:942–950

## Introduction

Livestock breeds have been formed by centuries of human and natural selection. Breeds have been selected to fit a wide range of environmental conditions and human needs. In the last 200 yr, herd book union restraints have led to genetic isolation of many cattle breeds. The selection of a few highly productive breeds has caused the decline of numerous other breeds. The genetic diversity found in domestic breeds allows farmers to develop new characteristics in response to changes in environment, diseases, or market condi-

tions. Indigenous breeds often possess gene combinations and special adaptations (such as disease resistance, adaptation to harsh conditions or poor-quality feeds, etc.) not found in other breeds. In the French Alps, several hardy cattle breeds have been selected for many centuries. Some of these mountain breeds are already extinct.

This study was conducted to determine the levels of genetic variation and relationships among seven French cattle breeds including four alpine breeds. We also compared the four alpine French cattle breeds with four alpine Swiss breeds (Original Brown Swiss, Hérens, Holstein, and Simmental) studied by Schmid et al. (1999). The impact of the geographic proximity on genetic differentiation was considered.

The usefulness of microsatellite markers for the estimation of genetic diversity and relationships among livestock breeds has been documented in numerous studies (e.g., Buchanan et al., 1994; Saitbekova et al., 1999; Schmid et al., 1999). The objectives of this study were 1) to quantify and compare levels of genetic variability within seven French breeds representing different uses and population sizes, 2) to determine the efficiency of microsatellite markers for estimating the relationships among breeds that are historically close geographically, and 3) to compare the performance of two assignment test methods to determine the breed of

<sup>1</sup>This work was supported by "Région Rhône-Alpes" in the Biotechnology Program, the CNRS, and the University Joseph Fourier. The authors wish to thank all biologists and veterinarians who helped collect samples, especially D. Gauthier and L. Giuliani from the "Laboratoire Départemental d'Analyses Vétérinaire de Savoie" and "Laboratoire Départemental d'Analyses Vétérinaire du Conseil Général" de l'Isère, and E. Maussi from the LIDAL. We also thank C. Gaillard and M. Schmid from the Institute of Animal Breeding, Berne, for sharing data.

<sup>2</sup>Correspondence: CNRS UMR5553, Université Joseph Fourier, BP53, 38041 Grenoble cedex 9, France (phone: 33 4 76 51 46 00; fax: 33 4 76 51 42 79; e-mail: cmaudet@ujf-grenoble.fr).

Received April 27, 2001.

Accepted November 8, 2001.

origin of individuals (or tissues) with a high certainty when using microsatellite genotype data.

## Materials and Methods

### Sampling

The Abondance and the Tarentaise are medium-sized dairy breeds (respectively 50,000 and 15,000 individuals) from the high valleys in the northern French Alps. They are known for their hardiness and their adaptation to mountain conditions. The Villard de Lans breed occurs in the western Alps and is considered as an endangered breed by the FAO (fewer than 500 individuals). The Montbéliarde is a widely represented dairy breed from the Jura Mountains (north of the Alps). We compared these mountain dairy breeds with three widely distributed French breeds ( $1 \times 10^6$  cows). Two are essentially beef breeds: the Limousin originated in the western of Massif Central (central France) and the Charolais originated in the northern Jura (Eastern France). The French Holstein was originated in northwestern Europe (The Netherlands) and was introduced into France as early as the 19th century. Currently, this breed represents 65% of the cattle in France. A total of 317 animals representing six French native breeds and one foreign breed were analyzed. Abondance (55), Charolais (32), Limousin (32), Montbéliarde (51), French Holstein (51), Tarentaise (50), and Villard de Lans (32) blood or semen samples were collected from French herds. Samples were collected from one or two unrelated animals (without common parents known) per herd. Especially for mountain breeds, herds were chosen in the entire area of repartition of the breed. In this way, we prevented the sampling of directly related animals and ensured an inbreeding coefficient of the analyzed samples comparable with the breed coefficient. Genomic DNA was extracted using the DNeasy Blood Kit (QIAGEN, Germany) following the manufacturer's procedure.

### Microsatellite Amplifications and Analysis

Twenty-three microsatellite markers were analyzed and correspond to the loci currently used in a large European project focusing on genetic diversity in European cattle breeds ([http://www.ri.bbsrc.ac.uk/cdiv\\_www/](http://www.ri.bbsrc.ac.uk/cdiv_www/)). Primer sequences and relevant references are available on the European project Web site. The PCR amplification and multiplexing conditions are given in Table 1. The 23 microsatellites were amplified alone or in multiplexes (two or three co-amplified loci) in 12 independent PCR reactions. All the PCR amplifications were performed in a 25- $\mu$ L reaction containing 2  $\mu$ L of the extracted DNA, 10 mM-Tris-HCl (pH 8.3), 50 mM-KCl, 1.5 mM-MgCl<sub>2</sub>, 25  $\mu$ M of each dNTP, 5 ng of BSA, 1 U of AmpliTaq Gold polymerase (PE Applied Biosystems, Foster City, CA), and 0.1 to 0.5  $\mu$ M of primers. The amplification was carried out

in a thermocycler (GeneAmp PCR System 9700, PE Applied Biosystems) using the following conditions: an initial denaturation step at 95°C for 10 min followed by 35 cycles of 95°C for 30 s, 58, 55, 50 or 45°C for 30 s, and 72°C for 60 s. Amplified microsatellites were mixed (according to their size and their fluorescent dye group: 6-FAM, HEX, or NED) in three loading multiplexes and were separated in a 6% denaturing gel using an ABI PRISM 377 automated sequencer (PE Applied Biosystems) following the manufacturer's procedures. All gels were analyzed using GENESCAN 2.0 and GENOTYPER 2.0 software.

### Genetic Diversity Analysis

Exact tests for deviations from Hardy-Weinberg equilibrium (**HWE**) and heterozygote deficiency were performed using the GENEPOP package (Raymond and Rousset, 1995). The program performed a probability test using a Markov chain (dememorization 5,000, batches 100, iterations per batch 1,000). Significant levels were calculated per locus, per population, and over all loci and populations combined. Genotypic linkage disequilibrium was estimated between all locus pairs with GENEPOP (Markov chain using dememorization 5,000, batches 100, iterations per batch 500). Genetic diversity within population was measured as the mean number of alleles (**NA**) per locus, the number of private alleles (**PA**, alleles found in only one breed), the observed heterozygosity (**H<sub>o</sub>**), and the expected heterozygosity (**H<sub>e</sub>**) under HWE that were calculated using the GENETIX software package (available at <http://www.univ-montp2.fr/~genetix/genetix.htm>). Significant differences in the number of alleles and the expected heterozygosity between two breeds were tested for using Wilcoxon's signed ranks test. The significance of breed differences was tested using the exact test of population differentiations in GENEPOP software based on allele frequencies (dememorization 5,000, batches 100, iterations per batch 1,000). Genetic differentiation between breeds was also estimated using the  $F_{ST}$  of Wright (Nei, 1987) computed by GENEPOP.

Principal components analysis (**PCA**) was performed according to the procedures described by Cavalli-Sforza et al. (1994) from allele frequencies for each population. A PCA was carried out from the seven French breeds and the 23 loci. Allele frequencies from 20 mutual loci from four Swiss breeds considered by Schmid et al. (1999) (Original Brown Swiss, Hérens, Swiss Holstein, and Simmental) were used to perform another PCA. We calculated two genetic distances from allele frequencies using GENETIX software. Nei et al. (1983) recommended the modified Cavalli-Sforza distance (**D<sub>A</sub>**) for use with closely related populations where genetic drift is the primary factor of evolutionary differentiation. However, standard genetic distance of Nei (1972) (**D<sub>S</sub>**), the more frequently used distance, was calculated to compare with other stud-

**Table 1.** Chromosomal location and amplification parameters for 23 microsatellite loci; exponent letters indicate multiplex PCR and exponent numbers indicate loci multiplex loaded together in one well

Locus	Annealing temp., °C	Primer conc., $\mu\text{M}$	Chromosome
TGLA227 <sup>1a</sup>	50	0.1	18
ILSTS006 <sup>1a</sup>	50	0.5	7
CSRM60 <sup>1b</sup>	55	0.1	10
BM2113 <sup>1b</sup>	55	0.4	2
HEL13 <sup>1c</sup>	55	0.4	11
TGLA122 <sup>1d</sup>	55	0.5	21
ETH152 <sup>1d</sup>	55	0.2	5
ILSTS005 <sup>2e</sup>	55	0.2	10
BM1818 <sup>2e</sup>	55	0.5	23
ETH3 <sup>2f</sup>	58	0.2	19
HAUT24 <sup>2g</sup>	50	0.4	22
INRA005 <sup>2g</sup>	50	0.5	12
HEL5 <sup>2g</sup>	50	0.4	21
CSSM66 <sup>2h</sup>	45	0.1	14
HAUT27 <sup>2h</sup>	45	0.4	26
INRA037 <sup>3i</sup>	55	0.2	11
ETH10 <sup>3i</sup>	55	0.4	5
BM1824 <sup>3j</sup>	50	0.2	1
INRA063 <sup>3j</sup>	50	0.3	18
TGLA126 <sup>3k</sup>	50	0.5	20
ETH225 <sup>3k</sup>	50	0.1	9
MM12 <sup>3l</sup>	50	0.2	7
HEL9 <sup>3l</sup>	50	0.4	8

ies. The neighbor-joining tree topology was obtained with the PHYLIP software (<http://evolution.genetics.washington.edu/phylip.html>) using the Cavalli-Sforza distance. Bootstrap values were computed over 1,000 replicates. To test for correlations between genetic and geographical distances (isolation by distance) a Mantel test was performed using the R Package 4.0 software. The center of origin of each breed was used as the geographical localization of the breeds and the Cavalli-Sforza distance ( $D_A$ ) was used as the genetic distance.

### Population Assignments

Several studies have shown that microsatellites can be used to identify the population of origin of an individual (e.g., Paetkau et al., 1995; Rannala and Mountain, 1997; Cornuet et al., 1999). Numerous approaches have recently been proposed for identifying the origin of individuals using molecular makers (e.g., Paetkau et al., 1995; Cornuet et al., 1996; Banks and Eichert, 2000). We compare two approaches using observed population allele frequencies to assign individuals to a breed: 1) a “direct” method similar to the one commonly used in livestock studies (Buchanan et al., 1994) and 2) a simulation-based method that provides a level of certainty ( $P$ -value) for each animal assigned (Cornuet et al., 1999).

Buchanan et al. (1994), MacHugh et al. (1998), and Diez-Tascon et al. (2000) using Buchanan et al.’s (1994) method have shown the accuracy of assignment tests for domestic breed assignment. Buchanan et al.’s

(1994) method consisted of simulating 1,000 multilocus genotypes for each breed using the allele frequencies observed in each breed sample (and assuming HWE and linkage equilibrium). The probability that a simulated genotype originated from a breed was calculated as the expected frequency of the genotype in the breed. In the same way, the frequency method (first presented by Jamieson [1965] using protein data and applied to microsatellite data by Paetkau et al. [1995]), also directly assigns an individual to the population in which the individual’s genotype is most likely to occur (the “direct” approach). But, unlike Buchanan et al.’s (1994) method, the method of Paetkau et al. (1995) does not simulate individuals from allele frequencies, but removes one real individual from the actual sample, recalculates allelic frequencies in each population, and assigns the real individual to the population most probable (in which the expected genotype frequency is highest assuming HWE and linkage equilibrium).

Two limitations of the previously used assignment methods (Buchanan et al., 1994; Paetkau et al., 1995) are that 1) they do not provide a  $P$ -value for measuring the confidence that the individual truly belongs to a given population and 2) they always designate a single population as the probable source of an individual (if the true population of origin of the individual is not represented in the set of reference populations, these methods could designate a wrong population of origin; Cornuet et al., 1999). Cornuet et al. (1999) suggested a new population assignment approach (simulation-exclusion approach) that calculates for each real indi-

vidual the probability originating from each sampled population.

The simulation-exclusion approach proposed by Cornuet et al. (1999) computes a probability that the individual belongs to a population by simulating 10,000 genotypes (using the sample allele frequencies from a population) and then by calculating the probability that the removed individual belongs to the population. For example, if an individual's genotype was observed one time in 10,000 simulated genotypes, the probability that the individual belongs to this population is  $P = 0.001$ . This approach allows the exclusion of populations as origins of individuals. A threshold  $P$ -value is fixed according to the required certainty of exclusion (generally between 0.05 and 0.001). If the individual's probability is lower than the decided threshold in a population, it considers that the individual does not originate from the population. Thus, an individual was considered as correctly assigned to a population when it was excluded from all of the non-origin population (e.g.,  $P < 0.001$ ), but not from the true population of origin.

The assignment tests were carried out using the GENECLASS software program (available at <http://www.ensam.inra.fr/URLB>). We used the "Frequency" option (and not the "Bayesian" option) to compute the population allele frequencies, because it is more similar to the Buchanan et al. (1994) method, which employs the allele frequencies observed in the population sample to calculate assignment probabilities. Direct and exclusion-simulation tests were carried out among all seven breeds (mean  $F_{ST} = 0.08$ ), between the most differentiated pair of breeds (Abondance and Holstein: pairwise  $F_{ST} = 0.112$ ), and between the least differentiated pair of breeds (Montbéliarde and Villard de Lans: pairwise  $F_{ST} = 0.043$ ).

## Results

### Microsatellite Markers

In total, 215 alleles were observed from the 23 loci surveyed. The number of alleles per locus ranged from 3 (ILSTS005) to 19 (TGLA122) with a mean of 9.4. Expected heterozygosity across all the breeds varied between 0.40 (ILSTS005) and 0.82 (TGLA227). No locus had a significant ( $P < 0.05$ ) deviation from HWE (in the same direction) in more than three populations. Significant linkage disequilibrium ( $P < 0.0001$ ) was found between CSRM60 and INRA037 in six of the seven breeds. Gametic disequilibrium can be due to a variety of factors, including physical linkage, epistatic selection, and genetic hitchhiking. Because CSRM60 and INRA037 loci have been mapped to different chromosomes (chromosomes 10 and 11), physical linkage was excluded. All the following analyses have been performed both with and without the CSRM60 locus but no difference was observed in the results.

### Genetic Variability

The mean number of alleles (NA) per breed varied from  $5.61 \pm 0.9$  in the Villard de Lans breed to  $6.52 \pm 0.9$  in the Abondance breed (Table 2). Mean number of alleles per locus (NA) differences observed between breeds might be explained by the variations of the studied sample sizes (about 30 individuals for Villard de Lans, Charolais, and Limousin and about 50 individuals for Abondance, Tarentaise, Montbéliarde, and Holstein breeds). However, the Abondance and the Tarentaise mean number of alleles per locus were significantly higher than Holstein NA ( $P < 0.05$ ). The expected heterozygosity ( $H_e$ ) ranged from  $0.65 \pm 0.12$  in the Abondance breed to  $0.70 \pm 0.11$  in the Tarentaise breed, but no significant differences were found. Consistent with Moazami-Goudarzi et al. (1997), we found a relatively low genetic and allelic variability in the French Holstein. The exact test for Hardy-Weinberg disequilibrium within breeds showed a significant deviation in the French Holstein and the Tarentaise breeds ( $P < 0.0001$ ). These two breeds present, respectively, five and eight loci with a significant heterozygote deficit ( $P < 0.01$ ) but no loci with significant heterozygote excess. If we remove the most deviant locus for each of the two breeds (CSSRM66 and TGLA126, respectively), the heterozygote deficit remains significant for the two breeds ( $P < 0.03$ ). Interestingly, in the case of the Holstein breed, the samples came from two different origins: approximately half the individuals (26) were sampled on local farms and the other half (25) are from French bulls widely used for insemination (including some originating from North America). The deficit of heterozygotes was found only in the insemination bull population ( $P < 0.0001$ ), whereas no departure from HWE was observed in the local farm population ( $P > 0.1$ ). The exact test for population differentiation based on allele frequency variations showed that all breeds tested were significantly different from each other ( $P < 0.0001$ ).

### Breed Relationships

Principal components analysis (PCA) was performed including all populations and loci using allele frequencies to summarize breed relationships (Figure 1). A total of 62% of the variance accounted for the first three dimensions of the PCA. The analysis indicated a grouping of the four alpine breeds (Abondance, Tarentaise, Villard de Lans, and Montbéliarde). The other breeds formed two groups: the Charolais and the Limousin breeds (meat breeds) and the French Holstein breed. The same grouping was also obtained using a principal coordinate analysis (PCoA) using the  $D_A$  distance (data not shown).

Allele frequencies were used to generate the genetic distance Cavalli-Sforza distance ( $D_A$ ) and Nei distance ( $D_S$ ) for each pair of breeds. The  $D_A$  genetic distance ranged from 0.102 to 0.208, and the  $D_S$  distance ranged

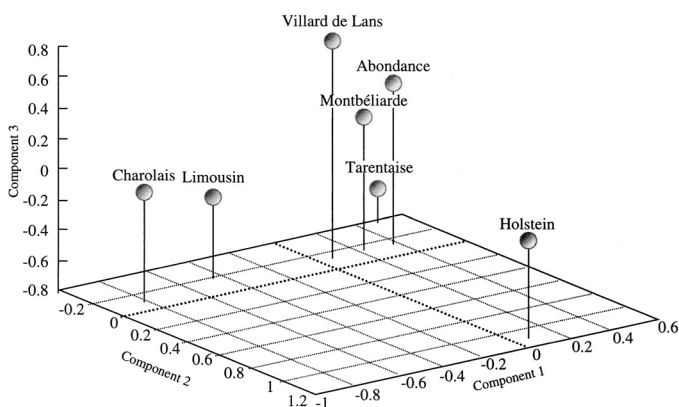
**Table 2.** Mean observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity, mean number of alleles per locus (NA), and total number of private alleles (PA) observed from the 23 microsatellites in seven French breeds

Breed	Mean heterozygosity			PA
	$H_o$	$H_e$	NA	
Abondance	0.643	0.650	6.52	5
Charolais	0.640	0.661	6.00	6
Limousin	0.674	0.675	5.78	3
Montbéliarde	0.675	0.670	6.39	3
Holstein	0.669	0.686	5.83	9
Tarentaise	0.685	0.699	6.48	6
Villard de Lans	0.693	0.676	5.61	2

from 0.125 to 0.325 (Table 3). As expected, the most divergent breed was the Holstein breed. Consistent with results obtained from the PCA and PCoA, the neighbor-joining tree constructed from  $D_A$  distance showed three groups (Figure 2). The two most robust features of the topology were the Abondance/Tarentaise and the Charolais/Limousin grouping (in 70 and 76% of the bootstrap trees). These groupings are consistent with previous genetic and biochemical studies of French cattle breeds (Moazami-Goudarzi et al., 1997; Grosclaude et al., 1990).

A PCA including the Swiss cattle breeds (Schmid et al., 1999) was also carried out (data not shown). Only 44% of the variance was accounted for in the first three components, and five dimensions were necessary to account for 65% of the information. Three groups were observed. The first one included the four strictly alpine dairy breeds (Tarentaise, Abondance, Original Swiss Brown, and Hérens breeds), the Montbéliarde, and the Swiss Simmental. The second one included the Villard de Lans, Limousin, and Charolais breeds. The last one contained the two Holstein breeds, the French and the Swiss Holstein.

Using the geographic center of origin of each breed as geographical localization, a significant correlation

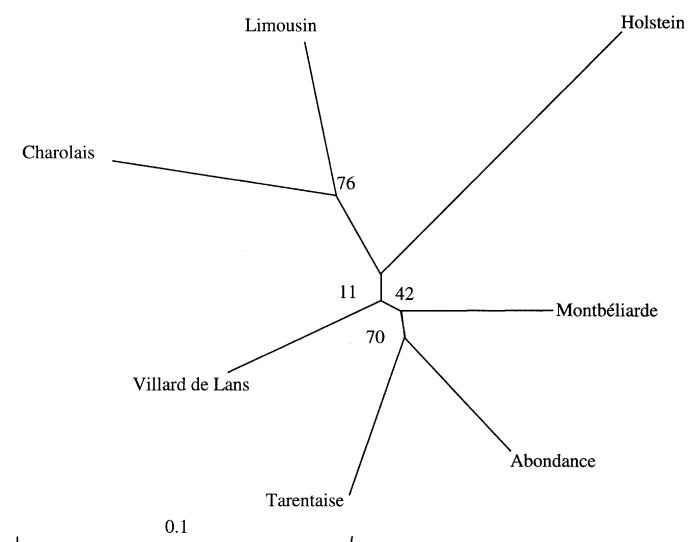


**Figure 1.** Principal components analysis based on allele frequencies of 23 loci showing the first three principal components that accounted for 62% of the variance (respectively 25.5, 21.4, and 15.1% of the total variance).

between genetic and geographical distance (isolation by distance) was found in the French breeds ( $r = 0.70$ ,  $P = 0.012$ ). However, the Holstein breed is the most genetically and geographically distant of the breeds and is the major source of the correlation. Removing the Holstein breed, the correlation was not significant ( $r = 0.47$ ,  $P = 0.067$ ). In the same way, including the Swiss cattle, a positive correlation was found between genetic and geographic distance ( $r = 0.62$ ,  $P = 0.04$ ), but when the Holstein breeds were removed no correlation was found.

#### Breed Assignment

Results on breed assignment are shown in Table 4. The direct method of Paetkau et al. (1995) allows the correct assignment of 93 to 100% of the real individuals to their breed of origin (using the 23 loci and seven breeds). The exclusion-simulation significance test of Cornuet et al. (1999) correctly assigned 67% of individ-



**Figure 2.** Phylogenetic tree constructed from  $D_A$  by the neighbor-joining method showing genetic relationships among seven cattle breeds. Numbers represent the percentage of times that a node occurred in 10,000 bootstraps replicates.

**Table 3.** Genetic distances between cow breeds; above the diagonal is Nei's distance (Nei, 1972)  $D_S$  and below is the modified Cavalli-Sforza distance, Nei et al. (1983)  $D_A$  computed using allele frequencies of the 23 microsatellite makers

Breed	Abondance	Charolais	Limousin	Montbéliarde	Holstein	Tarentaise	Villard
Abondance	—	0.272	0.209	0.134	0.293	0.125	0.113
Charolais	0.186	—	0.138	0.251	0.315	0.280	0.228
Limousin	0.155	0.127	—	0.199	0.307	0.176	0.186
Montbéliarde	0.106	0.171	0.134	—	0.259	0.136	0.099
Holstein	0.172	0.186	0.208	0.168	—	0.325	0.199
Tarentaise	0.102	0.181	0.141	0.105	0.196	—	0.149
Villard	0.108	0.154	0.149	0.105	0.161	0.129	—

uals when using a threshold  $P$ -value of 0.05 for each individual assignment. The performance of this method is strongly influenced by the chosen level of confidence ( $P$ -value). For example, when the  $P$ -value was reduced to 0.001, the percentage of correctly assigned individuals dropped to only 33%. Individuals misassigned with the exclusion-simulation method might be due to the following: 1) more than one population was not excluded as the origin (because the  $P$ -value for more than one population is greater than the chosen threshold) or 2) the  $P$ -value for all the populations was smaller than the threshold (the genotype was improbable in all the sampled populations and thus all populations were excluded). Within the two most genetically differentiated breeds (Abondance and Holstein) about 93% of the individuals could be correctly assigned with a very high certainty ( $P < 0.001$ ), whereas only 41% accuracy was achieved within the two least differentiated breeds (Montbéliarde and Villard de Lans).

## Discussion

### Breed Variability

One interesting result was the heterozygote deficit observed in the Holstein sample of bulls. In domestic species, heterozygote deficiencies can be explained by several factors such as the presence of unamplified alleles ("null" alleles), selection against heterozygotes, population subdivision (Wahlund's effects), or inbreeding. The majority of the loci used for this study have already been analyzed in many European cattle breeds (e.g., Moazami-Goudarzi et al., 1997; MacHugh et al., 1998; Schmid et al., 1999), and no loci were found to often deviate from Hardy-Weinberg proportions. Similarly, only a few breeds studied using these loci have shown a significant heterozygote deficit (Martin-Burriel et al., 1999; Schmid et al., 1999). Thus, the hypothesis of null alleles or selection against heterozygotes is improbable. Hanslik et al. (2000)

**Table 4.** Breed assignment test results, showing the percentage of individuals correctly assigned using the "direct" approach (without  $P$ -value) or the "exclusion-simulation" approach using 23 loci and all seven breeds (mean  $F_{ST} = 0.08$ ) or only two with different  $F_{ST}$

Item	Percentage of individuals correctly assigned			
	Direct	Exclusion-simulation (reject population probability) <sup>a</sup>		
		$P < 0.05$	$P < 0.01$	$P < 0.001$
Assignment among the seven breeds				
Abondance	92.7	52.7	34.5	10.9
Charolais	100	68.8	62.5	59.4
Limousin	96.9	68.8	53.1	21.9
Montbéliarde	93.9	65.2	43.9	18.2
Holstein	100	90.2	86.3	78.4
Tarentaise	96.0	72.0	68.0	38.0
Villard de Lans	96.9	51.6	28.1	6.5
Total	96.2	67.5	54.3	33.1
Assignment between two breeds				
Abondance-Holstein	100	95.3	93.4	92.5
Montbéliarde-Villard de Lans	95.1	70.7	68.3	40.2

<sup>a</sup>With the "exclusion-simulation" approach an individual is considered as correctly assigned to a population when it was excluded from all but the correct population (probability of belonging to a population under the threshold  $P < 0.05$ , 0.01, or 0.001).

showed that the Holstein breed is highly structured between the Old World and the New World populations. However, a substantial introgression from the New World Holstein into the European Holstein has been reported (Hanslik et al., 2000). Currently, in France all the Holstein inseminations are carried out using artificial insemination (AI) ("BRG: Bureau des Ressources Génétiques" data available at <http://www.brg.prd.fr/>). Five percent of the semen used in France comes from the United States and 5% from Canada. Thus, the heterozygote deficit observed in the French Holstein bulls could be explained by the subdivision of the breed at the intercontinental scale and by the recent use of New World semen for French Holstein reproduction. Moreover, the low number of alleles and the low heterozygosity observed in the Holstein could be explained by the intensive human selection accelerated by AI and a reduced number of reproducers (about 1 male for 1,500 females in French Holstein; in Tarentaise the ratio is 1 male for 35 females [BRG data]). Boichard et al. (1996) estimated that the French Holstein bulls used for AI show an inbreeding coefficient of 3% (this coefficient is about 1% for many dairy breeds).

In the same way, a significant heterozygote deficit was found in the Tarentaise breed. The Tarentaise is a medium-sized breed (currently 35,000 animals) located in mountainous areas in a restricted region. Unlike most French cow breeds, the Tarentaise Herd-Book union has never organized admixture from other breeds for the improvement of the breed (UPRA Tarentaise, personal communication). The heterozygote deficiency observed in the Tarentaise breed might be explained by inbreeding due to a small number of reproducers, genetic drift, and (or) structuration. Boichard et al. (1996), studying the pedigrees, indicated that the Tarentaise breed shows an inbreeding coefficient superior to that of many French dairy breeds. Nevertheless, the allele numbers are usually reduced faster than the heterozygosity during inbreeding or a bottleneck period (Nei et al., 1975). Thus, small inbred populations should exhibit reduced allele numbers. However, the Tarentaise breed has a high number of alleles (significantly higher than the Holstein breed). This contradicts the bottleneck hypotheses. The Tarentaise breed was officially created in the 1890s and descends from an old alpine cattle group, more precisely from two (or three) morphologically distinguishable groups of alpine cows (UPRA Tarentaise, personal communication). The lack of interconnections among alpine regions during the 19th century could lead to a fast genetic differentiation of the cows on a small scale (e.g., different valleys) due to drift. The combining of the two cow types should lead to a novel breed with an increased number of alleles. Thus, a mixed origin of the Tarentaise followed by inbreeding or substructure could explain the genetic pattern encountered in the Tarentaise breed.

Not surprising, the Villard de Lans showed the lowest mean number of alleles per locus (NA) and the lowest number of private alleles (PA). Indeed, this breed is considered as endangered by the FAO and currently less than 500 individuals exist (the minimum historical population size was several dozen in the 1970s). However, this breed has a heterozygosity ( $H_e$ ) comparable to that of other French breeds. The low allelic diversity yet relatively high heterozygosity is not surprising because alleles are lost faster than heterozygosity during population declines (Nei et al., 1975). Crossings between French Blonde d'Aquitaine or German Gelbvieh and Villard de Lans have been carried out since the 1970s to maintain genetic diversity and account now for about 10% of the reproduction (BRG data). These admixtures might also explain the maintenance of the heterozygosity in the Villard de Lans breed. Finally, despite a medium-sized population (50,000 individuals), the high number of alleles observed in Abondance compared to non-alpine breeds might be explained by recent admixtures with U.S. Red Holstein (concerning 5% of the annual servings: BRG data) and by a relatively high ratio between male and female breeders (1 for 55).

#### *Relationships Among Breeds*

Both the N-J tree and the PCA analysis showed that the French Holstein breed is highly distinct from the other breeds analyzed. The French Holstein breed was created recently (beginning of the 19th century), whereas all other breeds are endemic to France. Consistent with Moazami-Goudarzi et al. (1997), the Limousin and Charolais breeds clustered together and were clearly differentiated from the dairy alpine breeds, suggesting a possible common origin or recent gene flow between these two breeds. The grouping of the three dairy alpine breeds (Abondance, Tarentaise, and Montbéliarde) was consistent with previous biochemical and morphological studies (Grosclaude et al., 1990). The close kinship between Abondance and Tarentaise might suggest some past crossing between these two geographically close breeds. Likewise, the close genetic relationship among French and Swiss dairy alpine cows, and their geographic proximity, suggests the possibility of admixture between these breeds.

#### *Breed Assignment*

Assignment methods can be useful in population genetic studies to identify immigrants or their descendants and in wildlife management or forensic science to identify the origin of illegally killed animals (Paetkau et al., 1995; Cornuet et al., 1996; Manel et al., 2002). They can also be very useful in agriculture for the traceability of animals or animal products (e.g., assignment of a carcass, an embryo, sperm to a breed, or milk sample), for breed confirmation, or for hybrid-

ization detection. The aim of this study was not to compare all of the numerous assignment tests currently available (e.g., Paetkau et al., 1995; Prichard et al., 2000; Banks and Eichert, 2000), but rather to compare two assignment tests: one that computes a confidence level ( $P$ -value) for each animal to be assigned (Cornuet et al., 1999) and one that does not (Paetkau et al., 1995). The test that does not provide a  $P$ -value (Paetkau et al., 1995) is very similar to one previously used in livestock studies (Buchanan et al., 1994; MacHugh et al., 1998; Diez-Tascon et al., 2000). However, unlike the test of Paetkau et al. (1995) the previously used tests do not assign real individuals; they assign only simulated individuals (genotypes simulated from allele frequencies). The test of Cornuet et al. (1999) provides a  $P$ -value for measuring the confidence that each real individual truly belongs to a given population. For animal traceability or breed confirmation, high assignment certainty (e.g., 99.99%) will often be necessary for each individual assignment.

Previous authors (Buchanan et al., 1994; MacHugh et al., 1998; Diez-Tascon et al., 2000) were able to assign a breed designation to simulated genotypes with accuracy between 80 and 100% using allele frequencies from 5 to 21 microsatellites from 5 to 10 domestic breeds (either cows or sheep). Using the method of Paetkau et al. (1995) we found fairly similar performances: between 93 to 100% of individuals were correctly assigned. However, both these methods are of limited usefulness because they designate as breed of origin the most probable population without providing a confidence level ( $P$ -value) for each animal assigned.

The performance of Cornuet et al.'s (1999) test is strongly influenced by the threshold probability criterion used to assign individuals to a breed. Unfortunately, assignment tests perform less well as the  $P$ -value decreases (e.g., from 0.05 to 0.001). This result is not surprising: an increased stringency of assignment of individuals logically leads to a decrease in the accuracy of the test. The accuracy of the assignment was also strongly influenced by the population differentiation, as has been observed in previous studies (Cornuet et al., 1999; Manel et al., 2002). The performance of assignment tests increases with the level of population genetic differentiation. For example, using two divergent breeds ( $F_{ST} > 0.1$ ) almost 95% of the individuals could be correctly assigned with a very high certainty ( $P < 0.001$ ), but only 40% could be assigned ( $P < 0.001$ ) when  $F_{ST} < 0.05$ . Unfortunately, the generally low differentiation in allele frequencies observed between many domestic breeds ( $F_{ST} < 0.1$ ) might lead to a low assignment performance, especially if a high certainty is required. Clearly, the Cornuet et al. (1999) exclusion-simulation significance test can reliably identify the breed of origin of individuals (or tissues) with a high certainty (99.9%), at least when using only a few breeds that are substantially genetically differentiated.

## Implications

We showed that even with a much smaller population size the Abondance and Tarentaise breeds showed a higher genetic variability than the French Holstein breed. Intercontinental genetic structure and intensive human selection might be responsible for the heterozygote deficit and the lower number of alleles observed in the French Holstein breed. French and Swiss alpine dairy breeds seem to be genetically related, suggesting a common origin or historical crossing between breeds. Because breed assignment for animal or meat traceability requires a high certainty, extreme care must be taken when using assignment tests. Although breed assignment from microsatellite allele frequencies is useful between genetically differentiated breeds, it is often of limited usefulness between relatively undifferentiated breeds ( $F_{ST} < 0.1$ ). However, additional studies are needed to evaluate the performance of other published assignment tests for breed assignment.

## Literature Cited

- Banks, M. A., and W. Eichert. 2000. WHICHRUN (Version 3.2): A computer program for population assignment of individuals based on multilocus genotype data. *J. Hered.* 91:87–89.
- Boichard, D., L. Maignel, and E. Verrier. 1996. Analyse généalogique des races bovines laitières françaises. *INRA Prod. Anim.* 9:323–335.
- Buchanan, F. C., L. J. Adams, R. P. Littlejohn, J. F. Maddox, and A. M. Crawford. 1994. Determination of evolutionary relationships among sheep breeds using microsatellites. *Genomics* 22:397–403.
- Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1994. *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ.
- Cornuet, J. M., S. Aulagnier, S. Lek, S. Franck, and M. Solignac. 1996. Classifying individuals among infra-specific taxa using microsatellite data and neural networks. *C. R. Acad. Sci. III* 319:1167–1177.
- Cornuet, J.-M., S. Piry, G. Luikart, A. Estoup, and M. Solignac. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153:1989–2000.
- Diez-Tascon, C., R. P. Littlejohn, P. A. R. Almeida, and A. M. Crawford. 2000. Genetic variation within the Merino sheep breed: analysis of closely related populations using microsatellites. *Anim. Genet.* 31:243–251.
- Grosclaude, F., R. Y. Aupetit, J. Lefebvre, and J. C. Mériaux. 1990. Essai d'analyse des relations génétiques entre les races bovines françaises à l'aide du polymorphisme biochimique. *Genet. Selec. Evol.* 22:317–338.
- Hanslik, S., B. Harr, G. Brem, and C. Schlötterer. 2000. Microsatellite analysis reveals substantial genetic differentiation between contemporary New World and Old World Holstein Friesian populations. *Anim. Genet.* 31:31–38.
- Jamieson, A. 1965. The genetic of transferrins in cattle. *Heredity* 20:419–440.
- MacHugh, D. E. E., R. T. Loftus, P. Cunningham, and D. G. Bradley. 1998. Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim. Genet.* 29:333–340.
- Manel, S., P. Berthier, and G. Luikart. 2002. Detecting poaching: Identifying the origin of individuals using bayesian assignment tests and multi-locus genotypes. *Cons. Biol.* (In press).
- Martin-Burriel, I., E. Garcia-Muro, and P. Zaragoza. 1999. Genetic diversity analysis of six Spanish native cattle breeds using microsatellites. *Anim. Genet.* 30:177–182.

- Moazami-Goudarzi, K., D. Laloë, J. P. Furet, and F. Grosclaude. 1997. Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites. *Anim. Genet.* 28:338–345.
- Nei, M. 1972. Genetic distance between populations. *Am. Nat.* 106:283–292.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29:1–10.
- Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. *J. Mol. Evol.* 19:153–170.
- Paetkau, D., W. Calvert, I. Stirling, and C. Stobek. 1995. Variation in genetic diversity across the range of North American brown bears. *Conserv. Biol.* 12:418–429.
- Prichard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Rannala, B., and J. Mountain. 1997. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. USA* 94:9197–9201.
- Raymond, M., and F. Rousset. 1995. Genepop (version 1.2): Population genetics software for exact test and ecumenicism. *J. Hered.* 86:248–249.
- Saitbekova, N., C. Gaillard, G. Obexer-Ruff, and G. Dolf. 1999. Genetic diversity in Swiss goat breeds based on microsatellite analysis. *Anim. Genet.* 30:36–41.
- Schmid, M., N. Saitbekova, C. Gaillard, and G. Dolf. 1999. Genetic diversity in Swiss cattle breeds. *J. Anim. Breed. Genet.* 116:1–8.

**Citations**

This article has been cited by 5 HighWire-hosted articles:  
<http://jas.fass.org#otherarticles>