www.nature.com/hdy

Unfolding of population structure in Baltic sheep breeds using microsatellite analysis

I Tapio[1,5], M Tapio[1], Z Grislis[2], L-E Holm[3], S Jeppsson[4], J Kantanen[1], I Miceikiene[5], I Olsaker[6], H Viinalass[7] and E Eythorsdottir[8]

[1]MTT Agrifood Research Finland, Animal Breeding Research, Jokioinen FI-31600, Finland; [2]Department of Animal Science, Latvia University of Agriculture, 2 Liela Street, Jelgava LV-3001, Latvia; [3]Department of Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, Blichers Alle, PO Box 50, Tjele DK-8830, Denmark; [4]Swedish Board of Agriculture, Jönköping S-55182, Sweden; [5]Lithuanian Veterinary Academy, K Janusauskas Laboratory of Animal Genetics, Tilzes 18, Kaunas LT-47181, Lithuania; [6]Department of Basic Sciences and Aquatic Medicine, Norwegian School of Veterinary Science, PO Box 8146 Dep., Oslo N-0033, Norway; [7]Laboratory of Genetics, Institute of Animal Science, Estonian Agricultural University, Kreutzwaldi 1, Tartu EE-51014, Estonia; [8]Agricultural Research Institute, Keldnaholti, Reykjavik IS-112, Iceland

Studies of domestic animals are performed on breeds, but a breed does not necessarily equate to a genetically defined population. The division of sheep from three native and four modern Baltic sheep breeds was studied using 21 microsatellite loci and applying a Bayesian clustering method. A traditional breed-wise approach was compared to that relying on the pattern of molecular diversity. In this study, a breed was found to be inconsistent with a distinct genetic population for three reasons: (i) a lack of differentiation between modern Baltic breeds, since the majority of the studied sheep formed a single population; (ii) the presence of individuals of foreign ancestry within the breed; and (iii) an undefined local Saaremaa sheep was referred to as a breed, but was shown to consist of separate populations. In the breed-wise approach, only the clearly distinct Ruhnu sheep demonstrated low within-breed variation, although the newly identified Saaremaa populations also have low variability. Providing adequate management recommendations for the Saaremaa sheep is not possible without further studies, but the potential harmful effects of inbreeding in the Ruhnu sheep could be reduced through the use of two genetically related Saaremaa populations. In other breeds, excessive crossing appears to be a larger concern than inbreeding. Assigning individuals into populations based on the pattern of genetic diversity offers potentially unbiased means of elucidating the genetic population structure of species. Combining these genetic populations with phenotypic and aetiological data will enable formulation of the most informed recommendations for gene resource management.
Heredity (2005) **94,** 448–456. doi:10.1038/sj.hdy.6800640
Published online 26 January 2005

Keywords: *Ovis aries*; gene resources; diversity; admixture analysis

Introduction

Studies of genetic diversity in domestic animals are based on an evaluation of the genetic variation within breeds and genetic relationships among them, since the breed is the management unit for which factors such as inbreeding are controlled. The definition of a breed, as applied by the Food and Agriculture Organisation of the United Nations (FAO), is based on the homogeneity of external characteristics, or on a generally accepted identity of animals of a geographically or culturally separated group (FAO, 1998). However, the applied classification may not always reflect the underlying genetic population structure. In old, recognised, isolated native breeds, the uniqueness of the ancestry and the phenotype can be assumed to correspond. However, modern breeds with distinct selected external characteristics may have become genetically similar through gene flow, typically taking place in the form of male-mediated crossbreeding (Bradley *et al*, 1994) and the use of a few commercial sheep breeds as the basis for sheep breed development (Maijala and Terrill, 1991). In addition, closely related populations may be defined as separate breeds, for example, due to administrative borders. Thus, a molecular genetic study of population structure may improve the understanding of present-day genetic resources. This information could be used together with phenotypic or demographic data to guide management efforts and to define management units (Moritz, 1994).

To evaluate genetic diversity of domestic animal breeds, statistical measures derived from Wright's *F*-statistics (Wright, 1951) or phylogenetic techniques based on genetic distances estimated from polymorphic microsatellite markers (Hall and Bradley, 1995) have been the methods of choice. Recently, Bayesian model-based clustering methods have been proposed, which allow for the inference of population structure and the assignment of individuals to populations (Pritchard *et al*, 2000; Corander *et al*, 2003). These methods are also useful for the definition of management units and have been applied to ascertain population structure in 20 chicken

breeds (Rosenberg *et al*, 2001), and to study the genetic relationships in 85 domestic dog breeds (Parker *et al*, 2004).

The Baltic countries have old traditional local sheep populations, as well as several modern-type sheep breeds created less than 100 years ago by upgrading local populations with European breeds. The historical information of the old native Baltic sheep breeds is limited. The Estonian Ruhnu population is thought to descend from sheep left on the Ruhnu island by Swedish-speaking inhabitants who fled from the island before the 2nd World War. The current Ruhnu sheep population consists of a single semimanaged flock. The Estonian Saaremaa sheep may descend from ancient local sheep, but the extent of genetic influence from the more recent Estonian sheep breeds remains uncertain, since there is no breeding programme in place. None of the Estonian local sheep types are officially recognised as a breed. The Lithuanian Native Coarsewooled breed was created at the end of the 19th century by the crossing of Pomeranian, Polish long-tailed and Northern short-tailed sheep. In the middle of the 20th century, this breed became almost extinct. Currently, three flocks are maintained. Owing to the variable population types, the Baltic sheep breeds constitute an excellent set of populations to study the applicability of detailed population structure analysis for the conservation of domestic animals.

In this study, a genetic clustering approach (Corander *et al*, 2003) was applied on microsatellite data (21 loci) from three indigenous and four modern Baltic sheep breeds. This information was used to compare differences between the accepted traditional breed definitions and inferred population structure based entirely on molecular data. In the first case, sheep grouping will be referred to as a breed, whereas grouping in the second case was referred as a panmictic population.

Methods

Sampled breeds

A total of 195 individuals, representing three local (Lithuanian Native Coarsewooled, Estonian Ruhnu and Estonian Saaremaa) and four modern (Lithuanian Black-face, Latvian Darkheaded, Estonian Whitehead and Estonian Blackhead) Baltic sheep breeds, were sampled. The number of sampled flocks per breed and the number of individuals per flock are documented in Table 1. Sheep known to be closely related were not sampled, except for the Estonian Ruhnu sheep, where almost all existing individuals, including progeny, were sampled due to the small population size.

DNA extraction and microsatellite analysis

Blood samples were collected in tubes containing EDTA and stored at -20°C. DNA was isolated by salt extraction according to Miller *et al* (1988), with the exception that an additional phenol–chloroform extraction was performed. Samples were genotyped for 21 microsatellite markers (Table 2). Polymerase chain reactions (PCR) were carried out in volumes of 25 μl using 10–50 ng of template DNA, 5′-end fluorescent-labelled primers (10 pmol) and Finn-zymes (Espoo, Finland) PCR reagents (0.2 mM dNTP each, 10 mM Tris-HCl (pH 8.8), 1.5 mM MgCl$_2$, 50 mM

KCl, 0.1% Triton® X-100 and 1 U DyNAzyme II DNA polymerase). All loci, except BM1818, were amplified using a common touchdown procedure of 31 cycles in total. Initial annealing temperature was 60°C, which was decreased every second cycle by 3°C, until 48°C was reached. For BM1818, a higher MgCl$_2$ concentration (3 mM) was used and a touchdown procedure of 38 cycles in total was applied. Initial annealing temperature was 62°C, which was decreased every third cycle by 2°C, to a final temperature of 54°C. The amplified products were separated on 6% denaturing polyacrylamide gel using automated laser detection on A.L.F. and A.L.F. Express DNA Sequencers (Pharmacia, Uppsala, Sweden). Each gel contained a reference individual to ensure constancy of allele sizing across the gels and internal and external size standards (Pharmacia, Uppsala, Sweden) to define microsatellite sizes. Sizing of bands and analysis of genetic variants was performed by the A.L.F. win™ Fragment Analyser 1.0 (Pharmacia, Uppsala, Sweden).

Data analysis

A neighbour-joining tree was constructed with the program NEIGHBOR in the PHYLIP 3.6a3 package (Felsenstein, 2002) using the allele sharing distance (Bowcock *et al*, 1994) calculated with MICROSAT 1.4d (Minch *et al*, 1995).

Clustering of individuals into panmictic populations

The population structure was unfolded using a Bayesian method implemented in BAPS v2.0 (Corander *et al*, 2003). Initially, each individual was defined as a separate population and then individuals were clustered into the most likely set of ideal populations that are in Hardy–Weinberg (HWE) and linkage equilibrium. The method of Corander *et al* (2003) treats both the allele frequencies and the number of populations as random variables, and pools populations if the data does not support a distinction to separate panmictic populations. Two repeated Markov chain Monte Carlo (MCMC) analyses were performed, each with three chains having a burn-in of 6 000 000 and additional 15 000 000 iterations used in clustering after thinning by recording only every 100th iteration. The initial number of clusters in the MCMC runs was set to 7, corresponding to the number of sampled breeds, but this setting did not affect the prior probability of the population number.

Within- and between population genetic variation

Genetic diversity within the populations was estimated as the mean expected unbiased heterozygosity (Nei, 1987) and mean allelic richness (El Mousadik and Petit, 1996), corresponding to the minimum actual sample size (21 diploid individuals) using FSTAT 2.93 (Goudet, 1995). Deviations from HWE were tested by an exact test (Guo and Thompson, 1992) in ARLEQUIN (Schneider *et al*, 2000) and by Weir and Cockerham's (1984) *f*, which corresponds to Wright's within-population inbreeding coefficient F_{IS}, calculated with FSTAT 2.93. The same software was applied to calculate Weir and Cockerham's overall locus-wise *F*-statistics. The significance of the θ and *f* estimates was determined by permuting genotypes within the total population, and alleles within samples using 10 000 permutations. Genotypic linkage disequili-

Table 1 Clustering of Baltic sheep into panmictic populations based on Bayesian analysis

Breed	Flock (n)	Panmictic population (n)[a]								
		P1	P2	P3	P4	P5	P6	P7	P8	P9
Latvian Darkheaded	1 (16)								14	2
	2 (16)								12	4
Lithuanian Native Coarsewooled	1 (30)							30		
Lithuanian Blackface	1 (30)									30
Estonian Whitehead	1 (1)									1
	2 (3)									3
	3 (2)									2
	4 (3)									3
	5 (3)									3
	6 (5)									5
	7 (2)									2
	8 (1)									1
	9 (2)									2
	10 (4)									4
	11 (1)									1
	12 (1)									1
	13 (1)								1	
	14 (1)					1				
Estonian Blackhead	1 (2)									2
	2 (3)									3
	3 (2)									2
	4 (2)									2
	5 (3)									3
	6 (2)									2
	7 (2)									2
	8 (2)									2
	9 (1)									1
	10 (1)									1
	11 (3)								1	2
	12 (5)								1	4
Estonian Ruhnu	1 (24)					24				
Estonian Saaremaa	1 (3)				3					
	2 (5)						5			
	3 (1)									1
	4 (1)									1
	5 (1)			1						
	6 (4)		3							1
	7 (6)	4								2
Sum	195	4	3	1	3	24	6	30	29	95
H_{exp}		0.496	0.468	0.667	0.417	0.530	0.563	0.739	0.681	0.757
SD (H_{exp})		0.189	0.229	0.483	0.233	0.207	0.178	0.102	0.137	0.110
R		—	—	—	—	3.501	—	6.507	6.208	7.896
SD (R)		—	—	—	—	1.203	—	1.438	1.897	2.490

Rows correspond to source flocks and columns to assigned panmictic populations. Number (n) of individuals in each flock is indicated within parentheses. Mean expected unbiased heterozygosity (H_{exp}), sample size corrected mean allelic richness (R) per population corresponding to a sample size of 21 diploid individuals and associated standard deviations (SD) are reported in the lower table.
[a]P1, P2, P3, P4 and P6 correspond to Saaremaa panmictic populations; P5 to Ruhnu-P5; P7 to Coarsewooled-P7; P8 to Latvian-Estonian-P8 and P9 corresponds to the Modern-P9 panmictic population.

brium between all pairs of microsatellite loci was estimated with GENEPOP 3.3 (Raymond and Rousset, 1995), performing a probability test using a Markov chain method of 50 000 iterations and 100 batches.

The genetic relationships among breeds and panmictic populations were analysed by the principal co-ordinates analysis (PCoA) as implemented in VISTA v6.4 (Young, 1996) using the Chord distance (Cavalli-Sforza and Edwards, 1967). In addition, θ values (Weir and Cockerham, 1984) between pairs of breeds and pairs of panmictic populations were calculated using FSTAT 2.93.

Results

Variability of microsatellite loci
A total of 240 alleles were detected at the 21 micro-satellite loci analysed. All loci were polymorphic with

Table 2 Microsatellite markers and chromosomal location (Chr)

Locus	Chr.	A_{tot}	H_{tot}	References of microsatellite
BM0757	9	9	0.753	Bishop et al (1994)
BM1314	22	12	0.781	Bishop et al (1994)
BM1818	20	18	0.907	Bishop et al (1994)
BM4621	6	14	0.836	Bishop et al (1994)
BM6506	1	10	0.599	Bishop et al (1994)
BM6526	26	13	0.824	Bishop et al (1994)
BM8125	17	6	0.601	Bishop et al (1994)
CSSM31	23	17	0.865	Moore and Byrne (1994)
INRA023	1	13	0.877	Vaiman et al (1994)
MAF214	16	10	0.585	Buchanan and Crawford (1992)
MAF36	22	15	0.877	Swarbrick et al (1991)
MAF48	Unknown	8	0.740	Bishop et al (1994)
MAF65	15	9	0.755	Buchanan et al (1992)
MCM527	5	9	0.738	Hulme et al (1994)
OarCP20	21	12	0.805	Ede et al (1994a)
OarCP34	3	9	0.796	Ede et al (1994b)
OarFCB128	2	8	0.653	Buchanan and Crawford (1993)
OarFCB304	19	14	0.738	Buchanan and Crawford (1993)
OarFCB48	17	14	0.808	Buchanan et al (1994)
OarHH47	18	13	0.826	Henry et al (1993)
OarVH72	25	7	0.752	Pierson et al (1993)
Overall estim.		240	0.767	

Total number of alleles (A_{tot}) and genetic diversity in the total sampled population (H_{tot}).

the number of alleles per locus ranging from 6 to 18. Average expected heterozygosity for all loci was 0.767, varying between 0.585 and 0.907 for individual loci (Table 2).

Unfolding of population structure

The neighbour-joining tree did not show a clear separation of the sheep breeds (Figure 1). Only the Estonian Ruhnu sheep formed a distinct cluster. Both the Lithuanian Native Coarsewooled and the Latvian Darkheaded sheep tended to cluster with individuals from the same breed. In contrast, sheep from the Estonian modern, the Estonian Saaremaa and the Lithuanian Blackface breeds demonstrated little breed-wise clustering. Animals from these breeds and the remaining sheep from the Lithuanian Native Coarsewooled and Latvian Darkheaded populations were located as small groups distributed throughout the tree (Figure 1).

The clustering of individuals using the Bayesian method of Corander et al (2003) suggested nine panmictic populations (P) ($P = 0.82$) (Table 1), but there were three possible partitionings with posterior probability values greater than 0.1. The likelihoods of the two best supported partitionings were almost equal ($P = 0.31$ and 0.28), while the third highest partitioning had a lower posterior probability ($P = 0.12$). Differences between partitionings were only observed for the grouping of two Latvian Darkheaded individuals, either to the Latvian-Estonian-P8 or to the Modern-P9 panmictic population. The results presented are based on the most likely partitioning ($P = 0.31$). In this scenario, the Ruhnu-P5 and the Coarsewooled-P7 corresponded to the predefined breed categorisation (Table 1). The majority of animals from the modern Estonian sheep breeds, all Lithuanian Blackface sheep, and a few animals from the Latvian Darkheaded and Estonian Saaremaa breeds
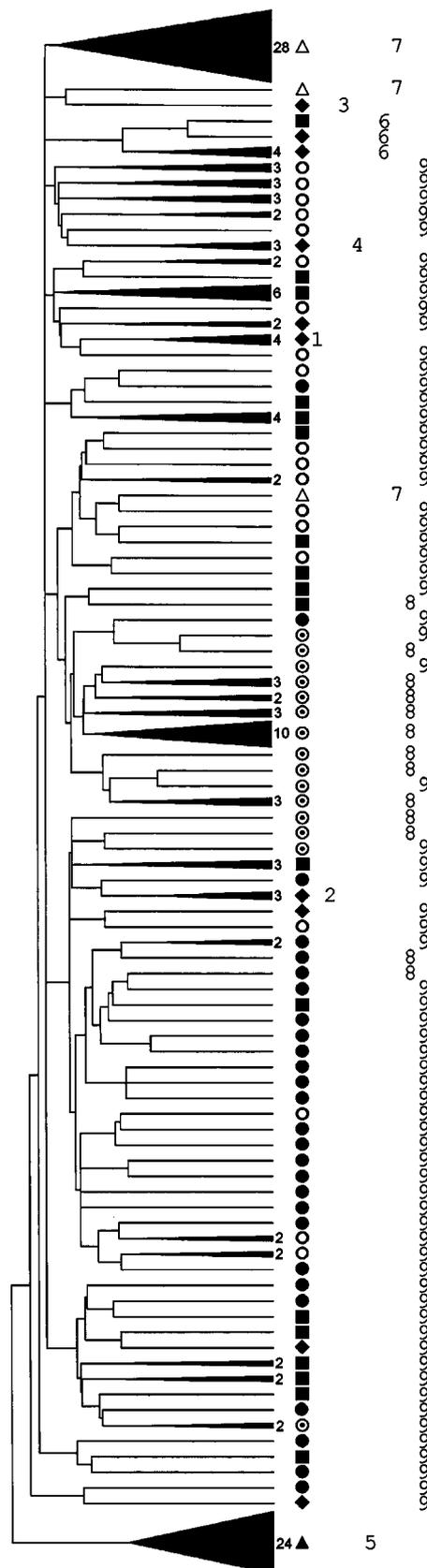
formed the largest common Modern-P9 panmictic population. Further 26 animals from the Latvian Darkheaded breed, one individual Estonian Whitehead and two Estonian Blackhead sheep formed a panmictic population (Latvian-Estonian-P8). The Estonian Saaremaa breed separated into four pure Saaremaa panmictic populations and a population containing five Saaremaa and one Estonian Whitehead sheep (Saaremaa-P6; Table 1).

Genetic variation within and among breeds and panmictic populations

The mean allelic richness observed per breed ranged from 3.501 in the Estonian Ruhnu to 7.539 in the Estonian Whitehead (Table 3). Among the panmictic populations, excluding populations of less than 21 sheep (Saaremaa-P1, -P2, -P3, -P4 and -P6), the allelic richness varied from 3.501 to 7.896 (Table 1). Within-breed mean expected heterozygosity ranged from 0.530 to 0.772, with Estonian Ruhnu as the least variable and Estonian Whitehead as the most variable breed (Table 3). Across all breeds, expected heterozygosity averaged 0.71. Among panmictic populations, the expected heterozygosity varied from 0.417 (Saaremaa-P4) to 0.757 (Modern-P9; Table 1), with a mean of 0.60 across populations.

The overall estimate of Weir and Cockerham's (1984) θ was 0.088 among breeds and 0.126 among the panmictic populations, being significantly ($P < 0.05$) different from zero in both cases. The overall estimate of Weir and Cockerham's (1984) f was 0.032 and 0.015 in breeds and panmictic populations, respectively. However, the f value only deviated significantly ($P < 0.05$) from zero in the breed estimator due to substructure within the Estonian Saaremaa breed. This substructure was not only detectable in the clustering of individuals described above but also as a substantially positive f value observed in the

Estonian Saaremaa breed (Table 3) and as a significant locus-wise deviation from HWE for two loci (INRA023 and MAF36) after Bonferroni correction. Furthermore,

Figure 1 Neighbour-joining tree constructed from allele sharing distances among 195 animals representing seven Baltic sheep breeds (◉, Latvian Darkheaded; △, Lithuanian Native Coarse-wooled; ○, Lithuanian Blackface; ■, Estonian Whitehead; ●, Estonian Blackhead; ▲, Estonian Ruhnu and ◆, Estonian Saaremaa). Animals representing the same breed and the same panmictic population were merged where possible. Numbers at branch tips indicate the number of individuals that merged. Number on the right denotes membership in a particular panmictic population (Table 1).

the test of allele associations between pairs of loci in each breed indicated that out of a possible 1470 comparisons, only one locus pair in the Estonian Saaremaa breed exhibited significant linkage disequilibrium at the 5% level after sequential Bonferroni correction was applied. As expected from the assumptions of the clustering process, no significant locus-wise or population-wise deviations from HWE were detected among the panmictic populations. Similarly, no significant deviations from linkage disequilibrium were observed.

In the pair-wise θ comparison, a significant differentiation between all breeds with values ranging from 0.019 (Estonian Blackhead–Estonian Whitehead) to 0.212 (Estonian Ruhnu–Lithuanian Native Coarsewooled) was detected. The comparisons involving Ruhnu sheep were higher than other pair-wise values (Table 4). Pair-wise θ values between the panmictic populations ranged from 0.045 (Modern-P9–Latvian-Estonian-P8) to 0.456 (Saaremaa-P1–Saaremaa-P4). However, estimates between the Saaremaa panmictic populations were not statistically significant (Table 4).

In the PCoA plot for breeds, the Estonian Ruhnu was separated from the other breeds on axis I, which explained 45% of the variation (Figure 2a). On axis II, two local sheep breeds (Lithuanian Native Coarsewooled and Estonian Saaremaa) were separated from the modern breeds, which accounted for 17% of the variation (Figure 2a). In the PCoA plot for the panmictic populations (Figure 2b), individuals from the modern Estonian and Lithuanian sheep breeds that were located closely to each other in the plot (2a) were merged forming the core of the Modern-P9 panmictic population. On axis I (30% of the variation), Modern-P9, Coarsewooled-P7 and Latvian-Estonian-P8 panmictic populations were grouped independently from the local Estonian sheep panmictic populations (Figure 2b). Axis II (16% of the variation) further divided the Estonian local panmictic populations into two groups: Saaremaa-P1 and -P4 were located relatively close to the Ruhnu-P5 sheep, while the three remaining Saaremaa panmictic populations (Saaremaa-P2, -P3 and -P6) were divergent (Figure 2b).

Discussion

Population structure

The population structure of Baltic sheep, based solely on microsatellite variation using a Bayesian clustering method (Corander *et al*, 2003), demonstrated that a traditional breed, defined as geographically separated groups with homogeneous external characteristics (FAO, 1998), may not necessarily equate to a genetic population, but can be narrower or wider than a breed. This

Table 3 Sample size (*n*), sample size corrected mean allelic richness (*R*) per breed corresponding to a minimum sample size of 21 diploid individuals

Breeds	n	R	H_{exp}	f 95% CI
Latvian Darkheaded	32	6.551 (1.967)	0.697 (0.128)	(−0.017, 0.064)
Lithuanian Native Coarsewooled	30	6.507 (1.438)	0.739 (0.102)	(−0.031, 0.059)
Lithuanian Blackface	30	6.585 (2.273)	0.724 (0.140)	(−0.015, 0.056)
Estonian Whitehead	30	7.539 (2.275)	0.772 (0.081)	(−0.008, 0.077)
Estonian Blackhead	28	7.467 (2.386)	0.722 (0.146)	(−0.037, 0.039)
Estonian Ruhnu	24	3.501 (1.203)	0.530 (0.207)	(−0.090, 0.030)
Estonian Saaremaa	21	7.095 (2.071)	0.755 (0.089)	(0.121, 0.235)
Mean		6.464 (1.376)	0.705 (0.081)	

Mean expected unbiased heterozygosity (H_{exp}) and 95% confidence intervals for within-population inbreeding coefficients (*f*), and associated standard deviations within parentheses.

analysis was motivated by the lack of clear distinct population boundaries among seven Baltic sheep breeds in the neighbour-joining tree based on simple allele sharing distances between individuals (Figure 1), which has been shown to allow grouping of individuals according to geographical origin (Bowcock *et al*, 1994; MacHugh *et al*, 1998; Bjørnstad and Røed, 2001).

A modern genetic population may extend over several breeds. From the four modern Baltic sheep breeds studied, only the Latvian Darkheaded appeared to be isolated and formed a separate panmictic population. The main production stock consisting of the Lithuanian Blackface, most of the Estonian Blackhead and the Estonian Whitehead, and some individuals from the Latvian Darkheaded and the Estonian Saaremaa formed a single panmictic population that included half of the studied Baltic sheep. Extensive introgression of genetic material from a relatively similar set of international breeds (http://www.nordgen.org/husdyrdatabase/en_sok.asp) and subsequent gene flow across the Baltic countries has contrived to homogenise the gene pool of the modern Baltic sheep breeds. Other popular production breeds are also likely to belong to common population owing to the similar background of many recently created European sheep breeds (Maijala and Terrill, 1991).

In some cases, an individual attributed to a breed can be genetically atypical. Three Estonian modern sheep were likely to be of Latvian ancestry, while five Latvian Darkheaded rams and one ewe appeared to be genetically different from the sampled population and were assigned to the large modern stock. Similarly, five Saaremaa sheep appeared to belong to the common production stock. These observations may indicate a lack of accurate information in breed records, or may reflect a continual upgrading or reduction of inbreeding in the modern Baltic sheep breeds. It is notable that the within-breed inbreeding coefficient (*f*) gave no indication of structures in the Latvian Darkheaded, the Estonian Whitehead or the Estonian Blackhead breeds, which emerged from the Bayesian clustering analysis.

Contrary to modern breeds, the old undefined local populations can consist of genetically differentiated flocks as previously reported in sheep (Petit *et al*, 1997; Tapio *et al*, 2003). The subdivision, as supported by the deviations from HWE and genotypic linkage disequilibrium, was evident for the Estonian Saaremaa sheep.

Apart from the contribution of individuals from the production stock, the Saaremaa sheep appeared to consist of genetically isolated flocks (Table 1).

Distribution of genetic variation
The range of variability within Baltic sheep breeds appeared larger than previously reported for native sheep (Arranz *et al*, 1998; Tapio *et al*, 2003) and was due to low variation in the Estonian Ruhnu sheep. Examining pure Saaremaa panmictic populations revealed their genetic diversity to be low, comparable to that of Estonian Ruhnu sheep (data not shown). Furthermore, the Latvian Darkheaded demonstrated a lower genetic diversity, which may arise from the small number of Latvian Darkheaded rams in use.
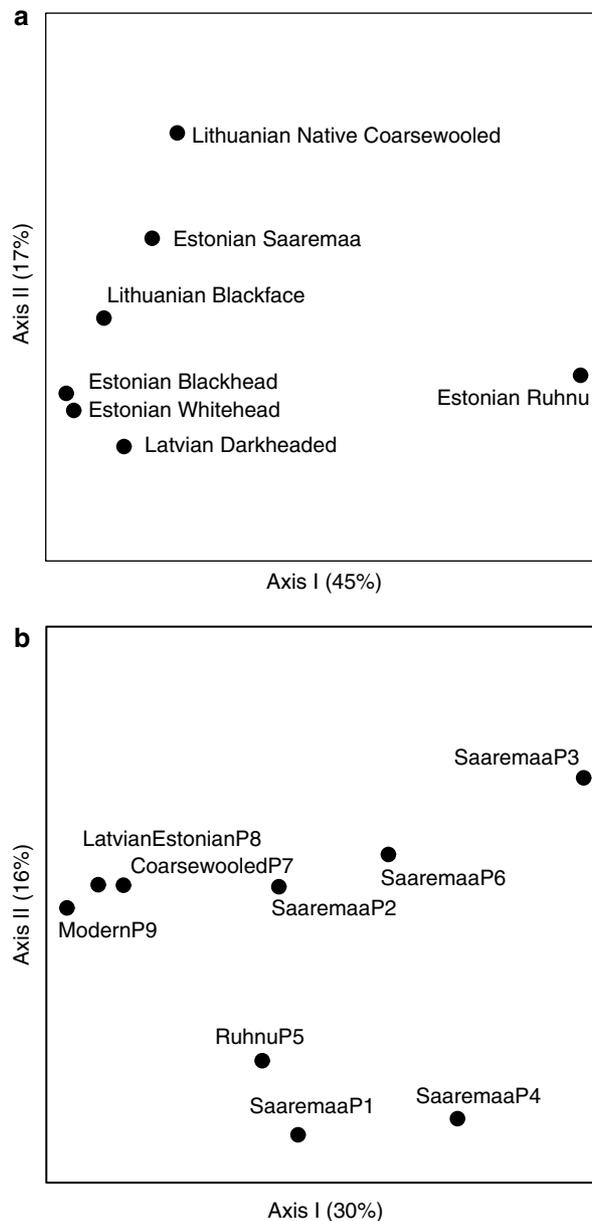
The observed differentiation among the Baltic breeds and panmictic populations is comparable to the genetic variation reported for Spanish sheep (Arranz *et al*, 1998) and that observed in other domestic species (Kantanen *et al*, 2000; Laval *et al*, 2000). However, despite the significant differentiation among all Baltic sheep breeds based on pair-wise θ, the majority of modern breeds merged into a single panmictic population. Hedrick (1999) noted that with the statistical power afforded by highly polymorphic loci, a statistically significant differentiation does not necessarily implicate a biologically important distinction. Data from this study further suggest that statistically significant differentiation may not always denote a real distinction between populations, if the boundaries between populations are unknown.

The most significant pattern of the PCoA plot is the separation of native Estonian sheep from other Baltic sheep on the first axis. The grouping of the Saaremaa with the native Estonian sheep was not evident in the PCoA for breeds, when the Saaremaa sheep included individuals, belonging rather to the modern sheep types. The low coefficient of variation and rapid increase in the first generations of isolation (Takezaki and Nei, 1996), makes the Chord distance (Cavalli-Sforza and Edwards, 1967) a suitable measure for resolving differences among closely related populations. Differences in the polymorphism of microsatellites (Table 2) does not result in loci making a variable contribution to the Chord distance (Landry *et al*, 2002). Furthermore, the Chord distance has been shown to yield similar findings for microsatellite

Table 4 Significant θ values for pairs of breeds above the diagonal and for pairs of panmictic populations below the diagonal (the Saaremaa-P3 is excluded)

		Latvian Darkheaded	Lithuanian Native Coarsewooled	Lithuanian Blackface	Estonian Whitehead	Estonian Blackhead	Estonian Ruhnu	Estonian Saaremaa
Latvian-Estonian-P8	0.049	—	0.091	0.071	0.041	0.044	0.204	0.084
Coarsewooled-P7	0.056	0.100	—	0.059	0.068	0.072	0.212	0.069
Saaremaa-P6	0.153	0.210	0.188	—	0.031	0.033	0.180	0.051
Ruhnu-P5	0.144	0.218	0.212	0.338	—	0.019	0.157	0.031
Saaremaa-P4	0.204	0.289	NS	NS	0.367	—	0.173	0.047
Saaremaa-P2	0.115	0.152	NS	NS	0.330	NS	—	0.173
Saaremaa-P1	0.130	0.211	0.174	NS	0.307	NS	NS	—
	Modern-P9	Latvian-Estonian-P8	Coarsewooled-P7	Saaremaa-P6	Ruhnu-P5	Saaremaa-P4	Saaremaa-P2	Saaremaa-P1

NS = not significant in the permutation test.

a

Axis II (17%)

- Lithuanian Native Coarsewooled
- Estonian Saaremaa
- Lithuanian Blackface
- Estonian Blackhead
- Estonian Whitehead
- Latvian Darkheaded
- Estonian Ruhnu

Axis I (45%)

b

Axis II (16%)

- SaaremaaP3
- LatvianEstonianP8
- CoarsewooledP7
- SaaremaaP6
- SaaremaaP2
- ModernP9
- RuhnuP5
- SaaremaaP1
- SaaremaaP4

Axis I (30%)

Figure 2 Principal coordinate plots for (**a**) breeds and (**b**) panmictic populations constructed using Chord distance.

and protein data in sheep (Tapio et al, 2003). The pure Saaremaa panmictic populations appeared unique. The strong effect of subdivision into flocks differs from earlier studies in cattle and mouflon (Petit et al, 1997; Casellas et al, 2004). Although the small number of studied individuals in these populations may exaggerate genetic distances (Paetkau et al, 1999), this bias is unlikely to affect strongly the grouping on the first axes of the PCoA. This together with the lack of assumption of tree hierarchy and an estimate of the proportion of explained differences makes the PCoA an attractive distance summary.

Management implications
The modern Baltic sheep breeds exhibit considerable within-breed variation, but the extensive introgression

has resulted in actively used national production breeds becoming genetically similar. This stock would probably be regarded as a single management unit (Moritz, 1994) if this was a wild species. The genetic variation would be more efficiently maintained in genetically separate populations rather than in large panmictic stocks (Hall and Bradley, 1995).

The dissection of the Estonian Saaremaa sheep population structure demonstrates the difficulty in evaluating local unimproved 'breeds'. While some animals were of a common type and should therefore be excluded from the founding generations of future Saaremaa breed, the rest of the sample consists of either several distinct local types or isolated flocks of a single type. Even though the detected differentiation could justify a recategorisation of sheep into separate types, genetic differentiation alone is not sufficient to determine the question. A detailed study of phenotypic similarities alongside population structure analysis could provide an appropriate basis on which to distinguish specific breeds.

According to the Bayesian analysis, the Estonian Ruhnu and the Lithuanian Native Coarsewooled formed separate populations consistent with the pre-existing definition of these breeds. Adaptation of the semimanaged Estonian Ruhnu sheep to the seaside pasture suggests that these sheep have had a sufficiently long period to adapt to this environment, making this breed particularly interesting for conservation. However, the Ruhnu breed was found to be less variable than other Baltic sheep breeds, indicating the need for a breed management plan to prevent further loss of genetic variation. In this study, all adults and juveniles of the Ruhnu population have been sampled and there are no other Ruhnu flocks in existence. A controlled introgression from the two closest (Saaremaa-P1 and -P4) Saaremaa panmictic populations (Figure 2b, Table 1) could be used to relieve inbreeding, if deleterious phenotypic effects arising from mating within such a small population appear. The Lithuanian Native Coarsewooled sheep is subject to a coordinated breeding program. Rescued from extinction less than 20 years ago, these sheep have been improved following the introduction of Lithuanian Blackface sheep (B Zapasnikiene, personal communication), which has resulted in genetic variation being as abundant as for the modern sheep breeds. The PCoA plot of the panmictic populations supported the idea of introgression, since the Lithuanian Native Coarsewooled breed was grouped close to the modern panmictic populations. Caution should be exercised in order to prevent the loss of genetic uniqueness through excessive crossing.

Conclusions

Although the variation within and between breeds can be discerned on the phenotypic level, external differences do not necessarily provide the same results as molecular data (Casellas *et al*, 2004). A traditional breedwise molecular genetic study is appropriate when the breed boundaries are incontestable. In that case, the neutral genetic markers can help to clarify if phenotypically similar breeds are also genetically similar. In populations without clear boundaries, clustering of individuals based on neutral variation is useful for avoiding unintended hybridisation, although extensive sampling may be required to ensure reliable findings. In addition, the Bayesian clustering is valuable in revealing the population structure resulting from earlier management decisions, and therefore can provide more precise insights for management planning than a traditional study. This approach showed that the Baltic sheep constitute a more diverse group of populations than would be implied from traditional studies of breed-wise diversity. Combining molecular genetic information with physiological, ecological and aetiological data will allow the most informed gene resource management programs to be implemented.

Acknowledgements

References

Arranz JJ, Bayón Y, San Primitivo F (1998). Genetic relationships among Spanish sheep using microsatellites. *Anim Genet* **29**: 435–440.

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SL, Hawkins GA *et al* (1994). A genetic linkage map for cattle. *Genetics* **136**: 619–639.

Bjørnstad G, Røed KH (2001). Breed demarcation and potential for breed allocation of horses assessed by microsatellite markers. *Anim Genet* **32**: 59–65.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**: 455–457.

Bradley DG, MacHugh DE, Loftus RT, Sow RS, Hoste CH, Cunningham EP (1994). Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in West African trypano-tolerant cattle populations. *Anim Genet* **25**: 7–12.

Buchanan FC, Crawford AM (1992). Ovine dinucleotide repeat polymorphism at the MAF214 locus. *Anim Genet* **23**: 394.

Buchanan FC, Crawford AM (1993). Ovine microsatellites at the OARFCB11, OARFCB128, OARFCB193, OARFCB266 and OARFCB304 loci. *Anim Genet* **24**: 145.

Buchanan FC, Swarbrick PA, Crawford AM (1992). Ovine dinucleotide repeat polymorphism at the MAF65 locus. *Anim Genet* **23**: 85.

Buchanan FC, Galloway SM, Crawford AM (1994). Ovine microsatellites at the OARFCB5, OARFCB19, OARFCB20, OARFCB48 OARFCB129 and OARFCB226 loci. *Anim Genet* **25**: 60.

Casellas ByJ, Jimenez N, Fina M, Tarres J, Sanchez A, Piedrafita J (2004). Genetic diversity measures of the bovine Alberes breed using microsatellites: variability among herds and types of coat colour. *J Anim Breed Genet* **121**: 101–110.

Cavalli-Sforza LL, Edwards AWF (1967). Phylogenetic analysis: models and estimation procedures. *Evolution* **21**: 550–570.

Corander J, Waldmann P, Sillanpää J (2003). Bayesian analysis of genetic differentiation between populations. *Genetics* **163**: 367–374.

Ede AJ, Pierson CA, Crawford AM (1994a). Ovine microsatellites at the OarCP9, OarCP16, OarCP20, OarCP21, OarCP23 and OarCP26 loci. *Anim Genet* **26**: 129–130.

Ede AJ, Pierson CA, Crawford AM (1994b). Ovine microsatellites at the OARCP34, OARCP38, OARCP43, OARCP49, OARCP73, OARCP79 and OARCP99 loci. *Anim Genet* **26**: 130–131.

El Mousadik A, Petit RJ (1996). High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* (L.) Skeels) endemic of Morocco. *Theor Appl Genet* **92**: 832–839.

FAO (1998). Second guidelines for development of national farm animal genetic resources management plans. Measurement of domestic animal diversity (MoDAD): original working group report. URL: http://dad.fao.org/en/refer/library/guidelin/workgrp.pdf.

Felsenstein J (2002). *PHYLIP (Phylogeny Inference Package), Version 3.6a3* Distributed by the author. Department of Genome Sciences, University of Washington: Seattle, USA.

Goudet J (1995). FSTAT (version 1.2): a computer program to calculate F-statistics. *J Hered* **86**: 485–486.

Guo S, Thompson E (1992). Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48**: 361–372.

Hall SJG, Bradley DG (1995). Conserving livestock breed diversity. *Trends Ecol Evol* **10**: 267–270.

Hedrick PW (1999). Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.

Henry HM, Penty JM, Pierson CA, Crawford AM (1993). Ovine microsatellites at the OARHH35, OARHH41, OARHH44, OARHH47 and OARHH64 loci. *Anim Genet* **24**: 223.

Hulme DJ, Silk JP, Redwin JM, Barendse W, Beh KJ (1994). Ten polymorphic ovine microsatellites. *Anim Genet* **25**: 434.

Kantanen J, Olsaker I, Holm L-E, Lien S, Vilkki J, Brusgaard K *et al* (2000). Genetic diversity and population structure of 20 North European cattle breeds. *J Hered* **91**: 446–457.

Landry P-A, Koskinen MT, Primmer CR (2002). Deriving evolutionary relationships among populations using microsatellites and $(\delta\mu)^2$: all loci are equal, but some are more equal than others. *Genetics* **161**: 1339–1347.

Laval G, Iannuccelli N, Legault C, Milan D, Groenen MAM, Giuffra E *et al* (2000). Genetic diversity of eleven European pig breeds. *Genet Sel Evol* **32**: 187–203.

MacHugh DE, Loftus RT, Cunningham P, Bradley DG (1998). Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim Genet* **29**: 333–340.

Maijala K, Terrill CE (1991). Breed structure, dynamics, and new breed development in sheep. In: Maijala K (ed) *Genetic Resources of Pig, Sheep and Goat*. Elsevier Science Publishers BV: Amsterdam, pp 305–326.

Miller SA, Dykes DD, Polesky HF (1988). A simple salting procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**: 1215.

Minch E, Ruiz-Linares A, Goldstein DB, Feldman MW, Cavalli-Sforza LL (1995). *Microsat 1.4d: A Computer Program for Calculating Various Statistics on Microsatellite Allele Data*. http://hpgl.stanford.edu/projects/microsat/.

Moore SS, Byrne K (1994). Characterisation of 65 bovine microsatellites. *Mamm Genome* **5**: 84–90.

Moritz C (1994). Defining 'Evolutionary Significant Units' for conservation. *Trends Ecol Evol* **10**: 373–375.

Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.

Paetkau D, Amstrup SC, Born EW, Calvert W, Derocher AE, Garner GW *et al* (1999). Genetic structure of the world's polar bear populations. *Mol Ecol* **8**: 1571–1584.

Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB *et al* (2004). Genetic structure of the purebred domestic dog. *Science* **304**: 1160–1164.

Petit E, Aulagnier S, Bon R, Dubois M, Crouau-Roy B (1997). Genetic structure of populations of the Mediterranean mouflon (*Ovis Gmelini*). *J Mammal* **78**: 459–467.

Pierson CA, Hanrahan V, Ede AJ, Crawford AM (1993). Ovine microsatellites at the OARVH34, OARVH41, OARVH58, OARVH61 and OARVH72 loci. *Anim Genet* **24**: 224.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Raymond M, Rousset F (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MAM *et al* (2001). Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* **159**: 699–713.

Schneider S, Roessli D, Excoffier L (2000). *Arlequin, Version 2.000: A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Swarbrick PA, Buchanan FC, Crawford AM (1991). Ovine dinucleotide repeat polymorphism at the MAF36 locus. *Anim Genet* **22**: 377.

Takezaki N, Nei M (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**: 389–399.

Tapio M, Miceikiene I, Vilkki J, Kantanen J (2003). Comparison of microsatellite and blood protein diversity in sheep: inconsistencies in fragmented breeds. *Mol Ecol* **12**: 2045–2056.

Vaiman D, Mercier D, Moazami-Goudarzi K, Eggen A, Ciampolini R, Lepingle A *et al* (1994). A set of 99 cattle microsatellites: characterisation, synteny mapping, and polymorphism. *Mamm Genome* **5**: 288–297.

Weir BS, Cockerham CC (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Wright S (1951). The genetical structure of populations. *Ann Eugen* **15**: 323–354.

Young FW (1996). *ViSta: The Visual Statistics System. Research Memorandum 94-1 (B)*, 2nd edn. LL Thursone Psychometric Laboratory, University of North Carolina: Chapel Hill, NC, USA.