

Native breeds demonstrate high contributions to the molecular variation in northern European sheep

M. TAPIO,* I. TAPIO,* Z. GRISLIS,† L.-E. HOLM,‡ S. JEPPSSON,§ J. KANTANEN,* I. MICEIKIENE,¶ I. OLSAKER,** H. VIINALASS†† and E. EYTHORSDDOTTIR‡‡

*Animal Production Research, MTT Agrifood Research Finland, Jokioinen FI-31600, Finland, †Department of Animal Science, Latvia University of Agriculture, 2 Liela Street, Jelgava LV-3001, Latvia, ‡Department of Genetics and Biotechnology, Danish Institute of Agricultural Sciences, Blichers Alle, PO Box 50, Tjele DK-8830, Denmark, §Swedish Board of Agriculture, Jönköping S-55182, Sweden, ¶K Janusauskas Laboratory of Animal Genetics, Lithuanian Veterinary Academy, Tilzes 18, Kaunas LT-47181, Lithuania, **Department of Basic Sciences and Aquatic Medicine, Norwegian School of Veterinary Science, PO Box 8146 Dep., Oslo N-0033, Norway, ††Laboratory of Genetics, Institute of Animal Science, Estonian Agricultural University, Kreutzwaldi 1, Tartu EE-51014, Estonia, ‡‡Agricultural University of Iceland, Keldnaholt, 112 Reykjavik, Iceland

Abstract

Population contribution to genetic diversity can be estimated using neutral variation. However, population expansion or hybridization of diverged ancestries may weaken correlation between neutral and non-neutral variation. Microsatellite variation was studied at 25 loci in 20 native and 12 modern or imported northern European sheep breeds. Breed contributions to total gene diversity, allelic richness and mean allele-sharing distance between individuals were measured. Indications of changes in population size and admixtures of divergent ancestries were investigated and the extent of inbreeding was estimated. The northern European sheep demonstrated signs of reduction in effective population size. Many old, small populations made a substantial positive contribution to total molecular variation, but populations with several divergent major ancestries did not contribute substantially to molecular variation, with the exception of the Norwegian Rygja sheep. However, several diverged major ancestries may cause it to contribute less to non-neutral variation than expected from the microsatellite data. Breed uniqueness and within-breed variability generally had opposite effects on breed contributions to molecular diversity. The degree of inbreeding did not reflect the breed contribution to total gene diversity or allelic richness, but inbred populations increased the mean allele-sharing distance between individuals. Our study indicates breed conservation to be especially important in maintaining allelic variation in northern European sheep and supports the evolutionary importance of peripheral populations.

Keywords: conservation genetics, domestic animals, microsatellite diversity, *Ovis aries*, population history, speciation

Received 20 July 2005; revision accepted 4 August 2005

Introduction

The intensity and standardization of animal production are threatening the existence of many domestic animal breeds (Scherf 2000). In establishing priorities for the conservation of breeds on the basis of genetic arguments, traits, genes and gene combinations should all be considered. In addition,

attention should be paid to the uniqueness of individual features (Ruane 1999; Scherf 2000; Barker 2001). It is not possible to know which traits and alleles are important in the future. This equals to the situation in natural populations, where it is not known beforehand which geographic population initiates a speciation event or provides the crucial beneficial genotypes when the environment changes (Lesica & Allendorf 1995). Genetic breed conservation value can be understood in the respect of loss of genetic material from the species (e.g. Simianer 2005), but concentrating on

Correspondence: Miika Tapio, Fax: +358 34188 3618; E-mail: miika.tapio@mtt.fi

the loss of alleles treats species as a transiently subdivided entity. In contrast, a species can be considered as a potentially diversifying set of populations. Then the genetic conservation value is related to the current and potential uniqueness of the population rather than to the loss of genetic material. Since potential uniqueness depends on within-population variability, the sum of current and potential uniqueness can be estimated as the total contribution to molecular variation considering both divergence and variability of a population (Petit *et al.* 1998).

Simulations have indicated that increasing the number of conserved neutral marker alleles also increases the number of conserved non-neutral alleles (Bataillon *et al.* 1996). Selection on polygenic traits has only minor effects on the allele frequencies at a single influential locus, making it behave approximately as a neutral locus (Latta 1998). Large genetic distances also relate to higher expected heterosis in breed crosses (Graml & Pirchner 1984). However, there are two situations where the amount of neutral variation may not properly reflect the amount of non-neutral variation: in expanding populations, the adaptive variability accumulates much more slowly than the neutral genetic variation (e.g. at the microsatellite loci), and crosses of diverged, i.e. inbred, populations are likely to contain a relatively large amount of neutral variation, but little non-neutral variation (Hedrick 2001). Thus, the evaluation of population contributions to molecular variation can be based on neutral molecular diversity, but should be accompanied by information on admixtures between the breeds and variation of the effective population size.

In order to evaluate populations, a way of measuring diversity is needed. The most common measure of molecular variation is gene diversity (i.e. expected heterozygosity; Nei 1973). The additive genetic variance or the potential rate of immediate genetic change is proportional to gene diversity, but the selection limit is determined by the number of alleles, i.e. allelic richness (Allendorf 1986; Zeng & Cockerham 1990). Maintaining a high number of alleles is therefore an appropriate long-term conservation goal (Bataillon *et al.* 1996; El Mousadik & Petit 1996). El Mousadik & Petit (1996) noted that gene diversity and related divergence measures (e.g. *F*-statistics; Nei 1973) describe the distribution of the more common alleles and are therefore not ideal for guiding the conservation of large numbers of alleles. As a better alternative, they presented methods for analysing the distribution of allelic richness within and between populations. More recently, Barker (2001) discussed breed conservation in terms of conserving gene complexes and the related nonadditive variation. This would mean analysing loci simultaneously instead of averaging over separate single-locus analyses. In cattle, Ciampolini *et al.* (1995) used this kind of measure by computing distances between individuals based on the proportion of shared alleles over loci. However, they did not estimate the pro-

portion of total mean allele-sharing distance originating from breed divergence. This estimate is analogous to F_{ST} and is achievable if the individuals' multilocus genotypes are treated as alleles that differ from each other by some measurable amount, as in the case of DNA sequences (e.g. Pons & Petit 1996).

In the present study, total gene diversity, allelic richness and mean allele-sharing distance have been used to evaluate the genetic contributions by single sheep breeds in northern Europe, defined as the region consisting of Denmark, Iceland, Norway, Sweden, Finland, northwestern Russia, Estonia, Latvia, and Lithuania. Domestic sheep were introduced to southern Scandinavia, the British Isles and southwestern Russia \approx 6000 years ago or 4000 years after domestication (Ryder 1991). The last expansion of sheep before modern times was their introduction to Iceland, mainly from Norway, 1000 years ago (Adalsteinnsson 1981). The primary spread of sheep was completed when sheep were introduced to Greenland from Iceland, the Faeroe Islands and Scotland in 1906–1915. The extent of gene flow between the regions is not known. The local sheep populations were developed into landrace breeds in Europe in the period from the end of the 18th century to the First World War (Ruane 1999). The old northern European native breeds belong mainly to the Northern Short-tailed breed group, within which breeds vary greatly in fecundity, physical size, wool types and colours, and physical management. In northern Europe, more intensive meat production has led to the use of more standardized, newly created and imported long-tailed breeds along with a few popular native breeds in the 20th century. The present study included 32 breeds or strains from northern Europe: all the common native breeds and the recently created Scandinavian breeds, 11 rare native breeds with fewer than 1000 adult sheep, and six imported breeds (four British, one Dutch, and one Russian). The aim of the study was to evaluate the importance of each breed for gene diversity, allelic richness and mean allele-sharing distance between individuals by using microsatellite markers. The effects of breed divergence and within-breed variability (potential divergence) were separated. The breed histories were studied to ascertain the biological relevance of the marker-based estimates. The effect of within-breed inbreeding on the contribution of each breed to molecular variation was explored.

Materials and methods

Collection of data

Genetic variation at 25 microsatellite loci (Table 1) was studied in 924 sheep from 32 breeds in northern Europe (Table 2). Sample collection and genotyping using standard polymerase chain reaction (PCR) methods and semi-automatic fragment typing followed Tapio *et al.* (2003, 2005). Sampling

Table 1 The 25 microsatellite loci studied, their chromosomal location (*Chr*) and total number of alleles detected (A_T). The estimates of total gene diversity (h_T), and total within-population allelic richness (corresponding to the sample of 13 diploid individuals: r_T). The relative measures of differentiation are given based on gene diversity (G_{ST}), and allelic richness (ρ_{ST}). Deviation from Hardy–Weinberg expectations is also presented (F_{IS})

Locus	<i>Chr</i>	A_T	h_T	r_T	G_{ST}^\dagger	ρ_{ST}^\dagger	F_{IS}
BM0757	9	12	0.78	6.06	0.10	0.32	0.02
BM1314	22	18	0.79	8.60	0.13	0.41	0.03
BM1818	20	18	0.91	11.29	0.13	0.42	0.03
BM4621	6	21	0.87	9.23	0.14	0.34	0.03
BM6506	1	13	0.67	5.84	0.16	0.38	0.01
BM6526	26	15	0.77	7.18	0.14	0.39	0.04
BM8125	17	9	0.66	5.56	0.18	0.40	0.05
CSSM31	23	24	0.88	11.04	0.18	0.45	0.08
ILSTS002	14	11	0.80	6.43	0.15	0.31	0.16*
INRA023	1	15	0.89	9.64	0.13	0.35	0.03
MAF214	16	16	0.53	4.25	0.19	0.36	0.03
MAF36	22	16	0.87	9.55	0.17	0.42	0.03
MAF48	u.a.‡	12	0.81	6.84	0.15	0.31	0.06
MAF65	15	11	0.77	6.11	0.13	0.27	0.06
McM527	5	12	0.81	7.30	0.19	0.35	0.16
OarCP20	21	13	0.81	6.52	0.14	0.28	0.05
OarCP34	3	8	0.79	5.50	0.18	0.28	0.08
OarCP38	10	9	0.68	5.60	0.15	0.33	0.03
OarFCB11	2	11	0.78	5.70	0.08	0.18	0.05
OarFCB128	2	13	0.79	7.01	0.13	0.33	0.13
OarFCB304	19	23	0.76	7.99	0.14	0.41	0.08
OarFCB48	17	17	0.77	7.58	0.13	0.39	0.03
OarHH47	18	14	0.86	9.13	0.13	0.34	0.01
OarHH64	4	9	0.78	6.68	0.14	0.33	0.43*
OarVH72	25	9	0.75	6.00	0.15	0.31	0.04
Mean		14	0.78	7.31	0.14§	0.36§	0.07§

*Deviates significantly (after Bonferroni correction over loci) from HWE ($P < 0.05$). Based on 16 000 permutations of alleles within populations.

†Relative differentiation is the difference between mean within-population variation and total variation, divided by total variation.

‡Unassigned (u.a.).

§The means of numerator and denominator were taken separately.

aimed to cover the main breed stock by avoiding sampling of individuals, where pedigrees had overlapping in the two previous generations. In the rare breeds (the Ruhnú sheep in particular), avoiding all overlapping was impossible. In the Finnsheep and the Icelandic sheep, the rare subtypes (the Finnish Grey Landrace and the Icelandic Leader sheep) were sampled separately. The Norwegian Dala sheep included its three subtypes. Description of breeds, data collection and laboratory methods can be found at www.lbhi.is/northshed. Data previously described in Tapio *et al.* (2003) and Tapio *et al.* (2005), i.e. data on 11 microsatellites in the Grey Finnish Landrace, the Finnsheep, the Lithuanian Romanov and the Russian Viena sheep, and 21 microsatellites in the four Baltic breeds, were included in the data set.

Statistical analysis

Detection of anomalous loci. Population genetic inferences based on molecular variation require data on neutral and,

preferably, codominant loci. To test whether the loci studied were neutral, their properties were first evaluated using the method of Beaumont & Nichols (1996). In their approach, the observed population differentiation and gene diversity are used to guide the simulation of their expected joint distribution using a model of 100 populations with symmetric migration. The comparison of simulated loci to genotyped loci reveals outlier loci with significantly stronger or weaker differentiation than expected. The *FDIST* program (Beaumont & Nichols 1996) was used to simulate 20 000 (stepwise mutating) loci sampled from 32 populations. The median sample was 50 chromosomes per population. Next, the deviation from Hardy–Weinberg equilibrium (HWE) was measured as $F_{IS} (= 1 - H_O/H_E)$, where H_E is expected heterozygosity or gene diversity and H_O is observed heterozygosity, using the program *FSTAT* version 2.9.3.2 (Goudet 2001). The significance of deviations was tested using permutation tests. Deviation from HWE can be caused by population structure (i.e. the Wahlund effect),

Table 2 The breeds studied, their morphological classification based on tail length (Tail) and demographic breed classification (Breed type). All phenotypically short-tailed sheep belonged to the Northern Short-tailed breeds. 'Native' refers to old breeds that have been in the present country for more than 300 years; younger breeds are indicated as 'Modern'. In addition, known foreign introgression into native breeds is indicated with a plus sign and breeds with fewer than 1000 ewes are shown as 'Rare.' Table gives breed-wise information about sample size (n), gene diversity (h_k), allelic richness (corresponding to the sample of 13 diploid individuals (r_k), mean allele-sharing distance (s_k), relative breed uniqueness (divergence of the particular breed from the other breeds) based on gene diversity (G_{ST}), deviation from HWE (F_{IS}), autozygosity estimate (f) and imbalance index (β) and their mean values based on 22 microsatellite loci

Breed	Tail	Breed type	n	h_k	r_k	s_k	G_{ST}	F_{IS}	f	β
Åland sheep	Short	Rare native	25	0.66	4.81	0.59	0.15	0.043	0.16	2.63
Danish Landrace sheep	Long	Rare native+	21	0.60	4.07	0.54	0.18	0.034	0.22	3.05
Danish Texel	Long†	Imported	24	0.62	4.59	0.55	0.17	0.031	0.18	2.69
Danish Whiteheaded Marsh sheep	Long	Imported	28	0.63	4.85	0.56	0.16	0.037	0.14	3.00
Estonian Ruhnu sheep	–‡	Rare native	24	0.53	3.33	0.46	0.23	–0.007	0.34	8.84
Faeroe Island sheep	Short	Native	21	0.67	4.89	0.58	0.15	–0.026	0.15	2.88
Finnish Grey Landrace	Short	Rare native	30	0.72	5.30	0.62	0.12	0.011	0.12	2.17
Finnsheep	Short	Native	30	0.76	6.26	0.67	0.10	0.007	0.07	1.44
Greenland sheep	Short	Modern	17	0.71	5.35	0.64	0.14	0.035	0.14	2.32
Icelandic sheep	Short	Native	30	0.71	5.72	0.64	0.13	0.080*	0.09	2.13
Icelandic Leader sheep	Short	Rare native	35	0.65	4.70	0.56	0.16	0.063	0.13	2.92
Latvian Darkheaded	Long	Modern	32	0.70	5.81	0.62	0.13	0.034	0.08	2.06
Lithuanian Blackface	Long	Modern	30	0.72	5.89	0.65	0.10	0.039	0.08	1.73
Lithuanian Native Coarsewooled	–‡	Native+	30	0.74	5.83	0.66	0.13	0.028	0.09	1.77
Lithuanian Romanov	Short	Imported	31	0.68	5.24	0.60	0.15	0.025	0.12	2.35
Norwegian Cheviot sheep	Long	Imported	28	0.62	4.21	0.54	0.17	0.041	0.18	3.40
Norwegian Dala sheep	Long	Modern	55	0.73	6.08	0.65	0.11	0.121*	0.07	1.91
Norwegian Feral sheep	Short	Native	37	0.70	5.45	0.61	0.12	0.052	0.10	1.54
Norwegian Fuglestad sheep	Long	Imported	13	0.73	5.59	0.69	0.12	0.071	0.11	2.19
Norwegian Grey Troender sheep	–‡	Rare native	25	0.71	5.61	0.64	0.13	0.095*	0.11	2.24
Norwegian Rygia sheep	Long	Modern	28	0.70	5.28	0.63	0.14	0.091*	0.13	2.01
Norwegian Spael sheep	Short	Native+	30	0.71	5.43	0.63	0.13	0.091*	0.10	1.98
Norwegian Old Spael sheep	Short	Rare native	29	0.71	5.61	0.66	0.14	0.225*	0.11	1.79
Norwegian Steigar sheep	Long	Modern	30	0.72	5.73	0.64	0.10	0.068*	0.08	1.89
Russian Viena sheep	Short	Native	31	0.72	5.97	0.68	0.13	0.244*	0.08	2.24
Swedish Dala Fur sheep	Short	Rare native	29	0.56	3.76	0.52	0.22	0.191*	0.26	3.60
Swedish Finewool sheep	Short	Native+	32	0.73	5.97	0.65	0.11	0.082*	0.07	1.83
Swedish Gotland sheep	Short	Native	30	0.63	4.65	0.57	0.16	0.058	0.15	3.39
Swedish Gute sheep	Short	Native	20	0.60	4.28	0.52	0.21	0.067	0.21	3.97
Swedish Forest sheep	Short	Rare native	38	0.63	4.70	0.58	0.16	0.168*	0.16	2.98
Swedish Roslag sheep	Short	Rare native	30	0.38	2.62	0.33	0.34	0.009	0.41	5.72
Swedish Rya sheep	Short	Native	31	0.66	5.37	0.59	0.16	0.087*	0.12	2.49
Mean			29	0.67	5.09	0.60	0.15§	0.070§	0.14	2.72

*Population estimate significantly ($P < 0.05$) larger than expected (after Bonferroni correction over populations) based on 16 000 permutations of alleles within populations.

†The Texel is a Marsh-type sheep, and these are short tailed. However, the Danish Texel is long tailed.

‡The phenotype is unclear (variable or intermediate length).

§The means of numerator and denominator were taken separately.

which may give false positives in a search for loci with nonamplifying alleles. However, a consistent deviation at a locus can be taken as evidence of the occurrence of non-amplifying alleles or selection acting on the locus. The possibility of false positives was tested using the nonparametric Friedman two-way analysis of variance (ANOVA) test (Sokal & Rohlf 1981) to explore whether the ranks of F_{IS} values at loci were random.

Molecular variation. Molecular genetic variation was measured using gene diversity (h ; Pons & Petit 1995) and allele number corrected for sample size (allelic richness, r ; El Mousadik & Petit 1996). The CONTRIB program (Petit *et al.* 1998) was used. In addition, differences between individuals were studied using mean allele-sharing distance between individuals (s) and applying the analysis of Pons & Petit (1996), originally derived for alleles with quantifiable

differences (e.g. DNA sequences). This measure was calculated as follows: (i) The proportion (P) of common alleles (A) over L loci is $A/2L$. (ii) The empirical distance between individuals is $d = 1 - P$. (iii) The mean within-population distance for the population k (\hat{s}_k) was estimated using the empirical mean distance between individuals within the population (\bar{d}_{kk}):

$$\hat{s}_k = \frac{n_k}{n_k - 1} \bar{d}_{kk},$$

where n_k is the number of individuals studied in population k . The overall mean within-population distance (\hat{s}_G) is the arithmetic mean over population-wise values. The mean distance in the total population (\hat{s}_T) was estimated as

$$\hat{s}_T = \left(1 - \frac{1}{n_p(n_p - 1)} \right) \bar{d}_{kl},$$

where n_p is the number of studied populations and \bar{d}_{kl} is the empirical mean allele-sharing distance between individuals from different populations.

The differentiation (the proportion of total variation due to subdivision) was estimated as described by Pons & Petit (1995) for gene diversity and as described by El Mousadik & Petit (1996) for allelic richness using the CONTRIB program (Petit *et al.* 1998). The differentiation measure was estimated for the mean allele-sharing distance using the formula for gene diversity, where the gene diversity estimates were replaced with the mean allele-sharing distance estimates.

Breed contributions. A breed's contribution to total gene diversity (h_T), allelic richness (r_T) and mean allele-sharing distance (s_T) was estimated. This contribution has been defined as the difference between total diversity including all populations and diversity without one specified population (Petit *et al.* 1998). Breed contributions to h_T and r_T were obtained using the CONTRIB software (Petit *et al.* 1998). CONTRIB was also used to estimate the effects of the within-breed variability and breed uniqueness (i.e. differentiation between the breed and the breed pool) of each breed on h_T and r_T . The breed contributions and the effects of within-breed variability and breed uniqueness on s_T were estimated as described by Petit *et al.* (1998) for allelic richness, but the allelic richness-based estimates were replaced with the mean allele-sharing distance estimates.

Breed relationships. The phylogenetic origin of breeds and individuals was studied using independent components analysis (ICA; see Stone 2002 for an informal introduction). The ICA can be considered as an extension of principal components analysis (PCO). Cavalli-Sforza *et al.* (1994) described the principal component (PC) value as weighted mean allele frequency. The weighting in PCO maximizes the variance explained by the axis, which can therefore be

described as a weighting scheme. In ICA, the weighting aims to maximize the variance and all other higher statistical moments explained by the axes. These parameters give information on the deviation from normal distribution and detect structures in the data. ICA was carried out in two steps. First, PCO was performed for standardized allele frequencies according to Cavalli-Sforza *et al.* (1994) using the ADE-4 software (Thioulouse *et al.* 1997) to create a compacted data matrix. Second, this matrix was used as input for the FastICA algorithm (Hyvärinen 1999) as implemented in the fastICA library for the R language (Marchini *et al.* 2003). Since the algorithm uses observations on only the same number of PCs as the number of independent components (ICs) searched, only the appropriate columns were included. The algorithm was run using a logcosh (with $\alpha = 1.5$) approximation of neg-entropy (a measure of non-normality) and a parallel search of components. The search was limited by setting the convergence criteria as 10^{-7} for the un-mixing matrix and 3×10^6 for the maximum number of iterations. The IC values are in standard deviation units. In the ICA for individuals, positive and negative ranges were considered separately and the IC values were converted into the proportion of explained variation. This is equal to the square of the IC value multiplied by the total variance explained by the component. Explained variances for the components were obtained from the sum of squares for the corresponding parts of the mixing matrix, which describes conversion from ICs into PCs.

Inbreeding and changes in the effective population sizes. The amount of inbreeding was estimated using the coalescence method of Ciofi *et al.* (1999) applied in their 2MOD program. In estimating this autozygosity, that is the probability of two alleles sharing a common ancestor within the population, the possibility of new mutations was ignored. The method employs two alternative models: (i) a model of a founder population fragmented into a number of isolated populations and (ii) an island model of populations in immigration-drift equilibrium. The likelihood of the data under each model controls the use of the two models along the Markov chain Monte Carlo (MCMC) run. The model usage proportion estimates the posterior probability of the model. The results were based on 150 000 MCMC iterations, where the first 10 000 iterations were excluded as a burn-in phase. The robustness of the results was evaluated by running 20 shorter chains (10 000 iterations) initializing half of the chains with the founder-fragmentation model and half with the immigration-drift equilibrium model.

The inbreeding estimates described above have a simple biological meaning, but they depend on founder-population allele frequencies inferred using simplified assumptions (no mutations and star phylogeny or island model for the populations). Therefore the estimates need to be confirmed using an estimate independent of these assumptions.

Assuming that there was a single, not-too-ancient founder population, the differences between the populations with respect to inbreeding are caused by subsequent changes in the effective population size (N_e); a growing population might not experience any increase in inbreeding, whereas a decreasing population accumulates inbreeding rapidly. The microsatellite allele size variance responds more slowly to changes in population size than the microsatellite gene diversity does (Kimmel *et al.* 1998). The ratio of N_e estimates based on these two measures, called the imbalance index (β ; Kimmel *et al.* 1998), reveals changes in effective population size. The ratio is expected to be 1 for an equilibrium population of constant size, when the allele size changes occur according to the stepwise mutation model without constraints on the repeat number. Multistep mutations may increase and size constraints may decrease the equilibrium value, but the value will still become smaller in expanding population and larger in decreasing population. The imbalance index estimate for the founder population before fragmentation was based on the most likely allele frequencies estimated by the 2MOD software, as explained above. The imbalance index was calculated for 10 points on the long MCMC run. Starting from iteration number 60 000, the result from every 10 000 iterations was used and the imbalance index was taken as a mean of the 10 estimates. The imbalance index was estimated according to Kimmel *et al.* (1998, p. 1927).

Results

Loci

In total, 363 alleles were detected at the 25 microsatellite loci assessed in 924 sheep. The number of alleles ranged from 8 to 24 per locus (Table 1), and the number of breed-specific alleles per locus ranged from none to four with a mean of 2.0. The total locus-wise gene diversity varied between 0.53 and 0.91 and the total allelic richness (corrected for sample size) ranged from 4.25 to 11.29 (Table 1). Breed differentiation (G_{ST}) accounted for 8–19% of the total gene diversity at the loci. For allelic richness, the corresponding proportion (ρ_{ST}) was 18–45%. The difference between G_{ST} and ρ_{ST} is caused by the stronger influence of the rare alleles on ρ_{ST} .

The genetic variation at three of the microsatellite loci studied indicated either selection or the presence of nonamplifying alleles. The simulated data, which had the same mean differentiation as the actual data, indicated that over 95% of the loci with a gene diversity over 0.5 should show differentiation between 0.1 and 0.2. The breed differentiation at *OarFCB11* ($h_T = 0.78$, $G_{ST} = 0.08$; Table 1) was significantly outside the expected range. Only the Roslag sheep was as diverged at *OarFCB11* as at the other loci (not presented). The locus generally demonstrated a low allele

number compared to the gene diversity across the breeds, indicating a locus-specific bottleneck (test of Cornuet & Luikart 1996, not presented). The two observations indicated that directional selection favouring the same alleles in different breeds has affected the marker. There were two other loci that appeared anomalous: *OarHH64* and *ILSTS002* deviated significantly ($P < 0.05$) from HWE expectations (Table 1). The deviations from HWE are probably caused by nonamplifying alleles rather than by population structure, since the F_{IS} estimates for the two loci were nonrandomly ranked within the breeds (significant or nearly significant locus effect, $P < 0.052$, disappeared when both loci were excluded). For these reasons, *OarFCB11*, *OarHH64* and *ILSTS002* were excluded from the population analyses.

Breed variation and differentiation

Based on the results from the 22 microsatellite loci, within-breed gene diversity ranged from 0.38 for the Swedish Roslag sheep to 0.76 for the Finnsheep (Table 2). The same breeds also showed the extreme values for allelic richness, which varied from 2.62 to 6.26. The Swedish Roslag sheep demonstrated the smallest mean within-breed allele-sharing distance (0.33), while the largest mean allele-sharing distance (0.68) was observed for the Russian Viena sheep. The positive F_{IS} value (Table 2) suggested that the high estimate for the Viena sheep was partly a result of within-breed structure. Breed divergence accounted for 15% of the total gene diversity for all breeds ($h_T = 0.78$), and a slightly larger proportion (17%) of the total mean allele-sharing distance ($s_T = 0.72$). Breed differentiation had a substantially larger influence (37%) on total allelic richness ($r_T = 7.45$), which is more affected by the rare alleles. Breed uniqueness based on gene diversity (G_{ST} ; Table 2), i.e. the differentiation of a particular breed from the pool of other breeds, ranged from 0.10 to 0.34.

Breed contributions to total variation

Breed contributions to total variation and the influence of breed uniqueness and within-breed variability on this value are presented graphically in Fig. 1. For gene diversity, a positive contribution indicates an increase in the probability of two random alleles being different, while for allelic richness it indicates an increase in the expected allele number in a sample of 26 chromosomes. Similarly, a positive contribution to mean allele-sharing distance indicates an increase in the distance between two random individuals. The mean breed contribution to gene diversity and mean allele-sharing distance was zero, while the mean contribution to allelic richness was slightly negative (−0.13; Fig. 1b). This is due to similar allele frequency distributions. As a result, there are not many breed-specific alleles, and when they do

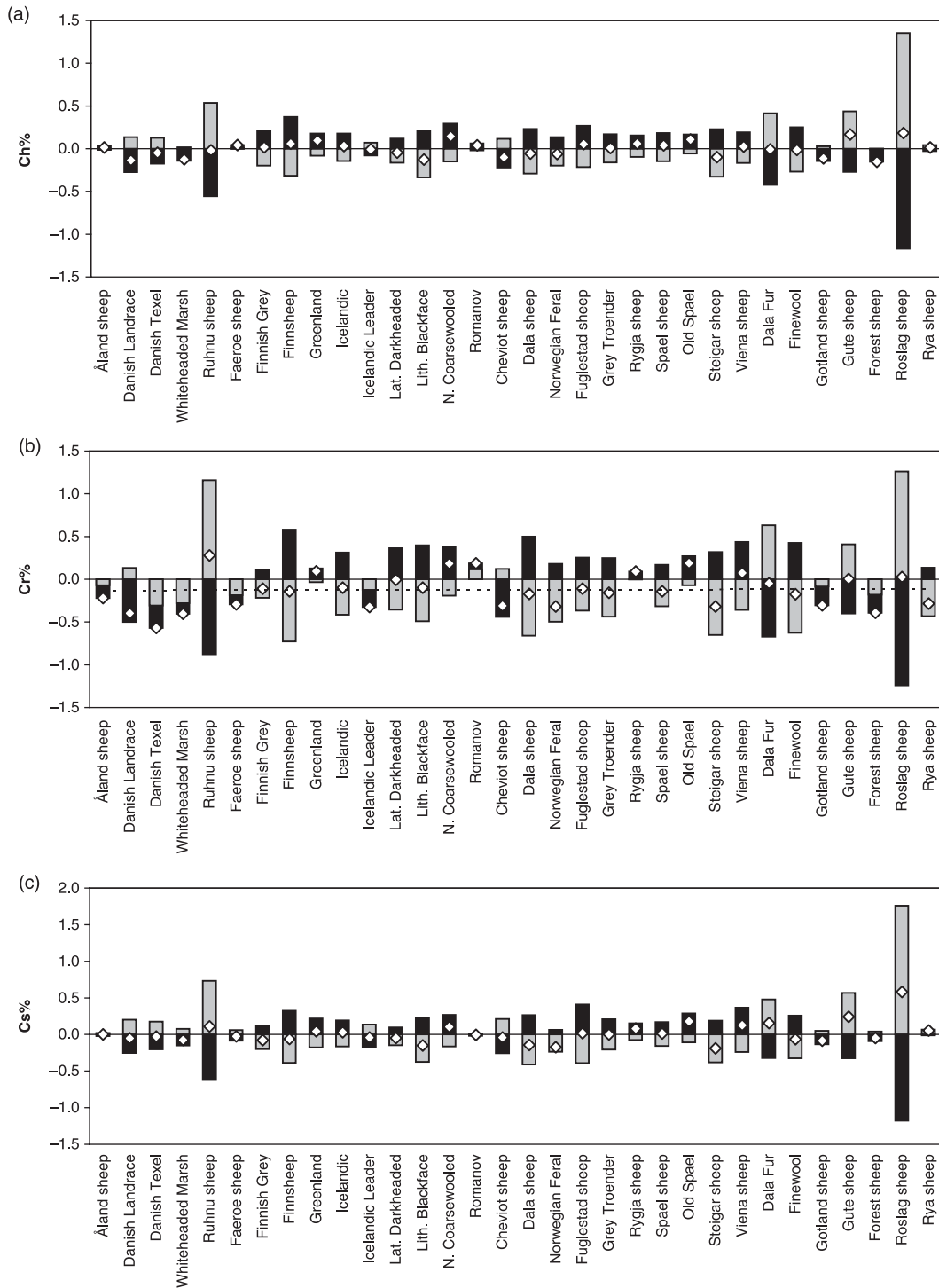


Fig. 1 Breed contributions to total gene diversity (a), total allelic richness (b) and total mean allele-sharing distance (c) plotted as diamonds. The mean contribution in (a) and (c) is zero, while in (b) it is -0.13 , which is indicated in (b) with horizontal broken line. The influence of within-breed diversity and breed uniqueness on the total breed contribution is presented as black and grey bars, respectively.

occur, their frequencies are low. Also, the alleles that are very common in related breeds are over-represented in the entire set. Excluding one of these related breeds equalizes the allele frequencies in the population set. Together, this

equalizing effect and the rarity of breed-specific alleles cause the expected number of alleles in a sample (i.e. allelic richness) to increase, though the total number of alleles decreases when a breed is excluded.

There was a significant but moderate positive correlation ($0.5 < r < 0.75$) between the breed contributions to the three types of diversity. The contributions of the Danish Texel were distinctively unequal, and the breed had a clearly more negative effect on allelic richness than on gene diversity or mean allele-sharing distance (Fig. 1). There was greater variation in the effect of within-breed diversity and breed uniqueness on breed contributions to molecular variation than in the breed contributions themselves. The way diversity was measured had less influence in evaluating these effects ($r > 0.9$ between the diversity measures) than in evaluating the breed contributions to molecular variation.

Of the 32 breeds studied, 21 made an above-average contribution with respect to at least one diversity measure. If maintaining a high number of alleles is considered to be the long-term goal, allelic richness is the most important measure of the three. The five Baltic breeds, the Finnish Grey Landrace, the Greenland sheep, the Icelandic sheep, the Norwegian Fuglestad, Rygja, and Old Spael sheep, the Russian Viena sheep and the Swedish Gute and Roslag sheep showed above-average contributions to total allelic richness (Fig. 1b). Of the 11 breeds that did not make an above-average contribution to any measure (Fig. 1a–c), six showed low within-breed variation (the Danish Landrace, Texel, and Whiteheaded Marsh sheep, the Icelandic Leader sheep, the Norwegian Cheviot sheep, and the Swedish Gotland and Forest sheep), while four were found not to be very divergent (the Norwegian Dala, Feral and Steigar sheep and the Swedish Finewool sheep) on the basis of the microsatellite data.

Single diversity measure may be insufficient to adequately evaluate the contributions to variation. The Greenland sheep, the Icelandic sheep, the Lithuanian Native Coarsewooled, the Norwegian Fuglestad, Rygja and Old Spael sheep, the Russian Viena sheep, and the Swedish Gute and Roslag sheep made above-average contributions to molecular variation according to all measures and are clearly important contributors. The means and variances of the three types of contributions (Fig. 1) were standardized before taking the average for the breed. In addition to the nine breeds listed above, the Ruhnu sheep, the Finnsheep, the Spael sheep, the Romanov, the Dala Fur sheep, and the Rya sheep appeared to be more important than the average breed. Constructing a significance test is difficult because the three analyses are not independent from each other, the mean contribution to allelic richness is not fixed at zero, and only h and r treat each locus as an independent sample. Nevertheless, these 15 breeds can be considered to be the most important contributors to genetic diversity in northern European sheep if within-breed variation and breed uniqueness are given the same weight and gene diversity, allelic richness and mean allele-sharing distance are regarded as equally important.

Breed relationships

The breed relationships of northern European sheep were studied using the ICA analysis. The phylogenetic structure was weak and 10 components would be needed to describe over 50% of the allele frequency differences between the breeds. There was a drop in explanatory power after the first two PCs, which together explained 16.1% of the differences. Therefore, two ICs were extracted. The first IC (IC1) separated the Northern Short-tailed breeds (Fig. 2) from the other breeds. The Grey Troender Sheep, the Lithuanian Native Coarsewooled and the Danish Landrace are usually considered as Northern Short-tailed breeds, but their tail is not clearly short as in the other northern breeds where IC1 was below 0.3 (Fig. 2, Table 2). Within the Northern Short-tailed group, three clusters were evident based on IC2: a northwestern cluster including the Icelandic sheep and related breeds, a northeastern cluster including the Finnsheep and related breeds, and a third heterogeneous cluster of short-tailed breeds including most Swedish and Norwegian breeds (Fig. 2). The Norwegian

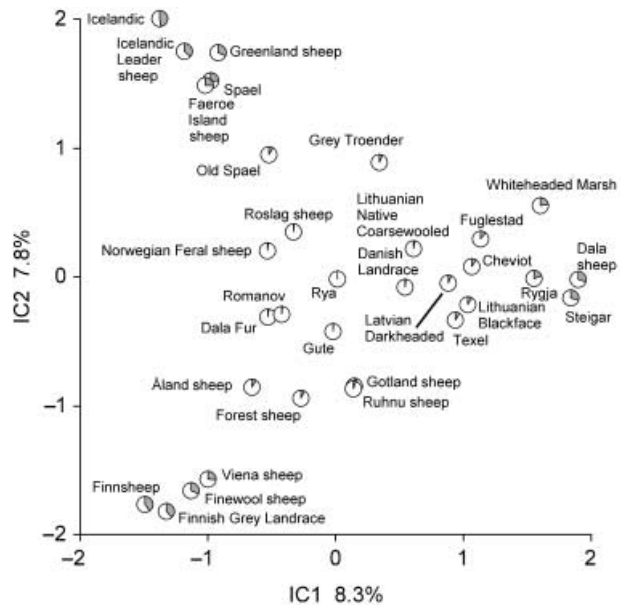


Fig. 2 Plot of two-component solution of ICA for the northern European sheep breeds. The location of the pie diagrams indicates the breed values. In each pie diagram, the dark proportion equals the proportion of the allele frequency variance within the population explained by the two components. For each population, the sum of the variation explained by IC1 and IC2 is identical to the variation that would be explained by the first two PCs together. The most important information for the centrally located breeds is that they do not belong to the three surrounding groups. The breeds with $IC1 \approx 0.5$ have intermediate or variable tail-length being intermediates between Northern Short-tailed ($IC1 < 0.3$) and modern long-tailed breeds ($IC1 > 0.75$; see text), but they are often categorized into the Northern Short-tailed breed group.

breeds appeared to be closer to the northwestern breeds and the Swedish breeds closer to the northeastern breeds, which is consistent with their geographical distribution.

The ICA was done without grouping the individuals into breeds before the analysis, in order to recognize breeds with several diverged ancestries. There was no distinct drop in the explanatory power of PCs here. The number of components investigated was determined on the basis of breed differentiation. With 14 components the cumulative explanatory power (21%) reached the level of breed differentiation (20% of the standardized allele frequencies between individuals). The breed averages for the explained within-individual variance varied between 6% (the Fuglestad sheep) and 59% (the Roslag sheep). Explained proportions reflect drift, which creates systematic differences between sheep in different populations. For the purebred populations, the mean explained proportion also reflects breed uniqueness (G_{ST} ; Table 1). Direct comparison with breed uniqueness was impossible, since the allele frequencies were standardized for ICA in order to equalize the allele effects. Most of the breeds studied were found to have multiple ancestries (Fig. 3). It is worth noting that each indicated ancestry does not necessarily indicate a separate hybridization. For instance, there were two ancient Finnish sheep types, an eastern type (from the same region as the

Finnish Grey Landrace) and a southwestern type (from the region nearest the Åland Islands), and the Finnsheep was mainly based on the eastern type at the beginning of 18th century (Majjala 1988). This agrees with the two ancestries for Finnsheep indicated in Fig. 3. Genetic material from the Finnsheep has been introgressed to the Swedish Finewool sheep recently. This introgression brought both of the two Finnsheep ancestries into the Finewool sheep at the same time. The Finewool also demonstrated an ancestry not observed in the Finnsheep, but observed, e.g. in the Swedish Gotland sheep (Fig. 3).

The aim of the study was to identify crossing of diverged ancestries, not admixtures only. There may be a larger difference between the amounts of neutral and non-neutral variation in the crossbred than in the purebred populations (Hedrick 2001). The breed mean values for explained within-individual variance were used to describe the effect of admixture and ancestral drift in a single number. An 'ancestry diversity index' was calculated by multiplying the breed mean values by each other and multiplying the result by two. The value of this ancestry diversity index is large when there are several major ancestries, but small in an isolated population or when the explained proportion is small. The Danish Texel and the Swedish Gotland sheep showed much higher values than the other breeds (Fig. 3),

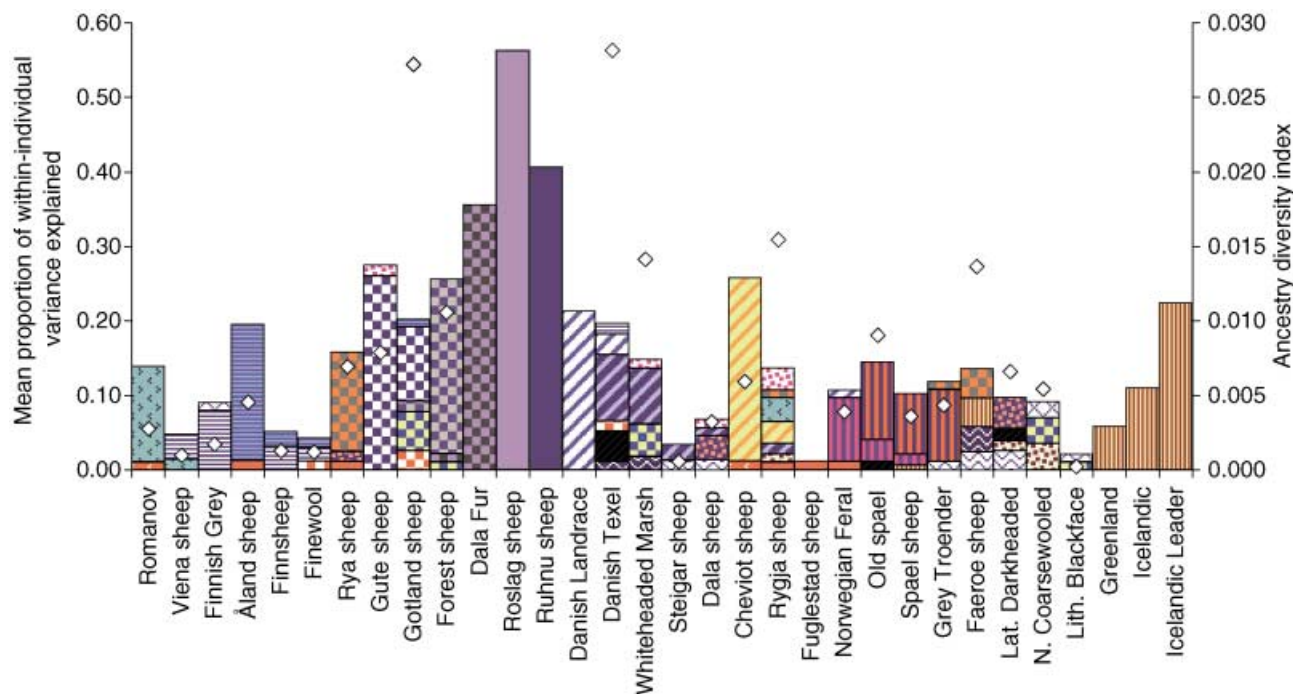


Fig. 3 The bars present the ICA results as mean proportion of explained within-individual variance (left axis). Positive and negative ranges of IC axes were treated separately when they occurred. Therefore, the total number of plotted ancestries is 26 and each is represented with a different colour pattern. The sizes of these partitions represent divergence in one dimension resulting from within-breed inbreeding. Admixed breeds show several partitions indicating the influence of ancestral divergence of source populations. This influence is quantified using 'ancestry diversity index', which is presented with diamonds (right axis) and was calculated as the explained proportions multiplied by each other times two. Only partitions > 1% are shown.

which suggest a larger difference between the amounts of neutral and non-neutral variation in these breeds.

Inbreeding and changes in effective population size

The founder-fragmentation model explained the molecular data better than the migration–drift equilibrium model. All but one of the 20 independent MCMC runs converged to the fragmentation model before the 1000th iteration, and the long MCMC run used for the estimation of autozygosity utilized this model alone after the burn-in phase. The average within-breed inbreeding was 0.14, and varied from 0.07 to 0.41 (Table 2). The inbreeding estimates were closely correlated ($r = 0.85$, $P < 0.001$) with the imbalance indices (β) of Kimmel *et al.* (1998; Table 2), reflecting changes in population size. This suggested that the simplified star-shaped phylogeny and lack of mutations assumed in the estimation of inbreeding did not greatly bias the inbreeding estimates. The only exception was the Estonian Ruhnu sheep, which showed a large imbalance in relation to the inbreeding estimate. This discrepancy is explained by the fact that the Ruhnu sample included nearly the entire population, violating the assumption of the standard coalescent model that population size greatly exceeds sample size (Wakeley & Takahashi 2003).

The imbalance index value for the inferred founder population ($\beta = 1.05 \pm 0.01$) was very close to one, which is expected for a constant-sized population, assuming that mutations always change the microsatellite sizes by one repeat unit. All the imbalance indices for the breeds were larger than one ($1.44 \leq \beta \leq 8.84$; Table 2) and suggested genetic bottlenecks. Likewise, the imbalance index for the current northern European sheep metapopulation ($\beta = 1.23$) suggested a bottleneck, although population structure may also increase the imbalance (Kimmel *et al.* 1998). Nevertheless, the inferred gene diversity for the founder population (0.82 ± 0.08) was greater than the total gene diversity for the current metapopulation.

Discussion

The present study investigated the history of northern European sheep breeds and their contributions to molecular variation. The study indicates that there was an initial founder population formed by local sheep types and that this was fragmented into isolated breeds. The effective population size of all the breeds has decreased during their history, although only the Grey Finnish Landrace and the Ruhnu sheep proved to have significantly fewer alleles than expected, indicating a recent reduction in effective population size [results not presented, test of Cornuet & Luikart (1996) performed as in Tapio *et al.* (2003)]. Thus most of the breed-wise reductions took place more than $2N_e$ to $4N_e$ generations ago (Luikart *et al.* 1998). A substantial

proportion of the breeds that were found to make an above-average contribution to genetic diversity in northern Europe are rare local breeds numbering fewer than 1000 ewes (Table 2). Considering that breed differentiation accounted for 18–45% of the total allelic richness at the microsatellite loci, it appears that conservation efforts will be very important in maintaining genetic diversity.

Phylogenetic structure was not taken into account in the evaluation of breeds, and this simplification was supported by the weakness of the structure (Fig. 2). The proportion of breed differences explained by the first two components is approximately two-thirds of that observed in corresponding European studies in cattle (Kantanen 1999; Cañón *et al.* 2001). Further, the overall breed differentiation, measured as G_{ST} , was 0.15, which exceeds previously reported estimates both in sheep (Arranz *et al.* 1998; Tapio *et al.* 2003, 2005; Álvarez *et al.* 2004) and in other domestic species (Kantanen *et al.* 2000; Laval *et al.* 2000; Cañón *et al.* 2001) and it would be considered as strong differentiation in any species. Thus, the breeds studied here are strongly differentiated and do not form very tight groups. This means that phylogenetic structure is relatively unimportant in evaluating the breed contributions. This was confirmed by the observation that even the two tightest breed groups from geographically peripheral regions, the northwestern and the northeastern breed groups, show above-average contributions to molecular variation. The proportion of within-population variation explained by phylogenetic structure was largest in these breeds (Fig. 2) and both breed uniqueness and breed contribution to molecular variation are more likely to have been undervalued for these breeds than for the others.

Speciation modelling and comparative sequencing studies have implied the selective sweeps to be able to maintain species as a unit despite of low migration rates (Riesenberg *et al.* 2003). Present observation of a selective sweep adds evidence from a strongly subdivided domestic species. Further, the old breeds correspond to natural peripheral or more isolated populations as the opposite of 'central' undifferentiated population (Tapio *et al.* 2005) formed by the modern breeds. Similar to the natural peripheral populations (Lesica & Allendorf 1995), the rare breeds were disproportionately important for genetic variation. For many of the sheep breeds studied here, the contribution to variation with respect to breed uniqueness outweighed the disadvantage of lack of within-breed variation. The uniqueness and within-population diversity were not independent. Generally, unique breeds were poor in internal variation, while very variable breeds were not as diverged. This differs from observation in the argan tree (Petit *et al.* 1998) and the divergence and variability likely become more independent when a large proportion of variation is lost, e.g. through population extinctions.

Inbreeding has been seen as an important step in speciation, although presently the isolation of the peripheral

populations is considered to be more important for emergence of novel adaptations (Riesenberg *et al.* 2003). Among the sheep breeds, the inbreeding estimates had no significant correlation with the breed contributions to h_T and r_T ($r < 0.2$). However, higher inbreeding levels co-occurred with larger breed contributions to total mean allele-sharing distance ($r = 0.7$, $P < 0.001$), although the four breeds demonstrating the most extreme values largely define the correlation. In other words, the inbred populations are not automatically better or worse sources of single-locus variation, but inbreeding generates some diversity by creating very divergent individuals. Apparent loss of within-population variability does not always cause reduced viability in harsh environments (Holm *et al.* 1999; Visscher *et al.* 2001; Aguilar *et al.* 2004).

The moderate correlations between the breed contributions to h_T , r_T and s_T indicate that the importance of a population can vary depending on the type of variation or the timescale considered. Imperfect correlation is expected for example because the sensitivity of gene diversity and allele numbers to population size changes is different (Cornuet & Luikart 1996). The 15 most important breeds were identified by considering within-breed variation to be as important as breed uniqueness and gene diversity, allelic richness and mean allele-sharing distance to be of equal importance. A case could be made for increasing the weight given to allelic richness in order to preserve rare alleles. Heavier weighting of allelic richness in the present analysis would be more favourable for the Finnish Grey Landrace, the Latvian Darkheaded and the Lithuanian Blackface, but their importance depends on how much additional weight is assigned to allelic richness. On the other hand, the real conservation value of the imported or newly created breeds, e.g. the Norwegian Fuglestad sheep and the Norwegian Rygja sheep, depends on the presence of similar sheep in other countries.

The different emergence rate of neutral and adaptive variation or hybridization of inbred populations, which was raised as complications by Hedrick (2001), appears not to be of great concern in northern European sheep. The effect of mutation rates should be small because effective population sizes have decreased in these breeds, and in addition, their evolutionary history is short. Second, the results did not typically suggest that breeds with a mixture of diverged ancestries would contribute much, with the exception of the Rygja sheep. Hybridization of diverged populations leads neutral within-population variation to increase in relation to the amount of non-neutral variation. This is explained by the fact that although exceptional, alleles may become fixed in inbred or diverged populations when neutral variation is considered, at functional genes selection may more often restrict fixation to the common allele (Hedrick 2001). However, the increase in within-breed variation appears to happen at the expense of breed unique-

ness, which limits the increase in contribution to molecular variation.

The effect of hybridization on gene diversity is different from its effect on allelic richness. Gene diversity measures the evenness of allele frequencies and can be considerably heightened by crossing with an inbred population, whereas the low allele number in an inbred population means that there is little effect on allelic richness. In this study, the Danish Texel demonstrated a substantially more negative contribution to allelic richness than to gene diversity or mean allele-sharing distance. This fits with the ancestry diversity index for the Danish Texel, which was higher than for any other breed (Fig. 3). The ancestry diversity index summarizes admixture and ancestral inbreeding. It was nearly as large for the Swedish Gotland sheep, for which the differences between contributions were less extreme. There might be differences between the divergences of ancestries. With more drift, fewer alleles show frequencies close to the founder values (Kimura 1955). Therefore, a more divergent component should reflect frequency differences more evenly across the alleles. It is likely that the ancestries of the Gotland sheep are more closely related than those of the Danish Texel. The variances of allele effects were not large since all the allele effects were below 5% (not presented), but the variances of the allele effects are larger in the Gotland sheep than in the Danish Texel. Similarly, the variance in the allele effects for most components in the Rygja sheep was relatively large (not presented).

The high importance found for native breeds agrees with phenotypic information, since these breeds harbour more phenotypic variation than the newly created or imported breeds, e.g. in body size, fecundity, wool types and colours (www.lbhi.is/northshed). Twelve of the 32 breeds studied have been reported to be especially adapted to marginal environments. Six of these 12 hardy breeds are not included in the set of 15 important breeds. Although the breeds did not form tight groups, the common origins of breeds suggest similarity in the underlying genetic construction of polygenic adaptations. The Icelandic Leader sheep is missing, but the set includes the related Icelandic sheep and Greenland sheep. Another breed that is omitted, the Faeroe Island sheep, is also related to these breeds (Fig. 2). Although the Norwegian Feral sheep, which survives Norwegian winters without additional feeding, is not included, the set includes other Norwegian short-tailed breeds. The Swedish Gotland sheep is not in the set of 15 important breeds, but the related Gute sheep is included. Similarly, the ancestry of the Åland sheep appears to be represented in other breeds. At least some, although not all, of the ancestries (Fig. 3) in the imported Whiteheaded Marsh sheep in Denmark are represented in the priority breeds. This comparison suggests that much of the polygenic adaptation to low-input husbandry systems

would be conserved in the set of 15 breeds. Marker information can be used to rank phenotypically similar breeds or to find breeds that might be more valuable than suggested by simple phenotypic evaluation. Phenotypic evaluations are important because selection at genes or gene combinations with a major effect might strongly affect their differentiation. The documented valuable traits, such as the leading instinct of the Icelandic Leader sheep (Dyrmondsson 2002), cannot be refuted on the basis of neutral markers.

The ICA results provide many details of the genetic structure of northern European sheep. The ICA method used in the present study can be viewed as an axis rotation of results from PCO. Direct use of PCO for individuals was not suitable because the direction of several principal component axes was determined solely (results not presented) by the sheep from the most divergent breeds (the Roslag sheep and the Ruhnu sheep; Table 2). The main difference between model-based clustering (Pritchard *et al.* 2000) and multivariate analysis is that the first infers panmictic populations that fit the observed data and describes individuals in relation to these populations, whereas the method used here aims only to detect systematic differences in the allele frequencies of individuals. The PCO step in the analysis ensures that the strongest data structures will be seen even if the amount of residual variation is high. In our case, the amount of explained variation was set to match breed differentiation. The choice of components affects the picture seen. Decreasing the number of components shifted the focus to coarser structures, which is similar to the decrease in the number of assumed populations in model-based clustering (Pritchard *et al.* 2000). For instance, it was noted that when fewer than 10 components were searched, one component explained a minute amount of variation in the Åland sheep but a large amount in the Gotland and Gute sheep. With a larger number of components, the explained within-individual variation in the Åland sheep accumulated in a single component, which was separate from the main component for the other two breeds. Thus, it is not certain that a single ICA analysis can always describe complex hierarchical data fully. However, ICA results as presented in Fig. 3 provide a brief summary of past genetic drift, and the ICA offers a powerful and relatively nonintensive computational method for analysing extensive population data that are strongly subdivided.

In conclusion, conservation decisions based on microsatellites are not greatly compromised by population expansion or hybridization of diverged ancestries in northern European sheep. The present study indicates that the domestic species resemble the wild species in the respect that the peripheral or isolated populations/breeds are important contributors, especially concerning allelic variation.

Acknowledgements

The authors would like to thank Professor Asko Mäki-Tanila, Zofia Kurowska and anonymous reviewers for commenting on the manuscript. We would also like to thank Bente Flügel (Denmark), Ole-Albert Guttersrud (Norway) and Anneli Virta (Finland) for their skilful technical assistance with genotyping. We are indebted to a number of people for their help in blood sampling. This work was financially supported by the Nordic Gene Bank for Farm Animals (NorthSheD project, www.lbhi.is/northshed) and the Department of Education in Finland (Finnish Graduate School in Population Genetics).

References

- Adalsteinsson S (1981) Origin and conservation of farm animal populations in Iceland. *Zeitschrift für Tierzüchtung und Züchtungsbiologie*, **98**, 258–264.
- Aguilar A, Roemer G, Debenham S *et al.* (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences, USA*, **101**, 3490–3494.
- Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*, **5**, 181–190.
- Álvarez I, Royo LJ, Fernández I *et al.* (2004) Genetic relationships and admixture among sheep breeds from northern Spain assessed using microsatellites. *Journal of Animal Science*, **83**, 2246–2252.
- Arranz JJ, Bayo'n Y, San Primitivo F (1998) Genetic relationships among Spanish sheep using microsatellites. *Animal Genetics*, **29**, 435–440.
- Barker JSF (2001) Conservation and management of genetic diversity: a domestic animal perspective. *Canadian Journal of Forest Research*, **31**, 588–595.
- Bataillon TM, David JL, Schoen DJ (1996) Neutral genetic markers and conservation genetics: simulated germplasm collections. *Genetics*, **144**, 409–417.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **263**, 1619–1626.
- Cañón J, Alexandrino P, Bessa I *et al.* (2001) Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genetics Selection Evolution*, **33**, 311–332.
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes*. Princeton University Press, Princeton, New Jersey.
- Ciampolini R, Moazami-Goudarzi K, Vaiman D *et al.* (1995) Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds. *Journal of Animal Science*, **73**, 3259–3268.
- Ciofi C, Beaumont MA, Swingland IR, Bruford MW (1999) Genetic divergence and units for conservation in the Komodo dragon *Varanus komodoensis*. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **266**, 2269–2274.
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.
- Dyrmondsson OR (2002) Leadersheep: the unique strain of Iceland sheep. *Animal Genetic Resources Information*, **32**, 45–48.
- El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the arvan tree [*Arvania*

- spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832–839.
- Goudet J (2001) *FSTAT, A program to estimate and test gene diversities and fixation indices*. Available from www.unil.ch/izea/software/fstat.html.
- Graml R, Pirchner F (1984) Relation of genetic distance between cattle breeds and heterosis of resulting crosses. *Animal Blood Groups and Biochemical Genetics*, **15**, 173–180.
- Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology & Evolution*, **16**, 629–636.
- Holm L-E, Forchhammer MC, Boomsma JJ (1999) Low genetic variation in muskoxen (*Ovibos moschatus*) from western Greenland using microsatellites. *Molecular Ecology*, **8**, 675–679.
- Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, **10**, 626–634.
- Kantanen J (1999) *Genetic diversity of domestic cattle (Bos taurus) in North Europe*. PhD Thesis, University of Joensuu Publications in Sciences, 52, University of Joensuu.
- Kantanen J, Olsaker I, Holm L-E *et al.* (2000) Genetic diversity and population structure of 20 North European cattle breeds. *Journal of Heredity*, **91**, 446–457.
- Kimmel M, Chakraborty R, King JP *et al.* (1998) Signatures of population expansion in microsatellite repeat data. *Genetics*, **148**, 1921–1930.
- Kimura M (1955) Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences, USA*, **41**, 144–150.
- Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *American Naturalist*, **151**, 283–292.
- Laval G, Iannuccelli N, Legault C *et al.* (2000) Genetic diversity of eleven European pig breeds. *Genetics Selection Evolution*, **32**, 187–203.
- Lesica P, Allendorf FW (1995) When are peripheral population valuable for conservation. *Conservation Biology*, **9**, 753–760.
- Luikart GL, Allendorf FW, Cornuet J-M, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**, 238–247.
- Majjala K (1988) History, recent development and uses of Finnish sheep. *Journal of Agricultural Science in Finland*, **60**, 449–454.
- Marchini JL, Heaton C, Ripley BD (2003) *FastICA algorithms to perform ICA and projection pursuit*. Available from www.stats.ox.ac.uk/~marchini/software.html.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA*, **70**, 3321–3323.
- Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.
- Pons O, Petit RJ (1995) Estimation, variance and optimal sampling of gene diversity. I. Haploid locus. *Theoretical and Applied Genetics*, **90**, 462–470.
- Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144**, 1237–1245.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Riesenberg LH, Church SA, Morjan CL (2003) Integration of populations and differentiation of species. *New Phytologist*, **161**, 59–69.
- Ruane J (1999) Selecting breeds for conservation. In: *Genebanks and the Conservation of Farm Animal Genetic Resources* (ed. Oldenbroek JK), pp. 59–73. DLO Institute for Animal Science and Health, Lelystad, The Netherlands.
- Ryder ML (1991) Domestication, history and breed evolution in sheep. In: *World Animal Science, B 8. Genetic Resources of Pig, Sheep and Goat* (ed. Majjala K), pp. 157–177. Elsevier, Amsterdam.
- Scherf BD (2000) *World Watch List for Domestic Animal Diversity*, 3rd edn. Food and Agriculture Organization of the United Nations, Rome.
- Simianer H (2005) Using expected allele number as objective function to design between and within breed conservation of farm animal biodiversity. *Journal of Animal Breeding and Genetics*, **122**, 177–187.
- Sokal RR, Rohlf FJ (1981) *Biometry*, 2nd edn. Freeman, San Francisco, California.
- Stone JV (2002) Independent component analysis: an introduction. *Trends in Cognitive Sciences*, **6**, 59–64.
- Tapio M, Miceikiene I, Vilkki J, Kantanen J (2003) Comparison of microsatellite and blood protein diversity in sheep: inconsistencies in fragmented breeds. *Molecular Ecology*, **12**, 2045–2056.
- Tapio I, Tapio M, Grisli Z *et al.* (2005) Unfolding of population structure in Baltic sheep breeds using microsatellite analysis. *Heredity*, **94**, 448–456.
- Thioulouse J, Chessel D, Dolédec S, Olivier JM (1997) ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing*, **7**, 75–83.
- Visscher PM, Smith D, Hall SJG, Williams JA (2001) A viable herd of genetically uniform cattle. *Nature*, **409**, 303.
- Wakeley J, Takahashi T (2003) Gene genealogies when the sample size exceeds the effective size of the population. *Molecular Biology and Evolution*, **20**, 208–213.
- Zeng Z-B, Cockerham CC (1990) Long-term response to artificial selection with multiple alleles — study by simulations. *Theoretical Population Biology*, **37**, 254–272.

This study is part of the collaborative North-SheD project (www.lbhi.is/northshed). The authors' interests are to investigate the conservation value and the origin of the different northern European sheep breeds using molecular genetic methods. This study is also part of the work conducted by Miika Tapio for his PhD thesis.
