Heredity (2007) 99, 620–631

www.nature.com/hdy

ORIGINAL ARTICLE

Genetic structure of European sheep breeds

L-J Lawson Handley[1,3], K Byrne[2], F Santucci[1,4], S Townsend[2], M Taylor[1,5], MW Bruford[2,6] and GM Hewitt[1]

[1]*Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich, UK and* [2]*Institute of Zoology, Zoological Society of London, Regents Park, London, UK*

Large-scale evaluations of genetic diversity in domestic livestock populations are necessary so that region-specific conservation measures can be implemented. We performed the first such survey in European sheep by analysing 820 individuals from 29 geographically and phenotypically diverse breeds and a closely related wild species at 23 microsatellite loci. In contrast to most other domestic species, we found evidence of widespread heterozygote deficit within breeds, even after removing loci with potentially high frequency of null alleles. This is most likely due to subdivision among flocks (Wahlund effect) and use of a small number of rams for breeding. Levels of heterozygosity were slightly higher in southern than in northern breeds, consistent with declining diversity with distance from the Near Eastern centre of domestication. Our results highlight the importance of isolation in terms of both geography and management in augmenting genetic differentiation through genetic drift, with isolated northern European breeds showing the greatest divergence and hence being obvious targets for conservation. Finally, using a Bayesian cluster analysis, we uncovered evidence of admixture between breeds, which has important implications for breed management.
Heredity (2007) **99,** 620–631; doi:10.1038/sj.hdy.6801039; published online 15 August 2007

Keywords: genetic diversity; domestic breeds; Bayesian cluster analysis; livestock conservation; null alleles; heterozygote deficit

Introduction

Sheep (*Ovis aries*, L.) were domesticated from at least three ancestral subspecies of the wild Mouflon (*O. gmelini*, Gmelin 1774) approximately 9000 years before present (YBP) in Southwest Asia, and by 5000 YBP, sheep had been transported throughout Europe (Ryder, 1981; Hiendleder *et al*., 1998, 2002; Guo *et al*., 2005; Pedrosa *et al*., 2005; Bruford and Townsend, 2006). Today, over 850 sheep breeds are recognized worldwide, and Europe supports a greater number of breeds than any other continent (United Nations Food and Agriculture Organization, FAO, 2000; Rege and Gibson, 2003).

Over the past few decades though, it has become apparent that many of these breeds are at risk of extinction. Europe faces the greatest threat, with 18% of its livestock breeds lost in the last century alone and a further 40% at risk of becoming extinct over the next 20 years (compared to 32% at risk worldwide, FAO, 2000; Signorello and Pappalardo, 2003). Intensive production and increased commercial demands, particularly since the end of the Second World War, have significantly contributed to the threats facing European sheep breeds. First, artificial insemination and improved transportation have reduced the number of breeding rams, leading to a reduction in the effective population size (N_e) of many breeds. Second, production has focused on only a few breeds, to the detriment of rare or minority breeds, which are likely to be important genetic resources because of their local adaptation, disease resistance, high fertility and unique product qualities (Mendelsohn, 2003). Minority breeds have also been lost by introgression into large commercial populations.

The loss of diversity in domestic species has important economic, ecological and scientific implications as well as social considerations, and in response to these threats, the FAO, through its Domestic Animal Diversity Information System (DAD-IS), initiated a programme to document the State of the World's Animal Genetic Resources (SoWAnGR, FAO, 2000). An understanding of the evolutionary history of domestic breeds and data on genetic variation within and among breeds is vital to these initiatives to provide critically important data for the decision-making process (Rege and Gibson, 2003). Information on both within- and among-breed diversity is important. The former provides information for management at the breed level. The latter helps to identify divergent breeds that may harbour distinct genotypes and are, therefore, worthy of conservation efforts even if their within-breed diversity is relatively high.

To date, most studies of genetic diversity in livestock species have been carried out at local geographic (national) scales (Arranz *et al*., 1998; Martín-Burriel *et al*., 1999; Saitbekova *et al*., 1999; Martinez *et al*., 2000;

Correspondence: Dr L-J Lawson Handley, Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK. E-mail: ljl27@cam.ac.uk
[3]*Current address: Department of Genetics, University of Cambridge, Cambridge, UK.*
[4]*Current address: Dipartimento di Medicina Sperimentale e Scienze Biochimiche Microbiologia, University of Perugia, Perugia, Italy.*
[5]*Current address: School of Biological Sciences, University of Wales, Bangor, Gwynedd, UK.*
[6]*Current address: Cardiff School of Biosciences, Cardiff, UK.*
Received 7 July 2006; revised 1 June 2007; accepted 18 June 2007; published online 15 August 2007

Mateus *et al*., 2004), or on a small number of breeds (MacHugh *et al*., 1998; Diez-Tascón *et al*., 2000; Pariset *et al*., 2003). While such studies are essential for region- or breed-specific management and conservation programmes, it is also important to assess how genetic diversity is partitioned at larger geographic scales to better implement region-specific conservation measures (Bruford *et al*., 2003).

In the present study, we use a panel of 23 autosomal microsatellite markers to evaluate how genetic diversity is partitioned within and among a diverse sample of 30 European sheep breeds. We consider the role of geography and breed type in determining diversity and differentiation, and examine the extent of admixture among breeds in relation to conservation and management.

Materials and methods

Sampling procedure and DNA extraction

A total of 820 individuals from 29 breeds of domestic sheep (*O. aries*) and one feral, putatively ancestral type, the European Mouflon (*O. gmelini musimon*), were analysed to ensure sampling of morphologically and geographically divergent types. Our aim was to collect at least 30 samples from a minimum of two separate flocks, although fulfilment of these objectives was not possible for all breeds (Table 1).

Samples (from blood, tissue and hair) and pre-extracted DNA were obtained from farmers and collaborators. Blood samples were diluted with five volumes of lysis buffer (Scott and Densmore, 1991) and DNA extracted following Sambrook *et al*. (1989). Tissue samples (such as heart and muscle collected as by-products from animals that had been slaughtered) were extracted using a sodium dodecyl sulphate–ethylenediaminetetraacetic acid (SDS–EDTA) procedure, as described by Hoelzel and Dover (1991). Hair samples (collected from live animals) were extracted using an SDS–EDTA procedure or a Chelex 100-based method (Bio-Rad, Walsh *et al*., 1991).

Microsatellite loci

Twenty-three microsatellite loci (Table 2) were chosen from the literature following three criteria: (1) a similar proportion of markers were derived from sheep, cattle and goat (to demonstrate cross-species utility of markers for the purposes of the European Union); (2) loci were distributed evenly throughout the autosomes and preferably unlinked; and (3) loci exhibited Mendelian inheritance and at least four alleles. For four of the breeds (Zeeland, Exmoor Horn, Llanwenog and Bizet), data were analysed for a subset of 10 of these loci (OarAE54, OarAE129, OarFCB20, OarFCB304, JMP29, JMP58, MAF65, MAF70, MAF209 and BM1824) that were typed in a collaborative study (K Byrne, unpublished results).

Polymerase chain reaction (PCR) amplifications were carried out in 10 μl total volumes, containing 2–4 pM of each primer (with the forward primer labelled with 5-FAM, TET or HEX), 0.275 mM dNTPs (Pharmacia), $1 \times NH_4$ buffer, 2–4.0 mM $MgCl_2$ (Table 2), 0.25 U of Amplitaq Gold DNA polymerase (Applied Biosystems, Foster City, CA, USA) and $\sim$50 ng DNA. PCRs were performed in a PE9700 thermocycler (Applied Biosystems), with the following cycling parameters: 94°C for 2 min followed by 30 cycles of 94°C for 30 s, anneal (Table 2) for 30 s, and 72°C for 30 s, and a final step at 72°C for 10 min. For some loci (identified by two annealing temperatures separated by '/' in Table 2), a two-step cycling procedure was used, which consisted of 94°C for 6 min, followed by seven cycles of 94°C for 30 s, anneal (Table 2) for 30 s, and 72°C for 1 min, then 23 cycles of 92°C for 30 s, anneal (Table 2) for 30 s, and 72°C for 1 min, and final extension at 72°C for 10 min.

Following PCR, products were diluted 1/10–1/20 and 0.8 μl was mixed with an internal standard (TAMRA, Applied Biosystems) according to the manufacturer's instructions. Products were electrophoresed on a 4.25% polyacrylamide gel in an ABI377 automated sequencer (Applied Biosystems). Fragment analysis was carried out using GENESCAN v3.1 and GENOTYPER v2.0 software (Applied Biosystems). Genotyping was repeated once when individuals failed to amplify.

General statistical analyses

Allelic richness was estimated over all samples for each locus (R_t), without correction for null alleles, using the rarefaction technique of El Mousadik and Petit (1996) in FSTAT (v2.9.3; Goudet, 2000). Genotypic linkage disequilibrium (LD) was assessed using the exact test in GENEPOP (v3.4; Raymond and Rousset, 1995; Rousset and Raymond, 1995), using a Markov chain with 10 000 dememorization steps, 200 batches and 2000 iterations per batch to produce an unbiased estimate of the exact *P*-value.

Null alleles—alleles that consistently fail to amplify during PCR due, for example, to priming site mutations, differential amplification of size variants or inconsistent DNA quality—can decrease estimates of genetic diversity and inflate genetic differentiation, particularly when loci characterized in one species are then analysed in a different species, as is the case here (Chakraborty *et al*., 1992; Paetkau and Strobeck, 1995; Slatkin, 1995; Dakin and Avise, 2004). To investigate whether null alleles were present in our data set, we first recorded the number of times amplification failed for just one locus per individual, which could signify a null homozygote. We then estimated the potential frequency of null alleles (r) for each locus in each breed using the EM algorithm of Dempster *et al*. (1977) in the software FreeNA (Chapuis and Estoup, 2007). Note that this method, along with all other methods for detecting null alleles, assumes that deviations from Hardy–Weinberg (HW) equilibrium do not result from other causes (such as a Wahlund effect). Values of $r \leqslant 0.2$ are not expected to cause significant problems in analyses (Dakin and Avise, 2004; Chapuis and Estoup, 2007); therefore, we only considered loci with $r \geqslant 0.2$ to be potentially problematic for our calculations.

Genetic diversity within and among breeds

Observed and expected heterozygosities (H_o and H_e, respectively) were calculated for each breed and locus combination in ARLEQUIN (v2.0; Schneider *et al*., 2000) and then averaged over loci to obtain an estimate per breed. Deviation from HW equilibrium was assessed using the method of Guo and Thompson (1992) for each

Table 1 Breed and sample information

Breed	*Code*	N	*Flocks*	*Origin*	*Locality*	*North or south*	*Breed type*
Awassi	AW	24	2	Israel and Jordan	Widespread Near East	S	Fat-tail
Cyprus fat-tail	CF	30	3	Cyprus	Throughout Cyprus	S	Fat-tail
Lesvos	LS	32	3	Greece	Island of Lesvos	S	Zackel × fat-tail
Chios	CS	30	3	Greece	Island of Chios	S	Semi fat-tail
Argos	AS	32	3	Greece	Peloponnese	S	Zackel × fat-tail
Istrian	IS	13	1	Croatia	Eastern Istria and Croatian coast	S	Zackel
Comisana	CO	30	3	Italy	Sicily	S	Mountain
Leccese	LC	25	3	Italy	Lecce, southern Apulia	S	Zackel
Sarda	SA	21	5	Italy	Sardinia	S	Coarse wool
Massese	MA	15	1	Italy	Massa, Tuscany	S	Mountain
Spanish Merino	ME	39	3	Spain	Andalucia and Estredamura	S	Ancient fine wool[a]
Andalucian Churro	CH	16	1	Spain	Guadalquivir marshes	S	Coarse wool
Rasa Aragonesa	AR	20	1	Spain	Northeastern Spain	S	Entrefino[a]
Thônes-Marthod	TM	23	3	France	Arly and Arc valleys, Savoie	S	Mountain
Bizet	BI	24	5	France	North Haute Loire and southeastern Puy-de-Dôme	S	Mountain
Turcana	TU	30	3	Romania	Carpathian Mountains	S	Zackel
Tsigai	TS	30	3	Romania	Foothills of Transylvania	S	Zackel[b]
Racka	RA	28	1	Hungary	Hortobagy	S	Zackel
Šumavka	SM	30	1	Czech Republic	Bohemia—mountainous areas	S	Zackel
Coburg	CB	30	4	Germany	Northern Bavaria—Highlands	N	Native hill[a]
Skudde	SK	30	7	Germany	East Prussia and Baltics	N	Heath
Heidschnucke	HS	14	1	Germany	Lüneburg Heath Hanover	N	Heath
Friesland	FL	32	19	The Netherlands	Coastal provinces	N	Marsh rat-tail
Zeeland	ZL	32	Undefined	The Netherlands	Zealand	N	Marsh rat-tail
Exmoor Horn	EH	32	6	England	Exmoor, southwest England	N	Southwestern horned[a]
Llanwenog	LW	30	10	Wales	Cardigan	N	Shropshire × blackface[a]
North Ronaldsay	NR	30	1	Scotland	Orkney islands	N	Northern short-tail
Soay	SO	28	1	Scotland	Hirta, St Kilda	N	Northern short-tail
Icelandic	IC	50	3	Iceland	Throughout	N	Northern short-tail
Mouflon	MO	18	2	France/Italy	Corsica, Sardinia	NA	Feral, subspecies[a]
		2	1	Cyprus	Troodos Mountains		

'*n*' number of individuals; 'flocks' refers to the number of flocks sampled.
[a]Breeds not included in the analysis by breed type.
[b]Tsigai may be classified as more than one type. Since this breed is believed to be strongly influenced by Zackel, it was included in this group for the analysis by breed type.

Table 2 Locus information

Locus	*Origin*	*Chr*	*NCBI*	*Size (bp)*	*Tn*	*Mg*	A_t	R_t	*Ref*
OarAE54	Ovine	25	L11048	119–157	55	2.5	20	11.0	a
OarAE129	Ovine	5	L11051	104–178	55	2.5	24	6.8	a
OarFCB20	Ovine	2	L20004	86–128	55	2.5	21	10.1	b
OarFCB48	Ovine	17	M82875	125-177	60	2.0	20	7.2	b
OarFCB304	Ovine	19	L01535	137–193	55	2.5	27	10.2	c
OarJMP29	Ovine	24	U30893	115–161	55	2.5	23	9.8	d
OarJMP58	Ovine	26	U35058	106–190	55	2.5	23	9.3	d
MAF65	Ovine	15	M67437	119–143	55	2.5	13	7.1	e
MAF70	Ovine	4	M77200	127–169	60	2.5	22	11.8	f
MAF209	Ovine	17	M80358	102–142	60	2.5	21	11.1	g
MAF214	Ovine	16	M88160	174–282	60	2.0	33	5.3	h
BM1824	Bovine	1	G18394	163–181	58	2.5	10	4.6	i
ETH225	Bovine	9	Z14043	131–157	55	2.0	14	4.6	j
ILSTS005	Bovine	7	L23481	174–218	58/55	2.0	18	5.8	k
ILSTS011	Bovine	9	L23485	256–294	58/55	2.0	19	6.4	k
ILSTS028	Bovine	3	L37211	105–177	58/55	2.0	22	7.6	l
INRA063	Bovine	14	X71507	151–197	55	2.5	22	7.2	m
SRCRSP1	Caprine	(13)	L22192	116–148	50/55	2.0	16	5.5	n
SRCRSP3	Caprine	(10)	L22195	111–197	50/55	2.0	20	6.2	n
SRCRSP5	Caprine	(21)	L22197	119–151	55/58	2.5	11	4.2	n
SRCRSP7	Caprine	(6)	L22199	158–210	50/55	2.0	21	8.1	o
SRCRSP8	Caprine	(?)	L22193	183–249	58	2.0	23	7.0	o
SRCRSP9	Caprine	(12)	L22200	99–135	55	2.0	15	4.8	o
Mean (s.d.)							19.9 (5.2)	7.5 (2.3)	

'Chr', assignment to ovine chromosome, parantheses refer to assignments not confirmed in sheep by linkage analysis; 'bp', base pairs; 'Tn', annealing temperature. In the case of two temperatures separated by a '/', the annealing temperatures differed in the two steps of the cycling reaction (see Materials and methods). 'Mg', concentration of $MgCl_2$ (mM) used in the PCR reaction; 'A_t', total number of alleles; 'R_t', allelic richness over all samples; 'Ref', references: (a) Penty J, Henry H, Ede A, Crawford A (1993) Ovine microsatellites at the OarAE16, OarAE54, OarAE57, OarAE119 and OarAE119 loci, unpublished, (b) Buchanan *et al*. (1994), (c) Buchanan and Crawford (1993), (d) Crawford *et al*. (1995), (e) Buchanan *et al*. (1992), (f) Buchanan and Crawford (1992a), (g) Buchanan and Crawford (1992b), (h) Buchanan and Crawford (1992c), (i) Bishop *et al*. (1994), (j) Steffen *et al*. (1993), (k) Brezinsky *et al*. (1993), (l) Kemp *et al*. (1995), (m) Vaiman *et al*. (1994), (n) Arevalo *et al*. (1994) and (o) Bhebhe *et al*. (1994).

locus and breed combination in ARLEQUIN using a Markov chain of 100 000 steps and 1000 dememorization steps. GENEPOP was used to assess global deviations from HW equilibrium (over all loci per breed and after Bonferroni correction) using the same Markov chain parameters as for the genotypic disequilibrium test. Weir and Cockerham's (1984) estimate of F_{is} was calculated per locus and over all loci per breed using FSTAT and *P*-values obtained based on 2300 randomizations (number of loci, $nl \times 100$). H_o, H_e and F_{is} were obtained, and deviations from HW equilibrium were assessed before and after removal of loci (on a per-breed basis) that showed potentially high null allele frequencies ($r \geqslant 0.2$).

Since heterozygote deficit was widespread even after removing loci with $r \geqslant 0.2$ (see Results and Table 4), and since excluding these loci greatly reduced the amount of available data for pairwise comparisons among breeds, all loci were retained for analyses described below. Pairwise genetic distance between breeds was estimated using Weir and Cockerham's (1984) estimate of Wright's F_{st} (θ, Wright, 1951, 1977) and *P*-values obtained using 435 000 permutations (number of tests $\times$ 100) and standard Bonferroni correction in FSTAT.

Genetic variation was compared among groups, among breeds within groups and within breeds using a hierarchical Analysis of MOlecular VAriance (AMOVA; Excoffier *et al*., 1992) in ARLEQUIN. Significance levels were determined after 1000 permutations. Breeds were partitioned in three ways: by main geographic region; by north versus south; and by breed type. The main geographic regions were (a) Near East (Israel and Jordan), (b) Greece and Cyprus, (c) Italy and Croatia, (d) Spain and France, (e) Romania, Hungary and Czech Republic, (f) Germany and The Netherlands and (g) The UK and Iceland (Table 1). Southern breeds comprised groups (a–e) from the main geographic regions above and northern breeds comprised groups (f–g) (Table 1). Breed types were fat-tail, Zackel, coarse wool, mountain, heath, marsh island rat-tail and northern short-tail (Table 1). Breeds of unique type or sole representatives of their type were excluded from the analysis (Table 1). Several other AMOVAs were performed using different regional group definitions (for example, mainland versus island breeds), but among group differences were not significant and they are, therefore, not presented.

Relationships among breeds and admixture

A principal component analysis (PCA) was performed to visualize pairwise differentiation among breeds (F_{st}) using the software PCAGEN (J Goudet, personal communication, available at http://www2.unil.ch/popgen/softwares/pcagen.htm). In total, 1000 randomizations of genotypes were performed to test for significance of axes.

Admixture was investigated using the Bayesian clustering algorithm implemented in STRUCTURE (Pritchard *et al*., 2000). The full data set of 30 breeds was run with $K = 1$–30, where K is the potential number of genetic clusters. Breeds that shared the same genetic

cluster were then analysed independently with $K = 1$–6. A total of 100 000 iterations of the Gibbs sampler were run, after a burn-in of 50 000 iterations using the admixture option. The program DISTRUCT (Rosenberg, 2004) was used to graphically display the membership coefficient of an individual for a sub-population (i.e. breed), which represents the fraction of its genome that has ancestry in the sub-population, (Rosenberg, 2004).

Results

The number of alleles per locus ranged from 10 for BM1824 to 33 for MAF214 (Table 2). Loci derived from sheep had more alleles and higher overall allelic richness (R_t) than cattle or goat loci (mean number of alleles 22.45, 17.50 and 17.67 respectively, R_t 9.07, 6.02, 5.94). In the case of R_t, but not allele number, this difference was significant (Mann–Whitney U score test, sheep/cattle $U = 56$, $P = 0.02$, sheep/goat $U = 58$, $P = 0.01$). Significant LD was found in comparisons of OarAE129/ILSTS028, OarJMP58/INRA063, BM1824/SRCRSP3 and ILSTS005/SRCRSP7 and SRCRSP3, and ILSTS028/INRA063; however, in each case, the loci were situated on different chromosomes.

Twenty-five breeds had at least one individual that failed to amplify at only one locus, which could signify a null homozygote. Of the 15 loci that were affected, SRCRSP5 was the worst offender (being the sole non-amplifying locus in 25/702 individuals in which it was typed, which accounts for 3.6% of genotypes) followed by ILSTS11, ILSTS28, SRCRSP8 and OarFCB48 (accounting for 2.4, 2.4, 2.2 and 2.0% of genotypes, respectively; see Supplementary Table 1 for details). Using the EM method (Dempster *et al*., 1977; Chapuis and Estoup, 2007), we identified 12 loci that had potential null alleles at high frequency ($r \geqslant 0.2$) in at least one breed (Supplementary Table 2, summarized in Table 3). OarAE54 had $r \geqslant 0.2$ in four breeds (Lesvos, Argos, Comisana and Mouflon), OarAE129 and SRCRSP5 had $r \geqslant 0.2$ in three breeds (Argos, Skudde, Heidschnuke and Comisana, Lecesse, Sarda, respectively), and the remaining nine loci had $r \geqslant 0.2$ in one or two breeds only (Table 3). Note that OarAE129 and SRCRSP5 have been shown to produce null alleles in other studies (Peter *et al*., 2005, 2007).

Genetic diversity within and among breeds

After exclusion of the loci with $r \geqslant 0.2$, observed heterozygosity ranged from 0.495 to 0.711 (mean $H_o = 0.620$, s.d. 0.060; Table 3), with the lowest values found in the North Ronaldsay, Friesland and Soay, and the highest in the Istrian, Leccese and Tsigai. Breeds from southern Europe had significantly higher levels of heterozygosity (mean $H_o = 0.641$, s.d. 0.044) than northern breeds (mean $H_o = 0.589$, s.d. 0.073, Mann–Whitney $U = 50$, $P = 0.04$). Out of a possible 613 locus × breed combinations (again excluding the loci with $r \geqslant 0.2$), heterozygote excess was found on only 16 occasions, compared to 199 instances of heterozygote deficit (Table 4 and Supplementary Table 3). On average, roughly nine loci per breed showed heterozygote deficit (from the total number of loci per breed with heterozygote deficit divided by the total number of locus × breed combinations, multiplied by the number of breeds, to the nearest integer: $(183/613) \times 30 = 9$). This translates to significant global heterozygote deficit in all breeds (Table 3, unbiased P-values for tests of global heterozygote deficit are $P < 0.000$ for all breeds except Soay and Churro, where $P = 0.002$) except two with small sample size ($n < 15$, Istrian and Heidschnuke, where $P = 0.541$ and 0.449, respectively). With the exception of these two, most breeds retained positive, significant values of F_{is} after correction for null alleles (Table 3). Excluding loci with $r \geqslant 0.2$ increases H_o and decreases values of F_{is}, but does not change the significance of any comparison (Table 3). Since it therefore seems unlikely that null alleles are the only cause for deviations from HW equilibrium, we based the following results on all loci.

Global F_{st} was 0.131 (95% confidence interval 0.116–0.147), while pairwise values ranged from 0.01 to 0.275 (Turcana/Tsigai and Soay/Friesland, respectively; Table 6). All comparisons except Turcana/Tsigai, Thônes-Marthod/Leccese and Thônes-Marthod/Istrian were significant after Bonferroni correction ($P < 0.05$; Table 6). Northern European breeds such as the Soay, Friesland and Skudde tended to show the greatest divergence ($F_{st} > 0.20$).

Relationships among breeds and admixture analysis

Our hierarchical AMOVA with hypothetical population structure revealed that the greatest variation among groups corresponds to breed type (2.7%, $P < 0.0001$; Table 5) as opposed to region (0.97%, $P < 0.05$), or north versus south (1.16%, $P < 0.0001$); however, most of the variation was consistently explained within breeds ($\sim 87\%$, $P < 0.0001$).

The PCA shown in Figure 1 illustrates that most breeds cluster within a central group, with little correspondence to breed type (except for Zackel breeds; Turcana, Tsigai, Racka and Šumavka). The most distant outliers in the PCA (Friesland, North Ronaldsay, Icelandic, Soay and Skudde) are all from northern Europe.

Our data set of 30 breeds was best described by $K = 25$ genetic clusters (likelihood Ln values declined with $K > 25$, and $K = 25$ had a significantly higher Ln than $K = 24$, two-tailed t-test, $N_1 = 28$, $N_2 = 19$, $t = -2.475$, $P = 0.017$). DISTRUCT (Rosenberg, 2004) plots constructed for $K = 25$ are shown in Figure 2a. The Mouflon, Soay, Icelandic, North Ronaldsay, Heidschnuke, Skudde, Coburg, Šumavka, Racka and Bizet remained the most genetically pure breeds (that is free from admixture) from $K = 23$ to $K = 30$. Clear evidence of admixture was found between two Greek breeds, with several Argos individuals having Lesvos multilocus genotypes (Figure 2a).

Three pairs of breeds shared clusters when $K = 25$ (Turcana and Tsigai, Friesland and Zeeland, and Cyprus fat-tail and Chios; Figure 2a), and were, therefore, analysed independently. Turcana and Tsigai were run with $K = 1$–6, but genetic clusters were not identified within these breeds (not shown). Friesland and Zeeland were run with $K = 1$–5, and the highest Ln was found for $K = 3$ (Figure 2b). Two clusters were found in the Zeeland, and all Friesland individuals except for one, which has a Zeeland genotype, were members of the same genetic cluster, whereas four Zeeland individuals have genotypes corresponding more to the Friesland

Table 3 Genetic diversity measured within breed before and after correction by removing loci with high frequency of potential null alleles

Breed	*Before correction*				*Loci with* $r \geqslant 0.2$[a]	*After correction*			
	H_o	H_e	F_{is}	P (F_{is})		H_o	H_e	F_{is}	P (F_{is})
Awassi	0.646	0.757	0.134	0.0000	None				
Cyprus fat-tail	0.579	0.677	0.130	0.0000	None				
Lesvos	0.613	0.691	0.104	0.0000	OarAE54	0.633	0.692	0.081	0.0005
Chios	0.520	0.641	0.165	0.0000	ILSTS05	0.525	0.628	0.145	0.0023
Argos	0.571	0.672	0.130	0.0000	OarAE54, OarAE129	0.592	0.659	0.092	0.0005
Istrian	0.711	0.729	0.006	0.4172	None				
Comisana	0.597	0.706	0.139	0.0000	OarAE54, SRCRSP5	0.624	0.695	0.093	0.0005
Leccese	0.680	0.757	0.113	0.0000	SRCRSP5	0.696	0.770	0.090	0.0005
Sarda	0.641	0.711	0.095	0.0001	SRCRSP5	0.658	0.721	0.078	0.0014
Massese	0.646	0.718	0.078	0.0036	None				
Merino	0.585	0.696	0.146	0.0000	MAF70, SRCRSP7	0.613	0.690	0.102	0.0005
Churro	0.632	0.689	0.060	0.0353	None				
Aragonesa	0.648	0.745	0.121	0.0000	None				
Thônes-Marthod	0.654	0.74	0.090	0.0003	None				
Bizet	0.679	0.738	0.063	0.0478	None				
Turcana	0.666	0.743	0.089	0.0000	None				
Tsigai	0.695	0.750	0.068	0.0003	None				
Racka	0.617	0.699	0.113	0.0000	None				
Šumavka	0.624	0.713	0.113	0.0000	BM1824, SRCRSP3	0.656	0.715	0.073	0.0010
Coburg	0.608	0.667	0.090	0.0000	None				
Skudde	0.619	0.681	0.078	0.0007	OarAE129	0.631	0.678	0.058	0.0082
Heidschnuke	0.661	0.684	0.000	0.5062	OarAE129	0.685	0.688	−0.027	0.8055
Friesland	0.488	0.543	0.116	0.0000	INRA063	0.496	0.553	0.093	0.0009
Zeeland	0.571	0.753	0.253	0.0000	OarAE54, MAF209	0.664	0.788	0.152	0.0013
Exmoor Horn	0.620	0.807	0.230	0.0000	MAF209	0.615	0.796	0.226	0.0010
Llanwenog	0.630	0.769	0.174	0.0000	None				
North Ronaldsay	0.474	0.624	0.222	0.0000	BM1824	0.495	0.633	0.203	0.0005
Soay	0.497	0.538	0.049	0.0388	None				
Icelandic	0.537	0.640	0.143	0.0000	None				
Mouflon	0.490	0.743	0.318	0.0000	OarAE54, JMP29, MAF209, ILSTS028, INRA063, SRCRSP7	0.562	0.713	0.196	0.0006
Mean[b] (s.d.)	0.606 (0.063)	0.695 (0.060)	0.121 (0.069)			0.620 (0.060)	0.700 (0.061)	0.102 (0.055)	

'H_o' and 'H_e', mean observed and expected heterozygosities, respectively, calculated per locus and then averaged over loci within breed; 'F_{is}', Weir and Cockerham's (1984) estimator of F_{is} calculated over all loci within breed, with *P*-values obtained by randomization (see text).

[a]Loci with high potential null allele frequency ($r \geqslant 0.2$) identified using the EM logarithm (Dempster *et al*., 1977).

[b]Mean estimates 'after correction' include those breeds that did not have loci with $r \geqslant 0.2$.

Table 4 Loci per breed that show significant deviation from Hardy–Weinberg equilibrium (where $P<0.05$, and excludes those with $r\geqslant 0.2$)

Breed	*Heterozygote deficit*	*Heterozygote excess*
Awassi	OarAE54, OarFCB304, MAF70, BM1824, ETH225, SRCRSP7	INRA063
Cyprus fat-tail	OarAE129, OarFCB304, OarJMP29, MAF214, BM1824, ETH225, INRA063, SRCRSP5, SRCRSP7	
Lesvos	OarAE129, MAF209, ETH225, ILSTS28, INRA063	MAF65
Chios	OarAE54, OarJMP58, MAF209, INRA063, SRCRSP5	
Argos	OarJMP58, MAF65, MAF209, INRA063, SRCRSP8	
Istrian	OarJMP29	MAF70
Comisana	OarAE129, OarFCB304, OarJMP29, OarJMP58, MAF65, MAF70, BM1824, INRA063, SRCRSP8	
Leccese	OarAE54, OarFCB20, OarFCB304, MAF65, BM1824, INRA063	OarJMP58
Sarda	OarAE129, OarFCB304, OarJMP29, MAF209, ILSTS11, SRCRSP3	
Massese	MAF70, BM1824, INRA063	
Merino	OarAE54, OarAE129, OarFCB304, OarJMP29, SRCRSP5	
Churro	OarJMP58, MAF70, ILSTS11, SRCRSP3	OarFCB304, MAF209
Aragonesa	OarJMP29, OarJMP58, MAF70, MAF214, BM1824, ILSTS28, SRCRSP1, SRCRSP5, SRCRSP7	
Thônes-Marthod	OarAE54, OarAE129, OarFCB304, MAF65, MAF209, MAF214	
Bizet	OarAE54, OarJMP58	OarFCB304, BM1824
Turcana	OarAE129, MAF70, BM1824, ETH225, SRCRSP3, SRCRSP5	
Tsigai	OarAE129, OarJMP29, MAF70, ETH225, SRCRSP5	
Ŗacka	OarAE129, OarFCB20, MAF65, MAF70, MAF209, INRA063, SRCRSP5, SRCRSP7	
Šumavka	OarAE54, OarAE129, MAF65, MAF209, MAF214, SRCRSP7	
Coburg	OarAE54, OarFCB48, OarFCB304, OarJMP58, ILSTS05, SRCRSP5, SRCRSP8	OarAE129, MAF65
Skudde	OarFCB48, MAF70, MAF209, BM1824, ETH225, ILSTS11, SRCRSP3, SRCRSP7	MAF65
Heidschnuke	OarAE54, SRCRSP8	ILSTS28
Friesland	OarAE129, MAF209, BM1824, SRCRSP7	
Zeeland	OarAE129, OarFCB20, OarFCB304, OarJMP29, OarJMP58, MAF70, MAF214	MAF65
Exmoor Horn	OarAE54, OarAE129, OarFCB20, OarJMP29, OarJMP58, MAF70, BM1824	
Llanwenog	OarAE54, OarAE129, OarFCB304, OarJMP29, OarJMP58, MAF209	OarFCB20, MAF70
North Ronaldsay	OarFCB48, OarFCB304, OarJMP29, MAF214, ETH225, ILSTS05, ILSTS28, INRA063, SRCRSP1, SRCRSP3, SRCRSP5	ILSTS11
Soay	OarAE129, ETH225, ILSTS28, SRCRSP5	
Icelandic	OarFCB48, OarFCB304, MAF209, MAF214, ETH225, ILSTS05, ILSTS11, ILSTS28, SRCRSP3, SRCRSP5, SRCRSP8	
Mouflon	OarFCB304, OarJMP58, MAF65, MAF70, MAF214, ILSTS05, SRCRSP1, SRCRSP2, SRCRSP8	

Table 5 Hierarchical AMOVA performed by grouping breeds according to their geographic origin or breed type

Comparison	*Variance components*	V	*% total*	P	*Φ statistics*
Between region ($n=7$)	Among regions σ_a^2	0.045	0.97	0.025	$\Phi_{CT}=0.010$
	Among breeds within region σ_b^2	0.537	11.67	<0.0001	$\Phi_{SC}=0.128$
	Within breeds σ_c^2	4.016	87.35	<0.0001	$\Phi_{ST}=0.126$
Northern versus southern breeds ($n=2$)	Among regions σ_a^2	0.054	1.16	<0.0001	$\Phi_{CT}=0.012$
	Among breeds within region σ_b^2	0.554	12.00	<0.0001	$\Phi_{SC}=0.121$
	Within breeds σ_c^2	4.014	86.84	<0.0001	$\Phi_{ST}=0.132$
Between main breed types ($n=7$)	Among breed types σ_a^2	0.124	2.70	<0.0001	$\Phi_{CT}=0.027$
	Among breeds within type σ_b^2	0.506	10.99	<0.0001	$\Phi_{SC}=0.113$
	Within breeds σ_c^2	3.978	86.31	<0.0001	$\Phi_{ST}=0.137$
All breeds	Among all breeds σ_a^2	0.671	13.46	<0.0001	$\Phi_{ST}=0.136$

Abbreviation: AMOVA, Analysis of MOlecular VAriance.
Variance components, σ_a^2, σ_b^2 and σ_c^2, are as described in Excoffier *et al*. (1992, Equation 10a). '*V*', variance; '*P*', significance level determined via 1000 permutations. Φ_{ST} is the correlation of random haplotypes within breeds relative to that of random pairs of haplotypes drawn from the whole sample, Φ_{CT} the correlation of random haplotypes within a group of breeds relative to that of random pairs of haplotypes drawn from the whole sample and Φ_{SC} the correlation of the molecular diversity of random haplotypes within breeds relative to that of random pairs of haplotypes drawn from the region or breed type (Excoffier *et al*., 1992). See text and Table 1 for definition of groups.

(Figure 2b). Cyprus fat-tail and Chios were run with $K=1$–5. While these two breeds were separated into two genetic clusters when $K=2$, the highest likelihood was found for $K=4$. Under $K=4$, Chios remained a single cluster, whereas Cyprus fat-tail individuals were split into several groups (Figure 2c).

Discussion

Our survey of 23 microsatellite markers in European sheep breeds revealed that 87% of the variation is found within rather than among breeds. This situation mirrors other domestic species such as cattle (Wiener *et al*., 2004), and levels of within-breed diversity (Table 3) are also comparable to those from other domestic species, for example cattle (Katanen *et al*., 2000; Kim *et al*., 2002; Mateus *et al*., 2004), horses (Aberle *et al*., 2004) and goats (Saitbekova *et al*., 1999; Cañon *et al*., 2006). In contrast to most other domestic species however, we found extensive evidence for heterozygote deficit in European sheep breeds. Even after removing the most extreme loci in terms of null alleles, almost 30% of breed × locus

Table 6 Pairwise F_{st}

	AW	*CF*	*LS*	*CS*	*AS*	*IS*	*CO*	*LC*	*SA*	*MA*	*ME*	*CH*	*AR*	*TM*	*BI*	*TU*	*TS*	*RA*	*SM*	*CB*	*SK*	*HS*	*FL*	*ZL*	*EH*	*LW*	*NR*	*SO*	*IC*	*MO*
AW	0	0.094	0.068	0.129	0.100	0.051	0.126	0.067	0.050	0.098	0.085	0.131	0.039	0.062	0.121	0.053	0.042	0.085	0.086	0.089	0.135	0.098	0.192	0.113	0.082	0.071	0.130	0.159	0.104	0.094
CF	**	0	0.097	0.148	0.145	0.113	0.156	0.103	0.105	0.133	0.134	0.142	0.088	0.097	0.161	0.086	0.091	0.129	0.132	0.140	0.193	0.150	0.208	0.134	0.125	0.122	0.151	0.195	0.159	0.139
LS	**	**	0	0.150	0.069	0.063	0.082	0.067	0.043	0.064	0.106	0.139	0.034	0.077	0.152	0.052	0.030	0.077	0.097	0.102	0.165	0.091	0.181	0.109	0.095	0.096	0.135	0.152	0.140	0.125
CS	**	**	**	0	0.197	0.157	0.200	0.117	0.130	0.203	0.191	0.230	0.123	0.137	0.210	0.134	0.118	0.117	0.134	0.196	0.237	0.181	0.251	0.168	0.157	0.145	0.222	0.242	0.210	0.179
AS	**	**	**	**	0	0.113	0.132	0.113	0.097	0.066	0.139	0.176	0.070	0.114	0.166	0.083	0.079	0.124	0.134	0.135	0.191	0.139	0.208	0.147	0.121	0.133	0.144	0.192	0.154	0.148
IS	**	**	**	**	**	0	0.119	0.040	0.038	0.086	0.072	0.154	0.036	0.060	0.145	0.038	0.033	0.092	0.079	0.074	0.139	0.083	0.185	0.103	0.099	0.081	0.135	0.166	0.099	0.109
CO	**	**	**	**	**	**	0	0.099	0.103	0.107	0.140	0.173	0.086	0.141	0.184	0.108	0.088	0.127	0.136	0.158	0.203	0.138	0.237	0.158	0.143	0.131	0.176	0.181	0.177	0.146
LC	**	**	**	**	**	**	**	0	0.059	0.086	0.079	0.129	0.045	0.076	0.128	0.056	0.040	0.082	0.058	0.099	0.158	0.094	0.171	0.084	0.082	0.073	0.138	0.175	0.107	0.096
SA	**	**	**	**	**	**	**	**	0	0.066	0.079	0.134	0.034	0.042	0.143	0.042	0.027	0.054	0.075	0.085	0.148	0.085	0.169	0.091	0.086	0.076	0.147	0.122	0.116	0.092
MA	**	**	**	**	**	**	**	**	**	0	0.103	0.158	0.067	0.113	0.160	0.080	0.058	0.100	0.119	0.127	0.184	0.127	0.218	0.141	0.121	0.115	0.179	0.186	0.153	0.158
ME	**	**	**	**	**	**	**	**	**	**	0	0.110	0.073	0.106	0.144	0.078	0.072	0.101	0.094	0.124	0.117	0.125	0.217	0.145	0.107	0.101	0.150	0.163	0.106	0.126
CH	**	**	**	**	**	**	**	**	**	**	**	0	0.108	0.151	0.140	0.109	0.113	0.138	0.121	0.172	0.184	0.171	0.260	0.172	0.130	0.137	0.190	0.230	0.169	0.165
AR	**	**	**	**	**	**	**	**	**	**	**	**	0	0.028	0.118	0.028	0.012	0.064	0.060	0.076	0.117	0.067	0.149	0.088	0.065	0.070	0.100	0.123	0.108	0.076
TM	**	*	**	**	**	NS	**	NS	**	*	*	*	*	0	0.127	0.038	0.043	0.074	0.067	0.074	0.154	0.114	0.142	0.081	0.074	0.090	0.116	0.159	0.114	0.089
BI	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.094	0.111	0.130	0.141	0.145	0.212	0.186	0.242	0.152	0.104	0.131	0.173	0.241	0.116	0.179
TU	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.010	0.059	0.062	0.056	0.142	0.084	0.177	0.094	0.052	0.062	0.108	0.142	0.088	0.094
TS	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	NS	0	0.055	0.061	0.062	0.145	0.073	0.168	0.092	0.066	0.059	0.111	0.122	0.099	0.089
RA	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.084	0.097	0.163	0.114	0.172	0.097	0.074	0.080	0.139	0.133	0.095	0.115
SM	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.112	0.172	0.126	0.191	0.108	0.075	0.075	0.142	0.183	0.136	0.085
CB	**	**	**	**	**	*	**	**	**	*	**	**	**	**	**	**	**	**	**	0	0.166	0.127	0.199	0.126	0.087	0.090	0.144	0.172	0.084	0.137
SK	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.177	0.271	0.197	0.146	0.177	0.210	0.229	0.172	0.186
HS	**	**	**	**	**	**	**	**	**	**	**	**	**	*	**	**	**	**	**	**	**	0	0.220	0.138	0.106	0.092	0.163	0.186	0.176	0.180
FL	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.057	0.180	0.199	0.227	0.275	0.177	0.236
ZL	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.096	0.123	0.155	0.206	0.119	0.145
EH	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.067	0.107	0.171	0.102	0.120
LW	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.127	0.165	0.108	0.118
NR	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.173	0.150	0.150
SO	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.187	0.160
IC	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0	0.148
MO	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	0

Abbreviation: NS, nonsignificant. See Table 1 for breed codes.
All pairwise F_{st}'s are significant at $P<0.05$ after 435 000 permutations and standard Bonferroni correction except for the comparisons between Thônes-Marthod and Leccese, Thônes-Marthod and Istrian, and Turcana and Tsigai.
*$P<0.05$, **$P<0.01$.

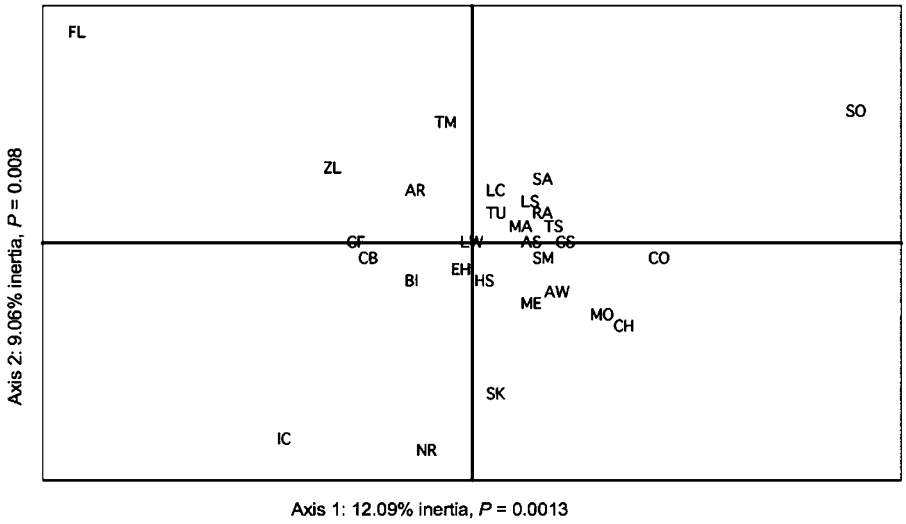

Figure 1 Principal component analysis (PCA) based on pairwise F_{st}. Both axes are significant. Corresponding F_{st} values are: axis 1, 0.02; axis 2, 0.015. See Table 1 for breed codes.

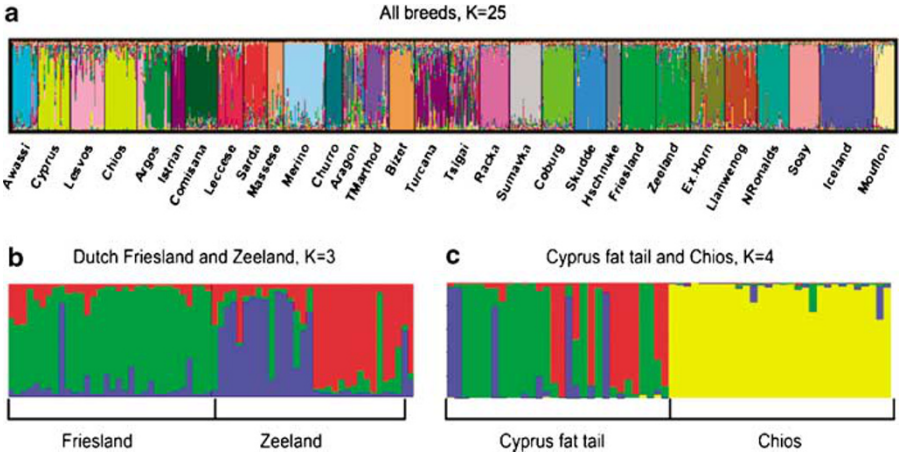

Figure 2 Individual assignment based on Bayesian cluster analysis. Plots are constructed using the program DISTRUCT (Rosenberg, 2004). The width of each segment relates to the sample size of the breed (Table 1). Individuals of different breeds are separated by black segments. Each individual is represented by a vertical line corresponding to its membership coefficient, that is a solid line corresponds to a membership coefficient of one (indicating no admixture). (**a**) Whole data set, $K=25$. Note the obvious admixture of Lesvos genotypes into Argos, the sharing of clusters between Turcana and Tsigai, Friesland and Zeeland, and Cyprus fat-tail and Chios. (**b**) Friesland and Zeeland investigated in more detail. The highest likelihood was found when $K=3$. Note the presence of two clusters in the Zeeland. (**c**) Cyprus fat-tail and Chios shown when $K=4$. Cyprus fat-tail individuals are split into several groups.

comparisons still show heterozygote deficit. Other studies have also reported heterozygote deficit in domestic sheep, although generally not to the extent described here (for example, Tapio *et al*., 2003, 2005; Calvo *et al*., 2006; Peter *et al*., 2007; Santucci *et al*., 2007). For other domestic species such as cattle (refs), pigs (refs), horses (refs) and goats (refs), deviation from HW equilibrium seems to be an exception rather than a rule.

Other than the occurrence of null alleles, there are two potential explanations for this widespread heterozygote deficit in European sheep: (1) subdivision among flocks leading to a Wahlund effect and (2) nonrandom mating due to inbreeding. First, subdivision is highly likely because although flocks are not closed units, gene flow may be restricted, especially between flocks that are geographically distant. If there is subdivision among flocks, sampling several flocks will lead to positive estimates of F_{is}. It is possible that our sampling strategy exacerbated this effect by including a small number of individuals from more than one flock. This is particularly illustrated by the case of the wild Mouflon, which was sampled from Corsica, Sardinia and Cyprus, and has the highest F_{is}. Subdivision caused by geographical isolation, confounded by sampling strategy, has therefore certainly contributed to heterozygote deficit. However, breeds for which just one flock was sampled also have positive F_{is} values (for example, Racka, Šumavka, Aragonesa, North Ronaldsay, Soay), suggesting that subdivision does not entirely account for HW deviations.

Second, nonrandom mating due to inbreeding is also likely, since a small number of rams are typically used for breeding purposes. Interestingly, the breed with the lowest F_{is} (excluding those with $n<15$) is the Soay, the only breed that is feral and unmanaged, with the potential for random mating and only fine-scale population structuring. Low values of F_{is} for the Soay have been reported previously and interpreted as due to inbreeding avoidance resulting from strict female philopatry and male-biased dispersal (Coltman *et al*., 1999, 2003).

Despite significant heterozygote deficit, levels of heterozygosity are relatively high in commercially managed breeds compared to the primitive, feral or wild sheep such as the Californian bighorn sheep, where $H_o = 0.28$–0.36 (Whittaker *et al*., 2004). A similar situation is found in domestic goats, which have substantially higher heterozygosity than their wild relatives (Saitbekova *et al*., 1999). This could suggest that selective breeding regimens are generally successful in maintaining diversity, but note that low diversity in wild species could be explained by successive bottleneck events, habitat fragmentation, isolation (Sage and Wolff, 1986) and strong fine-scale population structure (Coltman *et al*., 2003). Moreover, primitive breeds are often both genetically and geographically isolated, so genetic drift is likely to be a key factor in determining their diversity.

Although most of the genetic diversity is explained within breeds, the global estimate of F_{st} (13%) is quite high compared to the 7–11% reported for European cattle breeds (MacHugh *et al*., 1998; Katanen *et al*., 2000; Cañon *et al*., 2001; Mateus *et al*., 2004), or 8% for horses (Cañon *et al*., 2000). With the exception of only three comparisons (Turcana versus Tsigai, Thônes-Marthod versus Leccese and Thônes-Marthod versus Istrian), all breeds are significantly genetically differentiated. We identified several breeds that are very distinct and the Friesland and Soay breeds have particularly high F_{st} values. These breeds may harbour important disease resistance or uniquely adapted alleles and should, therefore, be given priority for conservation. Global estimates of F_{is} (12.3%) and F_{st} are similar; therefore, the variance explained by difference among breeds can be partially accounted for by correlations of allele frequencies within breeds.

Importance of geography and breed type

In the absence of recurrent introgression, we might expect most genetic diversity to be concentrated near the Near Eastern centre of domestication and then to decline with distance, since early farmers transported only a small sample of their diverse original stock (Bruford *et al*., 2003; Bruford, 2004). Such a pattern is suggested from analysis of sheep and cattle mitochondrial DNA sequences (Troy *et al*., 2001; Bruford and Townsend, 2006) and also from a recent Bayesian cluster analysis of European sheep breeds that revealed a southeast to northwest cline (Peter *et al*., 2007). Moreover, southern breeds have larger population sizes and have been transported more extensively than their northern counterparts (Mason, 1967). We therefore predicted higher within-breed diversity and lower genetic differentiation in southern than in northern breeds. Consistent with this prediction, the main geographical component to the partitioning of within-breed diversity is the difference between northern and southern breeds; southern breeds tend to have higher levels of heterozygosity than northern breeds (see also Peter *et al*., 2007), and the most distinctive breeds are all from northern Europe (Friesland, Icelandic, North Ronaldsay, Skudde and Soay). The cluster analysis also revealed more widespread evidence for admixture in southern than in northern breeds.

Domestic breeds are often categorized into 'types' according to morphological similarities, ecology and origins. Breeds of the same type are expected to be genetically similar, but this was generally not found to be the case. Although breed type explains the largest component of variance among groups, it only explains 2.7% of the total variation, and there is no general tendency for breeds to cluster by type in the PCA. In fact, three of the most differentiated breeds in the PCA (Soay, North Ronaldsay and Icelandic) are all of the northern short-tailed type. Similarities between breeds of the same type were, however, identified in the Bayesian cluster analysis. For example, the Dutch Friesland and Zeeland breeds, which are both of the marsh rat-tail type, always shared a cluster when $K = 23$–30.

Application to conservation and management

According to Weitzman (1992), for a breed to become a priority for conservation, it must add new diversity elements to a pre-defined set of conservation priority breeds. Breeds that add the highest overall genetic distance to the remainder of the set should be given priority (Bruford *et al*., 2003). Those breeds that are clearly differentiated in our PCA (Friesland, Icelandic, North Ronaldsay, Soay and Skudde) or stand out from our cluster analysis as harbouring unique genotypes (Soay, Icelandic, North Ronaldsay, Heidschnuke, Skudde, Coburg, Šumavka, Racka and Bizet) are, therefore, obvious targets for conservation. This approach should be treated with caution though because it fails to account for the diversity and geographical structure that can be found within some breeds. It should also be noted that neutral genetic variation does not necessarily correlate with variation in phenotype or quantitative traits, therefore, planning based on microsatellite variation alone might not be sufficient. Genetic distance estimates vary according to the marker used and the recent demographic history of the breed. For instance a severely bottlenecked breed will have a large genetic distance but often little within-breed diversity relative to other breeds. The Friesland, which has been bred true to maintain its high milk productivity following bottlenecks in the 1890s and 1960s (J Lenstra, personal communication), is a good example of this, since it exhibits very low within-breed diversity but is highly differentiated from other breeds.

Our analyses illustrate the utility of Bayesian cluster analysis to identify breeds that have or have not been bred true, which is potentially useful for breed management. Clear evidence of admixture was found between two Greek breeds, which are both Zackel × fat-tail type, with introgression of Lesvos genotypes into the Argos breed. A more fine-scale analysis clearly identified admixed individuals between the Friesland and Zeeland breeds, despite the efforts of breeders to keep the Friesland completely 'pure'.

In summary, genetic diversity in European domestic sheep breeds is characterized by extensive heterozygote

deficit, most likely due to subdivision within breeds and nonrandom mating due to inbreeding. Isolation in terms of geography or breed management has been crucially important in reducing within-breed variation and augmenting among-breed differentiation. We identified breeds that possess highly distinct genotypes, indicative of long histories of isolation, a distinct origin or the effects of selection. Finally, we illustrated that Bayesian clustering methods are valuable tools for detecting unrecorded admixture among breeds. This method could be valuable when breeders wish to establish current practices ongoing in their breeds.

Acknowledgements

This work was funded by the European Commission under contract BIO4961189 in the Biotech 2 programme of the Fourth Framework programme. The views expressed in this article do not represent those of the European Commission. We thank all farmers and collaborators who provided the many samples described in this study, including J Lenstra, LH Alderson, A Valentini, J Pemberton, A Knowles, A Reeb, Athalassa ARI Farm, M Ferrano, F Rosatini, M Albrecht, V Tauber, R Falge, M Reichel, L Head, D Ridd, G Leeves, O Dyrmundsson, M Hughes, I Bodo, C Ehling, F Pilla, A Alberti, GM Levrino, C Papachristoforou, E Rogdakis, P Vellema and J Futerova. LLH also thanks François Balloux and Jerome Goudet for helpful discussion.

References

Aberle KS, Hamann H, Drogemuller C, Distl O (2004). Genetic diversity in German draught horse breeds compared with a group of primitive, riding and wild horses by means of microsatellite DNA markers. *Anim Genet* **35**: 270–277.

Arevalo E, Holder D, Derr JN, Bhebe E, Linn RA, Taylor JF *et al.* (1994). Caprine microsatellite dinucleotide repeat polymorphisms at the SRCRSP1, SRCRSP2, SRCRSP3, SRCRSP4 and SRCRSP5 loci. *Anim Genet* **25**: 202.

Arranz J, Bayon Y, Primitivo FS (1998). Genetic relationships among Spanish sheep using microsatellites. *Anim Genet* **29**: 435–440.

Bhebhe E, Kogi J, Holder D, Arevalo E, Derr JN, Linn RA *et al.* (1994). Caprine microsatellite dinucleotide repeat polymorphisms at the SRCRSP6, SRCRSP7, SRCRSP8, SRCRSP9 and SRCRSP10 loci. *Anim Genet* **25**: 203.

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

Brezinsky L, Kemp SJ, Teale AJ (1993). Five polymorphic bovine microsatellites (ILSTS010-014). *Anim Genet* **24**: 75–76.

Bruford MW (2004). Conservation genetics of UK livestock: from molecules to management. In: Simm G, Villanueva B, Townsend S (eds). *Farm Animal Genetic Resources*. Nottingham University Press: Nottingham, UK, pp 151–169.

Bruford MW, Bradley DG, Luikart G (2003). DNA markers reveal the complexity of livestock domestication. *Nat Rev Genet* **4**: 900–910.

Bruford MW, Townsend SJ (2006). Case studies in the genetics of animal domestication: sheep. In: Zeder M, Decker-Walters D, Bradley D, Smith BD (eds). *Documenting Domestication: New Genetic and Archaeological Paradigms*. California University Press: Berkeley, CA, USA.

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

Buchanan FC, Crawford AM (1992a). Ovine dinucleotide repeat polymorphism at the Maf70 locus. *Anim Genet* **23**: 185.

Buchanan FC, Crawford AM (1992b). Ovine dinucleotide repeat polymorphism at the Maf209 locus. *Anim Genet* **23**: 183.

Buchanan FC, Crawford AM (1992c). 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, Galloway S, Crawford AM (1994). Ovine microsatellites at the OarFCB5, OarFCB19, OarFCB20, OarFCB48, OarFCB129, and OarFCB226 loci. *Anim Genet* **25**: 60.

Calvo JH, Bouzada JA, Jurado JJ, Serrano M (2006). Genetic substructure of the Spanish Manchega sheep breed. *Small Rumin Res* **64**: 116–125.

Cañon J, Alexandrino P, Bessa I, Carleos C, Carretero Y, Dunner S *et al.* (2001). Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genet Sel Evol* **33**: 311–332.

Cañon J, Checa ML, Carleos C, Vega-Pla JL, Vallejo M, Dunner S (2000). The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. *Anim Genet* **31**: 39–48.

Cañon J, Garcia D, Garcia-Atance MA, Obexer-Ruff G, Lenstra JA, Ajmone-Marsan P *et al.* (2006). Geographical partitioning of goat diversity in Europe and the Middle East. *Anim Genet* **37**: 327–334.

Chakraborty R, De Andrade M, Daiger SP, Budowle B (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Ann Hum Genet* **56**: 45–57.

Chapuis M-P, Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* **24**: 621–631.

Coltman DW, Pilkington JG, Pemberton JM (2003). Fine-scale genetic structure in a free-living ungulate population. *Mol Ecol* **12**: 733–742.

Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999). Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* **53**: 1259–1267.

Crawford A, Dodds K, Ede A, Pierson CA, Montgomery GW, Garmonsway HG *et al.* (1995). An autosomal genetic linkage map of the sheep genome. *Genetics* **140**: 703–724.

Dakin EE, Avise JC (2004). Microsatellite null alleles in parentage analysis. *Heredity* **93**: 504–509.

Dempster AP, Laird NM, Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* **39**: 1–38.

Diez-Tascon C, Littlejohn RP, Almeida PAR, Crawford AM (2000). Genetic variation within the Merino sheep breed: analysis of closely related populations using microsatellites. *Anim Genet* **31**: 243–251.

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 to Morocco. *Theor Appl Genet* **92**: 832–839.

Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

FAO (2000). Domestic Animal Diversity Information System (DAD-IS 2.0):http://www.fao.org/dad-is/.

Goudet J (2000). FSTAT, a programme to estimate and test gene diversities and fixation indices (version 2.9.1). Available from http://www.unil.ch/izea/softwares/fstat.html (updated from Goudet 1995).

Guo J, Du LX, Ma YH, Guan WJ, Li HB, Zhao QJ *et al.* (2005). A novel maternal lineage revealed in sheep (*Ovis aries*). *Anim Genet* **36**: 331–336.

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

Hiendleder S, Kaupe B, Wassmuth R, Janke A (2002). Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proc R Soc London B* **269**: 893–904.

Hiendleder S, Mainz K, Plante Y, Lewalski H (1998). Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from Urial and Argali sheep. *J Hered* **89**: 113–120.

Hoelzel AR, Dover G (1991). *Molecular Genetic Ecology.* Information Press: Oxford.

Katanen J, Olsaker I, Holm LE, 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.

Kemp S, Hishida O, Wambugu J, Rink A, Longeri ML, Ma RZ *et al*. (1995). A panel of polymorphic bovine, ovine and caprine microsatellite markers. *Anim Genet* **26**: 299–306.

Kim KS, Yeo JS, Choi CB (2002). Genetic diversity of north-east Asian cattle based on microsatellite data. *Anim Genet* **33**: 201–204.

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

Martín-Burriel I, Garcia-Muro E, Zargoza P (1999). Genetic diversity analysis of six Spanish native cattle breeds using microsatellites. *Anim Genet* **30**: 177–182.

Martinez AM, Delgado JV, Rodero A, Vega-Pla JL (2000). Genetic structure of the Iberian pig breed using microsatellites. *Anim Genet* **31**: 295–301.

Mason I (1967). *Sheep Breeds of the Mediterranean*. CAB: Slough, UK.

Mateus JC, Penedo MCT, Alves VC, Ramos M, Rangel-Figueiredo T (2004). Genetic diversity and differentiation in Portuguese cattle breeds using microsatellites. *Anim Genet* **35**: 106–113.

Mendelsohn R (2003). The challenge of conserving indigenous domesticated animals. *Ecol Econ* **45**: 501–510.

Paetkau D, Strobeck C (1995). The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol Ecol* **4**: 519–520.

Pariset L, Savarese MC, Cappuccio I, Valentini A (2003). Use of microsatellites for genetic variation and inbreeding analysis in Sarda sheep flocks of central Italy. *J Anim Breed Genet* **120**: 425–432.

Pedrosa S, Uzun M, Arranz JJ, Gutierrez-Gill B, Primitivo FS, Bayon Y (2005). Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proc R Soc London B* **272**: 2211–2217.

Peter C, Bruford MW, Perez T, Dalamitra S, Hewitt GM, Erhardt G, The Econogene Consortium (2007). Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Anim Genet* **38**: 37–44.

Peter C, Prinzenberg EM, Erhardt G, The Econogene Consortium (2005). Null allele at the OarAE129 locus and corresponding allele frequencies in German sheep breeds. *Anim Genet* **36**: 92.

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.

Rege JEO, Gibson JP (2003). Animal genetic resources and economic development: issues in relation to economic valuation. *Ecol Econ* **45**: 319–330.

Rosenberg NA (2004). DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* **4**: 137–138.

Rousset F, Raymond M (1995). Testing heterozygote excess and deficiency. *Genetics* **140**: 1413–1419.

Ryder M (1981). Sheep. In: Mason I (ed). *Evolution of Domesticated Animals*. Longman: London and New York, pp 63–85.

Sage RD, Wolff JO (1986). Pleistocene glaciations, fluctuating ranges, and low genetic-variability in a large mammal (*Ovis dalli*). *Evolution* **40**: 1092–1095.

Saitbekova N, Gaillard C, Obexer-Ruff G, Dolf G (1999). Genetic diversity in Swiss goat breeds based on microsatellite analysis. *Anim Genet* **30**: 36–41.

Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Santucci F, Ibrahim KM, Bruzzone A, Hewitt GM (2007). Selection on MHC-linked microsatellite loci in sheep populations. *Heredity* [E-pub ahead of print, 23 July 2007; http://dx.doi.org/10.1038/sj.hdy.6801006].

Schneider S, Roessli D, Excoffier L (2000). *ARLEQUIN ver. 2.000: a Software for Population Genetic Data Analysis*. Genetics and Biometry Laboratory, University of Geneva: Switzerland.

Scott P, Densmore LL (1991). Mitochondrial DNA isolation. In: Hoelzel AR (ed). *Molecular Genetic Analysis of Populations: a Practical Approach*. IRL Press: Oxford, UK, pp 29–58.

Signorello G, Pappalardo G (2003). Domestic animal biodiversity conservation: a case study of rural development plans in the European Union. *Ecol Econ* **45**: 487–499.

Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.

Steffen P, Eggen A, Dietz A, Womack JE, Stranzinger G, Fries R (1993). Isolation and mapping of polymorphic microsatellites in cattle. *Anim Genet* **24**: 121–124.

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.

Tapio M, Tapio I, Grislis Z, Holm LE, Jeppsson S, Kantanen J *et al*. (2005). Native breeds demonstrate high contributions to the molecular variation in northern European sheep. *Mol Ecol* **14**: 3951–3963.

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P *et al*. (2001). Genetic evidence for near-eastern origins of European cattle. *Nature* **410**: 1088–1091.

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 Genom* **5**: 288–297.

Walsh P, Metzger DA, Higuchi R (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**: 506–513.

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

Weitzman ML (1992). On diversity. *Quart J Econ* **CVII**: 363–405.

Whittaker DG, Osterman SD, Boyce WM (2004). Genetic variability of reintroduced California bighorn sheep in Oregon. *J Wildlife Manage* **68**: 850–859.

Wiener P, Burton D, Williams JL (2004). Breed relationships and definition in British cattle: a genetic analysis. *Heredity* **93**: 597–602.

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

Wright S (1977). *Evolution and the Genetics of Populations. Vol. 3: Experimental Results and Evolutionary Deductions*. University of Chicago Press: Chicago.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)