

Globally dispersed Y chromosomal haplotypes in wild and domestic sheep

J. R. S. Meadows^{*,†}, O. Hanotte[‡], C. Drögemüller[§], J. Calvo[¶], R. Godfrey^{**}, D. Coltman^{††}, J. F. Maddox^{‡‡}, N. Marzanov^{§§}, J. Kantanen^{¶¶} and J. W. Kijas^{*}

^{*}CSIRO Livestock Industries, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia 4067, Australia. [†]School of Rural Science and Agriculture, University of New England, Armidale 2351, Australia. [‡]International Livestock Research Institute, PO Box 30709, Nairobi 00100, Kenya. [§]Institute of Animal Breeding and Genetics, University of Veterinary Medicine Hannover, Bünteweg 2, D-30559 Hannover, Germany. [¶]Unidad de Tecnología en Producción Animal, CITA, Zaragoza 50059, Spain. ^{**}Agricultural Experiment Station, University of the Virgin Islands, RR 2, PO Box 10 000 Kingshill, 00850 The US Virgin Islands. ^{††}Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada T6G 2E9. ^{‡‡}Faculty of Veterinary Science, The University of Melbourne, Victoria 3010, Australia. ^{§§}All-Russian Research Institute of Animal Husbandry, Russian Academy of Agricultural Sciences, 142132 Moscow Region, Dubrovitsy, Russia. ^{¶¶}Biotechnology and Food Research, MTT Agrifood Research Finland, FI-31600 Jokioinen, Finland

Summary

To date, investigations of genetic diversity and the origins of domestication in sheep have utilised autosomal microsatellites and variation in the mitochondrial genome. We present the first analysis of both domestic and wild sheep using genetic markers residing on the ovine Y chromosome. Analysis of a single nucleotide polymorphism (*oY1*) in the *SRY* promoter region revealed that allele *A-oY1* was present in all wild bighorn sheep (*Ovis canadensis*), two subspecies of thinhorn sheep (*Ovis dalli*), European Mouflon (*Ovis musimon*) and the Barbary (*Ammontrags lervia*). *A-oY1* also had the highest frequency (71.4%) within 458 domestic sheep drawn from 65 breeds sampled from Africa, Asia, Australia, the Caribbean, Europe, the Middle East and Central Asia. Sequence analysis of a second locus, microsatellite *SRYM18*, revealed a compound repeat array displaying fixed differences, which identified bighorn and thinhorn sheep as distinct from the European Mouflon and domestic animals. Combined genotypic data identified 11 male-specific haplotypes that represented at least two separate lineages. Investigation of the geographical distribution of each haplotype revealed that one (H6) was both very common and widespread in the global sample of domestic breeds. The remaining haplotypes each displayed more restricted and informative distributions. For example, H5 was likely founded following the domestication of European breeds and was used to trace the recent transportation of animals to both the Caribbean and Australia. A high rate of Y chromosomal dispersal appears to have taken place during the development of domestic sheep as only 12.9% of the total observed variation was partitioned between major geographical regions.

Keywords breed diversity, haplotype, sheep, Y chromosome.

Introduction

Mesolithic man began the process of sheep domestication approximately 8000–9000 years ago (Ryder 1984), and this ongoing process has resulted in the establishment of

more than 1400 breeds (Scherf 2000). These are defined by phenotypic variation in coat colour, environmental tolerance, wool characteristics and food production traits. To date, studies into the process of domestication and the structure and relationship between modern *Ovis* populations have relied on autosomal microsatellites and mitochondrial (mt) DNA. Allelic polymorphism at microsatellite loci has proven particularly successful for distinguishing closely related breeds (Arranz *et al.* 1998, 2001; Diez-Tascon *et al.* 2000) and for precisely assessing population structure even where breed classification is a poor predictor of genetic similarity (Tapio *et al.* 2005). Analysis of mtDNA

Address for correspondence

J. Kijas, CSIRO Livestock Industries, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia 4067, Australia.
E-mail: james.kijas@csiro.au

Accepted for publication 31 May 2006

has failed to establish well-resolved phylogenies between breeds (Meadows *et al.* 2005); however, it has proven particularly insightful concerning the process of domestication. Maternal haplotypes form at least three clades: A, B and C (Heindleder *et al.* 2002; Wu *et al.* 2003; Guo *et al.* 2005; Meadows *et al.* 2005; Pedrosa *et al.* 2005). Clade A is predominantly found in Asian breeds, B in European animals and C in Chinese and Turkish breeds, although high levels of introgression between major geographical regions have been observed (Meadows *et al.* 2005). Each clade arose either from independent and temporally distinct domestication events or from a smaller number of events where sampling was made from genetically divergent ancestral populations.

Investigations into the population expansion and migratory history of modern human populations have recently focused on variation within the male-specific region of the Y chromosome (MSY). Observation of haploid state alleles in the absence of X-Y recombination allows haplotypes that define individual male lineages to be identified. This has proven highly effective for the study of global dispersal (Kayser *et al.* 2000, 2001; Deng *et al.* 2004) and for dating the occurrence of human population expansions (Shen *et al.* 2000). Examination of male lineages is likely to be of particular importance in domesticated species where controlled matings result in a small number of males contributing a disproportionately large number of progeny to subsequent generations. To date, however, few phylogenetic surveys involving the Y chromosome have been reported in domestic species due to a paucity of MSY variation. Very low rates of nucleotide diversity have been reported within the MSY of horse (Lindgren *et al.* 2004), cattle (Hellborg & Ellegren 2004) and sheep (Meadows *et al.* 2004). The few studies that have utilised MSY variation have revealed intriguing aspects of population history. Bradley *et al.* (1994) developed a Y-specific probe that distinguished subspecies *Bos taurus* (taurine) and *Bos indicus* (zebu). Analysis across African populations revealed a complex pattern of male-mediated zebu introgression, which would have remained hidden in the absence of Y chromosomal data (MacHugh *et al.* 1997; Hanotte *et al.* 2000; Freeman *et al.* 2004). Analysis of Asian cattle populations using a single nucleotide polymorphism (SNP) in the bovine sex determining gene (Kikkawa *et al.* 2003) also revealed an unequal contribution of males and females during the foundation of a variety of breeds (Mannen *et al.* 2004).

The aim of this study was to develop the first set of Y haplotypes for the study of evolution, breed relatedness and domestication in sheep. To examine the global distribution of male lineages, 65 breeds were sampled from Australia, Africa, Asia, the Caribbean, Europe and the Middle East. A number of wild species were also investigated to determine the ancestral state of MSY loci and to provide insight into their relationship with domesticated animals.

Materials and methods

Animals

A total of 519 animals were sourced from seven major geographical regions representing 65 breeds of domestic sheep and four wild species. The number of animals sampled per population and three letter abbreviations for each breed are given in Table 1. Asian animals from Mongolia (MON), Tibet (TIB) and Indonesia (JTT) have been described previously (Meadows *et al.* 2005). African animals (ASB, BAL, PED, SAB, SUK, TSW and WDD) from six countries were collected as part of studies conducted at the International Livestock Research Institute to characterise livestock genetic resources. Two hair sheep breeds from the Caribbean (BAB and STC) were collected on the US Virgin Islands. Breed samples from Israel (AWI), Kazakhstan (KAK and EDB), Azerbaijan (BAZ) and southern Russia (TSG) were collected within their respective countries of origin and considered as a single major geographical region termed the Middle East and Central Asia (Table 1). Breeds from Northern Europe, Continental Europe and Western Russia (34 breeds) were considered as a separate geographical unit (Europe; Table 1). Breeds from Finland (GFS, WFS, BRF and BLF), Åland (ALD), Russian Karelia (VEP and VEN), Austria (TMS, TSS, CS and FS) and Spain (LAX, MAN and RAG) have been described previously (Tapio *et al.* 2003; Calvo *et al.* 2004; Meadows *et al.* 2005). Eighteen German breeds (Table 1) were collected as part of a national screening service for prion protein gene variants (Drogemüller *et al.* 2001, 2004). Animals from 24 breeds were collected in Australia. These were considered 'imported' (Table 1) because they were originally developed in either Europe (12 breeds) or Africa (NQA and RON) prior to their importation, continued selection and sampling in Australia. Three wild species were sampled from within their native range. Two subspecies of wild thinhorn sheep, Dall's sheep (*Ovis dalli dalli*, DAL) and Stone's sheep (*Ovis dalli stonei*, STN) were collected from Alaska, Yukon, the Northwest Territories and British Columbia (Worley *et al.* 2004); bighorn rams (*Ovis canadensis*, CAN) were sampled from Alberta (Coltman *et al.* 2003); and the European Mouflon (*Ovis musimon*, MFL) was sampled from the El Hosquillo National Wildlife Reserve in central Spain (Santiago-Moreno *et al.* 2000). African Barbary sheep (*Ammonotragus lervia*, BAR), phylogenetically a closer relative of the goat than of the domestic sheep, were sampled from a captive population that is located at Western Plains Zoo (New South Wales, Australia) and expected to be highly inbred.

Analysis of SNP oY1

Genetic marker AY604734.2:g.67A>G (oY1) is a SNP located in the 5'-promoter region of the ovine sex determining region Y (SRY) gene (Meadows *et al.* 2004).

Table 1 Geographic origin, breed code, number of individuals and haplotype distribution for 65 breeds of domestic sheep and four wild species.

Region	Breeds	Breed code	(n)	Haplotype										
				H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
Africa														
Botswana	Tswana	TSW	4				1		3					
Ethiopia	Arsi Bale	ASB	5						4			1		
Nigeria	Balami	BAL	9						5		3		1	
	West African Dwarf Djallonke	WDD	9						6		3			1
South Africa	Pedi	PED	5						4		1			
Tanzania	Sukuma	SUK	5				3		2					
Zimbabwe	Sabi	SAB	5						4		1			
							4	0	28	0	8	1	1	0
Asia														
Indonesia	Javanese Thin Tail	JTT	5						5					
Mongolia	Mongolian	MON	3						3					
Tibet	Tibetan	TIB	19				4	2	13					
							4	2	21	0	0	0	0	0
Caribbean														
US Virgin Islands	Barbados Blackbelly	BAB	6					6						
	St Croix	STC	8					7	1					
							0	13	1	0	0	0	0	0
Europe														
Austria	Carynthian sheep	CS	10					9			1			
	Forest sheep	FS	2						2					
	Tyrolean Mountain sheep	TMS	11								11			
	Tyrolean Stone sheep	TSS	12						10		2			
Finland	Aland	ALD	5					3	2					
	Black Finnsheep	BLS	6					2	1	3				
	Brown Finnsheep	BRS	12					4	6	2				
	Grey Finnsheep	GFS	7						3	4				
	White Finnsheep	WFS	6					1	1	4				
Germany	Bavarian Forest sheep	BFS	6					1	5					
	Bentheim	BEN	7						7					
	Black-headed Mutton	BHM	8					8						
	Black-brown Milk sheep	BBM	9						9					
	Brown Mountain sheep	BMS	5								5			
	Coburg	COB	9						9					
	East Friesian milk sheep	EFM	7						7					
	German Grey Heath	GGH	10						9		1			
	German Merino	GM	4					2	2					
	Leine	LEI	8					5	3					
	Merino Longwool	MLW	5						5					
	Mutton Merino	MM	9					1	8					
	Pomeranian Coarsewool	PCW	9					7	2					
	Skudde	SKD	9						8		1			
	White-headed Mutton	WHM	6					6						
	White-horned Heath	WHH	9						8	1				
	White Mountain sheep	WMS	4								4			
	White-polled Heath	WPH	9						9					
Russia	Kuibyshev	KUI	5					5						
	Vepsia	VEP	5						1	1	3			
	Viena	VEN	6							6				
Spain	Latxa	LAX	12						12					
	Manchega	MAN	10						10					
	Mouflon (<i>Ovis musimon</i>)	MFL	10						10					
	Rasa Aragonesa	RAG	10						9	1				
Ukraine	Carpathian Mountain sheep	CMS	2						2					

Table 1 (Continued)

Region	Breeds	(n)	Haplotype										
			H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
						0	54	160	22	28	0	0	0
Imported													
Australia	Black Suffolk	BSF	3				3						
	Border Leister	BOL	1				1						
	Coopworth	COO	4				3	1					
	English Leister	ENL	4				4						
	Lincoln	LCN	10				9	1					
	Merino	MER	24					24					
	Perendale	PER	1					1					
	Polwarth	POL	3					3					
	Romney	ROM	12				12						
	Texel	TEX	3				3						
	White Suffolk	WSF	4				4						
	Whitshire	WIS	1				1						
	Namaqua Afrikaner	NQA	8					8					
	Ronderib Afrikaner	RON	17			8		9					
	Barbary (<i>Ammotragus leivia</i>)	BAR	17										17
						8	40	47	0	0	0	0	17
Middle East/Central Asia													
Azerbaijan	Bozakh	BOZ	5					4		1			
Israel	Awassi	AWI	10			1		7		2			
Kazakhstan	Edilbaevskaya	EDB	2					2					
	Karakul	KAK	4							4			
Russia	Tsigai	TSG	5					4		1			
						1	0	17	0	8	0	0	0
North America													
Canada	Bighorn sheep (<i>Ovis canadensis</i>)	CAN	6	5	1								
Canada and Alaska	Thinhorn – Dall's (<i>Ovis dalli dalli</i>)	DAL	19	9	10								
	Thinhorn – Stone's (<i>Ovis dalli stonei</i>)	STN	9	4	3	2							
			13	18	3								
Total			519	13	18	3	17	109	274	22	44	1	17

Primer pair SRY-U1 (AGCTCCAGAATATTTCACTGACCT) and SRY-D1 (GAAGGCAAATGCAGAGACAA) amplify a 130-bp fragment spanning *oY1* (AY604734). SNP detection was performed using an Applied Biosystems (ABI, Scoresby, Australia) allelic discrimination TaqMan Assay. Polymerase chain reaction (PCR) were performed in volumes of 10 µl, which contained TaqMan 2X PCR Master Mix (ABI), 0.2 M of SRY-U1 and SRY-D1, 0.3 M of competing fluorogenic probes SRY-3G (VICAACCTGAGCAGCTTA) and SRY-3A (FAMAACCTGAGCAGCTTAG) and 2 ng of template genomic DNA. Cycling was performed in two stages using the ABI PRISM 7700: one two-step cycle of 50 °C for 2 min and 95 °C for 10 min, followed by 50 two-step cycles of 30 s at 95 °C and 1 min at 60 °C. Results were analysed and visualised using SDSv2.1 software (ABI).

PCR analysis of microsatellite loci

Four microsatellites were assayed for ovine Y-chromosomal specificity. Bovine-derived microsatellites *INRA189*,

INRA126 and *INRA124* were reported to be male-specific in Togo sheep, and primers for their amplification have been described previously (Edwards *et al.* 2000). Microsatellite *SRYM18* was obtained from bacterial artificial chromosome (BAC) clone 574 R7C5, which was isolated from an ovine BAC library (Gill *et al.* 1999). Partial digestion of the BAC clone was performed with *AluI* prior to being subcloned into pCRScript (Stratagene, Annandale, Australia) and transformed into JM109 cells (Promega, Chatewood, Australia). The clone containing microsatellite *SRYM18* was identified and sequenced, and the primers *SRYM18F* (GGCATCACAAACAGGATCAGCAAT) and *SRYM18R* (GTGATGGCAGTTCTCACAATCTCTCT) were synthesised for PCR amplification of *SRYM18*. PCR was performed using standard conditions (Meadows *et al.* 2004); the forward primer of each microsatellite set was fluorescently labelled (GeneWorks, Hindmarsh, Australia) to facilitate product detection. Genomic DNA samples from two male and female cattle and from two male and female sheep were used to test for sex-specific amplification. All

amplified products were detected using the ABI Prism 377 DNA Sequencer before fragment sizes were determined using GeneScan 3.1 and Genotyper 2.5 software (ABI).

Cloning and sequence analysis of SRYM18 alleles

Primers SRYM18F (unlabelled) and SRYM18R were used to amplify PCR products from 17 animals in preparation for cloning and sequencing (*O. dalli*, $n = 3$; *O. canadensis*, $n = 2$; *O. musimon*, $n = 2$; *O. aries*, $n = 8$ and *A. lervia*, $n = 2$). Amplified fragments (106–145 bp) were cloned, and a minimum of three clones were sequenced from each fragment. Sequences were deposited into GenBank with accession nos DQ272448–DQ272464.

Data analysis

Haplotypes were constructed by combining sequence data describing the repeat structure of SRYM18 with genotypic data from *oY1* (Table 2). Genotyping was then performed and haplotypes assigned based on fragment length (SRYM18) and SNP genotype (*oY1*). It should be noted that the assignment of haplotypes, in the presence of demonstrated microsatellite homoplasy, relied on fragment length for the majority of individuals. The relationship between haplotypes was investigated by constructing median networks using Network version 4.1.1.2 (<http://www.fluxus-engineering.com>). Haplotype components (Table 2) were weighted (w) as follows: SRYM18 [TG] $_n$, $w = 1$; SRYM18 Indel, $w = 2$; SRYM18 [TTTTG] $_m$, $w = 2$; and SNP *oY1*, $w = 3$. This generated a priority order where the component with the lowest expected mutation rate (SNP *oY1*) was assigned the highest ranking (Bandelt *et al.* 2000). Haplotypes 10 and 11 were excluded as neither contained the SRYM18 repeat array. A median joining network (to be most conservative, $\epsilon = 0$) was constructed from all animals

that carried H1–H9 ($n = 501$; Fig. 1a) before the frequency of haplotypes within each major geographical region was expressed using region-specific median networks (Fig. 1b). Assigned haplotypes (H1–H11) were used to interrogate global population structure following grouping of animals into geographical regions. Using analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) as implemented in Arlequin (<http://cmpg.unibe.ch/software/arlequin3>), total genetic variation was partitioned into (i) between region, (ii) between breeds within region and (iii) within breed. Two groupings were tested where, in each case, animals sampled from within Australia ($n = 112$) were omitted due to their assumed 'imported' status. The first grouping considered all remaining wild and domestic individuals ($n = 407$) as originating from six major geographical regions (Table 1). A second analysis considered the subset consisting only of domestic animals ($n = 373$) grouped into five geographical regions.

Results

Localisation of microsatellite loci to the male-specific region of the ovine Y chromosome

One ovine- (SRYM18) and three bovine- (INRA189, INRA126 and INRA124; Edwards *et al.* 2000) derived microsatellite primer sets were tested for amplification of fragments from the male-specific region of the sheep and cattle Y chromosome (MSY). Amplification was observed from both male and female ovine DNA (data not shown). This indicated that the three bovine-derived microsatellites reside outside the sheep MSY, and they were not studied further. Primers for SRYM18 amplified a single product from male and failed to amplify product from female samples using both sheep and cattle DNA. This located SRYM18 to the MSY region of the Y chromosome in both species.

Table 2 Comparison of sheep haplotypes generated through the combination of a compound microsatellite and single nucleotide polymorphism (SNP), both from the sex-determining region (SRY) of the ovine Y chromosome.

Haplotype	SRYM18 ¹			Allele (bp)	SRY SNP	Species observation
	[TTTTG] $_m$	Indel G/–	[TG] $_n$			
H1	4	–	14	145	A	<i>Ovis dalli</i>
H2	4	–	13	143	A	<i>O. dalli</i> / <i>O. canadensis</i>
H3	4	–	11	139	A	<i>O. dalli</i> / <i>O. canadensis</i>
H4	3	G	16	145	A	<i>O. aries</i>
H5	3	G	16	145	G	<i>O. aries</i>
H6	3	G	15	143	A	<i>O. aries</i> / <i>O. musimon</i>
H7	3	G	15	143	G	<i>O. aries</i>
H8	3	G	14	141	A	<i>O. aries</i>
H9	3	G	9	131	A	<i>O. aries</i>
H10	–	–	–	110	A	<i>O. aries</i>
H11	–	–	–	106	A	<i>Ammotragus lervia</i>

¹SRYM18 is a compound microsatellite comprised of a pentanucleotide [TTTTG] $_m$ and dinucleotide [TG] $_n$ repeats separated by an insertion/deletion (indel G/–). The total SRYM18 length is reported in base pairs (bp).

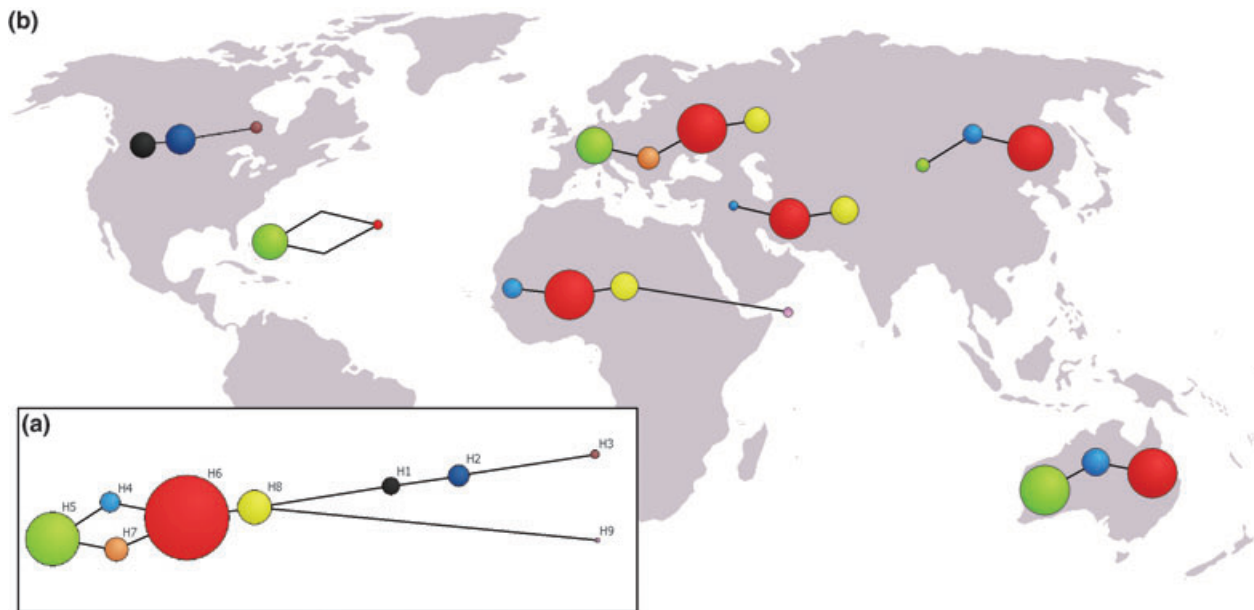


Figure 1 Median joining network diagrams show the approximate location of the seven major geographical regions from which sheep were sampled. (a) A median network diagram was constructed using haplotype data from 519 domestic and wild sheep. The diagram illustrates both the relationship between haplotypes and their observed frequency. The diameter of H9 represents one animal. (b) Breeds were grouped into seven major geographical regions as described in Table 1. The size of the individual network diagrams are region-specific and, therefore, are not directly proportional to each other.

Sequence architecture of microsatellite *SRYM18*

Amplification of *SRYM18* from 519 male domestic and wild sheep revealed a total of seven different length fragments with sizes ranging from 106 to 145 bp (Table 2). Wild and domestic sheep displayed the same length fragment for two of the seven products (143 and 145 bp; Table 2). This prompted cloning and sequence-based investigation of the *SRYM18* repeat motif within each length variant observed in domestic animals: the Mouflon, Bighorn, Thinhorn and Barbary sheep. The microsatellite array was found to be compound, comprising a pentanucleotide and a dinucleotide repeat separated by an indel ([TTTTG]_mG/-[TG]_n). The pentanucleotide repeat was fixed for a different array size in wild species (*O. dalli* ssp. and *O. canadensis*; four repeat units) compared with domestic sheep and the European Mouflon (*O. aries* and *O. musimon*; three repeat units; Table 2). In contrast, the dinucleotide repeat was polymorphic in both wild and domestic animals, displaying between nine and 16 tandem repeats (Table 2). This resulted in observed microsatellite homoplasy, where PCR products with the same fragment lengths contained a different repeat structure when compared between wild and domestic alleles (Table 2). Sequencing of the two shortest fragments (106 and 110 bp) revealed an allele that completely lacked the repeat array. Sequence alignment and comparison of the unique flanking regions (excluding primers) revealed that the 106-bp (*A. lervia*) and 110-bp fragments (*O. aries*) displayed 89.7% and 86.3% similarity compared with the longest length fragment (*O. aries*). This confirmed ampli-

cation of the *SRYM18* locus and allelism for all of the amplified products.

Determination of the ancestral allele of SNP *oY1*

The second locus that we investigated was an A>G SNP located upstream of *SRY* that has been previously described (Meadows *et al.* 2004). It was termed ovine Y SNP 1 (*oY1*) and was genotyped in 458 domestic and 61 wild individuals. In every case, a single allele (*A-oY1* or *G-oY1*) was detected per individual, which is consistent with the location of *oY1* on the male-specific portion of the Y chromosome. *A-oY1* was fixed in all of the European Mouflon (*O. musimon*, *n* = 10), thinhorn sheep (*O. dalli* ssp., *n* = 28), bighorn (*O. canadensis*, *n* = 6) and Barbary sheep tested (*A. lervia*, *n* = 17; Table 1). Investigation within 65 domestic breeds of sheep sampled from across Africa, Asia, the Caribbean, Europe, the Middle East/Central Asia and Australia revealed that *A-oY1* had the highest frequency (71.4%); however, *G-oY1* was segregating in 30 different breeds. Together these results indicate that *A-oY1* is the ancestral allele and that the mutational event giving rise to *G-oY1* likely occurred after the most recent speciation event that defines the lineage of the domestic sheep (*O. aries*).

Y chromosomal haplotypes and their distribution in both wild and domestic sheep

A total of 11 haplotypes (H1–H11) were defined by combining variation observed at the two Y loci (*SRYM18* and

oY1). The number of repeat units present at the pentanucleotide array of *SRYM18* grouped haplotypes H1–H3 separately from H4 to H9. Haplotypes within each of these two groups were further defined on the basis of the dinucleotide component of *SRYM18* and variation at *oY1* (Table 2). The final two observed haplotypes (H10 and H11) completely lack the microsatellite repeat and were distinguished based on *SRYM18* allele length. A median joining network was created using these molecular differences to represent the most likely relationship between haplotypes (Fig. 1a). Thinhorn and bighorn sheep haplotypes (H1–H3) were separated from those found in domestic animals, with H8 linking the two sections of the network. Haplotype H9 was separated from the torso of the network due to variability at the dinucleotide component of *SRYM18*. The distribution of haplotypes in animals sampled from different geographical regions was determined using fragment length analysis (*SRYM18*) and allele discrimination (*oY1*). The number and breed of animals observed carrying each haplotype is presented in Table 1, and the frequency observed within different geographical regions is presented in Fig. 1b. In the sample of domestic sheep, haplotype H6 (red-filled circle, Fig. 1) was the most prevalent (frequency $f = 0.53$) and was detected within every major geographical region from which *O. aries* were sampled. A total of 81% of breeds where five or more individuals were sampled contained an animal carrying H6 (39/48 breeds). Other haplotypes displayed more restricted geographical distributions. Haplotypes H4 (blue), H5 (green) and H8 (yellow) each showed a range of dispersal that overlapped in the Middle East/Central Asian region, but the combination was otherwise limited to a region. For example, haplotype H4 was present in animals from Africa, Middle East/Central Asia and Asia but not in European sheep, even though the latter represented the largest portion of individuals sampled. Haplotype H5 was present in Asia and Europe but was absent in animals sampled from Africa. Haplotype H7 (orange) was confined to 22 animals representing 8 European breeds (22/257), whilst H9 was observed in a single animal from the East African Arsi Bale (ASB; Table 1). Haplotype H10 was also observed in a single African animal (BAL), while H11 was observed in all of the Barbary sheep tested (*A. lervia* BAR, $n = 17$).

Y haplotypes reveal a weak phylogeographical structure in ovine patrines

The level of phylogeographical structure revealed by Y haplotypes (H1–H11) was explored by calculating the distribution of variance following imposition of two groupings. When the haplotypes observed in both wild and domestic animals ($n = 407$) were grouped into six regions (Africa, Asia, the Caribbean, Europe, the Middle East/Central Asia and North America; Table 1), over one-third of the observed variation (35.7%) was attributed to geographical specificity.

The remaining variation was present either within breeds (28.5%) or between breeds within regions (35.8%). Consideration only of domestic animals ($n = 373$) grouped into five geographical regions (Africa, Asia, the Caribbean, Europe and the Middle East/Central Asia) resulted in a reduced amount of variation being attributed to geographical origin (12.9%). This variation is approximately the same as that observed between geographical groupings of canine and human Y chromosomes (11.5% and 16.8% respectively; Kayser *et al.* 2001; Bannasch *et al.* 2005). The remainder of the variation in domestic sheep was partitioned either within breeds (37.0%) or between breeds within regions (50.1%).

Haplotype distribution within populations of imported Australian sheep

Individuals from breeds that were originally developed in Africa (NQA and RON) and Europe (POL, ROM, COO, TEX, LCN, BSF, WSF, ENL, PER, WIS, BOL and MER) were sampled from within Australia. The distribution of Y haplotypes observed in these imported breeds was considered separately from and compared against animals sampled from their continent of origin. Australian sheep of European descent ($n = 70$) carried only two haplotypes (H5, $f = 0.57$ and H6, $f = 0.43$; Table 1). A survey of 32 breeds sampled from continental Europe, western Russia and Scandinavia revealed these same two haplotypes were observed with the highest frequency (H5, $f = 0.20$ and H6, $f = 0.61$; Table 1). Australian sheep of African origin ($n = 25$) also contained only two haplotypes (H4, $f = 0.32$ and H6, $f = 0.68$), both of which were observed in breeds sampled in Africa (Table 1). Haplotypes H5 and H7 were absent from the imported African animals, which is consistent with the finding that these haplotypes were only observed in breeds sourced from Europe and Asia.

Discussion

The Y haplotypes defined in this study are comprised of a bi-allelic SNP and a multi-allelic microsatellite. The status of each was determined within wild sheep in order to establish their ancestral evolutionary state. The interpretation for SNP marker *oY1* was straightforward as each wild species was fixed for allele *A-oY1*; hence identifying it as the ancestral allele. The evolutionary history of *SRYM18* is more complex and difficult to unambiguously reconstruct. It contains a compound tandem array which displays non-uniform mutagenic properties. The pentanucleotide component appeared fixed within different species. This suggests it undergoes repeat expansion and contraction less frequently when compared with the dinucleotide component of the array, which displayed length variability within species (repeat range 11–14 in *O. dalli* ssp.; 9–16 in *O. aries*). The compound nature of the repeat means that care is required

during the interpretation of genetic similarity between animals that carry identical-by-state haplotypes. Inspection of a global sample of domestic breeds revealed that one haplotype (H6) was both very common and widely distributed. The same finding arose from analysis of Y haplotypes in phenotypically divergent dog breeds where one common haplotype was distributed across groups (Sundqvist *et al.* 2006). It is possible, however, that a significant portion of animals that carry H6 are not identical-by-descent due to homoplasy or the mutability and recurrent mutation associated with microsatellite loci (Heyer *et al.* 1997; de Knijff 2000). The observation that allele *G-oY1* was observed with only two of the seven *SRYM18* length variants suggests that the mutation rate at *SRYM18* is not sufficiently high to generate all of the theoretical or predicted haplotypes. Recurrent mutation, therefore, is unlikely to have acted to hide large amounts of population substructure within the group of animals that carry H6. The impact of homoplasy remains unclear, and the incorporation of additional Y markers is required before the observation of a common and geographically widespread haplotype can be conclusively assessed.

Two haplotypes (H5 and H7) were distinguished from all others by the minor SNP allele *G-oY1*. Given the clear evolutionary history of *oY1*, we propose that *G-oY1* defines haplotypes H5 and H7 as members of a single paternal lineage. This is supported by the distinct distribution of H5 and H7 (Fig. 1, green and orange), as both were completely absent from the African animals and those sampled from the Middle East and Central Asia. Haplotype H7 was unique to European breeds, while H5 had the second-highest frequency in European animals after H6 (Table 1). Interestingly, H5 was observed in a number of other major geographical regions, which likely reflects the documented transportation of animals during the last few hundred years. Almost all animals from two breeds of hair sheep developed in Barbados and the Virgin Islands (BAB and STC) carried haplotype H5. Anecdotally, these are assumed to have been derived from African hair breeds; however, the presence of H5 strongly suggests that the paternal origin of both breeds is European and not African. This is consistent with colonialist migration in the 17th century, where European animals were introduced into existing hair sheep populations across the Caribbean (Delgado *et al.* 2000). Similarly, a number of European breeds imported to Australia were fixed for haplotype H5, reflecting known transportations of English breeds during the foundation of the Australian colonies during the 18th century (Parsonson 1998). The presence of H5 within two Tibetan animals was unexpected based on historical accounts; however, it may not be surprising given the level of gene flow between European and Asian animals revealed using mtDNA (Meadows *et al.* 2005).

The geographical distribution of haplotype H4 (Fig. 1, blue) displayed a completely different pattern compared

with H5 and H7. This is preliminary evidence to suggest H4 is a member of a second paternal lineage. It was observed in animals sampled from Southern (TSW) and Eastern (SUK) Africa, the Middle East (AWI) and Asia (TIB) while being completely absent from any of the 34 European breeds tested. Archaeological data has implicated the Near East as a centre of ovine domestication (reviewed by Bar-Gal *et al.* 2003) before animals likely moved east and west into Africa following the route of zebu cattle (Ryder 1984). It is tempting to suggest this was reflected in the distribution of haplotype H4. It is clear that additional male-specific SNPs are required to define both the number and evolution of male lineages; however, this study indicates the presence of at least two in domestic sheep.

Analysis of the Y chromosome from wild sheep offered the opportunity to evaluate genetic relatedness with domestic animals using a component of the genome that has not been investigated to date. Previous studies based on mtDNA variation have shown that the mouflon is the most likely living relative linking domestic and ancient sheep, and that a number of other *Ovis* species are phylogenetically removed (Heindleder *et al.* 2002; Wu *et al.* 2003). Considering *Ovis* species from overlapping ranges can successfully hybridise and produce fertile offspring (reviewed in Heindleder *et al.* 2002), the potential paternal contribution of four species was investigated in this study. The number of fixed pentanucleotide repeats present at *SRYM18* clearly distinguished Thinhorn and Bighorn sheep from both the Mouflon and domestic sheep. This confirmed a phylogenetic division in the genus and indicates that no direct paternal contribution of mountain sheep (*O. canadensis* and *O. dalli*) occurred during the establishment of domestic breeds. In addition, the observation that the Mouflon was fixed for the most common haplotype observed in domestic sheep (H6) is consistent with it being a feral derivative of early domestic sheep.

The results also contribute to the understanding of the evolution of mountain sheep in North America. British Columbian Stone's sheep were found to share a Y-chromosome haplotype (H3) with bighorn that is not found in more northerly Stone's sheep or in Dall's sheep (Table 1). Together with microsatellite data suggesting that British Columbian Stone's sheep have undergone population expansion following a bottleneck (Worley *et al.* 2004), this suggests that they represent a distinct evolutionary lineage from other thinhorn sheep.

An unexpected finding arising from investigation of the Barbary sheep was the identification of an allele completely lacking the repeat array. It is unclear if the array had been lost, or if the allele represents an ancestral version prior to the development of polymorphism. Interestingly, a second allele lacking the repeat array was identified within a domestic animal of the African Balami breed. The relationship between the Barbary, a species originally from Saharan Africa, and the Balami breed is not clear. To date, no evidence has been reported indicating domestication of the

Barbary, and while it is theoretically possible for these animals to cross-breed and produce fertile offspring, the only reported mating used artificial insemination and failed to result in full-term pregnancy (Moore *et al.* 1981). Gene flow from the Barbary to domestic sheep, therefore, appears unlikely; however, additional MSY markers are required before it can be formally disproven. Whilst a more comprehensive sampling of domestic breeds and wild species from the Middle East and Asia are required to further our understanding of domestication and development of global sheep populations, it is clear from the data presented here that at least two paternal lineages are present in modern domestic sheep.

Acknowledgements

The authors thank Kui Li (Chinese Academy of Agricultural Sciences), Wolfgang Sipos (University of Veterinary Medicine, Vienna), Anthony Duffy (Rare Breeds Trust of Australia), Tamesha Gardner (Ida Vale Dorper Stud, Australia), Elisha Gootwine (Institute of Animal Science, Israel) and Belinda Norris (CSIRO) for contributing samples. North American mountain sheep DNA was extracted by Kirsty Worley from tissues provided by Alastair Veitch (Department of Resources, Wildlife & Economic Development, Government of the Northwest Territories), Jean Carey (Department of Environment, Government of Yukon), Helen Schwantje (Ministry of Environment, British Columbia), Steve Arthur (Alaska Department of Fish and Game) and Jon Jorgenson (Alberta Sustainable Resource Development). Russian sheep breeds were collected as part of the 'Russia in Flux' (funded by the Academy of Finland) and 'Sustainable Use of Natural Resources' (funded by the Academy of Finland and the Ministry of Agriculture and Forestry) projects. The authors also thank Cindy Bottema for SRY BAC isolation (University of Adelaide, Australia), Don Meadows for the electronic preparation of figure 1 and Rachel Hawken and Alan Brownlee for comments that improved the manuscript. J.F.M. was funded by Australian Wool Innovation, and J.R.S.M. was supported by a PhD scholarship from Meat & Livestock Australia.

References

Arranz J.J., Bayon Y. & San Primitivo F. (1998) Genetic relationships among Spanish sheep using microsatellites. *Animal Genetics* **29**, 435–40.

Arranz J.J., Bayon Y. & San Primitivo F. (2001) Differentiation among Spanish sheep breeds using microsatellites. *Genetics Selection Evolution* **33**, 529–42.

Bandelt H.-J., Forster P., Sykes B.C. & Richards M.B. (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Molecular Phylogenetics and Evolution* **16**, 8–28.

Bannasch D.L., Bannasch M.J., Ryun J.R., Famula T.R. & Pedersen N.C. (2005) Y chromosome haplotype analysis in purebred dogs. *Mammalian Genome* **16**, 273–80.

Bar-Gal G.K., Ducos P. & Horwitz L.K. (2003) The application of ancient DNA analysis to identify neolithic caprinae: a case study from the site in Hatoula, Israel. *International Journal of Osteoarchaeology* **13**, 120–31.

Bradley D.G., MacHugh D.E., Loftus R., Sow R.S., Hoste C.H. & Cunningham E.P. (1994) Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in Western African tyrantolerant cattle populations. *Animal Genetics* **15**, 7–12.

Calvo J.H., Marcos S., Beattie A.E., Gonzales C., Jurado J.J. & Serrano M. (2004) Ovine *alpha-amylase* genes: isolation, linkage mapping and association analysis with milk traits in Manchega breed sheep. *Animal Genetics* **35**, 329–32.

Coltman D.W., O'Donoghue P., Jorgenson J.T., Hogg J.T., Strobeck C. & Festa-Bianchet M. (2003) Undesirable evolutionary consequences of trophy hunting. *Nature* **426**, 655–8.

Delgado J.V., Perezgrovas R., Camacho M.E., Fresno M. & Barba C. (2000) The wool-less Canary sheep and their relationship with the present breeds in America. *Animal Genetic Resources Information* **28**, 27–34.

Deng W., Shi B., He X. *et al.* (2004) Evolution and migration history of the Chinese population inferred from Chinese Y-chromosome evidence. *Journal of Human Genetics* **49**, 339–48.

Diez-Tascon C., Littlejohn R.P., Almeida P.A. & Crawford A.M. (2000) Genetic variation within the Merino sheep breed: analysis of closely related populations using microsatellites. *Animal Genetics* **31**, 243–51.

Drogemuller C., Leeb T. & Distl O. (2001) PrP genotype frequencies in German breeding sheep and the potential to breed for resistance to scrapie. *Veterinary Record* **149**, 349–52.

Drogemuller C., de Vries F., Hamann H., Leeb T. & Distl O. (2004) Breeding German sheep for resistance to scrapie. *Veterinary Record* **154**, 257–60.

Edwards C.J., Gaillard C., Bradley D.G. & MacHugh D.E. (2000) Y-specific microsatellite polymorphisms in a range of bovid species. *Animal Genetics* **31**, 127–30.

Excoffier L., Smouse P.E. & Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–91.

Freeman A.R., Meghen C.M., MacHugh D.E., Loftus T., Achukwi M.D., Bado A., Sauveroché B. & Bradlet D.G. (2004) Admixture and diversity in West African cattle populations. *Molecular Ecology* **13**, 3477–87.

Gill C.A., Davis S.K., Taylor J.F., Cockett N.E. & Bottema C.D.K. (1999) Construction and characterization of an ovine bacterial artificial chromosome library. *Mammalian Genome* **10**, 1108–11.

Guo J., Du L.-X., Ma Y.-H., Guan W.-J., Li H.-B., Zhao Q.-J., Li X. & Rao S.G. (2005) A novel maternal lineage revealed in sheep (*Ovis aries*). *Animal Genetics* **36**, 331–6.

Hanotte O., Tawah C.L., Bradley D.G., Okomo M., Verjee Y., Ochieng J. & Rege J.E.O. (2000) Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Molecular Ecology* **9**, 387–96.

Heindleder 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. *Proceedings of the Royal Society of London Series B – Biological Sciences* **269**, 893–904.

- Hellborg L. & Ellegren H. (2004) Low levels of nucleotide diversity in mammalian Y chromosomes. *Molecular Biology and Evolution* **21**, 158–63.
- Heyer E., Puymirat J., Dieltjes P., Bakker E. & de Knijff P. (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Human Molecular Genetics* **6**, 799–803.
- Kayser M., Roewer L., Hedman M. *et al.* (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *American Journal of Human Genetics* **66**, 1580–8.
- Kayser M., Krawczak M., Excoffier L. *et al.* (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *American Journal of Human Genetics* **68**, 990–1018.
- Kikkawa Y., Takada T., Sutopo K., Nomura K., Namikawa T., Yonekawa H. & Amano T. (2003) Phylogenies using mtDNA and SRY provide evidence for male-mediated introgression in Asian domestic cattle. *Animal Genetics* **34**, 96–101.
- de Knijff P. (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *American Journal of Human Genetics* **67**, 1055–61.
- Lindgren G., Backstrom N., Swinburne J., Hellborg L., Einarsson A., Sandberg K., Cothran G., Vila C., Binns M. & Ellegren H. (2004) Limited number of patriline in horse domestication. *Nature Genetics* **36**, 335–6.
- MacHugh D.E., Shriver M.D., Loftus R.T., Cunningham P. & Bradley D.G. (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146**, 1071–86.
- Mannen H., Kohno M., Nagata Y. *et al.* (2004) Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. *Molecular Phylogenetics and Evolution* **32**, 539–44.
- Meadows J.R.S., Hawken R.J. & Kijas J.W. (2004) Nucleotide diversity of the ovine Y chromosome. *Animal Genetics* **35**, 379–85.
- Meadows J.R.S., Li K., Kantanen J. *et al.* (2005) Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. *Journal of Heredity* **96**, 494–501.
- Moore N.W., Halnan C.R., McKee J.J. & Watson J.I. (1981) Studies on hybridization between a barbary ram (*Ammotragus lervia*) and domestic ewes (*O. aries*) and nanny goats (*Capra hircus*). *Journal of Reproduction and Fertility* **61**, 79–82.
- Parsonson I.M. (1998) *The Australian Ark: A History of Domesticated Animals in Australia*. CSIRO Publishing, Victoria, Australia.
- Pedrosa S., Uzun M., Arranz J.J., Gutierrez-Gil B., San Primitivo F.S. & Bayon Y. (2005) Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proceedings of the Royal Society of London Series B – Biological Sciences* **272**, 2211–7.
- Ryder M.L. (1984) Sheep. In: *Evolution of Domestic Animals* (Ed. by I.L. Mason), pp. 63–84. Longman, New York.
- Santiago-Moreno J., Lopez-Sebastian A., Gonzalez-Bulnes A., Gomez-Brunet A. & Chemineau P. (2000) Seasonal changes in ovulatory activity, plasma prolactin, and melatonin concentrations, in mouflon (*Ovis gmelini musimon*) and Manchega (*Ovis aries*) ewes. *Reproduction Nutrition Development* **40**, 421–30.
- Scherf B.D. (2000) World Watch List for domestic animal diversity. In: *Food and Agriculture Organization of the United Nations* (Ed. by Scherf B.D.), pp. 58. Food and Agriculture Organisation of the United Nations, Rome.
- Shen P., Wang F., Underhill P.A. *et al.* (2000) Population genetic implications from sequence variation in four Y chromosome genes. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7354–9.
- Sundqvist A.K., Bjornerfeldt S., Leonard J.A., Hailer F., Hedhammar A., Ellegren H. & Vila C. (2006) Unequal contribution of sexes in the origin of dog breeds. *Genetics* **172**, 1121–8.
- 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**, 2045–56.
- Tapio I., Tapio M., Grisli Z., Holm L.E., Jeppsson S., Kantanen J., Miceikiene I., Olsaker I., Viinalass H. & Eythorsdottir E. (2005) Unfolding of population structure in Baltic sheep breeds using microsatellite analysis. *Heredity* **94**, 448–56.
- Worley K., Strobeck C., Arthur S., Carey J., Schwantje H., Veitch A. & Coltman D.W. (2004) Population genetic structure of North American thinhorn sheep (*Ovis dalli*). *Molecular Ecology* **13**, 2545–56.
- Wu C.H., Zhang Y.P., Bunch T.D., Wang S. & Wang W. (2003) Mitochondrial control region sequence variation within the argali wild sheep (*Ovis ammon*): evolution and conservation relevance. *Mammalia* **67**, 109–18.