Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds

The Bovine HapMap Consortium*

The emergence of modern civilization was accompanied by adaptation, assimilation, and interbreeding of captive animals. In cattle (*Bos taurus*), this resulted in the development of individual breeds differing in, for example, milk yield, meat quality, draft ability, and tolerance or resistance to disease and pests. However, despite mapping and diversity studies (1–5) and the identification of mutations affecting some quantitative phenotypes (6–8), the detailed genetic structure and history of cattle are not known.

The imprints of domestication and breed development on the genomes of livestock likely differ from those of companion animals. A deep draft sequence assembly of shotgun reads from a single Hereford female and comparative sequences sampled from six additional breeds were used to develop probes to interrogate 37,470 single-nucleotide polymorphisms (SNPs) in 497 cattle from 19 geographically and biologically diverse breeds. These data show that cattle have undergone a rapid recent decrease in effective population size from a very large ancestral population, possibly due to bottlenecks associated with domestication, selection, and breed formation. Domestication and artificial selection appear to have left detectable signatures of selection within the cattle genome, yet the current levels of diversity within breeds are at least as great as exists within humans.

Cattle occur as two major geographic types, the taurine (humpless—European, African, and Asian) and indicine (humped—South Asian, and East African), which diverged >250 thousand years ago (Kya) (3). We sampled individuals representing 14 taurine (*n* = 376), three indicine (*n* = 73) (table S1), and two hybrid breeds (*n* = 48), as well as two individuals each of *Bubalus quarlesi* and *Bubalus bubalis*, which diverged from *Bos taurus* ~1.25 to 2.0 Mya (9, 10). All breeds except Red Angus (*n* = 12) were represented by at least 24 individuals. We preferred individuals that were unrelated for ≥4 generations; however, each breed had one or two sire, dam, and progeny trios to allow assessment of genotype quality.

Single-nucleotide polymorphisms (SNPs) that were polymorphic in many populations were primarily derived by comparing whole-genome sequence reads representing five taurine and one indicine breed to the reference genome assembly obtained from a Hereford cow (10) (table S2). This led to the ascertainment of SNPs with high minor allele frequencies (MAFs) within the discovery breeds (table S5). Thus, as expected, with trio progeny removed, SNPs discovered within the taurine breeds had higher average MAFs.

---

Fig. 1. (A) Population structure assessed by InStruct. Bar plot, generated by DISTRUCT, depicts classifications with the highest probability under the model that assumes independent allele frequencies and inbreeding coefficients among assumed clusters. Each individual is represented by a vertical bar, often partitioned into colored segments with the length of each segment representing the proportion of the individual's genome from *K* = 2, 3, or 9 ancestral populations. Breeds are separated by black lines. NDA, N'Dama; SHK, Sheko; NEL, Nelore; BRM, Brahman; GIR, Gir; SGT, Santa Gertrudis; BMA, Beefmaster; ANG, Angus; RGG, Red Angus; HFD, Hereford; NRC, Norwegian Red; HOL, Holstein; LMS, Limousin; CHL, Charolais; BSW, Brown Swiss; JER, Jersey; GNS, Guernsey; PMT, Piedmontese; RMG, Romagnola. (B) Principal components PC1 and PC2 from all SNPs. Taurine breeds remain separated from indicine breeds, and admixed breeds are intermediate.
within the taurine than the indicine breeds, and vice versa (table S5); about 30% of SNPs had MAFs >0.3 within the taurine breeds, whereas only about 19% had MAFs >0.3 within the indicine breeds (table S4). The proportions of SNPs in intergenic, intronic, and exonic regions were 63.74, 34.9, and 1.35%, respectively, similar to their representation within the genome. We found that as few as 50 SNPs were necessary for parentage assignment and proof of identity (table S9). Additionally, when we compared ancestries based on pedigree and allele-sharing between individuals, we were able to predict accurately the extent of ancestry when the pedigree was not known (fig. S24), which could be a useful tool for the management of endangered bovine populations.

To examine relatedness among breeds, we analyzed SNP genotype frequencies with InSTRUCT (11) and performed principal component analysis (PCA) using Eigenstrat (12) (Fig. 1 and fig. S27). Varying the number of presumed ancestral populations (K) within InSTRUCT revealed clusters consistent with the known history of cattle breeds (Fig. 1A). The first level of clustering (K = 2) reflects the primary, predomestication division of taurine from indicine cattle. Consequently, breeds derived from indicine and taurine crosses (Beefmaster, Santa Gertrudis, and Sheko) show signatures of admixture with both approaches. At K = 3, the African breeds N’Dama and Sheko separate from the European breeds—a division that reflects an early, possibly predomestication, divergence. PCA recapitulated these findings (Fig. 1B). At higher levels of K, we observed clusters that identify single breeds as closed endogamous breeding units. For example, at K = 9, Jersey, Hereford, Romagnola, and Guernsey each form unique clusters.

If modern breeds arose from bottlenecks from a large ancestral population, we should detect bottleneck signatures within patterns of linkage disequilibrium (LD) and effective population size. We found that the decline of r^2 with genetic distance varied among breeds, although the decline was generally rapid (fig. S10). The extent of LD in cattle is greater than human (13) but less than dog (14). The Jersey and Hereford breeds had higher r^2 than other breeds across the range of distances separating loci. N’Dama had the highest r^2 values at short distances and the lowest r^2 at long distances, which suggested that they were derived from a relatively small ancestral population not subjected to very narrow bottlenecks. The indicine breeds had lower r^2 values at short distances and intermediate r^2 values at longer distances, which indicated that their ancestral popula-

---

**Fig. 2.** Effective population size in the past estimated from linkage disequilibrium data. Inset graph shows effective population size for the European humans over the same period; from (13). Breeds as in Fig. 1.

**Fig. 3.** Nucleotide diversity across five ENCODE regions resequenced in 47 animals from ANG, Angus; BRM, Brahman; and HOL, Holstein. (A) Watterson’s estimate (θ) of the population mutation rate per base pair (pooled across regions). (B) Average pairwise nucleotide distance (r) within breeds. (C and E) Nonparametric bootstrap estimates of diversity ratios among the three populations on the basis of θ. (D and F) Nonparametric bootstrap estimates of diversity ratios among the three populations on the basis of π.
tion was much larger than that from which taurine cattle were domesticated (Fig. 2). As the MAFs for utilized SNPs were generally high and the estimates of LD did not require phased chromosomes, these results should be robust.

When breeds were combined, the decline in LD was more rapid, which reflected a lack of conserved phase relations across breeds. We characterized the extent of haplotype-sharing among breeds between pairs of adjacent SNPs using the \( r \) statistic. A high correlation between \( r \) values between two breeds indicates that the same haplotypes tend to persist within both breeds. Correlations between \( r \) values for SNPs separated by 10 kb were high among the taurine and indicine breeds but were low between these groups (fig. S11). Once SNPs are separated by 100 to 250 kb, we found little haplotype sharing between breeds. Clearly, phase relations dissipated as populations diverged despite the relatively young origin of all breeds. Breeds known to have a recent shared ancestry, notably, Angus and Red Angus; Holstein and Norwegian Red; and Beefmaster and Santa Gertrudis, showed a high correlation among \( r \) values for SNPs separated by 100 to 250 kb.

Breeds were expected to differ for effective population sizes (\( N_e \)) on the basis of differences in the decline of \( r^2 \) with genetic distance (13). We estimated \( N_e \) at various times in each breed’s history by setting average \( r^2 \) values equal to their expectation (13) (Fig. 2 and table S1). \( N_e \) has recently declined for all breeds, which reflects bottlenecks associated with domestication, breed formation, and, in some breeds, recent intense selection for milk or beef production. In contrast, human \( N_e \) has expanded exponentially over the same period (inset to Fig. 2).

A smaller \( N_e \) suggests lower genetic diversity, which is of concern for species viability. To assess genetic diversity free from SNP ascertainment bias, we used the polymerase chain reaction to amplify and sequence 119 closely spaced fragments from five genomic regions on two chromosomes. Two of these regions were known to harbor quantitative trait loci (QTL). Following the amplification of these regions from 18 Angus, 16 Holstein, and 5 Brahman, the individual segments were Sanger-sequenced to detect SNPs. Of the 1201 discovered SNPs, only 258 were common to taurine and indicine breeds, consistent with their age of divergence. Remarkably, 569 SNP (47.4%) were unique to Brahmans, and 365 SNP (30.4%) were found only in Angus or Holstein, with 169 SNP (46.3%) common to both breeds. This suggests that breeds represent partly overlapping subsamples within the taurine diversity. However, seven times as many taurine animals had to be sequenced to uncover 75.3% as many SNPs as were discovered in indicine animals. Estimates of the unascertained genomeric distributions of SNPs by MAFs within taurine and indicine breeds are in fig. S19.

Diversities as measured by the population mutation rate (\( \theta \)) and pairwise nucleotide heterozygosity (\( \pi \)) were also estimated for the 119 fragments and compared between the three breeds (Fig. 3). Angus and Holstein have similar levels of nucleotide diversity measured by both statistics (\( 1.4 \times 10^{-3} \)) and have \( \approx 40\% \) more nucleotide variation than is found in human populations (\( 1.0 \times 10^{-3} \)). Brahmans had even higher, with average estimates of \( \theta \) and \( \pi \) of \( 3.35 \times 10^{-3} \) and \( 2.74 \times 10^{-3} \), respectively. These correspond to densities of 1 SNP every 714 bp for pairs of Angus or Holstein chromosomes and 1 SNP every 285 bp for pairs of Brahman chromosomes. These results demonstrate that genetic diversity in cattle is not low despite the decline in \( N_e \).

The lower genetic diversity within modern taurine cattle could reflect a lower diversity within the predominately ancestral population, and/or postdomestication effects of stronger bottlenecks at breed formation and stronger selection for docility and productivity. Selection is unlikely to be the primary cause, because the diversity distributions for \( \theta \) and \( \pi \) were similar for all five sequenced regions, and only one region revealed a signature of selection. On the other hand, Fig. 2 suggests that the predomination \( N_e \) of indicine cattle, which originated in southern Asia, a center of species diversity, was much larger than that of taurine cattle. Finally, the process of breed formation in European taurine cattle involved sequential limited migrations from the center of domestication in west Asia (5). Diversity declines with distance from primary sites of domestication (4) and ancient DNA from domesticated cattle and aurochs in Europe show that there was essentially no gene flow from the aurochs into domesticated cattle (5). Therefore, the evidence suggests that the current difference in diversity is mainly due to progenitor population diversity and bottleneck effects at, and before, breed formation rather than differences in the intensity of natural or artificial selection postdomestication.

Table 1. Genomic regions associated with extreme \( F_{ST} \) values with gene content consistent with domestication. \( F_{ST} \) values averaged over eight adjacent SNP. Gene functions from OMIM and NCBI Gene database, except for R3HDM1 described in (2).

<table>
<thead>
<tr>
<th>Genes</th>
<th>Index SNP</th>
<th>( F_{ST} )</th>
<th>BTA</th>
<th>Location</th>
<th>Effect or important phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>High values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZRANB3, R3HDM1</td>
<td>rs29021800</td>
<td>0.31</td>
<td>2</td>
<td></td>
<td>Feed efficiency</td>
</tr>
<tr>
<td>WIFI1</td>
<td>BTA-77454</td>
<td>0.29</td>
<td>5</td>
<td></td>
<td>Mammalian mesoderm segmentation</td>
</tr>
<tr>
<td>SPOCK1</td>
<td>BTA-142690</td>
<td>0.30</td>
<td>7</td>
<td></td>
<td>Proteoglycan—synaptic fields of the developing CNS</td>
</tr>
<tr>
<td>NBEA</td>
<td>BTA-153392</td>
<td>0.34</td>
<td>12</td>
<td></td>
<td>Human idiopathic autism</td>
</tr>
<tr>
<td>NMT1, DCAKD, C1QL1</td>
<td>BTA-45533</td>
<td>0.31</td>
<td>19</td>
<td></td>
<td>Activator of serum complement system</td>
</tr>
<tr>
<td>DACH2, CHM, POU3F4, BRWD3</td>
<td>BTA-161991</td>
<td>0.39</td>
<td>X</td>
<td>41471338...4478564</td>
<td>Human mental retardation</td>
</tr>
<tr>
<td>NLG3 to DGA126</td>
<td>BTA-164256</td>
<td>0.36</td>
<td>X</td>
<td>49279035...50192452</td>
<td>Severe combined immunodeficiency</td>
</tr>
<tr>
<td>Low values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPARC1A, DHX15, SOD3</td>
<td>BTC-039516</td>
<td>0.04</td>
<td>6</td>
<td></td>
<td>Antioxidative extracellular protection</td>
</tr>
<tr>
<td>No known gene</td>
<td>BTC-04723</td>
<td>0.05</td>
<td>14</td>
<td></td>
<td>Multisubunit molecular motor</td>
</tr>
<tr>
<td>DNAH9</td>
<td>rs29018632</td>
<td>0.05</td>
<td>19</td>
<td></td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>POUSF1, MHC</td>
<td>BTA-55856</td>
<td>0.05</td>
<td>23</td>
<td></td>
<td>Expressed in olfactory tissues</td>
</tr>
<tr>
<td>ZNF187</td>
<td>rs29024230</td>
<td>0.04</td>
<td>23</td>
<td></td>
<td>Human autism susceptibility candidate</td>
</tr>
<tr>
<td>AUTS2</td>
<td>BTC-074065</td>
<td>0.04</td>
<td>25</td>
<td></td>
<td>Stress- and exercise-induced sudden cardiac death</td>
</tr>
</tbody>
</table>

Cattle have been marked by selection during domestication, breed formation, and ongoing selection to enhance performance and productivity. We utilized three methods to detect genomic selection in cattle: (i) the iHS statistic, which identifies regions of increased local LD (16) suggestive of directional selection; (ii) the \( F_{ST} \) statistic, a measure of the degree of differentiation between subpopulations (17); and (iii) the composite likelihood ratio test (CLR) (18), which assumes a selective sweep model (10). The iHS method was limited by low SNP density and our inability to completely specify ancestral SNP allele states (10). However, despite these limitations, we found evidence for selective sweeps on chromosomes 2, 6, and 14 (table S8 and fig. S20). We identified selection near MSTN, in which mutations can cause double muscling (6). Similarly, high iHS values were found in the region near ABCG2 in which mutations cause differences in milk yield and composition (8). A peak in iHS values was also identified within a gene poor region of chromo-
some 14 adjacent to a region containing genes from \( KHDRBS3 \) to \( TG \), associated with intramuscular fat content in beef (19).

Calculation of \( F_{ST} \) across all populations for each SNP detected both balancing and divergent selection (fig. S20). Some of the highest and lowest average \( F_{ST} \) values were found in genes associated with behavior, the immune system, and feed efficiency (Table 1). Domestication most likely required the selection of smaller and more docile animals that could resist pathogens and adapt to a human-controlled environment (20). One region under selection contains \( R3HDMI \) and is associated with efficient food conversion and intramuscular fat content in some breeds (2). In addition to the \( R3HDMI \) gene (2), this region is also under selection in Europeans, most likely because it contains \( LCT \), mutations of which allow the digestion of lactose in adults (22). These results suggest that selection in this region may affect energy homeostasis. Furthermore, we detected selection between beef and dairy breeds with both CLR and iHS, represented by a broad, high \( F_{ST} \) peak across the region, centered on \( SPOCK1 \) (Table 1). As several QTL have been mapped to this region, multiple loci could be under divergent selection (J), although this peak does not encompass \( CAST \), which affects meat quality (23).

Our high resolution examination of cattle shows that unlike the dog—which has restricted diversity and high levels of inbreeding—domesticated cattle had a large ancestral population size and that more aurochs must have been domesticated than wolves; reducing the severity of the domestication bottleneck. SNP diversity within taurine breeds was similar to that of humans, but was more aurochs than wolves; reducing the severity of the domestication bottleneck.

The Bovine HapMap Consortium

Overall project leadership: Richard A. Gibbs,1,2,4 Jeremy F. Taylor,4 Curtis P. Van Tassell4

Hapmap project group leaders: William Barendse,5,6 Kelly A. Eversole,1,6 Richard A. Gibbs,1,2 Clare A. Gill,7 Ronnie D. Green,7 Debra L. Hamblin,10 Steven M. Kappes,9 Sigmijn Lien,11 Lakshmi K. Matukumalli,12,4 John C. McEwan,12,4 Lynne V. Nazareth,12 Robert D. Schnabel,12,4 Jeremy F. Taylor,12 Curtis P. Van Tassell12,4 George M. Weinstock,1,2 David A. Wheeler4

Breed champions: Paolo Ajmone-Marsan,14,18 William Barendse,5,6 Paul J. Boettcher,3 Alexandre R. Caetano,10 Jesus Fernandez-Garcia,10 Claire A. Gill,12 Ronnie D. Green,12,14 Oliver Hanotte,14 Sigmijn Lien,11 Paola Mariani,19 John C. McEwan,12 Loren C. Skow,20 Todd S. Sonstegard,12 Curtis P. Van Tassell12,4 John Williams12,21

Pedigree analysis and breed sampling: Alexandre R. Caetano,10 Boubacar Diallo,5 Ronnie D. Green,7 Lemecha Hallamenni,11 Olivier Hanotte,14 Mario L. Martineau,14 Chris A. Morris,15 Luiz D. C. Silva,10 Judith A. Spelman,10 Curtis P. Van Tassell12,4 (leader), Woodyoum Melata,5 Koyan Zhao29

Sample acquisition and DNA preparation: Colette A. Abbey,8 Clare A. Gill,7 Ronnie D. Green,7 Olivier Hanotte,14 Sigmijn Lien,11 Paola Mariani,19 John C. McEwan,12 Loren C. Skow,20 Todd S. Sonstegard,12 Curtis P. Van Tassell12,4 John Williams12,21

Genome assembly: Yu Yue,26 Xiang Qin,26 Kim C. Worley26 (leader)

SNP discovery: Chuan Gao,26 Clare A. Gill,7 Huaiyang Jiang,26 Yue Liu,26 Stephen S. Moore,26 Lynne V. Nazareth,26 Yann Ren,26 Xing-Zhi Song,26 David A. Wheeler26 (leader), Kim C. Worley26

ENCODE resequencing: Carlos A. Baptista,7,8 Ryan D. Hernandez,8 Donna M. Munn,8 Lynne V. Nazareth,26 Shobha Patil,2,3 Yannu Ren,2,3 Anthony San Lucas,2,3 David A. Wheeler26 (leader)

Genotyping: Oing Fu,2,3 Matthew P. Kent,2,3 Sigibjørn Lien,2,3 Stephen S. Moore,2,3 Lynne V. Nazareth26 (leader), Richard Vega,2,3 David A. Wheeler26 (leader)

Database & Web site development: Colette A. Abbey,7 Chuan Gao,26 Clare A. Gill,7 Ronnie D. Green,7 Lakshmi K. Matukumalli12,4 (leader), Aruna M. Prabhu,3 Renata B. de Magalhães,29 Keyan Zhao29

QA/QC: Colette A. Abbey,8 Clare A. Gill,7 Matthew P. Kent,11 Sigibjørn Lien,2,3 Lakshmi K. Matukumalli12,4 (leader), Robert D. Schnabel26 (leader), Gert Sceg,26 Jeremy F. Taylor26

Allele frequency analysis: Paolo Ajmone-Marsan,14 Katarzyna Bycz,26 Carlos D. Bustamante12 (leader), Junqun Chou,26 Hong Tong,14 John J. F. Grefenstette,7 Lakshmi K. Matukumalli12,4 (leader), Brenda Murdoch,26 Stephen S. Moore,32 Lynne V. Nazareth26 (leader), Alessandra Stella,39 Curtis P. Van Tassell12,4 Rafael Villa-Angulo26,12 David A. Wheeler26 (leader), Mark Wright26

Map data provision and analysis: Jan Aerts,26,38, Oliver Jann,38 Lakshmi K. Matukumalli,12,4 Ricardo Negrini,38 Tad S. Sonstegard26,38 John L. Williams12,21

Haploype estimation: Paolo Ajmone-Marsan,14 John J. F. Grefenstette7,26 Lakshmi K. Matukumalli,12,4,25 Curtis P. Van Tassell12,4 (leader), Rafael Villa-Angulo26,12 LD persistence across breeds: Mike E. Goddard,39,72 Ben J. Hayes39 (leader)

Selective sweeps: William Barendse,5,6,7,14 Daniel G. Bradley,39 Paul J. Boettcher,3 Carlos D. Bustamante12,4 Junqun Chou,26 Marcos Barbosa da Silva,5 Claire A. Gill,7 Lillian P. L. Liu,34 George E. Liu,34 David J. McVean,34,39 Mark Wright26

Applications: Carlos D. Bustamante12,4 (leader), Ken G. Dods,26 John C. McEwan26 (leader), Jeremy F. Taylor26 (leader), Curtis P. Van Tassell12

Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.
Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.
Division of Animal Sciences, University of Missouri, 920 East Campus Drive, Columbia, MO 65211–5300, USA.
Bovine Functional Genomics Laboratory, U.S. Department of Agriculture (USDA) Agricultural Research Service (USDA–ARS), Beltsville Agricultural Research Center (BARC), East–Beltsville, MD 20705, USA.
Commonwealth Scientific and Industrial Research Organization (CSIRO), Livestock Industries, 306 Carwood Road, St. Lucia, Queensland 4067, Australia.
Cooperative Research Center (CRC) for Beef Genetic Technologies, University of New England, Armidale, NSW 2351, Australia. 
Alliance for Animal Genome Research, 5207 Wyomond Road, Bethesda, MD 20816, USA. 
Department of Animal Science, Texas A&M University, 2471 TAMU, College Station, TX 77843–2471, USA. 
National Program Staff, USDA–ARS, 5601 Sunnyside Avenue, Beltsville, MD 20705, USA. 
USDA Cooperative State Research, Education, and Extension Service, 1400 Independence Avenue, SW, Washington, DC 20250–2220, USA. 
Centre for Integrative Genetics and Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Aaboretet 6, Ås 1432, Norway. 
Department of Bioinformatics and Computational Biology, George Mason University, 10900 University Boulevard, Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. 
Animal Genomic Genomics, AgResearch, Invermay, Post Box 50034, Mosgiel 9053, New Zealand. 
Istituto di Zootecnica, Università Cattolica del Sacro Cuore, via East Parmense, 84, Piacenza 29100, Italy. 
Joint United Nations Food and Agriculture Organization (FAO)–International Atomic Energy Agency (IAEA)–United Nations Food and Agriculture Organization (FAO)/IAEA Vienna, International Livestock Research Institute—Kenya, Italian Piedmontese Breeders—Parco Tecnologico Padano, Italian Romagnola Society—Università Cattolica del Sacro Cuore, Livestock Improvement Corporation, Meat and Wool New Zealand, North American Limousin Foundation, Red Angus Association of America, Roslin Institute for UK Guernsey, and Sygen (now Genus). See SOM for additional acknowledgments. The genome sequence (GCF0300000000) and SNPs (SOM) are available from NCBI.
Revealing the History of Sheep Domestication Using Retrovirus Integrations

Bernardo Chessa,1,2 Filipe Pereira,3 Frederick Arnaud,4 Antonio Amorim,5 Félix Goyache,6 Ingrid Mainland,5 Rowland R. Kao,7 Josephine M. Pemberton,6 Dario Beraldi,6 Michael J. Stear,1 Alberto Alberti,2 Marco Pittau,2 Leopoldo Iannuzzi,7 Lars-Erik Holm,1 Urmans Saarma,19 Sohail Ahmad,19 Nurbiy Marzanov,20 Emma Eythorsdottir,21 Martin J. Holland,22,23 Paolo Ajmone-Marsan,24 Michael W. Bruford,25 Juha Kantanen,26 Thomas E. Spencer,27 Massimo Palmarini28

The domestication of livestock represented a crucial step in human history. By using endogenous retroviruses as genetic markers, we found that sheep differentiated on the basis of their "retrotype" and morphological traits dispersed across Eurasia and Africa via separate migratory episodes. Relicts of the first migrations include the Mouflon, as well as breeds previously recognized as "primitive" on the basis of their morphology, such as the Orkney, Soay, and the Nordic short-tailed sheep now confined to the periphery of northwest Europe. A later migratory episode, involving sheep with improved production traits, shaped the great majority of present-day breeds. The ability to differentiate genetically primitive sheep from more modern breeds provides valuable insights into the history of sheep domestication.

The first agricultural systems, based on the cultivation of cereals, legumes, and the rearing of domesticated livestock, developed within Southwest Asia ~11,000 years before present (yr B.P.) (1, 2). By 6000 yr B.P., agro-pastoralism introduced by the Neolithic agricultural revolution became the main system of food production throughout prehistoric Europe, from the Mediterranean north to Britain, Ireland, and Scandinavia (3); south into North Africa (4); and east into West and Central Asia (5).

Sheep and goats were the first livestock species to be domesticated (6). Multiple domestication events, as inferred by multiple mitochondrial lineages, gave rise to domestic sheep and similarly other domestic species (7–10). Initially, sheep were reared mainly for meat but, during the fifth millennium B.P. in Southwest Asia and the fourth millennium B.P. in Europe, specialization for "secondary" products such as wool became apparent. Sheep selected for secondary products appear to have replaced more primitive domestic populations. Whether specialization for secondary products occurred first in Southwest Asia or occurred throughout Europe is not known with certainty, owing to the lack of definitive archaeological evidence for the beginning of wool production (6, 11, 12).

For this study, we used a family of endogenous retroviruses (ERVs) as genetic markers to examine the history of the domestic sheep. ERVs result from the stable integration of the retrovirus genome ("provirus") into the germline of the host (13) and are transmitted vertically from generation to generation in a Mendelian fashion. The sheep genome contains at least 27 copies of ERVs related to the exogenous and pathogenic Jaagsiekte sheep retrovirus (enJSRVs) (14–16). Most enJSRVs loci are fixed in domestic sheep, but some are differentially distributed between breeds and individuals (i.e., they are insertionally polymorphic) (14). enJSRVs can be used as highly informative genetic markers because of the presence of each endogenous retrovirus in the host
Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds
The Bovine HapMap Consortium

Science 324 (5926), 528-532.
DOI: 10.1126/science.1167936

Not Just Dinner on Legs
Several thousand years ago, human beings realized the virtues of domesticating wild animals as easy meat. Soon other possibilities became apparent, and as revealed in a series of papers in this issue, early pastoralists became selective about breeding for wool, leather, milk, and muscle power. In two papers, Gibbs et al. report on the bovine genome sequence (p. 522; see the cover, the Perspective by Lewin, and the Policy Forum by Roberts) and trace the diversity and genetic history of cattle (p. 528), while Chessa et al. (p. 532) survey the occurrence of endogenous retroviruses in sheep and map their distribution to historical waves of human selection and dispersal across Europe. Finally, Ludwig et al. (p. 485) note the origins of variation in the coat-color of horses and suggest that it is most likely to have been selected for by humans in need of good-looking transport.