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

The Bovine HapMap Consortium*

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.

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.

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.

*The full list of authors with their contributions and affiliations is included at the end of the manuscript.

Fig. 1. (A) Population structure assessed by InStruct. Bar plot, generated by DISTRICT, 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; RGU, 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 ($\theta$) 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 $\theta$. (D and F) Nonparametric bootstrap estimates of diversity ratios among the three populations on the basis of $\pi$. 

---

www.sciencemag.org SCIENCE VOL 324 24 APRIL 2009
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 and Santa Gertrudis, showed a high correlation.

S11). Once SNPs are separated by 100 to 250 kb, breeds but were low between these groups (fig. 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 and Santa Gertrudis, showed a high correlation.

Breeds were expected to differ for effective population sizes (Nₑ) on the basis of differences in the decline of r² with genetic distance (13). We estimated Nₑ at various times in each breed’s history by setting average r² values equal to their expectation (13) (Fig. 2 and table S1). Nₑ has recently declined for all breeds, which reflects bottleneckes associated with domestication, breed formation, and, in some breeds, recent intense selection for milk or beef production. In contrast, human Nₑ has expanded exponentially over the same period (inset to Fig. 2).

A smaller Nₑ 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 SNP, only 258 were common to taurine and indicine breeds, consistent with their age of divergence. Remarkably, 569 SNP (47.4%) were unique to Brahman, 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 sub-samples 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 genomic distributions of SNPs by MAFs within taurine and indicine breeds are in fig. S19.

Diversities as measured by the population mutation rate (θ) and pairwise nucleotide heterozygosity (π) 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 × 10⁻³) and have ~40% more nucleotide variation than is found in human populations (~1.0 × 10⁻³). Brahman variation was even higher, with average estimates of θ and π of 3.35 × 10⁻³ and 2.74 × 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 Nₑ.

The lower genetic diversity within modern taurine cattle could reflect a lower diversity within in the predomestication ancestral population, and/or post-domestication 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 θ and π 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 predominance Nₑ 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.

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 MSTMN, 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-

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

| Genes       | Index SNP   | FST | BTA     | Location | Effect or important phenotypes                      |
|-------------|-------------|-----|---------|----------|-----------------------------------------------------|
| High values |             |     |         |          |                                                     |
| ZRANB3, R3HDM1 | rs29021800 | 0.31| 2       | 64740286…64931017 | Feed efficiency                                      |
| WIFI1       | BTA-27454   | 0.29| 5       | 52696749…53098507 | Mammalian mesoderm segmentation                      |
| SPOCK1      | BTA-142690  | 0.30| 7       | 47501122…47899778 | Proteoglycan—synaptic fields of the developing CNS   |
| NBEA        | BTA-153392  | 0.34| 12      | 25884192…26189285 | Human idiopathic autism                              |
| NMT1, DCAKD, C1QL1 | BTA-45533 | 0.31| 19      | 46088946…46157261 | Activator of serum complement system                 |
| DACH2, CHIM, POU3F4, BRWD3 | BTA-161991 | 0.39| X       | 41471338…44478564 | Human mental retardation                             |
| NLGN3 to DGA716 | BTA-164256 | 0.36| X       | 49279035…50192452 | Severe combined immunodeficiency                     |
| Low values  |             |     |         |          |                                                     |
| PPARG1C1A, DHX15, SOD3 | BTA-039516 | 0.04| 6       | 45354707…45415844 | Antioxidative extracellular protection               |
| No known gene | BTA-049723 | 0.05| 14      | 4569804…5204473 | Multisubunit molecular motor                        |
| DNAH9       | rs29018632  | 0.05| 19      | 30943404…3122086 | Major histocompatibility complex                     |
| POUSF1, MHC | BTA-55856   | 0.05| 23      | 27895932…28145846 | Major histocompatibility complex                     |
| ZNF187      | rs29024230  | 0.04| 23      | 30241236…30502690 | Expressed in olfactory tissues                      |
| AUTS2       | BTA-074065  | 0.04| 25      | 31773107…32498861 | Human autism susceptibility candidate                |
| RYR2        | rs29011563  | 0.05| 28      | 8736599…8772178  | Stress- and exercise-induced sudden cardiac death   |
some 14 adjacent to a region containing genes from KHDRBS3 to TG, associated with intramuscular fat content in beef (19).

Calculation of FST across all populations for each SNP detected both balancing and divergent selection (fig. S20). Some of the highest and lowest average FST 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 R3HDM1 and is associated with efficient food conversion and intramuscular fat content in some breeds (2). In addition to the R3HDM1 gene (21), this region is also under selection in Europeans, most likely because it contains mutations of which allow the digestion of lactose in adults (22). These results suggest that mutations 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 FST 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 tauren breeds was similar to that of humans, but was significantly less than diversity within indicine breeds, which suggested that the Indian subcontinent was a major site of cattle domestication and predomestication diversity. Selection first for domestication and then for agricultural specialization have apparently reduced breed effective population sizes to relatively small numbers. The recent decline in diversity is sufficiently rapid that loss of diversity should be of concern to animal breeders. Despite this, population levels of LD are unexpectedly low considering the relatively small Ne, which indicates that effective population sizes were much larger in the very recent past.

References and Notes
1. S. M. Kappes et al., J. Anim. Sci. 78, 3053 (2000).
2. W. Barendse et al., Genetics 176, 1893 (2007).
3. R. J. Bradley et al., Proc. Natl. Acad. Sci. U.S.A. 93, 5132 (1996).
4. R. T. Loftus et al., Mol. Ecol. 18, 2015 (2019).
5. C. J. Edwards et al., Proc. R. Soc. London B. Biol. Sci. 274, 1377 (2007).
6. L. Grobet et al., Nat. Genet. 17, 71 (1997).
7. B. G. Gilbert et al., Genome Res. 12, 222 (2002).
8. M. Cohen-Zinder et al., Genome Res. 15, 936 (2005).
9. A. Schreiber et al., J. Hered. 90, 145 (1999).
10. Materials and methods are available as supporting materials on Science Online.
11. H. Gao, S. Williamson, C. D. Bustamante, Genetics 176, 1635 (2007).
12. N. Patterson, A. L. Price, D. Reich, PLoS Genet. 2, e190 (2006).
13. A. Tenesa et al., Genome Res. 17, 520 (2007).
14. K. Lindblad-Toh et al., Nature 457, 227 (2009).
15. J. A. Sved, Annu. Rev. Popul. Biol. 2, 125 (1971).
16. B. F. Voight, S. Kudaravalli, X. Wen, J. K. Pritchard, PLoS Biol. 4, e72 (2006).
17. J. M. Akey, G. Zhang, K. Zhang, L. Jin, M. D. Shriver, Genome Res. 12, 1805 (2002).
18. S. N. Eisen et al., Science 305, 1566 (2005).
19. W. Barendse et al., Aust. J. Exp. Agric. 44, 669 (2004).
20. J. Clutton-Brock, A Natural History of Domesticated Mammals (Cambridge Univ. Press, Cambridge, 1999), 238 pp.
21. P. C. Sabetti et al., Nature 449, 913 (2007).
22. S. A. Tishkoff et al., Nat. Genet. 39, 31 (2007).
23. W. Barendse et al., Genetics 176, 2601 (2007).
24. The Bovine HapMap Consortium funded by the National Human Genome Research Institute of the National Institutes of Health, U.S. Department of Agriculture’s Agricultural Research Service (USDA-ARS) and Cooperative State Research, Education, and Extension Service (USDA-CSREES), the Research Council of Norway, as well as: American Angus Association, American Hereford Association, American Jersey Cattle Association, AgResearch (New Zealand), Beef CRC and Meat and Livestock Australia for the Australian Brahman Breeders Association, Beef��me United, The Brazilian Agricultural Research Corporation (Embrapa), Brown Swiss Association, Commonwealth Scientific and Industrial Research Organization, Dairy InSight, GENO Breeding and Artificial Insemination Association—Norway, Herd Book/France Limousin Selection, Holstein Association USA, International Dairy Federation, Investigations in Genetics, Italy, New Zealand National Livestock Foundation, Foundation for 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 acknowledgements. The genome sequence (AAFC03000000) and SNPs (SOM) are available from NCBI.

The Bovine HapMap Consortium

Overall project leadership: Richard A. Gibbs,1,2 Jeremy F. Taylor,3,4 Curtis P. Van Tassell*4

Hapmap project group leaders: William Barendse,5,6 Kelly A. Eversole,7 Richard A. Gibbs,5,6 Clare A. Gill,8 Ronnie D. Green,9 Debra L. Hamrick,10 Steven M. Kappes,11 Siibjorn Lien,12 Lakshmi K. Matukumalli,5,6,12 John C. McEwan,12 Lynne V. Nazareth,1,2 Robert D. Schnabel,13 Jeremy F. Taylor,14 Curtis P. Van Tassell,15 George M. Weinstock,16 David A. Wheeler,17

Breed champions: Paolo Ajmone-Marsan,18 William Barendse,5,6 Paul J. Boettcher,15 Alexandre R. Caetano,16 Jose Fernando Garcia,17 Clare A. Gill,18 Ronnie D. Green (leader),17 Olvier Hanotte,18 Siibjorn Lien,12 Paola Mariani,19 John C. McEwan,12 Loren C. Snow,20 Tad S. Sonstegard,21 Curtis P. Van Tassell,19 John L. Williams22

Pedigree analysis and breed sampling: Alexandre R. Caetano,16 Boubacar Diallo,23 Ronnie D. Green,24 Lemecha Hallamierian,25 Olivier Hanotte,30 Marco L. Martinez,31 Chris A. Morris,31 Luis O. C. Silva,31 Ukrainian Livestock Institute, Kyiv, Ukraine; Agricultural Research Corporation (Embrapa), Brazil;legant Bovine 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), Beltville, East, 20705, USA. Commonwealth Scientific and Industrial Research Organization (CSIRO), Livestock Industries, 306 Carmony 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 Wyoming 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 Sunnyvale Avenue, Beltsville, MD 20705, USA. USDA Cooperative State Research, Education, and Extension Service, 1400 Independence Avenue, SW, Washington, DC 20525–2220, USA. Centre for Integrative Genetics and Department of Animal and Aquacultural Genomics, Norwegian University of Life Sciences, Arbevreien 6, As 1432, Norway. Department of Bioinformatics and Computational Biology, George Mason University, 10900 University Boulevard, 4400 College Park, Maryland, USA. Animal Genomics, AgResearch, Invermay, Post Box 50034, Mosgiel 9053, New Zealand. Istituto di Zootecnica, Universita Cattolica del Sacro Cuore, via East Parmese, 84, Piacenza 29100, Italy. Joint United Nations Food and Agriculture Organization (FAO)—International Atomic Energy Agency (IAEA), Division of Nuclear Power in Food and Agriculture, IAEA, Post Office Box 100, Wagramerstrasse 5, Vienna A1400, Austria. Embrapa Genetic Resources and Biotechnology Center, Final Avenida W/S Norte, Brasilia, DF 70770-900, Brazil. Universidade Estadual Paulista (UNESP), Department of Animal Production and Health, IAEA Collaborative Laboratory in Animal Breeding and Biotechnology, Aracatuba, SP 16050-680, Brazil. Animal Genetics Resources Sources Characterization, International Livestock Research Institute, Post Office Box 3079, Nairobi 00102, Kenya. Parco Tecnologico Padano, Via Einstein, Polo Universitario, Lodi 26900, Italy. Department of Veterinary Integrative Biological Sciences, College of Veterinary Medicine and
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 genome ($\lambda$) 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) ($17$). enJSRVs can be used as highly informative genetic markers because of the presence of each endogenous retrovirus in the host genome.

Bernardo Chessa,1,2 Filipe Pereira,3 Frederick Arnaud,4 Antonio Amorim,5 Félix Goyache,9 Ingrid Mainland,5 Rowland R. Kao,9 Josephine M. Pemberton,6 Dario Beraldi,6 Michael J. Stear,1 Alberto Alberti,2 Marco Pittau,2 Leopoldo Iannuzzi,7 Emma Eythorsdottir,21 Martin J. Holland,22,23 Paolo Ajmone-Marsan,24 Michael W. Bruford,25 Juha Kantanen,26 Thomas E. Spencer,27 Massimo Palmari*28

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

532 24 APRIL 2009 VOL 324 SCIENCE www.sciencemag.org

Downloaded from http://science.sciencemag.org/ on June 14, 2018
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.