Downloaded from www.sciencemag.org on April 23, 2009

The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution

The Bovine Genome Sequencing and Analysis Consortium,* Christine G. Elsik,[1] Ross L. Tellam,[2] Kim C. Worley[3]

To understand the biology and evolution of ruminants, the cattle genome was sequenced to about sevenfold coverage. The cattle genome contains a minimum of 22,000 genes, with a core set of 14,345 orthologs shared among seven mammalian species of which 1217 are absent or undetected in noneutherian (marsupial or monotreme) genomes. Cattle-specific evolutionary breakpoint regions in chromosomes have a higher density of segmental duplications, enrichment of repetitive elements, and species-specific variations in genes associated with lactation and immune responsiveness. Genes involved in metabolism are generally highly conserved, although five metabolic genes are deleted or extensively diverged from their human orthologs. The cattle genome sequence thus provides a resource for understanding mammalian evolution and accelerating livestock genetic improvement for milk and meat production.

Domesticated cattle (*Bos taurus* and *Bos taurus indicus*) provide a significant source of nutrition and livelihood to nearly 6.6 billion humans. Cattle belong to a clade phylogenetically distant from humans and rodents, the Cetartiodactyl order of eutherian mammals, which

first appeared ~60 million years ago (*1*). Cattle represent the Ruminantia, which occupy diverse terrestrial environments with their ability to efficiently convert low-quality forage into energy-dense fat, muscle, and milk. These biological processes have been exploited by humans since domestication, which began in the Near East some 8000 to 10,000 years ago (*2*). Since then, over 800 cattle breeds have been established, representing an important world heritage and a scientific resource for understanding the genetics of complex traits.

The cattle genome was assembled with methods similar to those used for the rat and sea urchin genomes (*3*, *4*). The most recent assemblies, Btau3.1 and Btau4.0, combined bacterial artificial chromosome (BAC) and whole-genome shotgun (WGS) sequences. Btau3.1 was used for gene-specific analyses. Btau4.0, which includes finished sequence data and used different mapping methods to place the sequence on chromosomes, was used for all global analyses other than gene prediction. The contig N50 (50% of the genome is in contigs of this size or greater) is 48.7 kb for both assemblies; the scaffold N50 for Btau4.0 is 1.9 Mb. In the Btau4.0 assembly, 90% of the total genome sequence was placed on the 29 autosomes and X chromosome and validated (*3*). Of 1.04 million expressed sequence tag (EST) sequences, 95.0% were contained in the assembled contigs. With an equivalent gene distribution in the remaining 5% of the genome, the estimated genome size is 2.87 Gbp. Comparison with 73 finished BACs and single-nucleotide polymorphism (SNP) linkage data (*5*, *6*) confirmed this assembly quality with greater than 92% genomic coverage, and fewer than 0.8% of SNPs were incorrectly positioned at the resolution of these maps (*3*, *4*).

We used the cattle genome to catalog protein-coding genes, microRNA (miRNA) genes, and ruminant-specific interspersed repeats, and we manually annotated over 4000 genes. The consensus protein-coding gene set for Btau3.1 (OGSv1), from six predicted gene sets (*4*), consists of 26,835 genes with a validation rate of 82% (*4*). On this basis, we estimate that the cattle genome contains at least 22,000 protein-coding genes. We identified 496 miRNA genes of which 135 were unpublished miRNAs (*4*). About half of the cattle miRNA occur in 60 genomic miRNA clusters, containing two to seven miRNA genes separated by less than 10 kbp (fig. S2). The overall GC content of the cattle genome is 41.7%, with an observed-to-expected CpG ratio of 0.234, similar to that of other mammals.

The cattle genome has transposable element classes like those of other mammals, as well as large numbers of ruminant-specific repeats (table S4) that compose 27% of its genome. The

[1]Department of Biology, 406 Reiss, Georgetown University, 37th and O Streets, NW, Washington, DC 20057, USA. E-mail: ce75@georgetown.edu [2]CSIRO Livestock Industries, 306 Carmody Road, St. Lucia, QLD 4067, Australia. E-mail: ross.tellam@csiro.au [3]Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, MS BCM226, One Baylor Plaza, Houston, TX 77030, USA. E-mail: kworley@bcm.edu

*All authors with their affiliations and contributions are listed at the end of this paper.

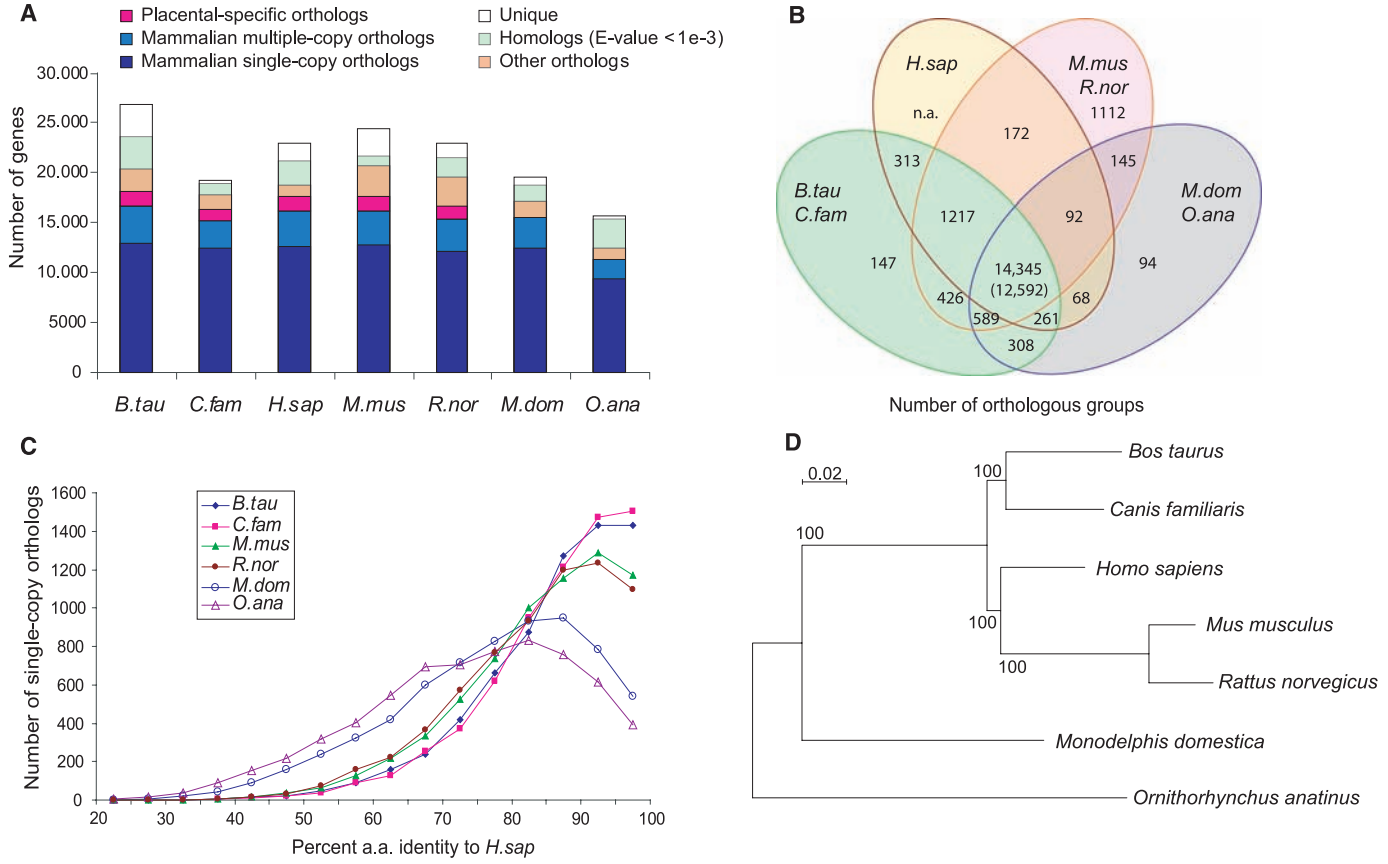

Fig. 1. Protein orthology comparison among genomes of cattle, dog, human, mouse, and rat (*Bos taurus*, *Canis familiaris*, *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, representing placental mammals), opossum (*Monodelphis domestica*, marsupial), and platypus (*Ornithorhynchus anatinus*, monotreme). (**A**) The majority of mammalian genes are orthologous, with more than half preserved as single copies (dark blue); a few thousand have species-specific duplications (blue); another few thousand have been lost in specific lineages (orange). We also show those lacking confident orthology assignment (green), and those that are apparently lineage specific [unique (white)]. Placental-specific orthologs are shown in pink. Single- or multiple-copy genes were defined on the basis of representatives in human, bovine, or dog; mouse or rat; and opossum or platypus. (**B**) Venn diagram showing shared orthologous groups (duplicated genes were counted as one) between laurasiatherians (cattle and dog), human, rodents (mouse and rat), and nonplacental mammals (opossum and platypus) on the basis of the presence of a representative gene in at least one of the grouped species [as in (A)]. (**C**) Distribution of ortholog protein identities between human and the other species for a subset of strictly conserved single-copy orthologs. (**D**) A maximum likelihood phylogenetic tree using all single-copy orthologs supports the accepted phylogeny and quantifies the relative rates of molecular evolution expressed as the branch lengths.

Downloaded from www.sciencemag.org on April 23, 2009

consensus sequence of Bov-B, a long interspersed nuclear element (LINE) lacked a functional open reading frame (ORF), which suggested that it was inactive (*7*). However, Bov-B repeats with intact ORF were identified in the genome, and their phylogeny (fig. S4) indicates that some are still actively expanding and evolving. Mapping chromosomal segments of high- and low-density ancient repeat content, L2/MIR [a LINE/SINE (short interspersed nuclear element) pair] and Bov-B, and more recent repeats, Bov-B/ART2A (Bov-B–derived SINE pair), revealed that the genome consists of ancient regions enriched for L2/MIR and recent regions enriched for Bov-B/ART2A (fig. S7). Exclusion of Bov-B/ART2A from contiguous blocks of ancient repeats suggests that evolution of the ruminant or cattle genome experienced invasions of new repeats into regions lacking ancient repeats. Alternatively, older repeats may have been destroyed by insertion of ruminant- or cattle-specific repeats. AGC trinucleotide repeats, the most common simple-sequence repeat (SSR) in artiodactyls (which include cattle, pigs, and sheep), are 90- and 142-fold overrepresented in cattle compared with human and dog, respectively (fig. S10). Of the AGC repeats in the cattle genome, 39% were associated with Bov-A2 SINE elements.

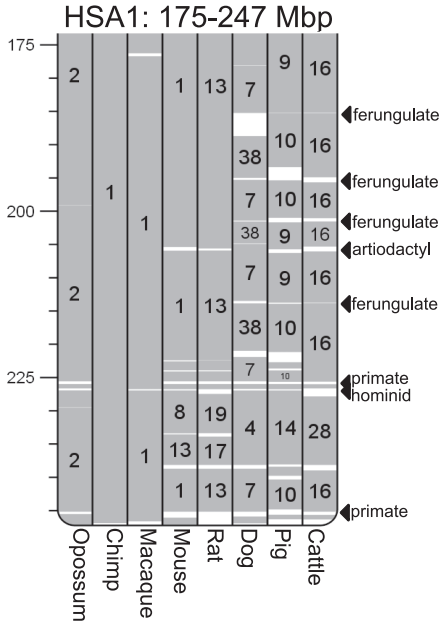

Fig. 2. Examples of EBRs. Ferungulate-, artiodactyl-, and primate-specific EBRs on HSA1 at 175 to 247 Mbp (other lineage-specific EBRs not shown). Homologous synteny blocks constructed for the macaque, chimp, cattle, dog, mouse, rat, and pig genomes were used for pairwise comparisons (*4*). White areas correspond to EBRs. Arrows to the right of the chromosome ideogram indicate positions of representative cattle-specific; artiodactyl-specific (specific to the chromosomes of pigs and cattle); ferungulate-specific (cattle, dog, and pig); primate-specific (human, macaque, and chimp); and hominoid-specific (human and chimp) rearrangements. Opossum is shown as an outgroup to the eutherian clade, which allows classification of ferungulate-specific EBRs.

A comparative analysis examined the rate of protein evolution and the conservation of gene repertoires among orthologs in the genomes of dog, human, mouse, and rat (representing placental mammals); opossum (marsupial); and platypus (monotreme). Orthology was resolved for >75% of cattle and >80% of human genes (Fig. 1A). There were 14,345 orthologous groups with representatives in human, cattle, or dog; mouse or rat; and opossum or platypus, which represent 16,749 cattle and 16,177 human genes, respectively, of which 12,592 are single-copy orthologs. We also identified 1217 placental mammal–specific orthologous groups with genes present in human, cattle, or dog; mouse or rat; but not opossum or platypus. About 1000 orthologs shared between rodents and laurasiatherians (cattle and dog), many of which encode G protein–coupled receptors, appear to have been lost or may be misannotated in the human genome (Fig. 1B). Gene repertoire conservation among these mammals correlates with conservation at the amino acid–sequence level (Fig. 1C). The elevated rate of evolution in rodents relative to other mammals (*8*) was supported by the higher amino acid sequence identity between human and dog or cattle proteins relative to that between human and rodent proteins. However, maximum-likelihood analysis of amino acid substitutions in single-copy orthologs supports the accepted sister lineage relation of primates and rodents (*1*) (Fig. 1D).

Table 1. Changes in the number of genes in innate immune gene families. Many of the β-defensin genes are present in unassigned scaffolds, i.e., they are not yet part of the current assembly. The exact number of β-defensin genes is uncertain. Interferon subfamily pseudogenes predicted on the basis of frame-shift mutations or stop codons within the first 100 amino acids of the coding sequence have been excluded from the table. The IFNX genes represent a newly discovered subfamily of IFN and are so named for convenience. BPI, Bactericidal and/or permeability-increasing; RNase, ribonuclease; LBP, lipopolysaccharide-binding protein; ULBP, UL16-binding protein.

Gene family	Bovine	Human	Murine
Cathelicidin	10	1	1
RNase	21	13	25
BPI-like	13	9	11
BPI/LBP	3	2	2
β-Defensin	~106	39	52
Interferon subfamilies			
IFNK	1	1	1
IFNE	1	1	1
IFNB	6	1	1
IFNA	13	13	14
IFNW	24	1	0
IFNT	3	0	0
IFNX	3	0	0
IFNL	0	3	2
IFNZ	0	0	2
C-type lysozyme	10	1	3
ULBP[1]	30	3	1

[1](*31*).

Alternative splicing is a major mechanism for transcript diversification (*9*), yet the extent of its evolutionary conservation and functional impact remain unclear. We used the cattle genome to analyze the conservation of the most common form of alternative splicing, exon skipping, defined as a triplet of exons in which the middle exon is absent in some transcripts, in a set of 1930 exon-skipping events across human, mouse, dog, and cattle (*4*). We examined 277 cases, with different conservation patterns between human and mouse, in 16 different cattle tissues with reverse transcription polymerase chain reaction (*4*). These splicing events were divided into a shared set (163 in both human and mouse) and a nonshared set (114 in human but not in mouse). Of the 277, we detected exon-skipping for 188 cases in cattle (table S5), which suggested that the majority of genes with exon-skipping in human were present and regulated in cattle and that, if an event is shared between human and mouse, it was more likely to be found in cattle. It was estimated that at most 40% of exon-skipping is conserved among mammals; thus, our data agree with the upper bound from previous analyses with human and rodents [e.g., (*10*)].

We constructed a cattle-human Oxford grid (fig. S12) (*4*) to conduct synteny-based chromosomal comparisons, which reinforced that human genome organization is more similar to cattle's than rodents' because most cattle chromosomes primarily correspond to part of one human chromosome, albeit with multiple rearrangements [e.g., (*11*)]. In contrast, the cattle-mouse Oxford grid shows poorer chromosomal correspondence. Lineage-specific evolutionary breakpoints were identified for cattle, artiodactyls, and ferungulates (a group encompassing artiodactyls and carnivores, represented by cattle, pig, and dog) and are shown with cattle (fig. S11) and human sequence coordinates (Fig. 2) (*4*). Primate, dog, rodent, mouse, and rat lineage-specific breakpoint positions were similarly identified. A total of 124 evolutionary breakpoint regions (EBRs) were identified in the cattle lineage, of which 100 were cattle- or ruminant-specific and 24 were artiodactyl-specific (e.g., Fig. 2). Nine additional EBRs represent presumptive ferungulate-specific rearrangements. *Bos taurus* chromosome 16 (BTA16) is populated with four ferungulate-specific EBRs, which suggests that this region was rearranged before the Artiodactyla and Carnivora divergence (Fig. 2). Such conserved regions demonstrate that many inversions that occurred before the divergence of the carnivores and artiodactyls have probably been retained in the ancestral form within the human genome. In contrast to the cattle genome, a pig physical map identified only 77 lineage-specific EBRs. Interchromosomal rearrangements and inversions characterize most of the lineage-specific rearrangements observed in the cattle, dog, and pig genomes.

Downloaded from www.sciencemag.org on April 23, 2009

An examination of repeat families and individual transposable elements within cattle-, artiodactyl- and ferungulate-specific EBRs showed a significantly higher density of LINE-L1 elements and the ruminant-specific LINE-RTE repeat family (*12*) in cattle-specific EBRs relative to the remainder of the cattle genome (table S6). In contrast, the SINE-BovA repeat family and the more ancient $tRNA^{Glu}$-derived SINE repeats (*13*) were present in lower density in cattle-specific EBRs, similar to other LINEs and SINEs (table S7). The differences in repeat densities were generally consistent in cattle-, artiodactyl- and ferungulate-specific EBRs, with the exception of the $tRNA^{Glu}$-derived and LTR-ERVL repeats, which are at higher densities in artiodactyl EBRs compared with the rest of the genome.

The $tRNA^{Glu}$-derived SINEs originated in the common ancestor of Suina (pigs and peccaries), Ruminantia, and Cetacea (whales) (*13*), which suggests that $tRNA^{Glu}$-derived SINEs were involved in ancestral artiodactyl chromosome rearrangements. Furthermore, the lower density of the more ancient repeat families in cattle-specific EBRs suggests that either more recently arising repeat elements were inserted into regions lacking ancient repeats or that older repeats were destroyed by this insertion (table S7). The repeat elements differing in density in EBRs were also found in regions of homologous synteny, which suggests that repeats may promote evolutionary rearrangements (see below). Differences in repeat density in cattle-specific EBRs are thus unlikely to be caused by the accumulation of repeats in EBRs after such rearrangements occur. We identified a cattle-specific EBR associated with a bidirectional promoter (figs. S14 and S15) that may affect control of the expression of the *CYB5R4* gene, which has been implicated in human diabetes and, therefore, may be important in the regulation of energy flow in cattle (*4*).

We identified 1020 segmental duplications (SDs) corresponding to 3.1% (94.4 Mbp) of the cattle genome (*4*). Duplications assigned to a chromosome showed a bipartite distribution with respect to length and percent identity (fig. S16), and interchromosomal duplications were shorter (median length 2.5 kbp) and more divergent (<94% identity) relative to intrachromosomal duplications (median length 20 kbp, ~97% identity) and tended to be locally clustered (fig. S17). Twenty-one of these duplications were >300 kbp and located in regions enriched for tandem duplications (e.g., BTA18) (fig. S18). This pattern is reminiscent of the duplication pattern of the dog, rat, and mouse but different from that of primate and great-ape genomes (*14*, *15*). On average, cattle SDs >10 kbp represent 11.7% of base pairs in 10-kbp intervals located within cattle-specific EBRs and 23.0% of base pairs located within the artiodactyl-specific EBRs. By contrast, in the remainder of the genome sequence assigned to chromosomes the fraction of SDs was 1.7% ($P < 1 \times 10^{-12}$). These data indicate that SDs play a role in promoting chromosome rearrangements by nonallelic homologous recombination [e.g., (*16*)] and suggest that either a significant fraction of the SDs observed in cattle occurred before the Ruminant-Suina split, and/or that the sites for accumulation of SDs are nonrandomly distributed in artiodactyl genomes.

SDs involving genic regions may give rise to new functional paralogs. Seventy-six percent (778 out of 1020) of the cattle SDs correspond to complete or partial gene duplications with high sequence identity (median 98.7%). This suggests that many of these gene duplications are specific to either the artiodactyla or the *Bos* lineage and tend to encode proteins that often interface with the external environment, particularly immune proteins and sensory and/or olfactory receptors. Several of these gene duplications are also duplicated in other mammalian lineages (e.g., cytochrome P-450, sulfotransferase, ribonuclease A, defensins, and pregnancy-associated glycoproteins). Paralogs located in segmental duplications that are present exclusively in cattle may have functional implications for the unique physiology, environment, and diet of cattle.

An overrepresentation of genes involved in reproduction in cattle SDs (tables S8 and S9) is associated with several gene families expressed in the ruminant placenta. These families encode the intercellular signaling proteins pregnancy-associated glycoproteins (on BTA29), trophoblast Kunitz domain proteins (on BTA13), and interferon tau (*IFNT*) (on BTA8). A gene family encoding prolactin-related proteins (on BTA23) was only identified in the assembly-dependent analysis of SDs. These genes regulate ruminant-specific aspects of fetal growth, maternal adaptations to pregnancy, and the coordination of parturition (*17*, *18*). Although type I interferon (IFN) genes are primarily involved in host defense (*19*), *IFNT* prevents regression of the corpus luteum during early pregnancy, which results in a uterine environment receptive to early conceptus development (*20*).

Signatures of positive selection (obtained by measurement of their rates of synonymous and nonsynonymous substitutions) identified 71 genes (*4*), including 10 immune-related genes (i.e., *IFNAR2*, *IFNG*, *CD34*, *TREM1*, *TREML1*, *FCER1A*, *IL23R*, *IL24*, *IL15*, and *LEAP2*). As previously mentioned, immune genes are overrepresented in SDs (see Table 1 and fig. S20). Examples of genes varying in cattle relative to mouse include a cluster of β-defensin genes, which encode antimicrobial peptides; the antimicrobial cathelicidin genes [which show increased sequence diversity of the mature cathelicidin peptides (*21*)]; and changes in the numbers of interferon genes (*22*) and the number and organization of genes involved in adaptive immune responses in cattle compared with human and mouse (*4*). This extensive duplication and divergence of genes involved in innate immunity may be because of the substantial load of microorganisms present in the rumen of cattle, which increases the risk of opportunistic infections at mucosal surfaces and positive selection for the traits that enabled stronger and more diversified innate immune responses at these locations. Another possibility is that immunity may have been under selection due to the herd structure, which can promote rapid disease transmission. Also, immune function–related duplicated genes have gained nonimmune functions, e.g., *IFNT* (see above), and the C-class lysozyme genes, which are involved in microbial degradation in the abomasum (see below).

There has been substantial reorganization of gene families encoding proteins present in milk. One such rearrangement affecting milk composition involves the histatherin (*HSTN*) gene within the casein gene cluster on BTA6 (fig. S21). In the cattle genome, *HSTN* is juxtaposed to a regulatory element (*BCE*) important (*23*) for β-casein (*CSN2*) expression, and as a probable consequence, *HSTN* is regulated like the casein genes during the lactation cycle. This rearrangement that led to the juxtaposition of *HSTN* next to the *BCE* is also the probable cause of deletion of one of the two copies of α-S2–like casein genes (*CSN1S2A*) present in other mammalian genomes (*24*). The biological implications of this change in casein gene copy number are not yet clear.

Additionally, the cattle serum amyloid A (*SAA*) gene cluster arose from both a laurasiatherian SD and a cattle-specific EBR, which resulted in two mammary gland–expressed *SAA3*-like genes, *SAA3.1* and *SAA3.2* on BTA29, and an *SAA3*-like gene on BTA15 (fig. S21). SAA3.2 has been shown to inhibit microbial growth (*25*). Two additional milk protein genes were associated with SDs: cathelicidin (*CATHL1*) and β_2-microglobulin (*B2M*)—part of the neonatal Fc receptor (FcRn) that transfers immunoglobulin IgG across epithelial cells of many tissues including the gut and mammary gland (*26*, *27*). IgG is the predominant immunoglobulin in cow's milk compared with IgA in human milk (*28*). Unlike humans, who acquire passive immunity from the mother via placental transfer of immunoglobulins during pregnancy, calves acquire passive immunity by ingestion of IgG in milk (*28*). *B2M* is also redistributed in epithelial cells upon calving, and it protects IgG from degradation (*26*). A genetic variant of *B2M* has negative effects on passive immune transfer (*29*). The additional copy of the gene encoding *B2M* might be associated with the abundance of IgG in cows' milk and an increased capacity for uptake in the neonatal gut. Considering that the passive transfer of immunity to the calf is one of the important functions of milk, it is striking that lactation-related genes affected by genomic rearrangements often encode immune-related proteins in milk.

Cattle metabolic pathways demonstrated a strong degree of conservation among the comprehensive set of genes involved in core mammalian metabolism (*4*) and permitted an examination of unique genetic events that may be related to ruminant-specific metabolic adaptations. However, among 1032 genes examined from the human metabolic pathways, five were deleted or extensively diverged in cattle: *PLA2G4C* (phospholipase A2, group IVC), *FAAH2* (fatty acid amide hydrolase 2), *IDI2* (isopentenyl-diphosphate delta isomerase 2), *GSTT2* (glutathione *S*-transferase

Downloaded from www.sciencemag.org on April 23, 2009

theta 2), and *TYMP* (thymidine phosphorylase), which may be adaptations that impact on fatty acid metabolism (*PLA2G4C* and *FAAH2)*; the mevalonate pathway (synthesis of dolichols, vitamins, steroid hormones, and cholesterol) (*IDI2*); detoxification (*GSTT2*); and pyrimidine metabolism (*TYMP*). Phylogenetic analysis shows that *PLA2G4C* was deleted ~87 to 97 million years ago in the laurasiatherian lineages (fig. S22). Strikingly, ~20% of the sequences from two abomasum (last chamber of the cattle stomach) EST libraries (a total of 2392 sequences) correspond to three C-type lysozyme genes. Lysozyme primarily functions in animals as an antibacterial protein, which suggests that they probably function in the abomasum (similar to the monogastric stomach) to degrade the cell walls of bacteria entering from the foregut (*30*). The cattle genome contains 10 C-type lysozyme genes (table S14 and fig. S23), and EST evidence (fig. S23) shows that six of the seven remaining C-type lysozyme genes are expressed primarily in the intestinal tract, which suggests additional roles for the encoded proteins in ruminant digestion.

In summary, the biological systems most affected by changes in the number and organization of genes in the cattle lineage include reproduction, immunity, lactation, and digestion. We highlighted the evolutionary activity associated with chromosomal breakpoint regions and their propensity for promoting gene birth and rearrangement. These changes in the cattle lineage probably reflect metabolic, physiologic, and immune adaptations due to microbial fermentation in the rumen, the herd environment and its influence on disease transmission, and the reproductive strategy of cattle. The cattle genome and associated resources will facilitate the identification of novel functions and regulatory systems of general importance in mammals and may provide an enabling tool for genetic improvement within the beef and dairy industries.

References and Notes

1. W. J. Murphy, P. A. Pevzner, S. J. O'Brien, *Trends Genet.* **20**, 631 (2004).
2. R. L. Willham, *J. Anim. Sci.* **62**, 1742 (1986).
3. Y. Liu *et al.*, *BMC Genomics* **10**, 180 (2009).
4. Materials, methods, and additional discussion are available on *Science* online.
5. H. Nilsen *et al.*, *Anim. Genet.* **39**, 97 (2008).
6. A. Prasad *et al.*, *BMC Genomics* **8**, 310 (2007).
7. H. S. Malik, T. H. Eickbush, *Mol. Biol. Evol.* **15**, 1123 (1998).
8. C. I. Wu, W. H. Li, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1741 (1985).
9. B. Modrek, C. J. Lee, *Nat. Genet.* **34**, 177 (2003).
10. R. Sorek, R. Shamir, G. Ast, *Trends Genet.* **20**, 68 (2004).
11. A. Everts-van der Wind *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18526 (2005).
12. D. Kordis, F. Gubensek, *Gene* **238**, 171 (1999).
13. M. Shimamura, H. Abe, M. Nikaido, K. Ohshima, N. Okada, *Mol. Biol. Evol.* **16**, 1046 (1999).
14. J. A. Bailey, E. E. Eichler, *Nat. Rev. Genet.* **7**, 552 (2006).
15. J. A. Bailey *et al.*, *Science* **297**, 1003 (2002).
16. W. J. Murphy *et al.*, *Science* **309**, 613 (2005).
17. K. Hashizume *et al.*, *Reprod. Fertil. Dev.* **19**, 79 (2007).
18. J. H. Larson *et al.*, *Physiol. Genomics* **25**, 405 (2006).
19. S. Y. Zhang *et al.*, *Immunol. Rev.* **226**, 29 (2008).
20. R. M. Roberts, Y. Chen, T. Ezashi, A. M. Walker, *Semin. Cell Dev. Biol.* **19**, 170 (2008).
21. M. Scocchi, S. Wang, M. Zanetti, *FEBS Lett.* **417**, 311 (1997).
22. M. G. Katze, Y. He, M. Gale Jr., *Nat. Rev. Immunol.* **2**, 675 (2002).
23. C. Schmidhauser *et al.*, *Mol. Biol. Cell* **3**, 699 (1992).
24. M. Rijnkels, L. Elnitski, W. Miller, J. M. Rosen, *Genomics* **82**, 417 (2003).
25. A. J. Molenaar *et al.*, *Biomarkers* **14**, 26 (2009).
26. B. Mayer *et al.*, *J. Dairy Res.* **72** (suppl. S1), 107 (2005).
27. D. C. Roopenian, S. Akilesh, *Nat. Rev. Immunol.* **7**, 715 (2007).
28. T. J. Newby, C. R. Stokes, F. J. Bourne, *Vet. Immunol. Immunopathol.* **3**, 67 (1982).
29. M. L. Clawson *et al.*, *Mamm. Genome* **15**, 227 (2004).
30. D. M. Irwin, *J. Mol. Evol.* **41**, 299 (1995).
31. J. H. Larson *et al.*, *BMC Genomics* **7**, 227 (2006).

32. Funded by the National Human Genome Research Institute (NHGRI U54 HG003273); the U.S. Department of Agriculture's Agricultural Research Service (USDA-ARS agreement no. 59-0790-3-196) and Cooperative State Research, Education, and Extension Service National Research Initiative (grant no. 2004-35216-14163); the state of Texas; Genome Canada through Genome British Columbia; the Alberta Science and Research Authority; the Commonwealth Scientific and Industrial Research Organization of Australia (CSIRO); Agritech Investments Ltd., Dairy Insight, Inc., and AgResearch Ltd., all of New Zealand; the Research Council of Norway; the Kleberg Foundation; and the National, Texas, and South Dakota Beef Check-off Funds. The master accession for this WGS sequencing project is AAFC03000000. The individual WGS sequences are AAFC03000001 to AAFC03131728, and the scaffold records are CM000177 to CM000206 (chromosomes) and DS490632 to DS495890 (unplaced scaffolds).

The Bovine Genome Sequencing and Analysis Consortium

Principal Investigator: Richard A. Gibbs[1]

Analysis project leadership: Christine G. Elsik,[2,3] Ross L. Tellam[4]

Sequencing project leadership: Richard A. Gibbs,[1] Donna M. Muzny,[1] George M. Weinstock[5,1]

Analysis group organization: David L. Adelson,[6] Evan E. Eichler,[7,8] Laura Elnitski,[9] Christine G. Elsik,[2,3] Roderic Guigó,[10] Debora L. Hamernik,[11] Steve M. Kappes,[12] Harris A. Lewin,[13,14] David J. Lynn,[15] Frank W. Nicholas,[16] Alexandre Reymond,[17] Monique Rijnkels,[18] Loren C. Skow,[19] Ross L. Tellam,[4] Kim C. Worley,[1] Evgeny M. Zdobnov[20,21,22]

Sequencing project white paper: Richard A. Gibbs,[1] Steve M. Kappes,[12] Lawrence Schook,[13] Loren C. Skow,[19] George M. Weinstock,[5,1] James Womack[23]

Gene prediction and consensus gene set: Tyler Alioto,[10] Stylianos E. Antonarakis,[20] Alex Astashyn,[24] Charles E. Chapple,[10] Hsiu-Chuan Chen,[24] Jacqueline Chrast,[17] Francisco Câmara,[10] Christine G. Elsik[2,3] (leader), Olga Ermolaeva,[24] Roderic Guigó,[10] Charlotte N. Henrichsen,[17] Wratko Hlavina,[24] Yuri Kapustin,[24] Boris Kiryutin,[24] Paul Kitts,[24] Felix Kokocinski,[25] Melissa Landrum,[24] Donna Maglott,[24] Kim Pruitt,[24] Alexandre Reymond,[17] Victor Sapojnikov,[24] Stephen M. Searle,[25] Victor Solovyev,[26] Alexandre Souvorov,[24] Catherine Ucla,[20] George M. Weinstock,[5,1] Carine Wyss[20]

Experimental validation of gene set: Tyler Alioto,[10] Stylianos E. Antonarakis,[20] Charles E. Chapple,[10] Jacqueline Chrast,[17] Francisco Câmara,[10] Roderic Guigó[10] (leader), Charlotte N. Henrichsen,[17] Alexandre Reymond,[17] Catherine Ucla,[20] Carine Wyss[20]

MicroRNA analysis: Juan M. Anzola,[3] Daniel Gerlach,[20,21] Evgeny M. Zdobnov[20,21,22] (leader)

GC composition analysis: Eran Elhaik,[27,28] Christine G. Elsik[2,3] (leader), Dan Graur,[27] Justin T. Reese[2]

Repeat analysis: David L. Adelson[6] (leader), Robert C. Edgar,[29] John C. McEwan,[30] Gemma M. Payne,[30] Joy M. Raison[31]

Protein ortholog analysis: Thomas Junier,[19,20] Evgenia V. Kriventseva,[32] Evgeny M. Zdobnov[20,21,22] (leader)

Exon-skipping analysis: Jacqueline Chrast,[17] Eduardo Eyras,[33,34] Charlotte N. Henrichsen,[17] Mireya Plass,[34] Alexandre Reymond[17] (leader)

Evolutionary breakpoint analysis and Oxford grid: Ravikiran Donthu,[13] Denis M. Larkin,[13,14] Harris A. Lewin[13,14] (leader), Frank W. Nicholas[16]

Bidirectional promoter analysis: Laura Elnitski[9] (leader), Denis M. Larkin,[13,14] Harris A. Lewin,[13,14] James Reecy,[35] Mary Q. Yang[9]

Segmental duplication analysis: David L. Adelson,[6] Lin Chen,[7] Ze Cheng,[7] Carol G. Chitko-McKown,[36] Evan E. Eichler[7,8] (leader), Laura Elnitski,[9] Christine G. Elsik,[2,3] George E. Liu,[37] Lakshmi K. Matukumalli,[38,37] Jiuzhou Song,[39] Bin Zhu[39]

Analysis of gene ontology in segmental duplications: Christine G. Elsik,[2,3] David J. Lynn[15] (leader), Justin T. Reese[2]

Adaptive evolution: Daniel G. Bradley,[40] Fiona S.L. Brinkman,[15] Lilian P.L. Lau,[40] David J. Lynn[15] (leader), Matthew D. Whiteside[15]

Innate immunity: Ross L. Tellam[4] (leader), Angela Walker,[41] Thomas T. Wheeler[42]

Lactation: Theresa Casey,[43] J. Bruce German,[44,45] Danielle G. Lemay,[45] David J. Lynn,[15] Nauman J. Maqbool,[46] Adrian J. Molenaar,[42] Monique Rijnkels[18] (leader)

Metabolism: Harris A. Lewin[13,14] (leader), Seongwon Seo,[47] Paul Stothard[48]

Adaptive immunity: Cynthia L. Baldwin,[49] Rebecca Baxter,[50] Candice L. Brinkmeyer-Langford,[19] Wendy C. Brown,[51] Christopher P. Childers,[2] Timothy Connelley,[52] Shirley A. Ellis,[53] Krista Fritz,[19] Elizabeth J. Glass,[50] Carolyn T.A. Herzig,[49] Antti Iivanainen,[54] Kevin K. Lahmers,[51] Loren C. Skow[19] (leader)

Annotation data management: Anna K. Bennett,[2] Christopher P. Childers,[2] C. Michael Dickens,[3] Christine G. Elsik[2,3] (leader), James G.R. Gilbert,[25] Darren E. Hagen,[2] Justin T. Reese,[2] Hanni Salih[3]

Manual annotation organization: Jan Aerts,[55] Alexandre R. Caetano,[56] Brian Dalrymple,[4] Christine G. Elsik,[2,3] Jose Fernando Garcia,[57] Richard A. Gibbs,[1] Clare A. Gill,[3,58] Debora L. Hamernik,[11] Stefan G. Hiendleder,[59] Erdogan Memili,[60] Frank W. Nicholas,[16] James Reecy,[35] Monique Rijnkels,[18] Loren C. Skow,[19] Diane Spurlock,[35] Paul Stothard,[48] Ross L. Tellam,[4] George M. Weinstock,[5,1] John L. Williams,[61] Kim C. Worley[1]

cDNA tissues, libraries, and sequencing: Lee Alexander,[62] Michael J. Brownstein,[63] Leluo Guan,[48] Robert A. Holt[64] (leader), Steven J.M. Jones[64] (leader), Marco A. Marra[64] (leader), Richard Moore,[64] Stephen S. Moore[48] (leader), Andy Roberts,[62] Masaaki Taniguchi,[65,48] Richard C. Waterman[62]

Genome sequence production: Joseph Chacko,[1] Mimi M. Chandrabose,[1] Andy Cree[1] (leader), Marvin Diep Dao,[1] Huyen H. Dinh[1] (leader), Ramatu Ayiesha Gabisi,[1] Sandra Hines,[1] Jennifer Hume[1] (leader), Shalini N. Jhangiani,[1] Vandita Joshi,[1] Christie L. Kovar[1] (leader), Lora R. Lewis,[1] Yih-shin Liu,[1] John Lopez,[1] Margaret B. Morgan,[1] Donna M. Muzny[1] (leader), Ngoc Bich Nguyen,[1] Geoffrey O. Okwuonu,[1] San Juana Ruiz,[1] Jireh Santibanez,[1] Rita A. Wright[1]

Sequence finishing: Christian Buhay[1] (leader), Yan Ding,[1] Shannon Dugan-Rocha[1] (leader), Judith Herdandez,[1] Michael Holder,[1] Aniko Sabo[1]

Automated BAC assembly: Amy Egan,[1] Jason Goodell,[1] Katarzyna Wilczek-Boney[1]

Sequence production informatics: Gerald R. Fowler[1] (leader), Matthew Edward Hitchens,[1] Ryan J. Lozado,[1] Charles Moen,[1] David Steffen,[66,1] James T. Warren,[1] Jingkun Zhang[1]

BAC mapping: Readman Chiu,[64] Steven J.M. Jones,[64] Marco A. Marra[64] (leader), Jacqueline E. Schein[64]

Genome assembly: K. James Durbin,[67,1] Paul Havlak,[68,1] Huaiyang Jiang,[1] Yue Liu,[1] Xiang Qin,[1] Yanru Ren,[1] Yufeng Shen,[1,69] Henry Song,[1] George M. Weinstock,[5,1] Kim C. Worley[1] (leader)

Sequence library production: Stephanie Nicole Bell,[1] Clay Davis,[1] Angela Jolivet Johnson,[1] Sandra Lee,[1] Lynne V. Nazareth[1] (leader), Bella Mayurkumar Patel,[1] Ling-Ling Pu,[1] Selina Vattathil,[1] Rex Lee Williams Jr.[1]

BAC production: Stacey Curry,[1] Cerissa Hamilton,[1] Erica Sodergren[5,1] (leader)

Sequence variation detection: Lynne V. Nazareth,[1] David A. Wheeler[1]

Markers and mapping: David L. Adelson,[6] Jan Aerts,[55] Wes Barris,[4] Gary L. Bennett,[36] Brian Dalrymple,[4] André Eggen,[70] Clare A. Gill,[3,58] Ronnie D. Green,[71] Gregory P. Harhay,[36] Matthew Hobbs,[72] Oliver Jann,[50] Steve M. Kappes[12] (leader), John W. Keele,[36] Matthew P. Kent,[73] Denis M. Larkin,[13,14] Harris A. Lewin,[13,14] Sigbjørn Lien,[73] John C. McEwan,[30] Stephanie D. McKay,[74] Sean McWilliam,[4] Stephen S. Moore,[48] Frank W. Nicholas,[16] Gemma M. Payne,[30] Abhirami Ratnakumar,[75,4] Hanni Salih,[3] Robert D. Schnabel,[74] Timothy Smith,[36] Warren M. Snelling,[36] Tad S. Sonstegard,[37] Roger T. Stone,[36] Yoshikazu Sugimoto,[76] Akiko Takasuga,[76] Jeremy F. Taylor,[74] Ross L. Tellam,[4] Curtis P. Van Tassell,[37] John L. Williams[61]

Genomic DNA: Michael D. MacNeil[62]

Manual annotation: Antonio R.R. Abatepaulo,[77] Colette A. Abbey,[3] Jan Aerts,[55] Virpi Ahola,[78] Iassudara G. Almeida,[57] Ariel F. Amadio,[79] Elen Anatriello,[77] Suria M. Bahadue,[2] Cynthia L. Baldwin,[49] Rebecca Baxter,[50] Anna K. Bennett,[2] Fernando H. Biase,[13] Clayton R. Boldt,[3] Candice L. Brinkmeyer-Langford,[19] Wendy C. Brown,[51] Alexandre R. Caetano,[56] Jeffery A. Carroll,[80] Wanessa A. Carvalho,[77] Theresa Casey,[43] Eliane P. Cervelatti,[57] Elsa Chacko,[81] Jennifer E. Chapin,[3] Ye Cheng,[35] Christopher P. Childers,[2] Jungwoo Choi,[3] Adam J. Colley,[82] Timothy Connelley,[52] Tatiana A. de Campos,[56] Marcos De Donato,[83]

Downloaded from www.sciencemag.org on April 23, 2009

Isabel K.F. de Miranda Santos,[56,77] Carlo J.F. de Oliveira,[77] Heather Deobald,[84] Eve Devinoy,[85] C. Michael Dickens,[3] Kaitlin E. Donohue,[2] Peter Dovc,[86] Annett Eberlein,[87] Shirley A. Ellis,[53] Carolyn J. Fitzsimmons,[59] Alessandra M. Franzin,[77] Krista Fritz,[19] Gustavo R. Garcia,[77] Jose Fernando Garcia,[57] Sem Genini,[61] J. Bruce German,[44,45] James G.R. Gilbert,[25] Clare A. Gill,[3,58] Cody J. Gladney,[3] Elizabeth J. Glass,[50] Jason R. Grant,[48] Marion L. Greaser,[88] Jonathan A. Green,[74] Darryl L. Hadsell,[18] Darren E. Hagen,[2] Hatam A. Hakimov,[89] Rob Halgren,[43] Jennifer L. Harrow,[25] Elizabeth A. Hart,[25] Nicola Hastings,[90,50] Marta Hernandez,[91] Carolyn T.A. Herzig,[49] Stefan G. Hiendleder,[59] Matthew Hobbs,[72] Zhi-Liang Hu,[35] Antti Iivanainen,[54] Aaron Ingham,[4] Terhi Iso-Touru,[78] Catherine Jamis,[2] Oliver Jann,[50] Kirsty Jensen,[50] Dimos Kapetis,[61] Tovah Kerr,[51] Sari S. Khalil,[2] Hasan Khatib,[92] Davood Kolbehdari,[48,93] Charu G. Kumar,[13] Dinesh Kumar,[94,35] Richard Leach,[50] Justin C-M Lee,[2] Danielle G. Lemay,[45] Changxi Li,[95,48] George E. Liu,[37] Krystin M. Logan,[96] Roberto Malinverni,[61] Nauman J. Maqbool,[46] Elisa Marques,[48] William F. Martin,[45] Natalia F. Martins,[56] Sandra R. Maruyama,[77] Raffaele Mazza,[97] Kim L. McLean,[84] Juan F. Medrano,[98] Erdogan Memili,[60] Adrian J. Molenaar,[42] Barbara T. Moreno,[57] Daniela D. Moré,[77] Carl T. Muntean,[3] Hari P. Nandakumar,[19] Marcelo F.G. Nogueira,[99] Ingrid Olsaker,[100] Sameer D. Pant,[82] Francesca Panzitta,[61] Rosemeire C.P. Pastor,[57] Mario A. Poli,[101] Nathan Poslusny,[2] Satyanarayana Rachagani,[35] Shoba Ranganathan,[81,102] Andrej Razpet,[86] James Reecy,[35] Penny K. Riggs,[3,58] Monique Rijnkels,[18] Gonzalo Rincon,[98] Nelida Rodriguez-Osorio,[60,103] Sandra L. Rodriguez-Zas,[13] Natasha E. Romero,[3] Anne Rosenwald,[2] Lillian Sando,[4] Sheila M. Schmutz,[84] Seongwon Seo,[47] Libing Shen,[2] Laura Sherman,[48] Loren C. Skow,[19] Bruce R. Southey,[104] Diane Spurlock,[35] Ylva Strandberg Lutzow,[4] Jonathan V. Sweedler,[104] Imke Tammen,[72] Masaaki Taniguchi,[65,48] Ross L. Tellam,[4] Bhanu Prakash V.L. Telugu,[74] Jennifer M. Urbanski,[2] Yuri T. Utsunomiya,[57] Chris P. Verschoor,[82] Ashley J. Waardenberg,[4,105] Angela Walker,[41] Zhiquan Wang,[48] Robert Ward,[106] Rosemarie Weikard,[87] Thomas H. Welsh Jr.,[3,58] Thomas T. Wheeler,[42] Stephen N. White,[51,107] John L. Williams,[61] Laurens G. Wilming,[25] Kris R. Wunderlich,[3] Jianqi Yang,[108] Feng-Qi Zhao[109]

[1]Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. [2]Department of Biology, 406 Reiss, Georgetown University, 37th & O Streets NW, Washington, DC 20057, USA. [3]Department of Animal Science, Texas A&M University, 2471 TAMU, College Station, TX 77843–2471, USA. [4]Livestock Industries, Commonwealth Scientific and Industrial Research Organization (CSIRO), 306 Carmody Road, St. Lucia, Queensland, 4067, Australia. [5]The Genome Center at Washington University, Washington University School of Medicine, 4444 Forest Park Avenue, St. Louis, MO 63108, USA. [6]School of Molecular and Biomedical Science, School of Agriculture, Food and Wine, The University of Adelaide, Adelaide, SA, 5005, Australia. [7]Department of Genome Sciences, University of Washington, 1705 NE Pacific Street, Seattle, WA 98195–5065, USA. [8]Howard Hughes Medical Institute, Seattle, WA 98195, USA. [9]National Human Genome Research Institute, National Institutes of Health, 5625 Fishers Lane, Rockville, MD 20878, USA. [10]Center for Genomic Regulation and Grup de Recerca en Informática Biomédica, Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra, 08003 Barcelona, Catalonia, Spain. [11]U.S. Department of Agriculture (USDA), Cooperative State Research, Education, & Extension Service, 1400 Independence Avenue SW, Stop 2220, Washington, DC 20250–2220, USA. [12]National Program Staff, USDA–Agricultural Research Service, 5601 Sunnyside Avenue, Beltsville, MD 20705, USA. [13]Department of Animal Sciences, University of Illinois at Urbana–Champaign, 1201 West Gregory Drive, Urbana, IL 61801, USA. [14]Institute for Genomic Biology, University of Illinois at Urbana–Champaign, 1201 West Gregory Drive, Urbana, IL 61801, USA. [15]Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada. [16]Faculty of Veterinary Science, University of Sydney, Sydney, NSW, 2006, Australia. [17]Center for Integrative Genomics, University of Lausanne, Lausanne, 1015, Switzerland. [18]Children's Nutrition Research Center, USDA–Agricultural Research Service, Department of Pediatrics–Nutrition, Baylor College of Medicine, 1100 Bates Street, Houston, TX 77030–2600, USA. [19]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA. [20]Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, Geneva, 1211, Switzerland. [21]Swiss Institute of Bioinformatics, 1 rue Michel-Servet, Geneva, 1211, Switzerland. [22]Division of Molecular Biosciences, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK. [23]Department of Veterinary Pathobiology, Texas A&M University, College Station, TX 77843, USA. [24]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA. [25]Informatics Department, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH, UK. [26]Department of Computer Science, University of London, Royal Holloway, Egham, Surrey, TW20 0EX, UK. [27]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA. [28]McKusick—Nathans Institute of Genetic Medicine, BRB 579, Johns Hopkins University School of Medicine, 733 North Broadway, Baltimore, MD 21205, USA. [29]45 Monterey Drive, Tiburon, CA 94920, USA. [30]Animal Genomics, AgResearch, Invermay, PB 50034, Mosgiel, 9053, New Zealand. [31]eResearch SA, University of Adelaide, North Terrace, Adelaide, SA, 5005, Australia. [32]Department of Structural Biology and Bioinformatics, University of Geneva Medical School, 1 rue Michel-Servet, Geneva, 1211, Switzerland. [33]Catalan Institution for Research and Advanced Studies, 08010 Barcelona, Catalonia, Spain. [34]Computational Genomics, Universitat Pompeu Fabra, 08003 Barcelona, Catalonia, Spain. [35]Department of Animal Science, Iowa State University, 2255 Kildee Hall, Ames, IA 50011–3150, USA. [36]Meat Animal Research Center, USDA–Agricultural Research Service, Clay Center, NE 68933, USA. [37]Bovine Functional Genomics Laboratory, USDA–Agricultural Research Service, Beltsville Agricultural Research Center (BARC)–East, Beltsville, MD 20705, USA. [38]Department of Bioinformatics and Computational Biology, George Mason University, 10900 University Blvd, Manassas, VA 20110, USA. [39]Department of Bioengineering, University of Maryland, College Park, MD 20742, USA. [40]Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. [41]Department of Veterinary Pathobiology, 245 Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA. [42]Dairy Science and Technology Section, AgResearch, Ruakura Research Centre, East Street, Private Bag 3123, Hamilton, 3240, New Zealand. [43]Department of Animal Science, Michigan State University, East Lansing, MI 48824–1225, USA. [44]Nestlé Research Centre, Vers chez les Blanc CH, Lausanne 26, 1000, Switzerland. [45]Department of Food Science and Technology, University of California–Davis, Davis, CA 95616, USA. [46]Bioinformatics, Mathematics and Statistics, AgResearch, Ruakura Research Centre, East Street, Private Bag 3123, Hamilton, 3240, New Zealand. [47]Division of Animal Science and Resource, Chungnam National University, Daejeon, 305-764, Korea. [48]Department of Agricultural, Food and Nutritional Science, University of Alberta, 410 AgFor Centre, Edmonton, AL, T6G 2P5, Canada. [49]Department of Veterinary and Animal Sciences, University of Massachusetts, Amherst, MA 01003, USA. [50]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin, Midlothian, EH25 9PS, UK. [51]Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA 99164, USA. [52]Division of Infection and Immunity, The Roslin Institute, Royal (Dick) School of Veterinary Science, University of Edinburgh, Roslin, Midlothian, EH25 9RG, UK. [53]Immunology Division, Institute for Animal Health, Compton, RG20 7NN, UK. [54]Department of Basic Veterinary Sciences, University of Helsinki, Post Office Box 66, Helsinki, FIN-00014, Finland. [55]Genome Dynamics and Evolution, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK. [56]Embrapa Recursos Genéticos e Biotecnologia, Final Avenida W/5 Norte, Brasilia, DF, 70770-900, Brazil. [57]Animal Production and Health Department, UNESP—Sao Paulo State University, Aracatuba, SP, 16050-680, Brazil. [58]Texas AgriLife Research, College Station, TX 77843, USA. [59]JS Davies Epigenetics and Genetics Group, School of Agriculture, Food & Wine and Research Centre for Reproductive Health, The University of Adelaide, Roseworthy Campus, Roseworthy, SA, 5371, Australia. [60]Department of Animal and Dairy Sciences, Mississippi Agricultural and Forestry Experiment Station, Mississippi State University, Mississippi State, MS 39762, USA. [61]Parco Tecnologico Padano, Via Einstein, Polo Universitario, Lodi, 26900, Italy. [62]Fort Keogh Livestock and Range Research Laboratory, USDA-Agricultural Research Service, Miles City, MT 59301, USA. [63]Laboratory of Genetics, National Institute of Mental Health, NIH, Building 49, B1EE16, 49 Convent Drive, Bethesda, MD 20892, USA. [64]Genome Sciences Centre, British Columbia Cancer Agency, 675 West 10th Avenue, Vancouver, BC, V5Z 1L3, Canada. [65]Division of Animal Sciences, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, 305-8602, Japan. [66]Bioinformatics Research Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. [67]Department of Biomolecular Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064, USA. [68]Department of Computer Science, University of Houston, Houston, TX 77204–3010, USA. [69]Department of Computer Science and Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA. [70]INRA, Animal Genetics and Integrative Biology, Bovine Genetics and Genomics, 78350 Jouy-en-Josas, France. [71]Pfizer Animal Genetics, Pfizer Animal Health, New York, NY 10017, USA. [72]Faculty of Veterinary Science, University of Sydney, Camden, NSW, 2570, Australia. [73]Centre for Integrative Genetics and Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Arboretveien 6, Ås, 1432, Norway. [74]Division of Animal Sciences, University of Missouri, 920 East Campus Drive, Columbia, MO 65211, USA. [75]Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala Biomedical Centre Husargatan 3, Uppsala, 75 123, Sweden. [76]Shirakawa Institute of Animal Genetics, Nishigo, Fukushima 961-8061, Japan. [77]Department of Biochemistry and Immunology, Ribeirão Preto Medical School, University of São Paulo, Av Bandeirantes 3900, Ribeirão Preto, SP, 14049-900, Brazil. [78]Biotechnology and Food Research, MTT Agrifood Research Finland, Jokioinen, FI-31600, Finland. [79]EEA Rafaela, Instituto Nacional de Tecnología Agropecuaria (INTA), Ruta 34 Km 227, Rafaela, Santa Fe, 2300, Argentina. [80]Livestock Issues Research Unit, USDA–Agricultural Research Service, Lubbock, TX 79403, USA. [81]Department of Chemistry and Biomolecular Sciences & ARC Centre of Excellence in Bioinformatics, Macquarie University, Sydney, 2109, NSW, Australia. [82]Department of Animal and Poultry Science, University of Guelph, Guelph, ON, N1G2W1, Canada. [83]Instituto de Investigaciones en Biomedicina y Ciencias Aplicadas, Universidad de Oriente, Avenida Universidad, Cumana, Sucre, 6101, Venezuela. [84]Department of Animal and Poultry Science, University of Saskatchewan, Saskatoon, SK, S7N 5A8, Canada. [85]INRA–UR1196, Génomique et Physiologie de la Lactation, F78352 Jouy-en-Josas, France. [86]Department of Animal Science, University of Ljubljana, Groblje 3, Domzale, SI-1230, Slovenia. [87]Research Unit Molecular Biology, Research Institute for the Biology of Farm Animals (FBN), Dummerstorf, 18196, Germany. [88]Department of Animal Sciences, University of Wisconsin–Madison, 1805 Linden Drive, Madison, WI 53706, USA. [89]Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, N1G 2W1, Canada. [90]Cell Biology and Biophysics, European Molecular Biology Laboratory (EMBL)–Heidelberg, Meyerhofstrasse 1, Heidelberg, Germany. [91]Laboratory of Molecular Biology, Instituto Tecnologico Agrario de Castilla y Leon (ITACyL), Carretera de Burgos km 119, Valladolid, 47071, Spain. [92]Department of Dairy Science, University of Wisconsin, Madison, WI 53706, USA. [93]Monsanto Company, 3302 SE Convenience Blvd, Ankeny, IA 50021, USA. [94]Genes & Genetic Resources Molecular Analysis Lab, National Bureau of Animal Genetic Resources, Baldi Bye Pass, Karnal, Haryana, 132001, India. [95]Lacombe Research Centre, Agriculture and Agri-Food Canada, Lacombe, AL, T4L 1W1, Canada. [96]Biomedical Sciences, University of Guelph, Guelph, ON, N1G 2W6, Canada. [97]Zootechnics Institute, Università Cattolica del Sacro Cuore, via Emilia Parmense 84, Piacenza, 29100, Italy. [98]Department of Animal Science, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA. [99]Departamento de Ciências Biológicas, Faculdade de Ciências e Letras, UNESP–São Paulo State University, Av Dom Antônio 2100, Vila Tênis Clube, Assis, SP, 19806-900, Brazil. [100]Department of Basic Sciences and Aquatic Medicine, Norwegian School of Veterinary Science, Post Office Box 8146 Dep, Oslo, NO-0033, Norway. [101]Instituto de Genética Ewald Favret, Instituto Nacional de Tecnología Agropecuaria (INTA), Las Cabañas y de Los Reseros s/n CC25, Castelar, Buenos Aires, B1712WAA, Argentina. [102]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore, 117597, Singapore. [103]Grupo CENTAURO, Universidad de Antioquia, Medellín, Colombia. [104]Department of Chemistry, University of Illinois, Urbana, IL 61801, USA. [105]Eskitis Institute for Cell and Molecular Therapies, Griffith University, Nathan, QLD, 4111, Australia.

Downloaded from www.sciencemag.org on April 23, 2009

[106]Nutrition and Food Sciences, Utah State University, Logan, UT 84322, USA. [107]Animal Disease Research Unit, USDA–Agricultural Research Service, Pullman, WA 99164, USA. [108]Department of Pharmacology, 2-344 BSB, University of Iowa, 51 Newton Road, Iowa City, IA 52242, USA. [109]Department of Animal Science, 211 Terrill, University of Vermont, 570 Main Street, Burlington, VT 05405, USA.

Supporting Online Material
www.sciencemag.org/cgi/content/full/324/5926/522/DC1
Materials and Methods
Figs. S1 to S23
Tables S1 to S14
References

10 December 2008; accepted 16 March 2009
10.1126/science.1169588

Genome-Wide Survey of SNP Variation Uncovers 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.

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

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

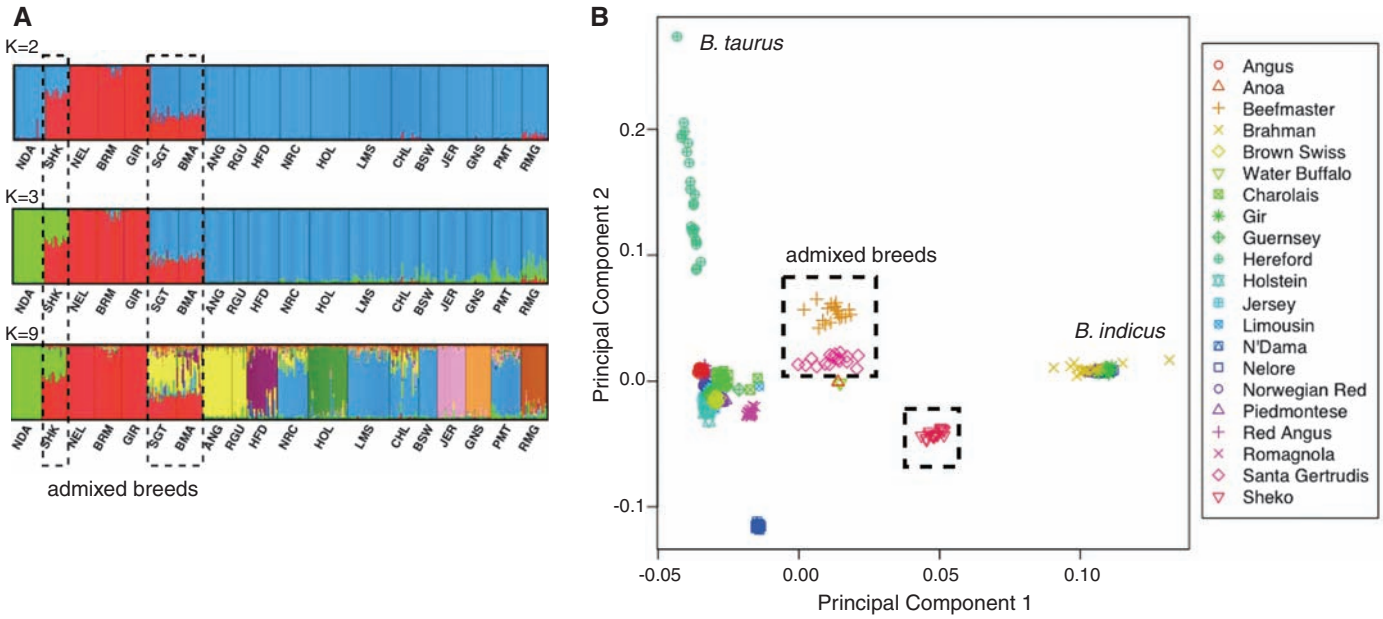

Fig. 1. (**A**) Population structure assessed by InStruct. Bar plot, generated by DISTRUCT, depicts classifications with the highest probability under the model that assumes independent allele frequencies and inbreeding coefficients among assumed clusters. Each individual is represented by a vertical 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; 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.

Downloaded from www.sciencemag.org on April 23, 2009