Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks

Ivica Medugorac[1], Alexander Graf[2], Cécile Grohs[3], Sophie Rothammer[1], Yondon Zagdsuren[4], Elena Gladyr[5], Natalia Zinovieva[5], Johanna Barbieri[6,7], Doris Seichter[8], Ingolf Russ[8], André Eggen[9], Garrett Hellenthal[10], Gottfried Brem[11], Helmut Blum[2], Stefan Krebs[2], Aurélien Capitan[3,12]

1 Chair of Animal Genetics and Husbandry, Ludwig-Maximilians-University Munich, Munich, Germany

2 Laboratory for Functional Genome Analysis, Gene Center, Ludwig-Maximilians-University Munich, Munich, Germany

3 GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

4 Mongolian Jak Society, Sum, Ulan Bator, Mongolia

5 Center of Biotechnology and Molecular Diagnostics of the L.K. Ernst Institute of Animal Husbandry, Podolsk District Moscow Region, Russia

6 Institut National de la Recherche Agronomique (INRA), UMR 1388 Génétique, Physiologie et Systèmes d'Elevage, GeT-PlaGe Genomic Facility, Castanet-Tolosan, France

7 Université de Toulouse INPT ENSAT, UMR1388 Génétique, Physiologie et Systèmes d'Elevage, Castanet-Tolosan, France

8 Tierzuchtforschung e.V. München, Grub, Germany

9 AgriGenomics, Illumina, San Diego, USA

10 Genetics Institute, Faculty of Life Sciences, University College London, UK

11 Institute of Animal Breeding and Genetics Department for Biomedical Sciences University of Veterinary Medicine, Vienna, Austria

12 ALLICE, Paris, France

Correspondence:

Ivica Medugorac
E-mail: ivica.medjugorac@gen.vetmed.uni-muenchen.de

Aurélien Capitan
E-mail: aurelien.capitan@inra.fr

The yak is remarkable for its adaptation to high altitude and occupies a central place in the economy of mountainous regions of Asia. At lower elevation, it is common to hybridize yaks with cattle to combine hardiness and productivity. Hybrid males are, however, sterile, preventing the establishment of stable hybrids but not a limited introgression after backcrossing several generations of females. Here we inferred bovine haplotypes in the genomes of 76 Mongolian yaks using high-density SNP genotyping and whole genome sequencing. These inherited ~1.3% of their genome from bovine ancestors after nearly continuous admixture over at least the last 1500 years. The introgressed regions were markedly enriched in genes involved in nervous system development and function, and particularly in glutamate metabolism and neurotransmission. A novel polled (i.e. hornless) mutation originating from Mongolian Turano cattle was also identified. Our results support that introgressive hybridization contributed to the improvement of yak management and breeding.

Hybridization is not unusual in nature. Although interspecific hybrids are rare at the population level, around 10% of animals and 25% of plants are known to occasionally hybridize with other species[1]. Evaluation of the genome-wide magnitude of this phenomenon has only recently become possible. The first results show that limited introgressions of the genome are widespread with a potentially important role in environmental adaptation, as suggested by incorporation of genetic material from local species into the genome of colonizing species (e.g. Neanderthal in non-African humans[2,3] or *Zea mays mexicana* in maize[4]). Successful analyses so far identified genes under selection but, because of some limitations, they were rarely able to determine precisely the nature of the selective pressure, to identify genes pathways under selection, and to pinpoint causative polymorphisms.

Yak and cattle diverged approximately 4.9 million years ago[5]. Despite anatomical and physiological differences, both species are raised in mixed herds in Central Asia and share similar husbandry practices. Recent studies have reported several examples of gene flow from cattle to yaks[6-8] and the existence of hornless (i.e. polled) animals in both species which do not carry the previously reported Celtic and Friesian *POLLED* mutations[9-12] (**Supplementary Table 1**) raises the question of a common origin for this phenotype. For this reason and because of the large genomic datasets available for cattle, analysis of bovine introgression in Mongolian yaks represents an appealing model to identify exchanges of traits of interest between domesticated species.

To get a first picture of bovine introgression in Mongolian yaks, we sequenced two individual genomes (YAK13, homozygous polled and YAK40, horned) and plotted the frequency of yak and bovine alleles for all positions of yak-specific SNPs in 70-kb sliding windows (**Supplementary Fig. 1**; see methods). Considering as yak-specific, the variants which were (i) homozygous for the alternate allele in the yak reference genome[5] (hereafter named YAKQIU) and (ii) absent from 235 bovine genomes (**Supplementary Table 2**)[13], we estimated that at least 1.73% and 1.22% of YAK13 and YAK40 genomes were of bovine origin (**Supplementary Fig. 1**). To identify introgression in YAKQIU itself, we used the number of yak-specific SNPs per 70-kb interval as an indicator and estimated a bovine proportion of 1.06% even in the yak reference sequence. To validate our WGS approach for inference of local-ancestry, two regions suggesting homozygous cattle introgression in YAKQIU were PCR amplified and sequenced in YAK13 and in 12 related species (**Supplementary Table 3**). Multiple alignment and phylogenetic analyses of 2,191-bp sequence data clustered YAK13 with gaur, banteng, bison and wisent in accordance with the phylogeny of the tribe Bovini[14] whereas YAKQIU clustered with cattle thus confirming cattle introgression in the yak reference genome too (**Fig. 1**; **Supplementary Table 4)**.

For a systematic analysis of cattle introgression in the Mongolian yak population, we investigated the Illumina BovineHD Beadchip genotyping data (777 k SNPs) of 76 animals originating from different localities (**Supplementary Table 1**). Analysis of SNPs mapping to mtDNA (N=245) identified two yaks with deviating matrilineal ancestors whereas analysis of SNPs mapping to BTAY (N=921) revealed an absence of bovine Y chromosome in this panel

(**Supplementary Fig. 2, 3**). We then applied a robust forward-backward algorithm (*RFMIX*)[2,15] to screen for the presence of cattle haplotypes in their autosomal genomes with the exception of the Major Histocompatibility locus (for reasons see **Supplementary Note**) using i) WGS data from three yaks to determine alleles present in yaks, ii) a six-bovini consensus to determine ancestral states for all SNPs, and iii) additional genotyping data from 384 cattle (**Supplementary Table 5**) as a reference panel assumed to harbour no yak ancestry (**Supplementary Fig. 4**).

The proportion of the genome inferred to be of cattle ancestry ranged between 0.67% and 2.82% (mean = 1.31 ±0.36; FDR=0.05) per animal (**Supplementary Table 6; Supplementary Note**), a result consistent with a severe restriction of introgression by the culling of most of the backcross calves and the persistence of hybrid male sterility up to the third or fourth generation of backcross[16]. In total, as high as 33.2% of the bovine genome was recovered from our panel of 76 yaks with noticeable variations between chromosomes (**Fig. 2a; Supplementary Tables 6 and 7**). In agreement with the "large X-effect" on hybrid male sterility (for a review see Presgraves[17]), BTAX was one of the least introgressed chromosomes and displayed the lowest medium and maximal sizes of introgressed segments.

Phylogenetic analysis revealed a close genetic relationship between the admixture source and the Turano-Mongolian cattle group (see Online Methods and **Supplementary Fig. 5**). Simulation results of one- and multiple-date admixture followed by segment retrieval by *RFMIX* supported nearly continuous admixture along the last 1500 years with a low proportion of cattle gametes (around 1/11000 per generation; **Fig. 2b; Supplementary Note**). While hybridization between yak and cattle was already a common practice 1800 years ago[16], we could not detect older admixture because of the limitation of the methods. Introgression was more intense during two periods (897-1121 and 1695-1828 CE), which coincides with the Medieval Climate Anomaly (900-1200 CE)[18] and the Dzungar–Qing Wars (1687–1758 CE)[19]. This is most probably because of increased mortality of livestock during these difficult times that have forced yak herders to breed all the females available to restore their herds, including backcross animals (**Supplementary Note**).

To identify phenotypes that have undergone positive selection, we next mined the gene content of 365 intervals defined as the smallest exogenous segment shared for each region showing introgression in at least 1% of the investigated haplotypes (**Supplementary Table 8**). Functional annotation of these 1311 transcripts using DAVID revealed a major enrichment for genes involved in sensory perception, cognition and neurological system processes (Benjamini corrected *P*-value <1.0E-8; **Supplementary Table 9**) which are known to be key domestication targets[20,21]. Furthermore, similar results were obtained with different thresholds on the percentage of introgression and size of intervals, indicating that selection on these genes, which most probably contributed to taming the ferocious temper of yaks, has been a regular and general process since the first hybridizations (**Supplementary Table 9**).

In total, we were able to retrieve 443 of such genes in 208 intervals after performing complementary gene set enrichment analyses and literature review (see Online Methods). These comprised genes related to nervous system development and function, synaptic transmission, sensory perception and a large variety of disorders impacting learning ability, social behavior, fear response and orientation in space in humans and animals (**Fig. 3a**; **Supplementary Tables 8-12**). Among them we should mention *ITGA9*, a susceptibility gene for bipolar affective disorder[22], which shows the highest level of introgression with 56% (85/152) of bovine alleles. We should also highlight the presence of nine genes from the glutamate receptor signaling canonical pathway including each of the four subtypes of receptors for this molecule which is the principal excitatory neurotransmitter in the brain[23] (**Fig. 3b,c**). Significantly enriched canonical pathways, according to Ingenuity Pathway Analysis, also include: (i) NAD biosynthesis from tryptophan and (ii) lysine degradation II & V which produce L-glutamate, (iii) the visual cycle involved in the sensory transduction of light

in the retina, (iv) Sphingosine-1-phosphate signaling which participates to neuromodulation[24], (v) neuropathic pain signaling in dorsal horn neurons and (vi) Huntington's disease (**Fig. 3b)**.

At the individual level, each yak carried numerous bovine genes involved in nervous system development and function (mean = 33.03 ±10.05; **Supplementary Table 6**), although most of them had moderate allele frequencies (median = 0.0461; **Supplementary Table 8**). Moreover, none of the genes we investigated exhibited deleterious mutations (**Supplementary Note**). These results are in line with previous studies which showed that affective disorders in humans and anxiety behaviours in different species have a polygenic basis and rely in part on the same genes (e.g.[25-29]). They further support our assumption that this specific gene enrichment observed in introgressed regions in yak is due to selection on behavioural traits.

Finally, with the exception of two regions encompassing *ABHD4* and *MYO6*, none of the 365 segments introgressed in our panel collocated with 182 recently reported signatures of domestication in yak[30] which confirms that introgressed segments constitute a source of favorable polymorphisms especially for genes which do not possess similar variants in yak. This is for example the case for a *KIT* duplication causing color-sidedness in cattle[7,8] which segregates in Mongolian yak (**Supplementary Fig. 6**) and presumably for a new polled mutation. To verify this hypothesis we modeled polledness as a quantitative trait in our panel (**Supplementary Fig. 7**) and mapped the locus to the beginning of chromosome 1 (P = 9.7-E9; 95% CI: 1.88-2.20-Mb; **Fig. 4a**) within a bovine introgressed segment (**Fig. 4b**; **Supplementary Table 8**). Between position 1,809,313 and 2,627,891-bp, we identified a total of 1,024 sequence variants which were homozygous in the homozygous polled YAK13 and absent from the horned YAK40. Nearly all of them were retrieved in the genome of one polled Turano Mongolian cattle (TM29), confirming the bovine origin of the polled mutation in yak.

Genotyping of twelve indels in 604 animals originating from two yak and 21 cattle subpopulations (**Supplementary Tables 1, 13, 14**) refined the polled locus interval to a 121-kb segment (1,889,854-2,010,574-bp) containing 238 variants. Contrasting these with the genomes of 234 bovines originating from Europe (**Supplementary Table 2**)[13], one horned Japanese Turano cattle[20], and TM29, we excluded all but two variants originating from the same microhomology-mediated break-induced replication event: a complex 219-bp duplication-insertion (P_{219ID}) beginning at 1,976,128-bp and a 7-bp deletion and 6-bp insertion (P_{1ID}) located 621-bp upstream of this position (**Supplementary Fig. 8,9**). This rearrangement results in the duplication of an 11-bp motif (AAAGAAGCAAA) which is entirely conserved among *Bovidae* (**Supplementary Fig. 10-13**) and which is also duplicated in the 80-kb duplication responsible for Friesian polledness[11]. Finally, genotyping of the P_{219ID}-P_{1ID} rearrangement in yaks and cattle revealed a perfect association with polledness of Turano-Mongolian origin, thus adding this polymorphism as third allele to the reported allelic heterogeneity at the *POLLED* locus (**Fig. 4c,d**)[9].

In conclusion, we present the first characterization of bovine introgression in yak at the genomic scale. We report (i) that Mongolian yaks inherited on average 1.31% of their genome from bovine ancestors after nearly continuous admixture over at least the last 1500 years and (ii) that these segments are significantly enriched in genes involved in nervous system development and function which most probably have contributed to the taming of yaks. We also show introgression of a new mutation which determines a phenotype of primary interest in bovine and yak husbandry: the genetic absence of horns. This study contributes to the emerging picture of the genes and pathways which have been the most affected by domestication and highlights the beneficial role played by introgressive hybridization in transferring favorable polymorphisms from one domestic species to another.

DATA AVAILABILITY:

Project accession codes (NCBI Sequence Read Archive (SRA)), PRJNA279385.

ACKOWLEDGMENTS

A.C. acknowledges P. Boudinot and D. Boichard for interesting discussions about the Major Histocompatibility Complex and the recombination rate in cattle respectively, M. Boussaha, Y. Djari and C. Klopp for introducing him to the use of *SAMTOOLS* and *PINDEL* softwares and M.-C. Deloche and C. Escouflaire for their punctual assistance. The authors acknowledge the Zoologischer Garten Berlin, Tierpark Cottbus and Tierpark Hellabrunn Munich (Germany), as well as the Jardin des Plantes de Paris (France) and all the breeders in Europe and Asia for generously providing samples and phenotypes. Apisgene is acknowledged for funding the AKELOS research project. LAFUGA is funded by the LMUexcellent program.

AUTHOR CONTRIBUTIONS

A.C. and I.M. conceived and coordinated the study. A.C., I.M. and S.K. designed the study. I.M. mapped the polled locus; performed introgression analysis using SNP Chip genotyping data; simulation analyses; and neighbor joining phylogenetic analyses. A.C. performed variant calling, annotation and screening for candidate mutations; analysis of sequence conservation; annotation of the gene content of the introgressed intervals; and gene set enrichment analyses. S.K., A.G. and I.M. performed introgression analysis based on WGS data; determination of ancestral alleles; genome and capture sequencing; and R-graphics. C.G., S.R., and A.C. performed PCR for Sanger sequencing and for genotyping by PCR and electrophoresis or PCR and Sanger sequencing. J.B. performed whole genome sequencing. D.S. and I.R. performed SNP chip genotyping and whole genome sequencing. Y.Z., E.G. and G.B. provided samples and phenotypes. H.B. provided sequencing and bioinformatics facilities. A.E. provided Illumina BovineHD SNP chip genotyping data. G.H. provided software and expertise in admixture analysis. A.C., I.M., S.K. and A.G. contributed to writing the manuscript.

COMPETING FINANCIAL INTEREST STATEMENTS:

The authors declare no competing financial interests.

REFERENCES

1. Mallet, J. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**, 229-237 (2005).
2. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354-357 (2014).
3. Vernot, B. & Akey, J.M. Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science* **343**, 1017-1021 (2014).
4. Hufford, M.B. *et al.* The Genomic Signature of Crop-Wild Introgression in Maize. *Plos Genetics* **9**, e1003477 (2013).
5. Qiu, Q. *et al.* The yak genome and adaptation to life at high altitude. *Nature Genetics* **44**, 946-949 (2012).
6. Qi, X.B., Jianlin, H., Wang, G., Rege, J.E.O. & Hanotte, O. Assessment of cattle genetic introgression into domestic yak populations using mitochondrial and microsatellite DNA markers. *Animal Genetics* **41**, 242-252 (2010).

7. Zhang, M.Q., Xu, X. & Luo, S.J. The genetics of brown coat color and white spotting in domestic yaks (Bos grunniens). *Animal Genetics* **45**, 652-659 (2014).
8. Durkin, K. *et al.* Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature* **482**, 81-84 (2012).
9. Medugorac, I. *et al.* Bovine polledness-an autosomal dominant trait with allelic heterogeneity. *PLoS ONE* **7**, e39477 (2012).
10. Allais-Bonnet, A. *et al.* Novel Insights into the Bovine Polled Phenotype and Horn Ontogenesis in Bovidae. *Plos One* **8**, e63512 (2013).
11. Rothammer, S. *et al.* The 80-kb DNA duplication on BTA1 is the only remaining candidate mutation for the polled phenotype of Friesian origin. *Genetics Selection Evolution* **46**, 44 (2014).
12. Liu, W.B. *et al.* Associations of single nucleotide polymorphisms in candidate genes with the polled trait in Datong domestic yaks. *Animal Genetics* **45**, 138-141 (2014).
13. Daetwyler, H.D. *et al.* Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet* **46**, 858-65 (2014).
14. Decker, J.E. *et al.* Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18644-18649 (2009).
15. Maples, B.K., Gravel, S., Kenny, E.E. & Bustamante, C.D. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *American Journal of Human Genetics* **93**, 278-288 (2013).
16. Wiener, G., Jianlin, H. & Ruijun, L. *The yak*, (FAO Regional Office for Asia and the Pacific, 2003).
17. Presgraves, D.C. Sex chromosomes and speciation in Drosophila. *Trends in Genetics* **24**, 336-343 (2008).
18. Adhikari, D. & Kumon, F. Climatic changes during the past 1300 years as deduced from the sediments of Lake Nakatsuna, central Japan. *Limnology* **2**, 157-168 (2001).
19. Perdue, P.C. *China marches west: the Qing conquest of central Eurasia*, (Harvard University Press, 2009).
20. Carneiro, M. *et al.* Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074-1079 (2014).
21. Schubert, M. *et al.* Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E5661-E5669 (2014).
22. Secolin, R. *et al.* Refinement of chromosome 3p22.3 region and identification of a susceptibility gene for bipolar affective disorder. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics* **162B**, 163-168 (2013).
23. Snyder, S.H. & Ferris, C.D. Novel neurotransmitters and their neuropsychiatric relevance. *American Journal of Psychiatry* **157**, 1738-1751 (2000).
24. Ghasemi, R., Dargahi, L. & Ahmadiani, A. Integrated Sphingosine-1 Phosphate Signaling in the Central Nervous System: From Physiological Equilibrium to Pathological Damage. *Pharmacological research* **104**, 156-164 (2016).
25. Domschke, K. & Reif, A. Behavioral Genetics of Affective and Anxiety Disorders. in *Behavioral Neurogenetics* (eds. Cryan, F.J. & Reif, A.) 463-502 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012).
26. Nurnberger, J.I. *et al.* Identification of Pathways for Bipolar Disorder A Meta-analysis. *JAMA psychiatry* **71**, 657-664 (2014).
27. Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* **38**, 879-887 (2006).
28. Johnsson, M., Williams, M.J., Jensen, P. & Wright, D. Genetical Genomics of Behavior: A Novel Chicken Genomic Model for Anxiety Behavior. *Genetics* **202**, 327-340 (2016).
29. Zapata, I., Serpell, J.A. & Alvarez, C.E. Genetic mapping of canine fear and aggression. *BMC Genomics* **17**, 572 (2016).

30. Qiu, Q. *et al.* Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nature Communications* **6**, 10283 (2015).

FIGURE LEGENDS

Figure 1. Phylogenetic analyses of sequence data on chromosomes 9 and 25 confirm bovine introgression in the individual sequenced for generating the yak reference genome. a) Bovine introgression plot based on WGS data. Blue and red dots show the relative frequencies of homozygous and heterozygous genotypes for yak-specific polymorphisms from the sequencing results used for generating the yak reference genome. Each dot represents the number of genotypes in a 70-kb sliding window divided by expected (genome-wide) number of polymorphisms in 70-kb windows. Background color of each interval is shaded according to read depth, ranging from 0 (dark gray) to >40 (white). Genotype frequencies from white shaded intervals are likely to be affected by artefacts from repeat expansions in yaks, and those from dark shaded intervals by poor mappability or deletions. Results with neutral grey background are regarded as more robust. Introgressed intervals are identified by a break in the red line (circular binary segmentation of mean homozygous genotype frequency) and a drop in the frequency of homozygous genotypes for yak-specific alleles. Heterozygous genotypes for yak-specific alleles and the yellow line (circular binary segmentation of mean heterozygous genotype frequency) serve as a control to distinguish homozygous versus heterozygous cattle introgression. These statistics suggest homozygous cattle introgression in two regions (BTA09:68.495-70.115-Mb and BTA25:17.345-19.995-Mb) in the individual sequenced for generating the yak reference genome. b) Details of five exons and flanking sequences from four genes (*L3MBTL3*, *SAMD3*, *ACSM2B* and *MGC134577*) which were sequenced in 14 *Bovini* animals for validation of our introgression analysis based on WGS data (**Supplementary Table 4**). c) Neighbor-joining phylogeny of 14 haplotypes representing yaks, cattle and 10 *Bovidae* species, supporting homozygous cattle introgression in the reference yak genome. This analysis was based on sequence data from the five regions of chromosome 9 and 25 presented above and totalizing 2,191 nucleotides. The reliability of the tree branches was tested by 1,000 bootstrap replicates.

Figure 2. Analysis of the size distribution of introgressed intervals reveals three major introgressions events. a) Minimal, maximal, average and median length of introgressed intervals on each of the 30 chromosomes (x-axis) over 76 yak animals genotyped with the Illumina BovineHD SNP chip. The genome-wide average and median lengths are presented by green and red dotted lines, respectively. b) Distribution of the size of the bovine DNA segments introgressed into the yak genomes as estimated by our *RFMIX* procedure. Absolute counts of fragments observed in (i) all 76 yak animals (green curve); (ii) 26 yaks sampled in Mongolia (black curve); (iii) 50 yaks of Mongolian descent sampled in Europe (blue curve); (iv) simulated three-date admixture in 76 de-introgressed yaks with a proportion of cattle DNA of 0.0005 at 250, 0.011 at 150 and 0.0045 at 37 generations ago (orange curve); and (v) continuous admixture with a proportion of cattle DNA of 0.00045 every 5 generations in a period between 40 to 220 generations ago (red curve) were divided by the number of considered haploids in each of the four groups. The size of the introgressed segments detected varied between 108-kb and 24.63 Mb with a median length of 601-kb. Ten Mb (Chr23:22.0 to 32.0 Mb) comprising the MHC region were not considered in this distribution. The figure presents intervals up to a maximal length of 5,000-kb. Longer intervals had frequencies of 0 or 1% and are not all shown here for reasons of clarity.

Figure 3. Bovine introgressed segments show a major enrichment for genes related to nervous system development and function. Word cloud illustrating the major enrichment for genes related to nervous system development and function, behavior, neurological diseases and psychological disorders revealed by the Ingenuity Pathway Analysis. A total of 1311 genes associated to 365 intervals showing at least 1% (i.e. 2 alleles) of bovine genome introgression in our panel of 76 yaks (**Supplementary Table 8**) were considered for the Ingenuity Pathway Analysis (IPA). A unique keyword was attributed to each significantly enriched pathway in the "Diseases and Bio Function analysis" according to the "Diseases or Functions Annotation" (see Online Methods). Keywords referring to ubiquitous cellular or organismal processes are not represented to not overload the cloud. Font size is proportional to the number of occurrence of the keywords. b) Venn-like diagram presenting the canonical pathways which are significantly enriched (P<0.01) in introgressed segments according to IPA and details on associated genes (**Supplementary Table 11**). These consist in five pathways related to nervous system development, function or pathologies and to two pathways resulting in the production of L-glutamate which is the principal excitatory neurotransmitter in the brain. c) Localization at the neuron synapse level of the main proteins belonging to the glutamate receptor signaling canonical pathway (adapted from IPA). Proteins encoded by genes listed in (b) are highlighted in pink. GRIA4: Glutamate Receptor, Ionotropic, AMPA 4; GRIK3: Glutamate Receptor, Ionotropic, Kainate 3; GRIN2A: Glutamate Receptor, Ionotropic, NMDA 2A; GRIN3A: Glutamate Receptor, Ionotropic, NMDA 3A; GRIP1: Glutamate Receptor Interacting Protein 1; GRM4: Glutamate Receptor, Metabotropic 4; CAMK4: Calcium/Calmodulin-Dependent Protein Kinase IV; DLG4: Discs, Large Homolog 4 (Drosophila). NMDA: N-methyl-D-aspartate; AMPA: α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid. CALM belonging to the Calcium/Calmodulin-Dependent Protein Kinase group and PSD-95, the postsynaptic density protein 95, are respectively encoded by CAMK4 and DLG4. EPSPs: Excitatory Postsynaptic Potentials. Note the presence of each of the four subtypes of glutamate receptors (Ionotropic, AMPA; Ionotropic, Kainate; Ionotropic, NMDA; and Metabotropic) among these proteins.

Figure 4. Introgression of a novel and complex mutation at the *POLLED* locus from bovines causes polledness in Mongolian yaks. a) Mapping of the *POLLED* locus with Illumina BovineHD Beadchip genotyping data from 36 polled and 40 horned animals and polledness modeled as a quantitative trait. **b**) Bovine introgression plot based on WGS data. Orange and turquoise dots show the mean frequency of bovine and yak alleles at positions of yak specific variants in a 70-kb sliding window. Background color of each interval is shaded according to read depth ranging from 0 (dark gray) to >40 (white). Introgressed intervals are identified by a break in the red line (circular binary segmentation of mean allele frequency) and a drop of yak-allele frequency below 0.5. Note that YAK13 is homozygous for a bovine introgressed segment encompassing the mapping interval of the polled locus (P = 9.7-E9; 95% CI: 1.88-2.20-Mb). This result is independently supported by a reduction of the divergence of YAK13 genome sequence from the UMD3.1 bovine reference sequence in the polled region (0.28 %) between position 1,809,313 and 2,627,891-bp as compared with the average divergence of 1.08 % exhibited by both yaks at the genome level. **c**) Scheme presenting the nature and location of the three different mutations identified at the polled locus in bovine as compared with the wild type allele. Red boxes represent the segments that are duplicated in the Celtic, Friesian and Mongolian alleles whereas light and dark grey boxes represent the original segments. Note that none of the three polled mutations affect coding regions and that the molecular mechanism underlying polledness remains unknown at the present time. **d**) Details of the complex Polled Mongolian mutation which results in the duplication of a 11-bp motif which is entirely conserved among *Bovidae* and well conserved among vertebrates (**Supplementary Fig. 10-13**). Boxes from different colors are used to show segmental duplications.

393

394

ONLINE METHODS

Animals

In total 120 yaks, 1025 cattle from a wide diversity of breeds originating from Eurasia and Africa, as well as representatives of nine other bovid species were considered in at least one of the analyses performed in this study. Briefly, they consist of animals used for mapping of POLLED locus in yak and Mongolian turano cattle (**Supplementary Table 1** and **Supplementary Fig. 7**); sets of whole genome sequences of yak and cattle (**Supplementary Table 2**) used for introgression analyses (**Supplementary Fig. 1**) and filtering of candidate mutations; bovid species used for target sequencing and phylogenetic analyses (**Supplementary Tables 3** and **4**); and sets of Illumina BovineHD chip genotypes used for admixture and mapping analyses (**Supplementary Table 5**).

Horn/polled phenotypes and derived genotypes

The polled phenotype is an autosomal dominant trait in cattle[31] and yak[12], readily measurable on any animal older than six months. Artificial dehorning of yak and cattle is not practiced in the sampling area in Central Asia. Therefore, any polled yak descending from one polled and one horned parent is necessarily heterozygous, i.e. Pp, at the underlying POLLED locus. One horned offspring with confirmed paternity is sufficient to declare a polled parent as Pp. Animals having two polled parents and consecutively ten or more polled offspring with horned mates are declared as homozygous polled PP. Similar animals having less than ten offspring (all polled) with horned mates are either PP or Pp and were declared as P•. Finally, all horned animals are pp. Derived genotypes of yak animals used for mapping are presented in **Supplementary Fig. 7**.

Whole genome sequencing of two Mongolian yaks and one Mongolian Turano cow.

The genomes of one heterozygous polled Turano-Mongolian cow (TM29), one homozygous polled yak (YAK13) and one horned yak (YAK40) were sequenced with Illumina technology. Paired-end libraries were generated according to the manufacturer's instructions using the Rapid DNA library system (NuGen, San Carlos, USA) for animal TM29 and the NEXTflex PCR-Free DNA Sequencing Kit (Biooscientific) for YAK13 and YAK40. Libraries were quantified using the KAPA Library Quantification Kit (Cliniscience), controlled on a High Sensitivity DNA Chip (Agilent) and sequenced on an Illumina HiSeq1500 with 2*110-bp read length (TM29) or on a HiSeq 2000 with 2*101-bp read length (YAK13 and YAK40). The average sequence coverage was 8.7, 13.4 and 14.9 x, respectively. Reads were mapped on the UMD3.1 bovine sequence assembly using BWA[32]. Reads with multiple alignments were removed. SNPs and small indels were called using *SAMTOOLS* pileup option[33]. Only variants with a quality score (QUAL) of >=30 and a mapping quality (MQ) score of >=30 were kept. Discovery of larger indels was achieved with *PINDEL*[34]. Variants supported by only one read or found in the homozygous state in the three animals were considered as possible artifacts and eliminated. Detection of Copy Number Variation was performed according to Medugorac *et al.*[9] by calculating coverage ratios between pairs of individuals in dynamical bin sizes of 5000 reads. YAK13 and YAK40 were each compared to three cattle WGS sequences in

order to call possible CNVs in introgressed regions. Significant CNV results were kept if they overlapped in all three comparisons of one YAK and three cattle WGS. Signals that were caused by mapping of apparently repetitive sequences getting very high coverage (>1000 fold) were manually removed. The reliability and borders of the retained CNVs were verified using the Integrative Genomics Viewer (IGV)[35] and paired-end information. In the end, only one polymorphism was considered as a true introgressed CNV: the mutation responsible for color-sidedness in bovines that is presented in **Supplementary Fig. 6**. For the latter, the log2 ratios of sequence coverages per 5000-bp window between the solid colored YAK40 and the color-sided YAK13 were plotted using R and the average ratios were segmented using the circular binary segmentation implementation in the DNAcopy package (v1.14.0) from Bioconductor.

Introgression analysis in WGS data of three yaks and one Mongolian Turano cow.

The detection of bovine genome segments in WGS data of YAK13, YAK40 and the reference genome[5] (YAKQIU) was conducted as follows. First, variants which were homozygous for the alternate allele in YAKQIU and absent from TM29 and 234 additional bovine genomes[13] were identified and considered as yak-specific. Then, the mean frequency of yak and bovine alleles for each of these variants was estimated for sliding windows of 70-kb along the genomes of YAK13, YAK40, and TM29. This window size corresponds to half of the expected mean size of segments which would have been introgressed from the earliest possible hybridization between cattle and yak (**Supplementary Note**). To detect introgression in the yak reference sequence itself, the number of homozygous and heterozygous genotypes for yak-specific polymorphisms were estimated in 70-kb windows and compared to the expected numbers based on genome-wide observations (see Figure 1). Introgressed intervals were identified by circular binary segmentation (CBS) of mean allele frequency and a drop of yak-allele frequency below 0.5. CBS is implemented in the R-package DNAcopy (v1.14.0) from Bioconductor. Frequencies of yak-specific alleles in both, TM29 and YAKQIU, served as a control.

Conventional Sanger sequencing of target genomic regions with cattle ancestry in reference yak genome.

Two regions suggesting homozygous cattle introgression in the yak reference genome[5] (Chr9:68,495,000-70,115,000 and Chr25:17,345,000-19,995,000; **Fig.1** and **Supplementary Fig. 1**) were selected to test the reliability of our approach. For each region, two PCR products were amplified in 13 animals representing 12 bovid species (**Supplementary Table 3 and 4)**. PCR was performed using the Go-Taq Flexi system (Promega) according to the manufacturer's instructions on a Mastercycler pro thermocycler (Eppendorf). Amplicons were purified and bidirectionally sequenced by Eurofins MWG using conventional Sanger sequencing. The resulting sequences were aligned with the corresponding sequences from the yak reference genome using the *CLUSTALW* (http://www.clustal.org) software[36] which is part of the *MEGA* software[37] package version 6.06 (http://www.megasoftware.net). Then they were trimmed to get equal lengths for most animals and fragments. Finally, phylogeny was inferred based on a total of 2191 nucleotides of sequence using the Neighbor-Joining methods[38] implemented in *MEGA* software[37]. The percentage of replicate trees in which the associated taxa clustered together was determined by the bootstrap test[39] (1,000 replicates).

A similar approach was used to study the MHC locus and to estimate the false discovery rate of the *RFMIX* analysis, as described in **Supplementary Note**.

Analysis of Illumina HD genotypes

General information. Illumina BovineHD BeadChip genotypes from 467 *Bovidae* animals were considered. These consisted of 76 yaks (36 polled and 40 horned; **Supplementary Fig. 7**), a panel of 384 individuals representative of the world wide diversity of cattle and assumed to harbor no yak ancestry (**Supplementary Note**, **Supplementary Table 5**), and representatives of six *bovini* species (two gaur, one wood bison, one European bison, one banteng, one water buffalo and one nilgai; **Supplementary Table 3**). Of note, the panel of 384 cattle comprised 11 polled and 14 horned turano cattle from Mongolian and Yakutian breeds (**Supplementary Tables 1** and **5**). A total of 697,172 SNP markers were successfully genotyped in three to six *bovini* species. Only 42,230 SNP (5.43%) were informative in yaks. Haplotypes were inferred and missing genotypes imputed using hidden Markov models (software package *BEAGLE*)[40] and three cohort types; namely trios (two parents, one offspring), pairs (one parent, one offspring) and unrelated animals. Marker order was based on release UMD3.1 of the *Bos taurus* genome (http://www.cbcb.umd.edu/research/bos_taurus_assembly.shtml).

Inferring maternal and paternal phylogenies. To avoid artifacts, only SNP from the mitochondrial genome and Y chromosome showing high call rates (>99%) and complete homozygosity within each single animal (n= 245/343 and 921/1224 markers respectively) were used. Moreover, only animals with less than 5% of missing genotypes for mitochondrial or Y chromosome markers were considered. Maternal and paternal phylogenies were constructed with the Neighbor-Joining methods[38] implemented in *MEGA* software[37] version 6.06. The percentage of replicate trees in which the associated taxa clustered together was determined by the bootstrap test[39] (1,000 replicates).

Introgression analysis in 76 Mongolian yaks. Yak-specific alleles were inferred from homozygous SNPs located in genomic regions of YAK13, YAK40 or YAKQIU that are free of cattle ancestry, based on previous analyses of WGS data (**Supplementary Fig. 1**). Since WGS data didn't provide clear introgression status in two specific region (Chr22:31,682,450-31,842,000 and Chr23:24,661,105-29,153,851; **Supplementary Fig. 1**), we used genotypes of 76 yaks to define the major allele (frequency≥0.90) as yak-specific. For all remaining SNPs (<1.00%) the major allele (frequency ≥0.75) in six *bovini* species was considered as ancestral and yak-specific. Then, a rapid and robust forward-backward algorithm implemented in the software package *RFMIX*[2,15] was used to screen for the presence of cattle haplotypes in 76 Mongolian yaks. This algorithm uses designated reference haplotypes to infer local ancestry in designated admixed haplotypes which supposes to include pure yak and pure cattle in the analysis. Since there is no genetic and historical support for introgression of yak into cattle, we considered the 384 cattle (**Supplementary Table 5**) as a reference panel assumed to harbor no yak ancestry.

On the other hand, we were not able to find a complete yak genome without cattle ancestry but we detected complete chromosomes or large chromosomal fragments with pure yak ancestry. These chromosomal regions as well as yak-specific alleles were used to create a

526 synthetic pure yak genome (YAKYAK) which served as reference yak in initial *RFMIX*
527 analyses.[2]

528 For each chromosome we started two rounds of *RFMIX* analyses. The first round used 384
529 cattle genomes as reference cattle population and only YAKYAK as reference yak
530 population. The admixed sample consisted of all 76 yak genomes. For each chromosome,
531 initial *RFMIX* analyses detected different subsets of yak haploids as pure. These pure yak
532 chromosomes supplemented the YAKYAK chromosome in the second round of *RFMIX*
533 analyses to produce final results for a specific chromosome.

534 The *RFMIX* program performs forward-backward analyses in non-overlapping windows of
535 predefined size. In some situations, like for short segments in an unfavorable location
536 (window transition), *RFMIX* occasionally detected signatures only in the more informative half
537 or even in none of the two windows. To deal with these problems, we set the window size at
538 0.2 cM and performed four overlapping *RFMIX* analyses (**Supplementary Fig. 4)**.

539 *Source, date and number of admixture events. CHROMOPAINTER*[41] was used to decompose
540 the chromosomes of each of the 76 Mongolian yaks as a series of haplotypic chunks inferred
541 to be shared with at least one of the 384 cattle representing 24 breeds. In theory, given a
542 single admixture event, ancestry chunks inherited from each source have an exponential size
543 distribution, resulting in an exponential decay of these coancestry curves[41,42]. The shape of
544 decay curve in different groups enable the estimation of admixture dates[42] and the
545 determination of recipient and donor groups involved in asymmetric admixture events.
546 Multiple admixture times result in a mixture of exponentials[42]; which can be tested by
547 comparing the fit of a single exponential decay rate versus a mixture of rates. Inferences of
548 the haplotypic makeup of admixing source groups as well as of the admixture date(s) were
549 carried out using the *GLOBETROTTER*[42,43] method and complemented by simulation studies
550 described in **Supplementary Note**.

551 The inference of the source of admixture was complemented by phylogenetic analyses of
552 pure and admixed haploids. For 139 chromosomal segments introgressed in ten or more
553 yaks (**Supplementary Table 8**), ten pure and ten introgressed haploids were randomly
554 selected to constitute two yak groups. Similarly, 20 cattle groups representing 20 breeds with
555 four or more animals genotyped with the BovineHD chip were constructed (**Supplementary**
556 **Table 5)**. These segments were divided into a total of 3076 non-overlapping blocks
557 comprising four SNPs for which the inter-marker distance was less than 25 kb with
558 neighbouring SNPs. Each block was considered as a multi-allelic marker in phylogenetic
559 analysis to reduce ascertainment bias [44]. The proportion of shared alleles between
560 individuals, *PS*[45], was converted to genetic distances ($D_{PS} = ln(PS)$). A neighbor-joining tree
561 (**Supplementary Figure 5**) reflects the averaged individual distances between groups and
562 was constructed with the *SPLITSTREE4* program[46].

563

564 **Annotation of the gene content of the introgressed segments**

565 For 365 regions showing a minimum of 2 introgressed segments among the 76 animals
566 studied, we defined the smallest portion shared by these segments. To assess the gene
567 content of the resulting intervals, we used the “Refseq Genes” track from the UCSC Genome
568 Browser (http://genome.ucsc.edu) as a primary resource. We also used the “Non-Cow
569 Refseq Genes”, “Cow mRNAs from Genbank”, and “Cow ESTs that have been spliced

tracks" to recover protein genes that may have been missed during annotation of the UMD3.1 bovine sequence assembly. These consisted in genes annotated in at least human and mouse with no bovine RNA alignments in the orthologous region or genes with bovine RNA alignments corresponding to at least one gene annotated in human or mouse in the orthologous region. Intervals which did not contain genes were attributed the name of the closest gene in 5'3' orientation and located at a maximum of 500kb downstream of its borders according to the same orientation.

Gene set enrichment analysis

Gene set enrichment analyses were carried out with five software using different methods and source of information, i.e. Gene Ontology classes for DAVID 6.7 (http://david.abcc.ncifcrf.gov/) and PANTHER (http://pantherdb.org/), bibliographic and experimental data for Genetrail2 (http://genetrail2.bioinf.uni-sb.de/) and Ingenuity Pathway Analysis (http://www.ingenuity.com/products/ipa/), and Mammalian Phenotype ontology (level 3) from Mouse Genome Informatics for the specific analysis we performed with Enrichr (http://amp.pharm.mssm.edu/Enrichr/)[47,48]. Since these analyses produced comparable results, and for the sake of simplicity, only two of them were selected to be presented in this study. To provide a first overview of the overrepresented groups of genes, and to test their reliability, we performed different Gene Ontology (GO) term enrichment analyses were performed with DAVID using different lists of genes located in chromosomal regions detected as introgressed from cattle to yaks by *RFMIX* analyses (results are presented in **Supplementary Table 8**). Then we used Ingenuity Pathway Analysis for the precision of its annotations. We focused on the "Top Canonical Pathways" and on the "Diseases and bio Functions". Only canonical pathways or annotations with a *P*-value $<10^{-2}$ were retained. Annotations related to cancer and drug metabolism which were not relevant for this study were not considered. In addition to the IPA annotations we attributed a unique keyword to each significantly enriched pathway according to the "Diseases or Functions Annotation" in order to draw a word cloud. A particular attention was paid to attribute keywords related to subcellular portions, cell types and organs rather than to general processes. Keywords which appeared only once were finally regrouped with higher order items (e.g. celltype changed for organ or process changed for the category defined by IPA) or with the predefined IPA "categories" (results are presented in **Supplementary Table 10**). Finally, whereas they are not presented in detail, results from the three other analyses were used to complete the list of genes involved in nervous system development and function presented in **Supplementary Tables 8 and 12**.

Mapping of the polled locus in yaks sampled in Europe and Mongolia.

Mapping of the polled locus was performed using a combined linkage disequilibrium and linkage analysis (*cLDLA*) with horn status modelled as a quantitative trait (*pp*=0, *Pp*=1, *PP*=2 and *P•*=1.5). Genomic relationship matrix (**G**)[49] was estimated and its inverse ($\mathbf{G}^{-1}$) used to correct for population structure and possible polygenic effects in the model of the later QTL mapping. Identical-by descent (IBD) probabilities between pairs of haplotypes[50] were estimated for sliding windows of 40 SNP and summarized into a diplotype relationship matrix ($\mathbf{D_{RM}}$) which is computed in a similar way to the additive genotype relationship matrix ($\mathbf{G_{RM}}$)[51]. *cLDLA* mapping of polledness was carried out with a procedure similar to that reported in

Meuwissen *et al.*[52], which considers random QTL and polygenic effects. Variance component analysis in the middle of each of the 40-SNP sliding windows was performed by the *ASREML* package (https://www.vsni.co.uk/downloads/asreml/release3/UserGuide.pdf) and a mixed linear model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z_1 u} + \mathbf{Z_2 q} + \mathbf{e}$$

where **y** is a vector of horn status, **ß** a vector of fixed effects (including overall mean μ), **u** is a vector of n random polygenic effects for each animal with $\mathbf{u} \sim N(0, \mathbf{G}\sigma_u^2)$, **q** a vector of random additive genetic effects due to *POLLED* locus with $\mathbf{q} \sim N(0, \mathbf{D_{RMp}}\sigma_q^2)$, where $\mathbf{D_{RMp}}$ is the diplotype relationship matrix at position *p* of the putative *POLLED* locus, and **e** a vector of random residual effects with $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where **I** is an identity matrix. The random effects **u**, **q** and **e** were assumed to be uncorrelated and normally distributed and their variances (σ_u^2, σ_q^2, σ_e^2) were simultaneously estimated using *ASREML*.

Using the logarithm of the likelihood estimated by *ASREML* for the model with ($logL_P$) and without *POLLED* locus effects ($logL_0$; corresponding to the null hypothesis), we calculated the likelihood ratio test statistic ($LRT = -2 (logL_0 - logL_P)$, which is known to be χ^2-distributed with one degree of freedom[53]. Appropriately, an *LRT*-value higher than 10.8 was considered statistically significant (equivalent to $P < 0.001$).

Fine mapping and identification of the Mongolian POLLED mutation.

The first step consisted of selecting sequence variants which were homozygous in the homozygous polled YAK13, absent from the horned YAK40, and located between positions 1,809,313 and 2,627,891-bp on chromosome 1. This region comprises the 95% confidence interval (1.88-2.20-Mb) obtained with the QTL mapping approach and corresponds to a bovine chromosomal segment introgressed in yak (**Fig. 4b**; **Supplementary Table 8**). Then, to narrow down the candidate region, 120 yaks and 484 Eurasian taurine cattle (**Supplementary Table 1**) were genotyped for twelve indels using standard PCR, and agarose gel or capillary (ABI PRISM® 377 and 3100 Genetic Analyzer, Applied Biosystems) electrophoresis (**Supplementary Tables 13 and 14**). Of note, the same animals were also genotyped for the Celtic (P_{202ID}) and Friesian (P_{80kbID}) polled mutations[9-11] and clearly excluded these as possible candidates for polledness of Mongolian origin. Genotyping for twelve indels (**Supplementary Table 13 and 14**) excluded all but two indels (LMP04 and LMP12) as candidate mutation, and haplotype analyses reduced the polled locus interval to a 121-kb segment (1,889,854-2,010,574bp) containing 238 variants. Considering that the Mongolian polled mutation occurred in Turano cattle and is absent even in European polled cattle (**Supplementary Note**), these variants were subsequently filtered to retain only those which were heterozygous in the heterozygous polled Mongolian Turano cattle TM29, and absent in the genomes of one horned Japanese Turano cattle[20] and 234 bovines originating from Europe (**Supplementary Table 2**)[13].

Finally, to ensure that we did not miss any candidate variants for polledness, we performed two independent verifications. We performed a new detection of structural variants in the refined 121-kb polled interval using DELLY[54] and a visual examination of the whole genome sequences of YAK13, YAK40 and TM29 in the same interval using IGV[35]. We did not detect

new candidate polymorphisms. Considering that there is no gap in the UMD3.1 bovine genome sequence assembly and in the WGS of the homozygous polled yak (YAK13) in this interval we can claim that we did not miss any candidate variant with our approach.

Analysis of sequence conservation around the Mongolian *POLLED* mutation in Mammals.

Regions orthologous to the segments duplicated in the Mongolian *POLLED* mutation were retrieved for 34 eutherian mammals using the EPO multiple alignment from ENSEMBL. A consensus sequence and a sequence logo were generated using *MULTALIN* (http://multalin.toulouse.inra.fr/multalin/)[55] and *WEBLOGO* (http://weblogo.berkeley.edu/)[56], respectively. After the identification of a well-conserved 11-bp motif, a novel consensus sequence and a novel sequence logo were generated. Details on the 11-bp orthologous sequences are presented in **Supplementary Fig. 11 and 12**.

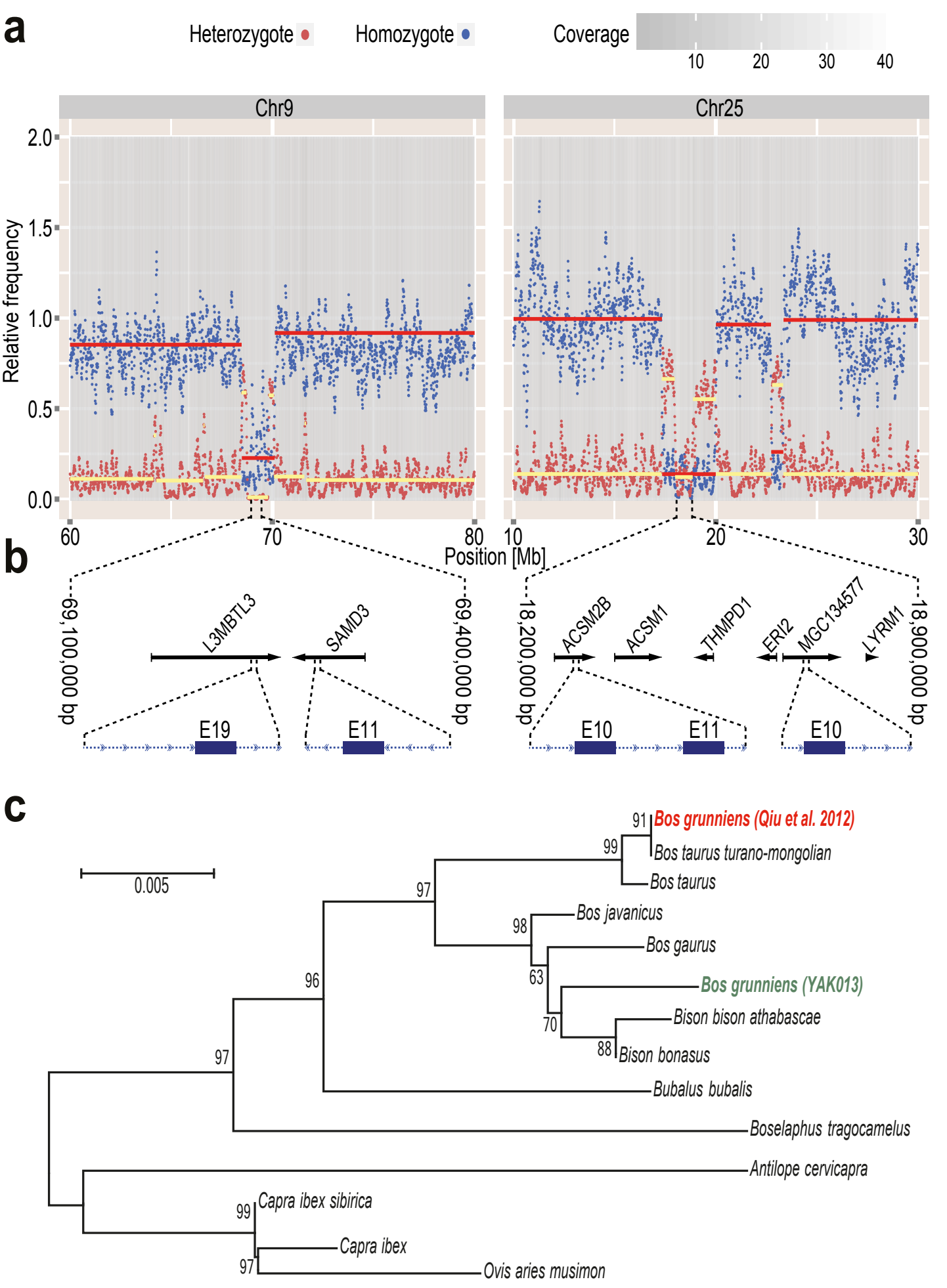
Analysis of sequence conservation around the Mongolian *POLLED* mutation in *Bovidae*.

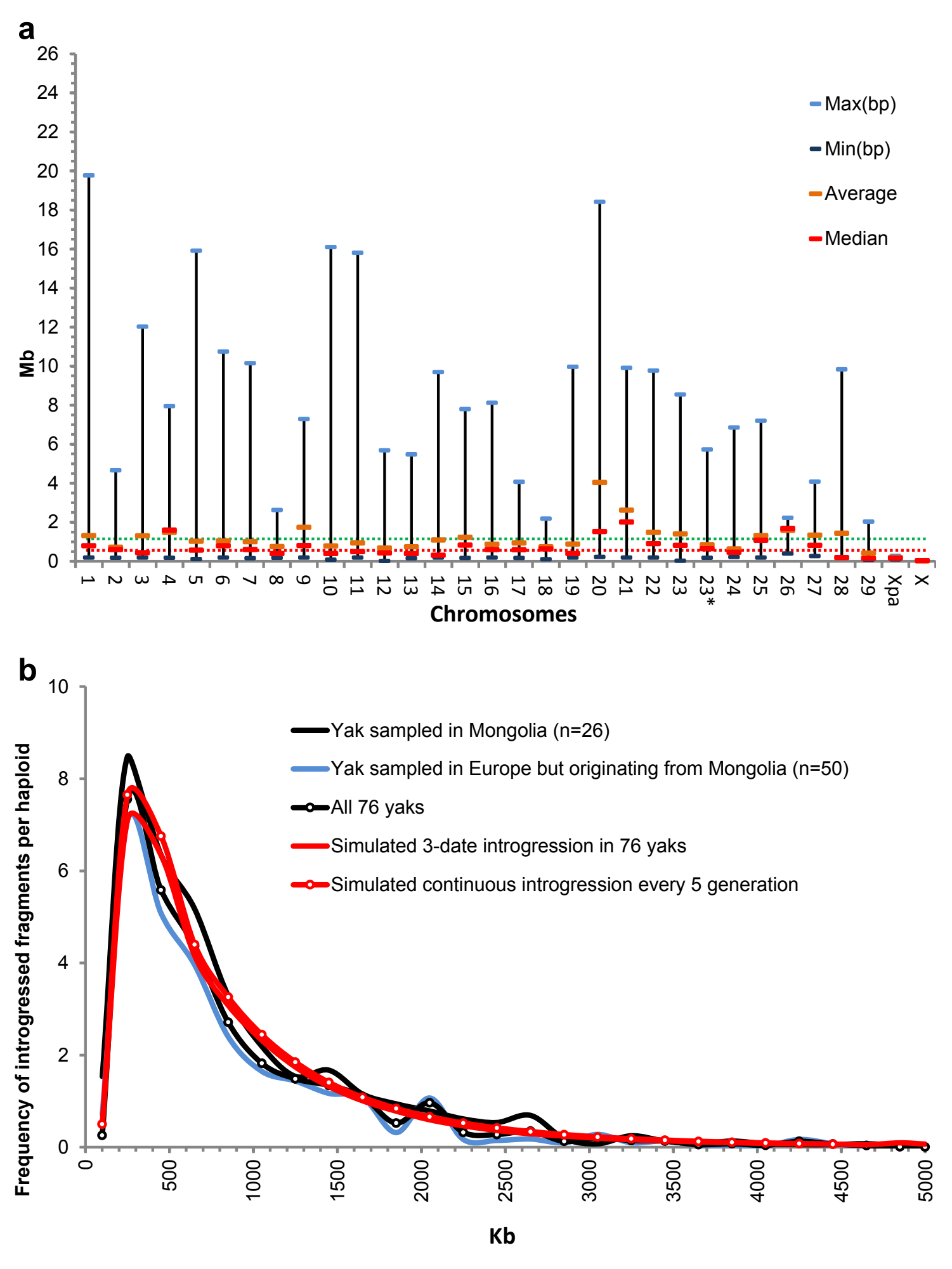
The region encompassing the Mongolian POLLED mutation was PCR amplified from genomic DNA samples of nine bovid species (**Supplementary Fig. 3** and **Supplementary Table 3**). Two individuals were used for each species. PCR primers were manually designed in regions which were conserved between bovine UMD3.1 and sheep Oar_v3.1 genome assemblies (**Supplementary Table 15**). PCR reactions and Sanger sequencing were performed as previously described. The corresponding regions in cattle and yak were obtained from the bovine UMD3.1 genome assembly and from YAK40 whole genome sequencing data, respectively. Multispecies alignment was generated with *CLUSTALW* software, version 2.0.1[36].

REFERENCES

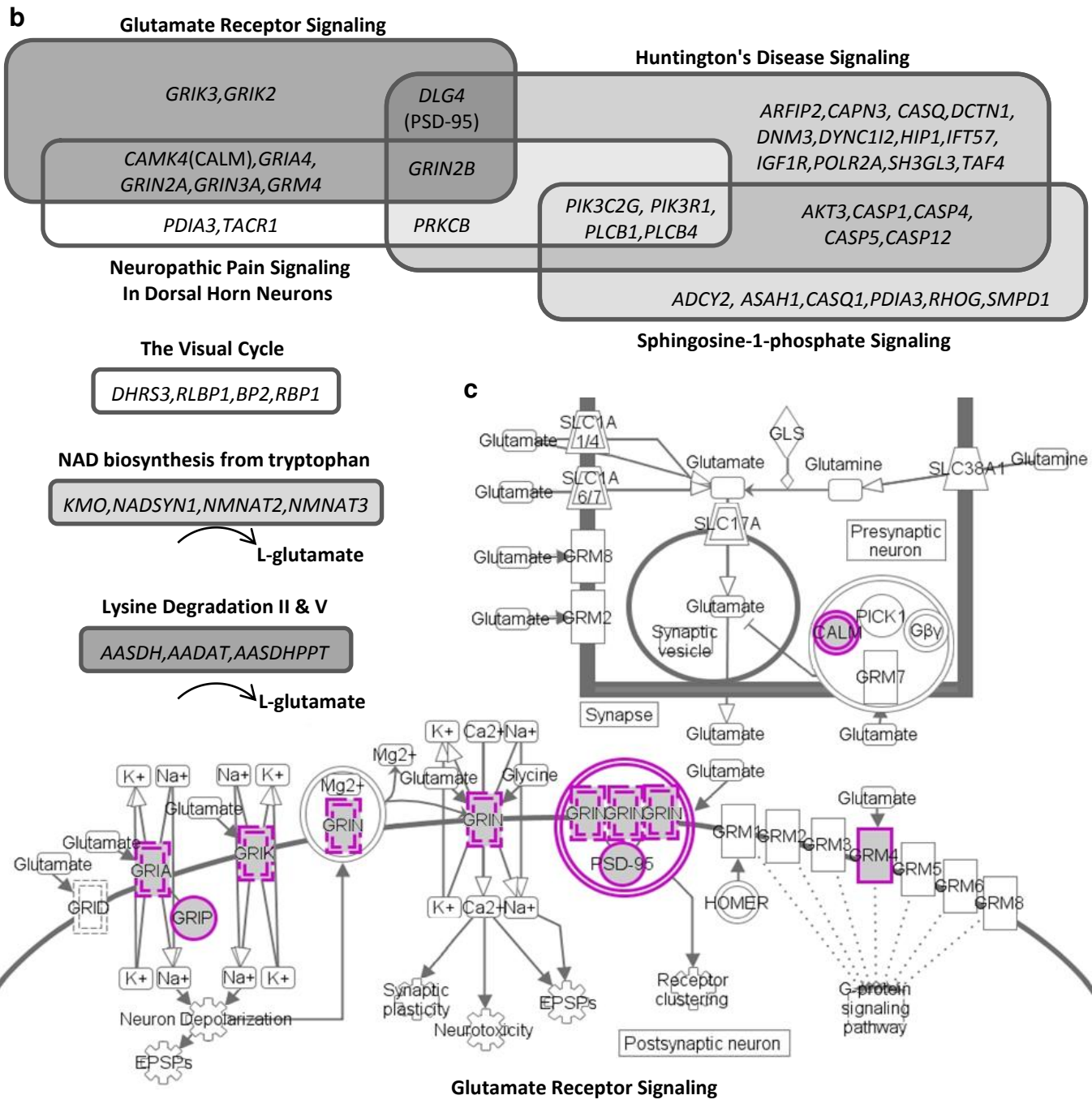
31. Dove, W.F. The physiology of horn growth: A study of the morphogenesis, the interaction of tissues, and the evolutionary processes of a mendelian recessive character by means of transplantation of tissues. *Journal of Experimental Zoology* **69**, 347-405 (1935).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
33. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
34. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z.M. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-2871 (2009).
35. Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178-192 (2013).
36. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680 (1994).

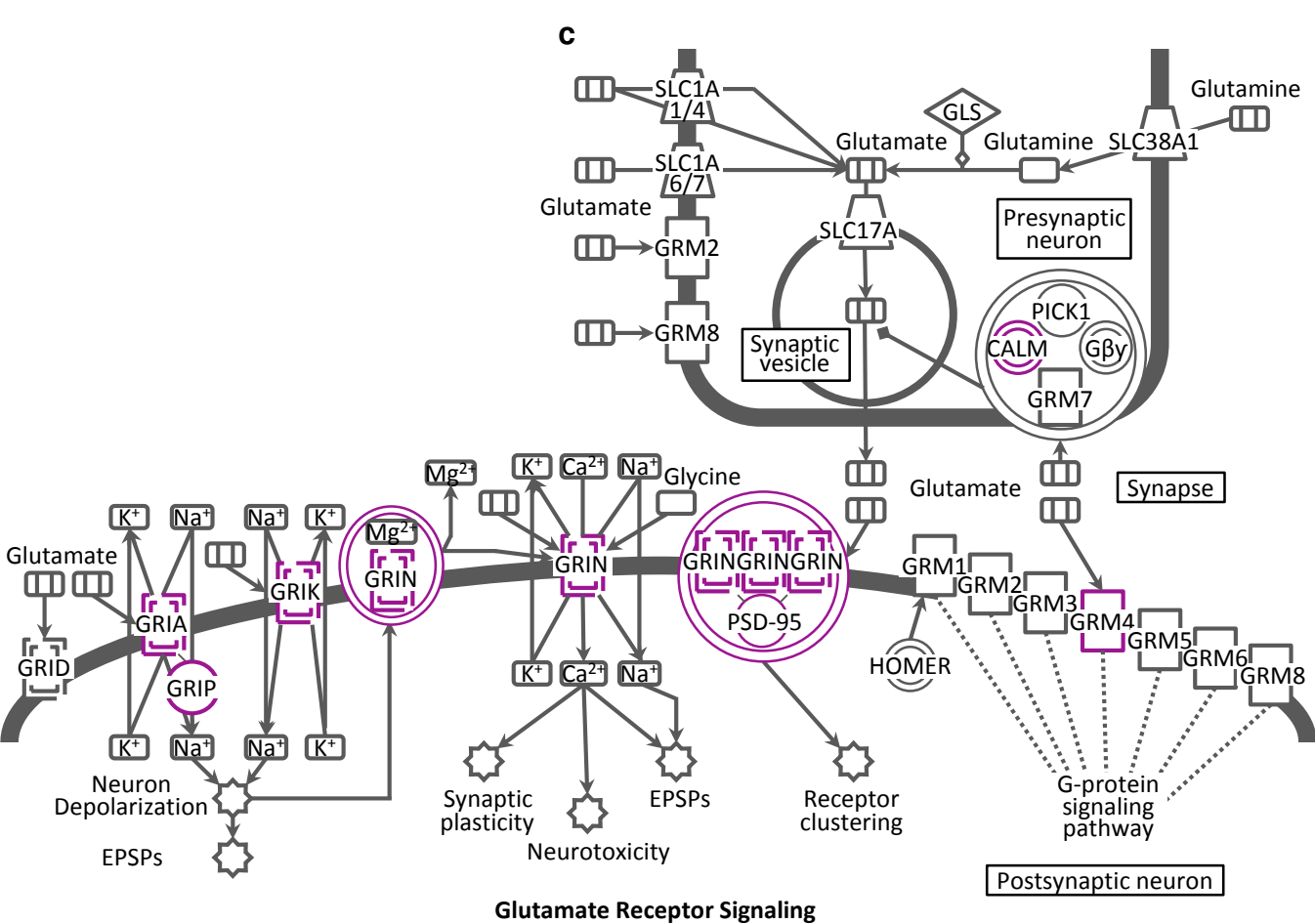
37. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729 (2013).
38. Saitou, N. & Nei, M. The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425 (1987).
39. Felsenstein, J. Confidence-limits on phylogenies - an approach using the bootstrap. *Evolution* **39**, 783-791 (1985).
40. Browning, B.L. & Browning, S.R. Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. *Genetic Epidemiology* **31**, 365-375 (2007).
41. Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. Inference of Population Structure using Dense Haplotype Data. *Plos Genetics* **8**, e1002453 (2012).
42. Hellenthal, G. *et al.* A genetic atlas of human admixture history. *Science (New York, N.Y.)* **343**, 747-751 (2014).
43. Leslie, S. *et al.* The fine scale genetic structure of the British population. *Nature* **519**, 309-314 (2015).
44. Simcic, M. *et al.* Recovery of Native Genetic Background in Admixed Populations Using Haplotypes, Phenotypes, and Pedigree Information - Using Cika Cattle as a Case Breed. *Plos One* **10**, e0123253 (2015).
45. Bowcock, A.M. *et al.* High-resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455-457 (1994).
46. Huson, D.H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-267 (2006).
47. Kuleshov, M.V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research* **44**, W90-W97 (2016).
48. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128-128 (2013).
49. Powell, J.E., Visscher, P.M. & Goddard, M.E. Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics* **11**, 800-805 (2010).
50. Meuwissen, T.H.E. & Goddard, M.E. Prediction of identity by descent probabilities from marker-haplotypes. *Genetics Selection Evolution* **33**, 605-634 (2001).
51. Lee, S.H. & Van der Werf, J.H.J. Using dominance relationship coefficients based on linkage disequilibrium and linkage with a general complex pedigree to increase mapping resolution. *Genetics* **174**, 1009-1016 (2006).
52. Meuwissen, T.H., Karlsen, A., Lien, S., Olsaker, I. & Goddard, M.E. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* **161**, 373-9 (2002).
53. Olsen, H.G. *et al.* Fine mapping of milk production QTL on BTA6 by combined linkage and linkage disequilibrium analysis. *J Dairy Sci* **87**, 690-8 (2004).
54. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333-i339 (2012).
55. Corpet, F. Multiple sequence alignment with hierarchical-clustering. *Nucleic Acids Research* **16**, 10881-10890 (1988).
56. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: A sequence logo generator. *Genome Research* **14**, 1188-1190 (2004).


a
Heterozygote
Homozygote
Coverage
10
20
30
40
Chr9
Chr25
2.0
1.5
1.0
0.5
0.0
Relative frequency
60
70
80
10
20
30
Position [Mb]
b
69,100,000 bp
69,400,000 bp
L3MBTL3
SAMD3
E19
E11
18,200,000 bp
18,900,000 bp
ACSM2B
ACSM1
THMPD1
ERI2
MGC134577
LYRM1
E10
E11
E10
c
0.005
Bos grunniens (Qiu et al. 2012)
Bos taurus turano-mongolian
Bos taurus
Bos javanicus
Bos gaurus
Bos grunniens (YAK013)
Bison bison athabascae
Bison bonasus
Bubalus bubalis
Boselaphus tragocamelus
Antilope cervicapra
Capra ibex sibirica
Capra ibex
Ovis aries musimon
91
99
97
98
63
96
70
88
97
99
97

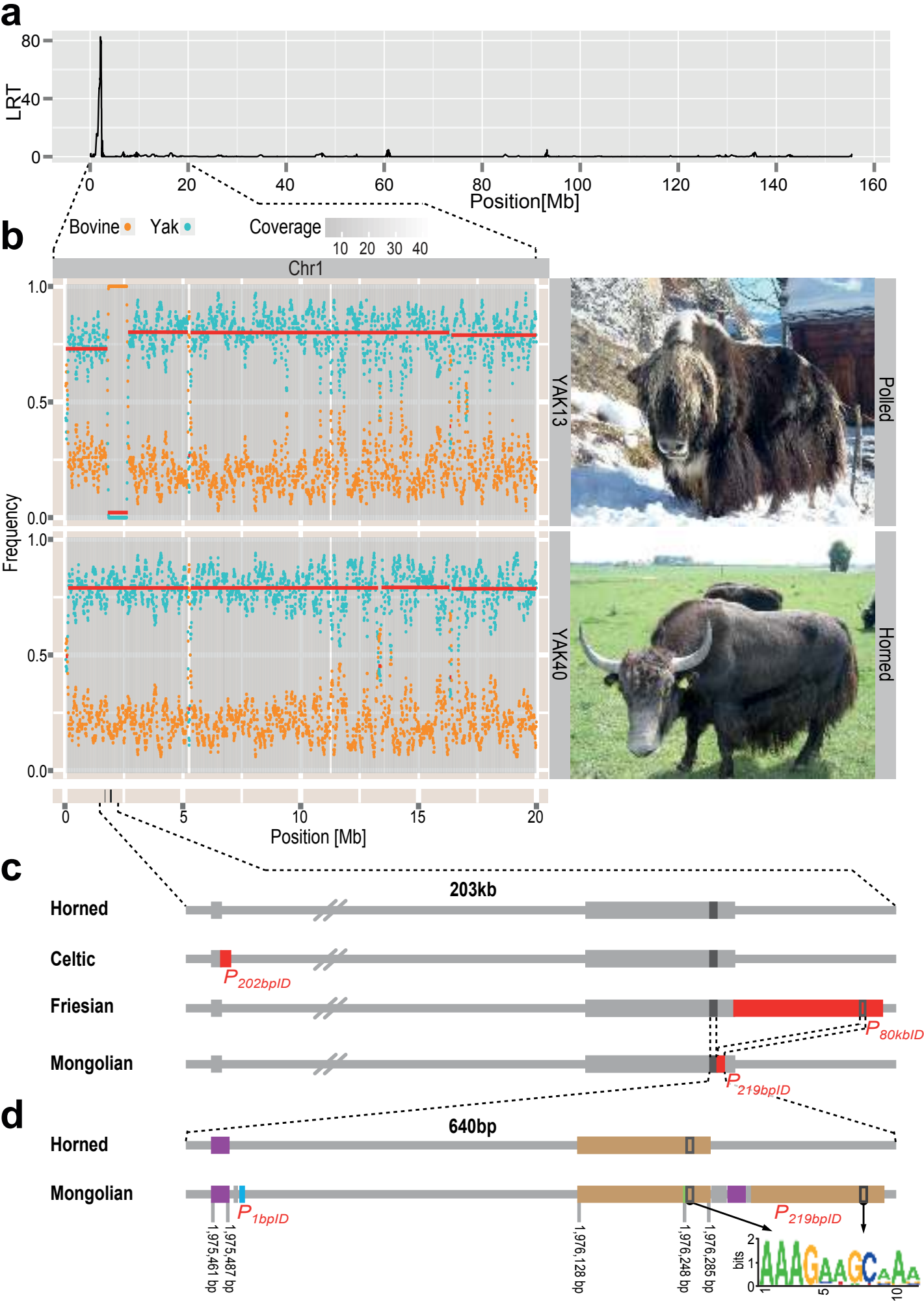

a
Mb
Max(bp)
Min(bp)
Average
Median
Chromosomes
b
Frequency of introgressed fragments per haploid
Yak sampled in Mongolia (n=26)
Yak sampled in Europe but originating from Mongolia (n=50)
All 76 yaks
Simulated 3-date introgression in 76 yaks
Simulated continuous introgression every 5 generation
Kb


a
Brain
Nervous System
Synapses
Arteries
Heart
Hematological Diseases
Neurons
Psychological Disorders
Hematopoiesis
Axons
Astrocytes
Cornea
Ear
Head
Blood Pressure
Neurites
Behavior
Blood Vessels
Neurological Diseases
Heart Beats
Gonads
Kidney
GermCells
Dermatological Diseases
Breast
Lymphocytes
Bones
Genital Organs
Gastrointestinal Diseases
Cilia
Leukocytes
Skeleton
Muscles
Infectious Diseases
Lymphatic System
Fibroblasts
Myoblasts
Metabolic Disease
Embryo
Body Cavity
Vitamin and Mineral Metabolism
Liver
Adipocytes
Lipid Metabolism
Carbohydrate Metabolism


b
Glutamate Receptor Signaling
GRIK3,GRIK2
DLG4 (PSD-95)
CAMK4(CALM),GRIA4, GRIN2A,GRIN3A,GRM4
GRIN2B
PDIA3,TACR1
PRKCB
Neuropathic Pain Signaling In Dorsal Horn Neurons
Huntington's Disease Signaling
ARFIP2,CAPN3, CASQ,DCTN1, DNM3,DYNC1I2,HIP1,IFT57, IGF1R,POLR2A,SH3GL3,TAF4
PIK3C2G, PIK3R1, PLCB1,PLCB4
AKT3,CASP1,CASP4, CASP5,CASP12
ADCY2, ASAH1,CASQ1,PDIA3,RHOG,SMPD1
Sphingosine-1-phosphate Signaling
The Visual Cycle
DHRS3,RLBP1,BP2,RBP1
NAD biosynthesis from tryptophan
KMO,NADSYN1,NMNAT2,NMNAT3
L-glutamate
Lysine Degradation II & V
AASDH,AADAT,AASDHPPT
L-glutamate
c
Glutamate
SLC1A 1/4
SLC1A 6/7
GLS
Glutamine
SLC38A1
SLC17A
Presynaptic neuron
GRM8
GRM2
Synaptic vesicle
PICK1
CALM
Gβγ
GRM7
Synapse
Mg2+
K+
Ca2+
Na+
Glycine
GRIN
GRIK
GRIA
GRID
GRIP
PSD-95
HOMER
GRM1
GRM2
GRM3
GRM4
GRM5
GRM6
GRM8
Neuron Depolarization
EPSPs
Synaptic plasticity
Neurotoxicity
Receptor clustering
G-protein signaling pathway
Postsynaptic neuron
Glutamate Receptor Signaling



Glutamate Receptor Signaling


a
LRT
80
40
0
0
20
40
60
80
100
120
140
160
Position[Mb]
b
Bovine
Yak
Coverage
10 20 30 40
Chr1
Frequency
1.0
0.5
0.0
YAK13
Polled
YAK40
Horned
0
5
10
15
20
Position [Mb]
c
203kb
Horned
Celtic
P202bpID
Friesian
P80kbID
Mongolian
P219bpID
d
640bp
Horned
Mongolian
P1bpID
1,975,461 bp
1,975,487 bp
1,976,128 bp
1,976,248 bp
1,976,285 bp
P219bpID
bits