Demographic history, cold adaptation, and recent NRAP recurrent convergent evolution at amino acid residue 100 in the world northernmost cattle from Russia

Laura Buggiotti¹, Andrey A. Yurchenko², Nikolay S. Yudin², Christy J. Vander Jagt³, Hans D. Daetwyler^{3,4}, Denis M. Larkin^{1,2,5}

¹Royal Veterinary College, University of London, London, UK.

²The Federal State Budgetary Institution of Science Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (ICG SB RAS), Novosibirsk, Russia.

³Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, 3083, Victoria, Australia.

⁴School of Applied Systems Biology, La Trobe University, Bundoora, 3083, Victoria, Australia.

⁵Kurchatov Genomics Center, the Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Science, Novosibirsk, Russia.

Abstract

Native cattle breeds represent an important cultural heritage. They are a reservoir of genetic variation useful for properly responding to agriculture needs in light of ongoing climate changes. Evolutionary processes that occur in response to extreme environmental conditions could also be better understood using adapted local populations. Herein, different evolutionary histories for two of the world northernmost native cattle breeds from Russia were investigated. They highlighted Kholmogory as a typical taurine cattle, while Yakut cattle separated from European taurines ~5,000 years ago and contain numerous ancestral and some novel genetic variants allowing their adaptation to harsh conditions of living above the Polar Circle. Scans for selection signatures pointed to several common gene pathways related to adaptation to harsh climates in both breeds. But genes affected by selection from these pathways were mostly different. A Yakut cattle breed-specific missense mutation, H100Q, in a highly conserved *NRAP* gene, represents a unique example of a young amino acid residue

27 convergent change shared with at least 16 species of hibernating/cold-adapted mammals from nine
 28 distinct phylogenetic orders. This suggests a convergent evolution event along the mammalian
 29 phylogenetic tree and fast fixation in a single isolated cattle population exposed to a harsh climate.
 30

Introduction

Cattle domestication occurred ~8,000–10,000 years ago as a result of at least two independent events in the Fertile Crescent and the Indus Valley from two main sources of ancient *Bos* subspecies, *B. taurus* and *B. indicus* (1, 2). Many contemporary breeds adapted to a vast variety of environmental conditions originate from ancient and/or recent interbreeding between the *B. taurus* (taurine) and *B. indicus* (indicine) populations (2) and admixture with several other bovine species, including yak, banteng, and gaur (3, 4). While adaptations of cattle breeds to hot climates are relatively well studied due to the economic needs in light of global warming (5–7), there is limited knowledge about the mechanisms of cattle adaptation to colder climates with most studies limited to genotyping datasets (8) rather than whole genome (re)sequencing. There are several northern cattle breeds that are of a particular interest for such studies. Two of them originate from, and are adapted to, the harsh climates of Russia: the Kholmogory and the Yakut cattle. The Kholmogory cattle breed was formed in the European part of Russia (Arkhangelsk District), about 300 years ago, from local taurine landraces which were crossed with an imported ‘Dutch cattle’ as early as in the 18th century (9). Nowadays, the Kholmogory cattle is a highly productive dairy breed, which can be found in different parts of Russia including the northern territories (Siberia). The Yakut cattle (*Sakha Ynaga*) have a different history. They were likely formed at the Baikal area of Siberia (10) and migrated together with Yakut people to the contemporary Yakutia region about 500–800 years ago (10). They now are the world northernmost cattle population which can be found up to 200 km above the Polar Circle, where they are exposed to long winters and low temperatures that occasionally can drop to below -70°C. At the territory of their original formation they were likely influenced by Asian cattle populations and belong therefore to the so-called ‘Turano-Mongolian’ cattle group. Yakut cattle represent the only extant pure commercial population from this group as most other Turano-Mongolian breeds were interbred with European taurines to improve performance (9). Indeed, a recent phylogenetic analysis placed the Yakut cattle together with other Turano-Mongolian breeds, including the Hanwoo and Japanese Black, but at the same time suggested possible historical admixture with indicine and African taurine

cattle (11). However, the origin of this Turano-Mongolian cattle group that is highly divergent from the European taurines remains unclear with some authors suggesting an independent domestication event in Eastern Asia (12, 13).

Studies of multiple cold adapted species show that they can temporarily slow their metabolisms in response to harsh environmental conditions, leading to torpor or, in extreme cases, to hibernation (14). Ability/inability to hibernate could be observed in species from the same phylogenetic node (e.g. there are hibernating and non-hibernating bats) suggesting that this could be an example of convergently evolving phenotype (15, 16). Other researches, however, propose that the ability to hibernate could be an ancestral mechanism lost in some lineages (17). Comparison of selected genes in multiple Arctic and Antarctic mammals pointed to a set of genes that could be independently selected in response to cold adaptation. These include genes involved in hypoxia, actin binding, extracellular matrix etc., highlighting common pathways of cold adaptation (18). The study of Arctic and Antarctic fishes provides an example of independent changes in antifreeze glycoproteins allowing these species to survive sub-zero temperatures (19). In Antarctic fishes the antifreeze protein evolved from trypsinogen-like serine protease gene, while the Arctic polar cod genome contains nearly identical proteins evolved from a different genomic locus, suggesting convergent evolution (19). A recent study of Russian cattle and sheep breed genomes (20) found putative selection in regions containing known candidate genes for cold adaptations (21), but failed to identify if there could be any convergent molecular changes in the recent evolution of artificially designed livestock populations because of the limited data resolution (~100,000 SNP loci in the genome). Previous studies have indicated that convergent amino acid changes contribute to appearance of similar features in phylogenetically distinct groups of animals (22), environmental and ecological adaptations (23, 24), but this phenomenon was not so far found in individual breeds of domesticated species.

The present study focuses on comparison of resequenced genomes from two cold-adapted cattle breeds from Russia, the Kholmogory and Yakut cattle. These distinct populations were compared to their phylogenetically closest breeds to identify signatures of selection and copy number variants that

could have influence on their biology. In addition, the demographic and admixture analysis of the two populations suggests that the Kholmogory and Yakut cattle indeed belong to two distinct branches of taurine cattle, which became separated about 4,000-5,000 years ago. Our data suggest that while the Kholmogory cattle are a typical European taurine cattle breed, the Yakut cattle genome contains a distinct fraction of a largely ancestral cattle variants, either being introgressed into the Yakut cattle genome or being lost in the European taurines. The two breeds demonstrate signatures of selection in the same gene pathways related to adaptation to harsh climates, but the genes affected by selection from these pathways are mostly different. All resequenced Yakut cattle individuals contain a missense mutation in a highly conserved *NRAP* gene. This change is shared with a majority of hibernating mammals but absent from all other cattle breeds tested and a majority of mammalian species, suggesting a recent convergent evolution event and fast selection in a single isolated cattle population exposed to a harsh climate.

Materials and Methods

Sample selection and genome sequencing

Twenty samples per breed from the *Yakut* and *Kholmogory* Russian native cattle breeds were selected based on their kinship coefficient (to exclude closely related animals) and maximum purity estimated by the plink software command –genome and ADMIXTURE software respectively from our previous genotyping of the same samples (20). Next generation sequencing (NGS) paired end reads (PE, 150bp) were generated using HiSeq sequencing platform at Novogene Co., Ltd. Hong Kong, China. Library preparation was carried out by Novogene and raw sequence data per sample, ranging from 41.1 to 66.8 Gb, were generated. In addition, we obtained raw sequence data of 153 cattle individuals from the NCBI SRA database, including 20 Holstein (25), 20 Hanwoo (26), 47 African cattle (Ankole – ten, Boran – nine, Kenana – nine, nDama – ten, Ogaden – nine), 36 Chinese cattle (Luxi – five, Nanyang – five, JaxianRed – five, Dianzhong – six, Wannan – five, Ji'an – four, Leiqiong – three, Yanbian - three), five Brahman, ten Yak (*Bos grunniens*), and 15 of other *Bos* species (five Gayal (*Bos frontalis*), five Banteng (*Bos javanicus*), three Gaur (*Bos gaurus*), two Bison (*Bison bison*); Supplementary Information 1) (27, 28). Reference Hereford cattle (*Bos taurus*) assembly (UMD3.1, *BosTau6*) was downloaded from NCBI.

Read mapping, and variant calling

All cleaned reads were mapped to the cattle reference assembly using BWA-MEM (29) with default parameters. The average mapping rate of the reads generated in this study was 99.68%, and the average sequencing coverage was approximately 11.43X (ranging from 9.7 to 15.1, Table 1). Alignment preprocessing steps and variant calling were done following the Genome Analysis Toolkit (GATK v. 3.8 (30)) pipeline. For each raw BAM file, we marked duplicate reads with Picard (v. 1.69) using the tool MarkDuplicates (<http://broadinstitute.github.io/picard/>). Next, we performed base quality score recalibration (using cattle known variants: dbSNP148). Variant calling was performed using the Haplotype Caller tool from GATK. Three joint VCF files were generated for: 1) the Yakut,

Kholmogory, Holstein, and Hanwoo breeds, 2) the Yakut, Kholmogory, Holstein, Hanwoo, Indian, Chinese and African breeds, and 3) the Yakut, Kholmogory, Holstein, Hanwoo, Indian, Chinese and African breeds and bovine species samples (see Supplementary Information 1 for samples). The filtering of SNPs for quality ("hard filtering") was applied with the following parameters: i) variant confidence/quality by depth < 2 ; ii) RMS mapping quality (MQ) < 40.0 ; iii) Phred-scaled p-value using Fisher's exact test to detect strand bias > 60 ; iv) Z-score from the Wilcoxon rank sum test of alternative vs. reference read MQs (MQRankSum) < -12.5 ; and (v) Z-score from the Wilcoxon rank sum test of alternative vs. reference read position bias (ReadPosRankSum) < -8 . The thresholds for these parameters were adopted from GATK Best Practices (30). Multi-allelic SNPs and INDELs were removed and the final filtered VCF file was used for further analysis.

Genomic annotation of SNPs

Whole genome hard filtered SNPs were annotated using the Ensembl Variant Effect Predictor (VEP) with Ensembl v94 (31). SNPs were annotated with the classification categories of impact HIGH, LOW, MODERATE, and MODIFIER (Supplementary Information 2). Moreover, dNdScv (32), used to estimate whole genome dN/dS, also classified SNPs based on the RefSeq Release 103 (33). We derived a subset of SNPs from the four breeds' VCF files with the following characteristics: 1) high frequency of alternative (non-reference) allele in Yakut cattle (Yakut alternative allele frequency of $\geq 70\%$ and $\leq 10\%$ in Holstein, Kholmogory, and Hanwoo), 2) SNPs present in at least 25% of the four breed sample set. NGS annotation pipeline (34) was used to assign classification (e.g., intronic, missense, synonymous, splice variant, stop gain, etc.) to each SNP in this subset of SNPs and to provide several fields of information describing the affected transcript and protein, if applicable. NGS-SNP also reports NCBI, Ensembl, and UniProt IDs for genes, transcripts, and proteins when applicable. For missense SNPs, an "alignment score change" was calculated by comparing the reference amino acid and the non-reference amino acid to an orthologous amino acid in a range of genomes. A positive value indicated that the alternative allele amino acid was more

common in the orthologous position in other species than the reference amino acid, whereas a negative value indicated that the reference amino acid was more common (Supplementary Information 3).

Demographic history reconstruction

To calculate historical demography and separation time between pairs of populations we used two complementary approaches: SMC++ (35) and GADMA (36), which use different population inference models. To remove spurious variants and regions with bad mapping quality from further consideration, we first computed the mappability score of the bovine reference genome using the GEM mappability program (37) with kmer size =100 bp allowing one mismatch. The regions with mappability < 1 were removed. Additionally, we removed all regions under selection identified in the bovine genome (38) and sex chromosomes.

SMC++ uses a sequential Markovian Coalescent approach to infer demographic history of separate populations and pairwise time of divergence between them on the unphased samples. The drawback of this method is an inability to implement migration into the model. For the SMC++ we analyzed demographic history using the cubic spline approach with mutation rate per bp per generation = 1.2e-8.

The approach implemented in the GADMA software is based on the comparisons of observed and modeled joint allele frequency spectrums of populations (39, 40) and allows the efficient modelling of demographic parameters and migration, given a divergence scenario with up to three populations, using a powerful genetic algorithm. We ran GADMA with five years as cattle generation time using 10 processes simultaneously and linked loci. The main algorithm chosen was moments (ordinary differential equations).

170 **Detection of genetic introgression between breeds**

171 We applied a robust forward-backward algorithm implemented in RFMix (41) to screen for the
 172 presence of putative taurine or indicine haplotypes in autosomes of the Yakut cattle. This algorithm
 173 uses designated reference haplotypes to infer local ancestry in designated admixed haplotypes, thus
 174 five genetic groups were selected as a reference panel: European taurine (Holstein), Russian taurine
 175 (Kholmogory), African indicine (Ogaden), Chinese indicine (Wannan, Ji'an, and Leiqiong), and
 176 Indian indicine (Brahman). Window size was set to three (-w 3) and the option "--reanalyze-
 177 reference" with three iterations was used to analyze the reference haplotypes as if they were query
 178 haplotypes (41).

179 We conducted TreeMix v1.12 (42) analyses to infer the relationships, divergence, and major mixtures
 180 among 18 cattle breeds and five bovine species (Supplementary Information 1). We applied the
 181 option "-root YAK", which sets Yak as the position of the root, option "-k 1000", which builds the
 182 tree using blocks of 1000 SNPs to account for linkage disequilibrium and used the option "-se" to
 183 calculate standard errors of migration proportions. We allowed up to 15 migration edges on the tree
 184 (m ranging from 0 to 15) and generated a residual heatmap to identify populations that were not well-
 185 modeled after adding each migration edge. The percentages of variation explained by the maximum
 186 likelihood trees were also calculated. Migration edges were considered until 99.9% of the variance in
 187 ancestry between populations was explained by the model. We also ensured that the incorporated
 188 migration edges were statistically significant. Residuals from the model were visualized using the R
 189 script implemented in TreeMix. Finally, in order to provide further support for a past admixture
 190 between populations, we calculated f_3 and D statistics using ADMIXTOOLS (v 5.1) with default
 191 parameters (43). We calculated f_3 statistics using qp3Pop of the form (X; A, B) where a negative
 192 value mean implies that X is admixed with populations close to A and B. We considered negative
 193 statistics with Z-score values below -2 as significant signals of admixture.

194 We computed D statistics using qpDstat. D statistics of the form D (A, B, X, Y) to test the null
 195 hypothesis of the unrooted tree topology ((A, B), (X, Y)) was used. A positive value indicates that

either A and X, or B and Y share more drift than expected under the null hypothesis. We quote D statistics as the Z score computed using default block jackknife parameters.

Detection of copy number variants (CNVs) and copy number variant regions (CNVRs).

cn.MOPS (Copy number estimation by a Mixture Of PoissonS; (44)) was used for CNV region detection in four cattle breeds (Kholmogory, Yakut, Hanwoo, and Holstein) and pure indicine breeds (Brahman, Ogaden, Wannan, Ji'an, and Leiquiong) chosen based on the structure analysis reported by Chen et al (28). A total of 75 samples of the four cattle breeds were used for CNV detection, and five samples (one from Holstein, two from Hanwoo, and two from Yakut) were removed from the analysis as they were problematic probably due to their lower sequence coverage. Based on the average sequence coverage of our data, window length was set to 700 and posterior probabilities estimated. The tool represents a CNV detection pipeline that models the depths of coverage across multiple samples at each genomic position. Using a Bayesian approach, it decomposes read variations across samples into integer CNVs and noise using mixture components and Poisson distributions, respectively. The multiple samples approach increases statistical power and decreases computational burden and the FDR in CNV detection. CNVs were then used to construct a set of copy number variable regions (CNVRs) for Kholmogory, Yakut, Hanwoo, and Holstein. Moreover, we applied the following filtration steps: i) cn.MOPS classes one, two, and three (CN1-CN3) were considered as normal and the rest (CN0, CN4-CN8) were considered as CNVs; ii) CNVRs were constructed by merging CNVs across samples of the same breed that exhibited at least 50% pairwise reciprocal overlap in their genomic coordinates; iii) unique CNVRs per breed were those having less than 10% overlap with CNVRs in other breeds. In addition, cn.MOPS was run a second time by using the four above mentioned breeds together with additional 26 individuals of Indian (5), Chinese (12), and African (9) indicine origin, to detect indicine CNVs. BEDTools (45) and BEDOPS (46) tools were used to calculate CNVR overlaps.

222 **Genome scan for selection signals**

223 *Identification of signatures of selection with hapFLK statistics.* We performed a genome scan for
 224 selective sweeps within the groups of phylogenetically closely related breeds: Yakut and Hanwoo
 225 and Kholmogory and Holstein using haplotype-based statistics hapFLK (47). Due to the hapFLK
 226 model assuming that selection acts on shared ancestral SNP allele frequencies, we excluded rare SNPs
 227 with low minor allele frequencies (MAFs) from breed groups ($MAF < 0.1$). We also excluded poorly
 228 genotyped individuals ($< 95\%$ of SNPs with genotypes), loci genotyped in $< 99\%$ of samples, SNPs
 229 without chromosomal assignments, and SNPs on sex chromosomes in PLINK, using the commands:
 230 `--maf 0.1, --mind 0.05, --geno 0.01, and --chr 1-29` prior to performing the genome selection scans.
 231 This resulted in 8,423,043 and 7,926,039 SNPs for Yakut-Hanwoo and Kholmogory-Holstein groups,
 232 respectively.

233 The hapFLK method takes the haplotype structure of the population into account. What was important
 234 for our dataset is that this method can account for population bottlenecks and migration. Reynolds
 235 distances and a kinship matrix were calculated by the hapFLK program v.1.4 (47). For the hapFLK
 236 analysis, the number of haplotype clusters for each breed group were estimated with fastPhase (48)
 237 and were set as $-K$ 10 for the Yakut-Hanwoo, and $-K$ 20 for Kholmogory-Holstein groups. The
 238 expected maximum number of iterations was set to 30 for both groups. We applied midpoint rooting
 239 to all sets of breeds.

240 *P-value calculation.* For hapFLK, the calculation of raw p-values was performed assuming that the
 241 selected regions represent only a small fraction of the genome (49). The genome-wide distribution of
 242 hapFLK statistics could be modelled relatively well with a normal distribution except for a small
 243 fraction of outliers from potentially selected regions (49). Robust estimations of the mean and
 244 variance of the hapFLK statistic were obtained using the R MASS package *rlm* function to eliminate
 245 influence of outlying regions following Biotard and co-workers (50). This has been done for each
 246 group (Yakut-Hanwoo, and Kholmogory-Holstein). The hapFLK values were Z-transformed using
 247 these parameter estimates, and p-values were calculated from the normal distribution in R. The R

248 *qvalue* package was used to correct p-values for multiple testing (51). To identify selected intervals,
 249 we scanned the hapFLK output for intervals containing q-values < 0.01 . The boundaries of the
 250 selected intervals were defined by the first marker up- and down-stream with a q-value > 0.1 . To
 251 assign a selection signal to a specific breed Local Reynolds distances were calculated for selected
 252 regions using *local_reynolds.py* script and local population trees were then built with the *local_trees*
 253 R script obtained from <https://forge-dga.jouy.inra.fr/projects/hapflk/wiki/LocalTrees>.
 254 *Identification of genomic regions under selection: dN/dS ratio.* Whole genome dN/dS estimates were
 255 obtained for the SNPs found in any of the four cattle breeds (Yakut, Kholmogory, Holstein, and
 256 Hanwoo) using the dNdScv (32). The background mutation rate of each gene was estimated by
 257 combining local information (synonymous mutations in the gene) and global information (variation
 258 of the mutation rate across genes) and controlling for the sequence composition of the gene and
 259 mutational signatures. Unlike traditional implementations of dN/dS, dNdScv uses trinucleotide
 260 context-dependent substitution matrices to avoid common mutation biases affecting dN/dS (52). The
 261 calculations were initially made using all hard-filtered SNPs (MAF in each breed ≥ 0.1) using the
 262 *Btau6* reference genome database generated and the “dndscv” function. The second run was done
 263 using hard filtered SNPs with MAF ≥ 0.6 in a breed to identify genes under selection which likely
 264 affected the whole breed. One sided p-values were calculated and adjusted for multiple testing using
 265 Benjamini and Hochberg’s false discovery rate to detect both positive and negative selection (q-value
 266 positive selection and q-value negative selection).

267

268 ***F_{ST}* calculations against the 1000 Bull Genome Dataset**

269 Fixation index of genetic differentiation (*F_{ST}*) between biallelic SNPs identified in our Yakut or
 270 Kholmogory cattle samples as part of their analysis by the 1000 Bull Genome Project Run 7 against
 271 the reference cattle ARS-USD1.2 genome was calculated against: a) the set of pure taurine breeds
 272 (2,003 animals, 18 breeds) as identified from the structure analysis reported by Chen et al (28) and
 273 b) against all cattle samples of taurine or indicine origins (3,362 animals) from the 1000 Bull Genome

274 Project Run 7. This was performed using *vcftools* (53) for each cattle autosome applying the following
 275 settings: *--fst-window-size 10000 --fst-window-step 5000*. All animals from the 1000 Bull Genome
 276 Project Run 7 have been filtered for a minimum of 6x mean genome coverage. One percent of the
 277 windows with the maximum mean F_{ST} values were extracted and overlapped with the gene set from
 278 Ensembl v.94.

279

280 **Functional enrichment analysis and protein modeling**

281 Cattle genes based on the Ensembl v.87 database were retrieved using the Ensembl BioMart online
 282 tool (54) and gene ontology (GO) analysis was performed using genes found in unique CNVRs in
 283 each breed: ‘indicine-like’ genes from the RFMix analysis, and the hapFLK selected intervals using
 284 an in house pipeline. Moreover, a David functional cluster analysis (55) and a Reactome (56) pathway
 285 enrichment analysis were performed.

286 Genomic Association Test – GAT (57) was used to estimate the significance of overlap between
 287 multiple sets of genomic intervals by using 10K permutation. The genomic territory was represented
 288 by the *Btau6* whole genome and the tool was used to assess the significance of unique CNVRs among
 289 the four cattle breeds, and indicine-like assigned genes from the RFMix analysis.

290 Protein sequences for several species and their alignments were retrieved from the UniProt database
 291 (58) and ESPript v3.0 (59) was used to gather sequence similarities and secondary structure
 292 information from aligned sequences. Next, Phyre2 (60) allowed to predict protein 3D structure and
 293 estimate mutation sensitivity; the “intense” modelling option was used.

294 **Results**

295 *Whole genome resequencing, mapping, and SNP detection*

296 Resequencing of the Yakut and Kholmogory cattle samples resulted in a total of 1,667.61 Gb of
 297 filtered sequence data. The read mapping rate against the reference genome for Kholmogory cattle
 298 was 99.62%, while for the Yakut cattle it was 99.59%, resulting in an average sequencing coverage
 299 of 11.68x and 11.18x for the Kholmogory and Yakut samples, respectively (Table 1). The total
 300 number of SNPs was 17.03 and 15.43 million for the Yakut and Kholmogory samples, respectively.
 301 To identify copy number variants (CNVs) and signatures of selection differentiating the Yakut and
 302 Kholmogory cattle breeds from phylogenetically most related breeds (20), 20 samples of Holstein
 303 and 20 samples of Hanwoo breeds were included in our dataset. The average sequencing coverage
 304 per sample was 10.60x for Holsetin and 10.21x for Hanwoo. A total of 31.08 million SNPs were
 305 called in the four-breed joint set of which 27.49 million passed our quality filtering criteria. As
 306 expected, the majority (21.69 million (78.90%)) of the SNPs were located in the intergenic and
 307 intronic regions. About 0.60%, 0.31%, and 0.56% of the SNPs overlapped mammalian, cetartiodactyl,
 308 and ruminant conserved non-coding elements (61) and the number of synonymous and missense
 309 SNPs were 109,774 and 64,221, respectively (Table 1). There were 35,919 and 31,089 missense
 310 variants in the Russian Yakut and Kholmogory breeds, respectively. We found 224 and 210 nonsense
 311 variants in the Yakut and Kholmogory breeds, of which 32.60% and 21.90% were found in a single
 312 breed, respectively. A total of 678 and 594 SNPs were classified as “essential splice variants” while
 313 17 and 15 were stop losses, in the Yakut and Kholmogory breeds, respectively (Supplementary
 314 Information 2).

315 To identify historical relationships and admixture between the Yakut cattle and other cattle breeds
 316 and species with potential historical admixture, SNPs from additional 113 individuals, representing
 317 indicine (African, Chinese, and Indian) cattle and five bovine species (yak, banteng, gayal, gaur,
 318 bison) were combined with the four breed SNP set, resulting in a total of 193 individuals
 319 (Supplementary Information 1) and 11.03 million high quality biallelic SNPs.

320 ***Demographic history and population divergence***

321 We estimated demographic and population history using two methodologically distinct approaches:
 322 coalescent-based SMC++ (35) and ordinary differential equations (moments) realised in the GADMA
 323 package (36). To understand demographic histories of the Kholmogory and Yakut cattle breeds we
 324 focused to the three breeds with largest sample data and the least possibility of a recent admixture:
 325 Yakut, Kholmogory, and Holstein. SMC++ algorithm estimated a stepwise decline in the population
 326 size of cattle breeds with Indicine lineage (Brahman) as the most divergent and a much less
 327 pronounced recent decline. Yakut cattle demonstrated much stronger population decline more
 328 recently (around 200 – 500 years ago) than Kholmogory and Holstein (Fig. 1D).

329 We then tried to estimate the time of pairwise divergence between the breeds using SMC++-based
 330 coalescent approach. As expected from the historical data, the divergence time of Holstein and
 331 Kholmogory was estimated to be around 500 years (Fig. 1A), while the divergence of Yakut lineage
 332 was much earlier with a 5,000 years estimation for Yakut-Holstein divergence (Fig. 1C) and 25,000
 333 years for Yakut-Kholmogory divergence (Fig. 1B). The later estimation is likely to be an
 334 overestimation (i.e. pre domestication) because the SMC++ does not consider any possible migration
 335 events between populations, but estimation of an early Yakut cattle divergence from the common
 336 ancestor of Holstein and Kholmogory is in line with an early separation of the Turano-Mongolian
 337 cattle from the European taurine breeds.

338 The second approach, based on the joint allele frequency spectrum (GADMA) returned a quite
 339 realistic demographic population model of Yakut, Kholmogory, and Holstein divergence (Fig. 1E).
 340 The estimated divergence between Kholmogory and Holstein was around 300 years ago in line with
 341 historical data, while the divergence of Yakut cattle from the common ancestor of Holstein and
 342 Kholmogory was estimated to be around 4,900 years ago. The best Log-likelihood model also
 343 indicated sharp decline of Yakut cattle population around 300 years ago and migrational events from
 344 Holstein, Kholmogory and their common ancestor to the Yakut lineage (Fig. 1E). Thus, the very
 345 different methods with contrasting assumptions (allowed and not migrational events) point out to a

very old (several thousands of years) time of divergence of the Yakut breed from the common ancestor of modern taurine breeds, which was accompanied also by a historically very recent bottleneck of Yakut cattle population. The reality of admixture between European taurines and Yakut cattle is not very clear because we did not observe significant amount of European taurine ancestry in the Yakut genetic pool with ADMIXTURE algorithm (11) but it is also impossible to completely exclude such events.

Yakut cattle: indicine introgression or ancestral taurine genetic variants

As it follows from our demography and previous studies (11, 20, 62) the Kholmogory and Yakut cattle have different histories, with the Kholmogory being a ‘pure’ taurine breed formed in European Russia in the XVII century as the result of selection applied to admixed population of imported European cattle and native populations (11). This is confirmed by similar N_e trajectories of the Kholmogory cattle with Holstein (Fig. 1E). The Yakut cattle were likely formed in Asia and demonstrated signatures of possible introgression from indicine cattle (11, 20). To identify genome intervals subjected to potential introgression into the Yakut cattle, a combination of several approaches has been used in this study. We used ‘pure’ taurine (Holstein, Kholmogory) and indicine (African, Chinese, and Indian) breeds as reference populations in the RFMix analysis, which suggested that the Yakut cattle contains ~86.5% of the ‘taurine-like’ and ~13.5% of the ‘indicine-like’ genome. On the other hand, the TreeMix analysis, using a total of 18 cattle breeds and five bovine species did not detect any introgression between the Yakut cattle and other populations used (Fig. 2). On the TreeMix tree the Yakut cattle was placed close to the Hanwoo and Yanbian breeds and the Kholmogory close to Holstein, which is consistent with previous studies (11, 20). We observed an introgression from banteng to Chinese indicine cattle breeds consistent with the previous report (28) as well as multiple additional introgression events. The f_3 statistics calculated on population triples using the Yakut cattle as a target, various indicine cattle breeds and five related species as source populations did not result in a significant Z-score (Supplementary Information 4a).

372 The *D* statistics (bovine species, indicine, Yakut, Yak) results, however, suggested admixture
 373 between the Yakut and indicine cattle as well as with taurine breeds (Supplementary Information 4b,
 374 Supplementary Fig. 1).

375 To investigate possible reasons for discrepancies between the RFMix and the TreeMix results we
 376 extracted four sets of SNPs with high derived allele frequencies in the Yakut, Kholmogory, Holstein,
 377 and Hanwoo genomes but with low frequency of the same allele in the other three breeds (one breed
 378 $MAF \geq 0.7$; other three breeds $MAF \leq 0.1$; autosomal). The number of such SNPs was 13,257 in the
 379 Yakut cattle but much lower in the other breeds: 758 for Hanwoo, 103 for Kholmogory, and 311 for
 380 Holstein. Of the 13,257 Yakut SNPs the majority (9,809 (73.92%)) overlapped the Yakut cattle
 381 genome intervals identified as ‘indicine-like’ by the RFMix analysis. To look into the evolutionary
 382 history of the high-frequency Yakut alleles we focused on all 48 missense mutations present in the
 383 13,257 SNP set (Table 2; Fig. 3). Of these, 37 Yakut cattle high-frequency derived alleles were absent
 384 from the European taurine cattle (Holstein and Kholmogory) but preferably had high frequencies in
 385 the indicine cattle breeds. Furthermore, 30/37 derived alleles were found in at least one additional
 386 bovine species suggesting that they could be present in the ancestral taurine genome and could be
 387 eliminated from at least two European taurine breed (Holstein and Kholmogory) genomes due to
 388 selection or drift. To check this hypothesis we looked into the orthologous positions in proteins of up
 389 to 82 evolutionary distinct animals (ranging from mammals to fish) and found that for 10/30 positions
 390 the Hereford reference (*Btau6*) amino acids matched the amino acid preferably found in the
 391 orthologous position of other animals, while for the 19/30 positions the Yakut cattle amino acid was
 392 overrepresented in other animal proteins. These observations coupled with the TreeMix and f_3
 393 statistics results support the hypothesis that most of the high-frequency Yakut cattle SNPs are either
 394 absent or with low frequencies in the European taurine cattle could represent alleles eliminated from
 395 the European taurine cattle. For the final check we looked for the presence of these 48 missense SNPs
 396 in the non-European taurine breeds and found 37/48 Yakut cattle alleles in Hanwoo, 46/48 in the

Chinese and 42/48 in the African (nDama) taurine breeds confirming our hypothesis that these variants could be present in the ancestral taurine genome.

We detected only one Yakut cattle missense derived allele not found in the indicine, taurine cattle or other bovine species, in the NRAP protein. This H100Q change was highly dissimilar to the orthologous amino acids found in other animals. The fact that all 48 SNPs represent alleles different from the reference genome (Hereford) suggests that they are likely to be absent from both the dairy and beef European taurines. Indeed, we found only 331 SNPs for which the Yakut cattle had a high-frequency allele identical to the reference Hereford genome while Kholmogory, and Holstein had high-frequency derived alleles. None of these 331 SNPs were missense mutations.

Putative ancestral taurine alleles absent or nearly absent in the European taurine breeds

We used the whole set of 13,257 high frequency Yakut cattle SNPs to identify putative regions of the ancestral taurine genome that are present in high frequency in the Yakut cattle but are absent from the European taurines. Among the 13,257 SNPs we found 263 alleles present in the Yakut cattle only, 4,866 alleles found in a low frequency (<10%) in the European taurines and 8,128 alleles absent from the European taurine set. Of the 8,128 SNPs, 7,546 (92.83%) were found in indicine breeds, however only 166 (2.00%) of them were indicine-specific, while 5,760 (70.87%) were also present in at least one additional bovine genome, 2,018 (24.82%) additional SNPs were present in either the African or Chinese taurines, and 184 (2.26%) were shared with Hanwoo only (Table 2; Fig. 4 and Supplementary Fig. 2). These results reassured us that the ‘indicine-like’ genome regions of the Yakut cattle to a large extent, if not all, represent ancestral taurine genome regions with alleles mostly eliminated or present in a low frequency in the European taurine genomes. As a final test we identified all intervals in the Yakut cattle genome containing 1.67 million ‘indicine-like’ alleles as assigned by the RFMix analysis, intersected them with bovine, African, Chinese taurine genomes and Hanwoo, then eliminated all intersected intervals except for those which contained indicine-specific alleles (155,542 SNPs) and repeated the RFMix analysis. It showed that the filtered set had 2.5% of the

423 ‘indicine-like’ genome suggesting that the removed intervals were responsible for the majority of the
 424 ‘indicine-like’ part of the Yakut genome.

425 To understand the role of genes found in the ancestral taurine genome segments, we extracted all
 426 genes found in the ancestral and European taurine genome intervals (gene total length overlap $\geq 60\%$
 427 for a gene to be assigned to an either set). A total of 1,639 (6.7%) genes were assigned to the ancestral
 428 and 19,205 (78.2%) genes to the European taurine segments. GO enrichment analysis of the genes in
 429 ancestral segments revealed the term “*response to stimulus*” (q-value = 0.0) and “*metabolic process*”
 430 (q-value = 0.0) among others (Fig. 5; Supplementary Information 5a). Moreover, the genes from the
 431 ancestral segments were enriched in the “*immune system process*” pathway (q-value = 0.0; Fig. 5)
 432 including the MHC class II antigens. The top DAVID cluster (enrichment score 3.68) included
 433 multiple MHC class II antigens, immune genes (e.g. *IL36A* and *IL36B*), heat shock proteins (*HSPA1A*
 434 and *HSPAIL*) and the insulin receptor (*INSR*; Supplementary Information 5b). Moreover, we selected
 435 four GO categories potentially involved in adaptation to extreme environments: response to stress,
 436 response to pain, neuropathy, and actin binding (Supplementary Information 6). The GO category
 437 response to pain resulted significantly enriched in genes found in ancestral segments identified by
 438 RFMix (12fold = 1.21, p-value = 5.00e-03, q-value = 1.00e-02; 10,000 permutation test).

439 We then looked closer into the 25 genes that had high-frequency derived missense alleles in the Yakut
 440 genomes (Fig 3). A gene with the largest number of such variants was the actin-binding *SPTBN5*
 441 gene containing 21 missense mutations. Not surprisingly, the region containing the *SPTBN5* also
 442 demonstrated the highest F_{ST} value between the Yakut cattle and all taurine breeds from the Run 7 of
 443 the 1000 Bull Genome project suggesting that this region is quite different between the taurine breeds
 444 and the Yakut cattle. When all breeds of the Run 7 were contrasted against the Yakut cattle *SPTBN5*
 445 was still on top of the list (top fourth gene position, with the BOLA-DQB MHC class II antigene
 446 being the number one; Supplementary Information 7). We found high-frequency derived alleles in
 447 other actin-binding genes, including the *MYO10*, *CLMN*, and the *NRAP*. Among these the *MYO10*

also contained a unique missense mutation in the Yakut horse (63), suggesting its probable contribution to the local adaptation of distinct domesticated species to harsh Siberian climates.

Yakut-specific NRAP missense mutation

The *NRAP* variant was the only Yakut-cattle specific missense mutation which was not observed in other cattle breeds nor the bovine species in our set. Out of 19 Yakut cattle individuals with high coverage in this region (no. reads >5) 12 were homozygous for the derived allele while seven were heterozygous. We checked this position in the complete 1000 Bull Genome Run 7 dataset (3,767 individuals, indicine and taurine breeds). The frequency of the reference (G) allele was 0.995089 while the frequency of the derived (T) allele was 0.00491107. The only four additional (to our Yakut cattle samples) animals which contained the same derived allele were four Yakut cattle samples obtained from the NCBI SRA archive, suggesting that this genetic variant is Yakut cattle-specific.

The nebulin-related-anchoring protein (NRAP) is a highly evolutionary conserved actin-binding cytoskeletal protein expressed exclusively in the skeletal and cardiac muscles, and is involved in myofibrillar assembly and force transmission in the heart (64). Our phylogenetic comparison shows that the H100Q variant observed in the Yakut cattle is rare in other mammals (found in 16/113 species), but is preferably present in species that either hibernate, e.g. little brown bat and other hibernating bats and mouse lemur or enter seasonal torpor e.g., cape golden mole and tree shrew. Our statistical analysis using phylogenetic logistic regression (65) considering phylogeny of the 113 species from nine mammalian orders (66) shows phylogeny-independent significant correlation (R^2 0.77, p-value = 2.72e-06, alpha 0.00011) between the presence of the glutamine (Gln, Q) variant and cold adaptation, hibernation or ability to enter torpor (Fig. 6, Table 3). The four species that don't hibernate or enter torpor but still have the glutamine variant are the walrus, sea lion and two seals. Interestingly seals and sea lions are capable of slowing heart rate to several beat per minute while diving (67, 68) while walrus is cold adapted suggesting that the same adaptation as in the Yakut cattle could be found in other large animals.

474 There are multiple independent lines of evidence suggesting that the highly evolutionarily conserved
 475 NRAP protein could contribute to climate adaptations in various animal groups. For instance, in the
 476 Australian bearded dragon *NRAP* is differentially expressed in the skeletal muscle and heart before
 477 and during hibernation (69) while in the alligator *NRAP* expression was downregulated in the heart
 478 in response to hypoxia (70). In the glass eel fixation of an alternative *NRAP* allele was associated
 479 with latitude and river water temperature in a landscape genomics study (71). Interestingly, in a recent
 480 environmental GWAS study (72) variants in the *NRAP* gene were associated with environmental
 481 adaptations in the US Simmental cattle.

482 Human and mice studies provide a key to likely mechanism of the *NRAP* contribution to cold climate
 483 adaptations. In mice *NRAP* is downregulated during dilated cardiomyopathy (DCM) (73); human
 484 patients with DCM express homozygous *NRAP* mutations (64). During DCM, pumping ability of the
 485 left ventricle of the heart increases leading to issues with pumping blood out of the heart (74)
 486 suggesting that the heart force transmission function of *NRAP* in the Yakut cattle could lead to more
 487 efficient blood pumping during the winter period therefore facilitating energy saving similar to
 488 species that either lower metabolic rates during the winter period, enter torpor or hibernate (Fig. 6
 489 and Fig. 7).

490

491 ***Copy number variant detection in the Yakut, Kholmogory, Holstein, and Hanwoo breeds***

492 Copy number variants (CNVs) together with SNPs are a major source of genomic variation
 493 contributing to evolution and adaptation (75). To identify CNVs and CNV regions (CNVRs) which
 494 could co-evolve with adaptation of the Russian cattle breeds to harsh climate conditions, we utilized
 495 sequences of four breeds (Yakut, Kholmogory, Holstein, and Hanwoo). A total of 60,338 autosomal
 496 CNVs were detected in the four-breed set, which were then merged into 17,736 CNVRs. The largest
 497 shared fraction of Kholmogory CNVRs was with Holstein (2.48 Mbp) and Yakut CNVRs with
 498 Hanwoo (2.51 Mbp), confirming known breed relations (Fig. 8A). A total of 588 CNVRs (total
 499 length: 2.42 Mbp) were breed-specific (had <10% length overlap with the CNVRs from other three

500 breeds) for the Yakut, 743 (3.73 Mbp) for the Kholmogory, 1,408 (5.42 Mbp) for the Hanwoo and
501 2,464 (8.97 Mbp) for Holstein cattle (Supplementary Information 8a, Fig. 8a). To check if the unique
502 CNVRs in the Russian (especially Yakut) cattle could result from ancestral taurine genome, we
503 additionally identified 9,960 CNVRs in 26 samples of pure indicine (African, Chinese, and Indian)
504 breeds. Overall, the Yakut and Kholmogory cattle had the same fraction of their unique CNVR
505 lengths overlapping the indicine CNVRs (34%), while Hanwoo had the largest fraction (51%) and
506 Holstein had the lowest (23%; Supplementary Information 8a).

507 To reveal possible contribution of CNVRs to the breed-specific biology and adaptations we looked
508 into 1,887 genes found in the four breed-specific CNVRs of which 1,261, 279, 208, and 139 genes
509 were found in the Holstein, Hanwoo, Kholmogory, and Yakut CNVRs, respectively (Supplementary
510 Information 9a). Gene Ontology (GO) enrichment analysis highlighted distinct pathways being
511 enriched in these gene sets, with the largest number (25) found in the Holstein cattle (Supplementary
512 Information 9a). The Yakut cattle had the second largest number of GO categories (four) of which
513 *regulation of actin filament-based processes*, and *antigen processing and presentation* could
514 potentially contribute to adaptation to local environments. The Kholmogory showed enrichment in
515 two GO categories, *G-protein coupled receptor signalling pathway* and *system process*, while
516 Hanwoo had only one GO category enrichment, the *ion binding*. There were 299 CNVRs shared
517 between the Yakut and Kholmogory breeds, of which the CNVR covering >90% of *ACTG1* (actin
518 gamma 1) gene was the most frequent (15% and 11% in Kholmogory and Yakut, respectively).
519 Another actin gene, *ACTA1* (alpha-skeletal actin) was initially reported in the Yakut-specific CNVR,
520 with one Kholmogory individual having CNV in the same region. A manual check of BAM files
521 revealed that more Kholmogory samples followed the same increased coverage/heterozygosity
522 pattern, suggesting that this CNV is more widely spread in these two breeds, while none of the
523 Holstein and Hanwoo samples showed this pattern and had actin genes in their CNVRs. The gene
524 network analysis (Supplementary Fig. 3) shows that *NRAP* affected by the Yakut-specific missense
525 mutation (see above) directly interacts with *ACTA1* gene supporting assembly of actin filaments in

both skeletal muscles and the heart. Additional Yakut cattle-specific CNVRs involved multiple olfactory receptor genes, *CYP4A11* (cytochrome P-450 4A11) involved in lipogenesis and growth traits. This CNVR has been previously reported in various Chinese native taurine cattle breeds (Jaxian, Quinchuan, Nanyang, Jinnan, Luxi, and Chinese Red Steppe (76)).

The Kholmogory cattle unique CNVRs covered 54 annotated genes among which 17 were olfactory receptors, *DEFB1* (defensin beta 1) antimicrobial peptide, *IFITM2* (interferon induced transmembrane protein) both involved in the bovine innate immunity, *PLA2G2D3* (phospholipase) involved in inflammation and immune response, *C8G* (complement C8 gamma chain) belonging to the lipocalin family and participating in the formation of the membrane attack complex, *MYLPF* (myosin light chain) associated with meat tenderness and intramuscular fat (77), *SHOX* (short stature homeobox) regulating elements of growth and development and its mutations are associated with skeletal atavism (78).

Among the genes in the breed-specific CNVRs overlapping indicine CNVRs (356 (1.45%)) 112, 109, 81, and 54 were present in the Holstein, Hanwoo, Kholmogory, and Yakut breeds, respectively (Supplementary Information 8b). GO enrichment analysis highlighted the term “*detection of stimulus*” in both Yakut and Kholmogory breeds, and the term “*MHC protein complex*” in the Yakut cattle (Supplementary Information 10a).

Genomic regions under selection in world northernmost cattle breeds

In an attempt to decipher genomic regions under selection in Russian cattle breeds and thus better understand their unique adaptations, we utilized two independent approaches: one focusing on the identification of haplotypes subjected to selection in a population (HapFLK) and another looking for the ratio of non-synonymous to synonymous substitutions within a gene (dN/dS) while considering the background genome mutation rate to differentiate between the signatures of positive and negative selection. In the latter approach, we first analyzed all SNPs in each breed (MAF ≤ 0.1) and then reanalyzed the dataset focusing on SNPs present in more than half chromosomes in each population

(MAF $\geq$ 0.6). In fact, we were interested in finding genes under selection with a likely effect on the whole population rather looking at all variation within a breed. Indeed, most genes found to be under selection when considering a more stringent allele frequency were also reported under selection when more SNPs were considered (Supplementary Information 11). In addition, we used information about 13,257 high-frequency SNPs with derived alleles in Yakut cattle to support our findings together with the top 1% fixation index of genetic differentiation (F_{ST}) intervals in Yakut and Kholmogory breeds calculated against the *pure taurine*, *all taurine* and *all breeds* from the 1000 Bull Genome Project. Calculation of functional category enrichments within these sets were independently estimated using the GO categories, DAVID functional clusters, and the Reactome. Overall, we identified 282 regions under selection in the Yakut and 219 in the Kholmogory cattle breeds using the HapFLK analyses (q-value <0.01 ; Supplementary Information 12), four and three genes under positive and 247 and 62 genes under negative selection using the dN/dS in the Yakut and Kholmogory breeds, respectively when applying the 60% MAF filter (Supplementary Information 11b). There were 14 and four genes in the Yakut and Kholmogory cattle, respectively from the HapFLK selected regions which were also picked up by the dN/dS analysis. Of these, genes *SPTBN5* was also found in the top 1% F_{ST} intervals in the Yakut cattle, while *HIDI* (involved in regulation of neuropeptide sorting) in the Kholmogory cattle.

As previously shown, adaptation to harsh environments requires proper responses to external stimuli (79). Both the Yakut and Kholmogory breeds demonstrated significant enrichment for the *response to stimulus* GO category in the HapFLK regions, supported further by the dN/dS analysis of the Yakut cattle genes. Interestingly, in the Yakut cattle this GO term was significantly overrepresented in genes containing high-frequency SNPs absent from the European taurines, suggesting that the genetic profile of selection between the two breeds is different. One of the strongest HapFLK signals in the Yakut cattle was observed for the gene *ZNF622*, which was recently established as an antiviral protein in humans (80). Glycophorins (*GYPB* and *GYP A*), which determine MN and Ss blood types in humans (81) and define resistance to parasites, including malaria (82), demonstrated high F_{ST} between

the Yakut cattle and the pure taurine cattle with the *GYPB* being on top of the F_{ST} gene list (Supplementary Information 7) and the positive selection list in Yakut cattle from the dN/dS analysis (MAF ≤ 0.1 ; Supplementary Information 11a). Our finding of a convergent *NRAP* amino-acid change in the Yakut cattle shared with multiple cold-adapted/hibernating species suggesting that the Yakut cattle could be capable of slowing metabolism during the coldest winter periods, was supported by the GO *metabolic process* (q-value = 0.00; Supplementary Information 13a) and *integration of energy metabolism* (q-value = 0.04; Supplementary Information 13c) Reactome terms enrichment in the HapFLK Yakut gene set. This adaptation would require changes to the heart's ability to distribute blood efficiently despite cold environmental temperatures. Indeed, the *cardiac conduction* Reactome pathway was overrepresented (q-value = 4.94E-05; Supplementary Information 14a and Fig. 9) in the dN/dS gene set supported by the DAVID overrepresented cluster (enrichment score 2.00; Supplementary Information 15b) combining terms *cardiac muscle contraction*, *dilated cardiomyopathy*, *hypertrophic cardiomyopathy*, etc. when the Yakut cattle F_{ST} was calculated against the pure taurine breeds from the 1000 Bull Genome Project dataset and the same cluster with enrichment score of 1.32 when compared to all breeds (Supplementary Information 15b). These categories included a common gene *RYR2*, which is central to myocardial excitation contraction coupling (83).

An important organ contributing to thermogenesis is the brown adipose tissue (84, 85). In the Yakut cattle the third strongest HapFLK signature of selection was in the area of the pleiotrophin (*PTN*; q-value = 4.42E-10; Supplementary Information 12), a key gene in preserving insulin sensitivity, driving adipose tissue lipid turnover and the regulator of energy metabolism and thermogenesis (86). In line with this observation we found that the *regulation of insulin secretion* term was enriched (q-value = 0.04; Supplementary Information 14c) in the Yakut dN/dS gene set (Fig. 9). This was further supported by the HapFLK analysis highlighting the Reactome categories *regulation of insulin secretion* (q-value = 0.01) and *integration of energy metabolism* (q-value = 0.04; Supplementary Information 13c). Intriguingly, we identified two out of three candidate genes for cold/diet adaptation

in native Siberian human populations (*PLA2G2A* and *ANGPTL8*) (87) involved in lipid metabolism in the Yakut cattle HapFLK set. Another mechanism of reaction to extreme cold temperatures in mammalian cells is the cytoskeletal stabilization and/or disassembly including microtubules and actin filaments. In hibernating mammals this is required for a rapid cytoskeleton reorganization during return from torpor to euthermia (88). We found multiple cytoskeleton proteins in selected intervals of the Yakut cattle genome, including highly-enriched DAVID clusters of *spectrin/alpha-actinin*, etc. (enrichment score 6.22); *myosin*, *actin-binding*, *microfilament motor activity*, *actin filament binding*, etc. (enrichment score 5.14) in the dN/dS set (Supplementary Information 14b). Genes with the highest selection signals included actin-binding *SPTBN5* and *MYO10*, both of which were highlighted by our high-frequency SNP analysis. The *SPTBN5* is among 12 genes under selection in multiple Arctic and Antarctic species (18), while *MYO10* contained Yakut horse specific mutations (63). Interestingly, the actin-binding genes were also enriched in selected intervals of Kholmogory and Hanwoo breeds, but at much lower enrichment rates (enrichment score 2.95 in Kholmogory and 1.92 in Hanwoo) and the gene sets were different to Yakut cattle (Supplementary Information 14b). Another related group of cytoskeleton-related proteins under selection in the Yakut cattle were *pleckstrin homology-like domain* proteins in the dN/dS (enrichment score 2.80; Supplementary Information 14b) and HapFLK (enrichment score 1.63; Supplementary Information 13b) analyses. These domains are found in proteins that link the cytoskeleton to the cell membrane (89). One of these genes, *FARPI* shows the highest HapFLK signal in the Yakut cattle (q-value = 0.0; Supplementary Information 12). An important organ of thermoregulation is the brain. It was shown (90) that the extracellular component proteins play role in axonal guidance, synaptogenesis (collagens) and were overexpressed in the cerebral cortex during deep hibernation of ground squirrel. A similar pattern was observed in hypothalamic transcriptome of Djungarian hamster during torpor (91). The Reactome analysis of the dN/dS gene set highlighted multiple extracellular matrix terms enriched in the Yakut cattle, including *elastic fibre formation* (enrichment 3.48; q-value = 0.01); *laminin interactions* (enrichment 7.79; q-value = 3.2E-06), *assembly of collagen fibrils and other*

630 *multimeric structures* (enrichment 5.05; q-value = 5.16E-04), *collagen formation* (enrichment 3.60;
631 q-value = 0.01), etc. (Fig. 9; Supplementary Information 14c).

632 *Gene enrichment for Yakut SNP set absent or in low frequency in the European taurine cattle.* To
633 understand probable contribution of the alleles, absent from European taurines, we looked for GO
634 enrichment in the 13,257 SNPs with high frequency in the Yakut cattle. The *actin cytoskeleton* (q-
635 value = 0.0) category was overrepresented in the set of SNPs that were found at low frequency
636 ($\leq 10\%$) in the European taurines, while the *response to stimulus* and *metabolic process* categories
637 were enriched (q-value = 0.04) in the set of SNPs absent in the European taurine breeds
638 (Supplementary Information 16a). We observed that SNPs absent in the European taurines but found
639 in bovine species were enriched in the GO category “*phospholipase A2 activity*” (q-value = 0.0), the
640 Reactome pathway *Acyl chain remodelling of PG* (enrichment score = 3.7; q-value = 0.05;
641 Supplementary Information 16c), and “*MHC class I peptide loading complex*” (q-value = 0.0).
642 Interestingly the genes containing Yakut-specific SNPs were enriched in *biological regulation* (q-
643 value = 0.0; Supplementary Information 16a).

644 Discussion

645 Native cattle breeds represent an important cultural heritage, often sharing common history with local
 646 human populations. Local breeds could also contain genetic variations useful to properly respond to
 647 agricultural needs in light of ongoing climatic changes. They could help understand evolutionary
 648 processes that occur in response to human led selection and adaptation of cattle to different
 649 (sometimes drastic) environmental conditions. In this study we looked into the genetic history of two
 650 unique world northernmost native cattle breeds from Russia: the Kholmogory, and the Yakut cattle.
 651 We revealed the molecular layout for distinct genetic profiles of European and Asian taurine cattle
 652 lineages, spotted a recent functional convergent evolutionary change in the Yakut cattle genome in
 653 response to exposure to climates of Siberia, found candidate genes and pathways which could
 654 contribute to cattle adaptations to cold climates, and offered a set of SNP markers for the industry to
 655 be tested for genomic selection of cattle in the Northern countries.

656 Comparison of demographic histories of the Yakut, Kholmogory, and Holstein breeds suggests an
 657 early separation of the Yakut cattle from European taurines (represented by Kholmogory and
 658 Holstein). The estimated separation time (4,900 years ago) agrees with Payne and Hodges (92) who
 659 estimated that domesticated cattle appeared 4,000-5,000 years ago in the Northern East Asia,
 660 Northern China, Korea and Japan (92). As this is a post-domestication event with the domestication
 661 itself dated to ~8,500-10,000 years ago (1, 2), our data supports a single taurine cattle domestication
 662 leading to both the European and Asian taurines. Our results, however, point to a much smaller N_e
 663 size of the ancestral Asian taurine population compared to a much larger N_e of the population leading
 664 to European taurine breeds. This could be due to known historical admixture of European cattle
 665 population with wild aurochs (93) until the European cattle's N_e reduced ~2,500 years ago and
 666 possibly could account for phenotypic differences between the extant Turano-Mongolian cattle and
 667 European taurines. Interestingly, there is a slight gradual increase of the N_e of the Yakut cattle lineage
 668 until its drastic decrease ~300 YA, which approximately matches the period of the Yakutian history
 669 when in 1620s they came in contact with the expanding Tsardom of Muscovy. The imposed fur tax

670 led to several suppressed Yakutian rebellions between 1634 and 1642, but more importantly, there is
671 evidence that the Yakutian population went down by ~70% between 1642 and 1682 due to smallpox
672 and other introduced infectious diseases (94). Our data suggest that the Yakut cattle population could
673 have gone through a severe bottleneck at nearly the same time corroborating the event in the human
674 history.

675 A high fraction of SNPs with high frequency in the Yakut cattle, absent from the European taurine
676 breeds, but present in the indicine cattle, could imply historical introgression of these alleles into the
677 Yakut cattle. However, the fact that vast majority of these alleles were also present in different
678 bovine species means that these variants could belong to the pool of ancestral *Bos primigenius* alleles
679 shared with other *Bos* species. Indeed, we identified a large fraction of alleles with high frequency in
680 the Yakut cattle, but with low frequency in European taurines and Hanwoo, suggesting that these
681 alleles were segregating in the ancestral domesticated population more than 5,000 years ago. The fact
682 that the Yakut and Kholmogory (a pure taurine) cattle had nearly the same fraction of CNVRs
683 overlapping CNVRs from indicine breeds provides another line of evidence against significant
684 introgression of Yakut and indicine cattle. Another possible scenario which could potentially explain
685 high genetic diversity of Yakut cattle at these positions is an ancient introgression with the East Asian
686 aurochs (*Bos primigenius*) that lived in the East Asian region during the arrival of taurine cattle
687 together with several other bovine species, including yak, banteng, and gaur (3, 4, 28). Kantanen et
688 al. (95) have indeed suggested the near-eastern origins of the ancestral population of the Yakut cattle
689 breed by studies on mtDNA and Y-chromosomal diversity. This scenario, however, cannot explain
690 large fraction of SNPs with a high frequency in the Yakut cattle, indicine cattle, and African taurines,
691 and with a low frequency in European taurines. The fact that nearly two-thirds of missense alleles
692 absent from the European taurines are more similar to orthologous amino acids found in majority of
693 other phylogenetically distinct animal species also implies that these variants could represent the
694 ancestral state. We hypothesise that these putatively ancestral variants could be a rich reservoir of
695 genetic variation important for adaptation of Yakut to extreme cold environment. This is indeed

696 supported by their enrichment in genes from *response to stimulus/pain* and *immune system* categories
697 potentially involved in adaptation to harsh environment. The fact that the *response to stimulus*
698 category is under selection in another Northern Russian breed, the Kholmogory, confirms similar
699 mechanisms of adaptation to cold climates in cattle breeds. Interestingly, the GO category response
700 to stimulus was found to be enriched in a recent whole genome resequencing study of the Finnish and
701 Yakut cattle breeds in the list of genes with large number of missense mutations (96). A large fraction
702 of genes in selected regions in the Yakut cattle genome belong to the actin-binding category; the same
703 category was found enriched in the selected regions of the Kholmogory cattle, suggesting that these
704 genes could contribute to cold adaptation as it was shown previously in (38, 88, 97). This is supported
705 by the fact that the *ACTA1* (alpha actin 1) was the most frequent Yakut CNVR and also present in
706 Kholmogory cattle, while this CNVR was absent from all Holstein and Hanwoo samples. Overall,
707 the CNVRs found in the Yakut cattle were significantly enriched for actin-binding genes, suggesting
708 that this group of genes has been evolved in the Yakut cattle. Another actin-binding gene *SPTBN5*,
709 which had the largest number of high-frequency missense mutations in Yakut cattle (21) compared
710 to other taurine breeds, was shown to be under positive selection in at least three Arctic and Antarctic
711 mammals including the woolly mammoth (18, 98) and in Chinese indicine cattle breeds (27); *MYO10*
712 gene which has unique missense mutation in the Yakut horse (63), suggesting its probable
713 contribution to local adaptation of distinct domesticated species to harsh Siberian climates. Our
714 findings are confirmed by the work of Weldenegodguad et al. (96) who identified genes with >5
715 missense mutations in the Yakut cattle genome from five samples, and both the *SPTBN5* and *MYO10*
716 were reported in their list; but the limited number of animals in their study was not sufficient to
717 estimate signatures of selection and frequencies of derived alleles in the Yakut cattle genome.
718 While the actin genes themselves and genes interacting with actin could contribute to multiple
719 biological processes related to adaptation to cold climates (97) the only high-frequency Yakut-cattle
720 specific missense mutation found in the actin-binding gene *NRAP* provides a clue for the likely
721 mechanism. The fact that this H100Q variant, found in every single resequenced Yakut cattle animal

was absent from all ~3,800 cattle animals from the 1000 Bull Genome Project, including ancient aurochs samples, suggests its Yakut cattle origin. Moreover, its absence from all bovine species used in our work as well, suggests its relatively recent appearance in the Yakut cattle dating back to maximum 5,000 years ago, after separation from the European taurine branch. However, as the Yakut people have migrated to Yakutia in the 13th century and other Asian, including the Turano-Mongolian breeds (Hanwoo), lack this variant as well it could form as late as ~800 years ago in Yakutia followed a very fast expansion in the Yakut cattle. While absent in cattle breeds and phylogenetically close bovine species, the exact same amino acid change has formed independently at least six times in mammalian evolution, in species lowering their metabolism in response to cold, including hibernating bats and lemurs, cold adapted species, like walrus and sea lions. Interestingly, a potential contribution of a highly evolutionary conserved NRAP protein (99) to climate adaptation is supported by findings in a range of species including fish and reptiles (69–71). It is expressed exclusively in the skeletal and cardiac muscles and is involved in myofibrillar assembly and force transmission in the heart (64), suggesting that the mechanism of Yakut cattle adaptation could be related to the function of the heart. In human and mice mutations in *NRAP* lead to dilated cardiomyopathy (73), a heart muscle genetic disorder, causing enlargement of ventricles and issues with pumping blood out of the heart. We hypothesise that the *NRAP* variant plays a crucial role in slowing the metabolism during the coldest periods in the Yakut cattle and other species, which would require enhanced heart ability to distribute blood at cold environmental temperatures. This is supported by strong enrichment for terms related to the cardiac contraction pathway in the Yakut cattle selected regions and the metabolic process and energy metabolism as well.

Scans for selection signatures pointed to multiple candidate genes and pathways for climate adaptations in Russian cattle breeds. A surprising result was that the Yakut cattle contains a large number of genes with signatures of negative selection and a very few with signatures of positive selection according to our dN/dS analysis. A possible explanation could be that the ancestral variants present in the Yakut cattle genome already make them a good fit into their environmental conditions,

748 therefore a very few novel functional changes (like one in the *NRAP* gene) are required. Alternatively,
749 it is possible that some additional functional changes could affect either regulatory gene regions or
750 be represented by a different type of mutations not studied in this work, e.g. by indels. Genes with
751 the highest dN/dS selection signals in the Yakut cattle included genes involved in the cytoskeleton
752 stabilization which plays an important role in the hibernating mammals during return from torpor to
753 euthermia (88), pointing to another possible common mechanism of reaction to extreme cold
754 temperatures in the Yakut cattle and other animals. One of the top ranked regions under selection in
755 the Yakut cattle contains the *PTN* a key gene in preserving insulin sensitivity, driving adipose tissue
756 lipid turnover and the regulator of energy metabolism and thermogenesis (86). Additionally, we
757 identified two candidate genes for cold/diet adaptation in native Siberian human populations
758 (*PLA2G2A* and *ANGPTL8*) involved in lipid metabolism (87) in selected intervals of the Yakut cattle
759 genome suggesting common mechanisms of adaptation to Siberian climates in mammals.

760 In conclusion, this work has revealed different evolutionary histories of two Russian northernmost
761 cattle breeds, of which one is a typical taurine cattle and another contains multiple ancestral genetic
762 variants, which allowed its adaptation to harsh conditions of living above the Polar Circle.
763 Surprisingly, very few truly novel high-frequency genetic variants were required for this extreme
764 adaptation, but the only detected missense change of this type represents a unique example of a young
765 amino acid residue convergent change limited to a single cattle breed but shared with at least 16
766 species of hibernating/cold adapted mammals from nine distinct phylogenetic orders.

Acknowledgements

This work was funded in part by the Russian Foundation for Basic Researches (RFBR) grant 18-016-00185 to DML. LB was funded by Marie Skłodowska-Curie grant agreement No. 703376. The identification and functional annotation of high-frequency breed-specific variants was supported by the Kurchatov Genomics Center of IC&G (075-15-2019-1662). The authors thank the participants of the 1000 Bull Genomes Project for providing sequence data, Michael Hiller and Emma Teeling for help with the 120-mammalian genome alignments.

References

1. R. T. Loftus, D. E. MacHugh, D. G. Bradley, P. M. Sharp, P. Cunningham, Evidence for two independent domestications of cattle. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 2757–2761 (1994).
2. D. M. Larkin, N. S. Yudin, The genomes and history of domestic animals. *Mol. Genet. Microbiol. Virol.* **31**, 197–202 (2016).
3. J. E. Decker, *et al.*, Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLOS Genet.* **10**, e1004254 (2014).
4. I. Medugorac, *et al.*, Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat. Genet.* **49**, 470–475 (2017).
5. R. Sinha, *et al.*, The Impact of Climate Change on Livestock Production and Reproduction: Ameliorative Management. *Int. J. Livest. Res.*, 1 (2017).
6. U. Bernabucci, *et al.*, The effects of heat stress in Italian Holstein dairy cattle. *J. Dairy Sci.* **97**, 471–486 (2014).
7. D. Renaudeau, *et al.*, Adaptation to hot climate and strategies to alleviate heat stress in livestock production. *Anim. Int. J. Anim. Biosci.* **6**, 707–728 (2012).
8. A. V. Igoshin, *et al.*, Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. *BMC Genet.* **20** (2019).
9. M. Felius, *et al.*, On the History of Cattle Genetic Resources. *Diversity* **6**, 705–750 (2014).
10. S. A. Crate, *Cows, Kin, and Globalization: An Ethnography of Sustainability* (Rowman Altamira, 2006).
11. A. Yurchenko, *et al.*, Genome-wide genotyping uncovers genetic profiles and history of the Russian cattle breeds. *Heredity* **120**, 125–137 (2018).
12. H. Mannen, *et al.*, Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. *Mol. Phylogenet. Evol.* **32**, 539–544 (2004).
13. M.-H. Li, *et al.*, The genetic structure of cattle populations (*Bos taurus*) in northern Eurasia and the neighbouring Near Eastern regions: implications for breeding strategies and conservation. *Mol. Ecol.* **16**, 3839–3853 (2007).
14. T. RUF, F. GEISER, Daily torpor and hibernation in birds and mammals. *Biol. Rev. Camb. Philos. Soc.* **90**, 891–926 (2015).
15. M. E. Lazzeroni, F. T. Burbrink, N. B. Simmons, Hibernation in bats (Mammalia: Chiroptera) did not evolve through positive selection of leptin. *Ecol. Evol.* **8**, 12576–12596 (2018).
16. J. G. Boyles, *et al.*, A global heterothermic continuum in mammals: Global patterns of heterothermy in mammals. *Glob. Ecol. Biogeogr.* **22**, 1029–1039 (2013).
17. B. G. Lovegrove, The evolution of endothermy in Cenozoic mammals: a plesiomorphic-apomorphic continuum. *Biol. Rev. Camb. Philos. Soc.* **87**, 128–162 (2012).
18. N. S. Yudin, D. M. Larkin, E. V. Ignatieva, A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. *BMC Genet.* **18** (2017).
19. C. Verde, E. Cocca, D. de Pascale, E. Parisi, G. di Frisco, Adaptations and lifestyle in polar marine environments: a biological challenge for the study of fish evolution. *Polar Res.* **23**, 3–10 (2004).
20. A. A. Yurchenko, *et al.*, Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Sci. Rep.* **8**, 1–16 (2018).
21. K. C. Wollenberg Valero, *et al.*, A candidate multimodal functional genetic network for thermal adaptation. *PeerJ* **2**, e578 (2014).
22. Y. Liu, *et al.*, Convergent sequence evolution between echolocating bats and dolphins. *Curr. Biol. CB* **20**, R53–54 (2010).
23. J. Hill, *et al.*, Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proc. Natl. Acad. Sci.* **116**, 18473–18478 (2019).

24. A. Martinez Barrio, *et al.*, The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *eLife* **5** (2016).
25. H. D. Daetwyler, *et al.*, Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **46**, 858–865 (2014).
26. D.-H. Shin, *et al.*, Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. *BMC Genomics* **15**, 240 (2014).
27. C. Mei, *et al.*, Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. *Mol. Biol. Evol.* (2017) <https://doi.org/10.1093/molbev/msx322>.
28. N. Chen, *et al.*, Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* **9**, 1–13 (2018).
29. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
30. M. A. DePristo, *et al.*, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
31. W. McLaren, *et al.*, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
32. I. Martincorena, *et al.*, Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029–1041.e21 (2017).
33. K. D. Pruitt, T. Tatusova, G. R. Brown, D. R. Maglott, NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* **40**, D130–D135 (2012).
34. J. R. Grant, A. S. Arantes, X. Liao, P. Stothard, In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. *Bioinforma. Oxf. Engl.* **27**, 2300–2301 (2011).
35. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
36. E. Noskova, V. Ulyantsev, K.-P. Koepfli, S. J. O’Brien, P. Dobrynin, GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. *GigaScience* **9** (2020).
37. T. Derrien, *et al.*, Fast Computation and Applications of Genome Mappability. *PLOS ONE* **7**, e30377 (2012).
38. A. A. Yurchenko, *et al.*, Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Sci. Rep.* **8**, 12984 (2018).
39. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genet.* **5**, e1000695 (2009).
40. J. Jouganous, W. Long, A. P. Ragsdale, S. Gravel, Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation. *Genetics* **206**, 1549–1567 (2017).
41. B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013).
42. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
43. N. Patterson, *et al.*, Ancient Admixture in Human History. *Genetics* **192**, 1065–1093 (2012).
44. G. Klambauer, *et al.*, cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* **40**, e69 (2012).
45. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
46. S. Neph, *et al.*, BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
47. M. I. Fariello, S. Boitard, H. Naya, M. SanCristobal, B. Servin, Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* **193**, 929–941 (2013).

48. P. Scheet, M. Stephens, A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78**, 629–644 (2006).
49. M.-I. Fariello, *et al.*, Selection signatures in worldwide sheep populations. *PloS One* **9**, e103813 (2014).
50. S. Boitard, M. Boussaha, A. Capitan, D. Rocha, B. Servin, Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. *Genetics* **203**, 433–450 (2016).
51. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).
52. C. Greenman, R. Wooster, P. A. Futreal, M. R. Stratton, D. F. Easton, Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics* **173**, 2187–2198 (2006).
53. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
54. D. Smedley, *et al.*, BioMart--biological queries made easy. *BMC Genomics* **10**, 22 (2009).
55. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
56. A. Fabregat, *et al.*, Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics* **18**, 142 (2017).
57. A. Heger, C. Webber, M. Goodson, C. P. Ponting, G. Lunter, GAT: a simulation framework for testing the association of genomic intervals. *Bioinformatics* **29**, 2046–2048 (2013).
58. , UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
59. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
60. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
61. M. Farré, *et al.*, Evolution of gene regulation in ruminants differs between evolutionary breakpoint regions and homologous synteny blocks. *Genome Res.* **29**, 576–589 (2019).
62. A. A. Sermyagin, *et al.*, Whole-genome SNP analysis elucidates the genetic structure of Russian cattle and its relationship with Eurasian taurine breeds. *Genet. Sel. Evol. GSE* **50** (2018).
63. P. Librado, *et al.*, Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc. Natl. Acad. Sci.* **112**, E6889–E6897 (2015).
64. G. T. Truszkowska, *et al.*, Homozygous truncating mutation in NRAP gene identified by whole exome sequencing in a patient with dilated cardiomyopathy. *Sci. Rep.* **7** (2017).
65. L. si Tung Ho, C. Ané, A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. *Syst. Biol.* **63**, 397–408 (2014).
66. N. Hecker, M. Hiller, A genome alignment of 120 mammals highlights ultraconserved element variability and placenta-associated enhancers. *GigaScience* **9** (2020).
67. T. M. Williams, *et al.*, Exercise at depth alters bradycardia and incidence of cardiac anomalies in deep-diving marine mammals. *Nat. Commun.* **6**, 6055 (2015).
68. J. C. McKnight, *et al.*, Shining new light on mammalian diving physiology using wearable near-infrared spectroscopy. *PLOS Biol.* **17**, e3000306 (2019).
69. A. Capraro, *et al.*, Waking the sleeping dragon: gene expression profiling reveals adaptive strategies of the hibernating reptile *Pogona vitticeps*. *BMC Genomics* **20**, 460 (2019).
70. S. L. Alderman, D. A. Crossley, R. M. Elsey, T. E. Gillis, Hypoxia-induced reprogramming of the cardiac phenotype in American alligators (*Alligator mississippiensis*) revealed by quantitative proteomics. *Sci. Rep.* **9** (2019).
71. P.-A. Gagnaire, E. Normandeau, C. Côté, M. Møller Hansen, L. Bernatchez, The genetic consequences of spatially varying selection in the panmictic American eel (*Anguilla rostrata*). *Genetics* **190**, 725–736 (2012).

72. T. N. Rowan, *et al.*, Powerful detection of polygenic selection and environmental adaptation in US beef cattle populations. *bioRxiv*, 2020.03.11.988121 (2020).
73. E. Ehler, *et al.*, Alterations at the Intercalated Disk Associated with the Absence of Muscle Lim Protein. *J. Cell Biol.* **153**, 763–772 (2001).
74. H.-P. Schultheiss, *et al.*, Dilated cardiomyopathy. *Nat. Rev. Dis. Primer* **5** (2019).
75. G. E. Liu, *et al.*, Analysis of copy number variations among diverse cattle breeds. *Genome Res.* **20**, 693–703 (2010).
76. M. Yang, *et al.*, Association study and expression analysis of CYP4A11 gene copy number variation in Chinese cattle. *Sci. Rep.* **7**, 46599 (2017).
77. L. H. P. Silva, *et al.*, Explaining meat quality of bulls and steers by differential proteome and phosphoproteome analysis of skeletal muscle. *J. Proteomics* **199**, 51–66 (2019).
78. M. Carneiro, *et al.*, Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074–1079 (2014).
79. B. M. Beker, B. M. Beker, Human Physiology in Extreme Heat and Cold <https://doi.org/10.23937/iacph-2017/1710001> (May 4, 2020).
80. K. Mun, T. Punga, Cellular Zinc Finger Protein 622 Hinders Human Adenovirus Lytic Growth and Limits Binding of the Viral pVII Protein to Virus DNA. *J. Virol.* **93** (2019).
81. O. O. Blumenfeld, C. H. Huang, Molecular genetics of the glycophorin gene family, the antigens for MNSs blood groups: multiple gene rearrangements and modulation of splice site usage result in extensive diversification. *Hum. Mutat.* **6**, 199–209 (1995).
82. Malaria Genomic Epidemiology Network, G. Band, K. A. Rockett, C. C. A. Spencer, D. P. Kwiatkowski, A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* **526**, 253–257 (2015).
83. H. Milting, *et al.*, Composite polymorphisms in the ryanodine receptor 2 gene associated with arrhythmogenic right ventricular cardiomyopathy. *Cardiovasc. Res.* **71**, 496–505 (2006).
84. K. Townsend, Y.-H. Tseng, Brown adipose tissue. *Adipocyte* **1**, 13–24 (2012).
85. A. Fenzl, F. W. Kiefer, Brown adipose tissue and thermogenesis. *Horm. Mol. Biol. Clin. Investig.* **19**, 25–37 (2014).
86. J. Sevillano, *et al.*, Pleiotrophin deletion alters glucose homeostasis, energy metabolism and brown fat thermogenic function in mice. *Diabetologia* **62**, 123–135 (2019).
87. B. Hallmark, *et al.*, Genomic Evidence of Local Adaptation to Climate and Diet in Indigenous Siberians. *Mol. Biol. Evol.* **36**, 315–327 (2019).
88. K. R. Dave, S. L. Christian, M. A. Perez-Pinzon, K. L. Drew, Neuroprotection: Lessons from hibernators. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **162**, 1–9 (2012).
89. M. Lenoir, I. Kufareva, R. Abagyan, M. Overduin, Membrane and Protein Interactions of the Pleckstrin Homology Domain Superfamily. *Membranes* **5**, 646–663 (2015).
90. C. Schwartz, M. Hampton, M. T. Andrews, Seasonal and Regional Differences in Gene Expression in the Brain of a Hibernating Mammal. *PLOS ONE* **8**, e58427 (2013).
91. C. Cubuk, J. Kemmling, A. Fabrizius, A. Herwig, Transcriptome Analysis of Hypothalamic Gene Expression during Daily Torpor in Djungarian Hamsters (*Phodopus sungorus*). *Front. Neurosci.* **11** (2017).
92. W. J. A. Payne, J. Hodges, Tropical cattle: origins, breeds and breeding policies (1997) (May 11, 2020).
93. M. R. Upadhyay, *et al.*, Genetic origin, admixture and population history of aurochs (*Bos primigenius*) and primitive European cattle. *Heredity* **118**, 169–176 (2017).
94. A. Wood, *The History of Siberia: From Russian Conquest to Revolution* (Routledge, 1991).
95. J. Kantanen, *et al.*, Maternal and paternal genealogy of Eurasian taurine cattle (*Bos taurus*). *Heredity* **103**, 404–415 (2009).
96. M. Weldenegodguad, *et al.*, Whole-Genome Sequencing of Three Native Cattle Breeds Originating From the Northernmost Cattle Farming Regions. *Front. Genet.* **9**, 728 (2018).

- 974 97. S. Banerjee, M. L. Gardel, U. S. Schwarz, The Actin Cytoskeleton as an Active Adaptive
975 Material. *Annu. Rev. Condens. Matter Phys.* **11**, 421–439 (2020).
- 976 98. Z. Lin, *et al.*, Biological adaptations in the Arctic cervid, the reindeer (*Rangifer tarandus*).
977 *Science* **364** (2019).
- 978 99. B. Utama, D. Kennedy, K. Ru, J. S. Mattick, Isolation and characterization of a new nucleolar
979 protein, Nrap, that is conserved from yeast to humans. *Genes Cells* **7**, 115–132 (2002).
- 980 100. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A Resource for Timelines,
981 Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
- 982

983 **Figure legends**

984 **Figure 1.** Demographic history of Yakut and Kholmogory cattle breeds. SMC++-inferred effective
985 population sizes (N_e) with respect to time (generations) for A) Holstein (HOL) and Kholmogory
986 (KHO), B) Yakut and KHO, C) Yakut and HOL, and D) Yakut, HOL, KHO, BRAH (Brahman), and
987 CIND (Chinese indicine). Blue lines highlight separation time. E) Demographic model for Holstein,
988 Kholmogory and Yakut cattle breeds using GADMA.

989

990 **Figure 2.** A) The maximum-likelihood tree generated by TreeMix (up to the migration 12). The scale
991 bar shows ten times the average standard error of the entries in the sample covariance matrix.
992 Abbreviations of the populations are as follow: BIS – Bison, GAU – Gaurus, FRO – Gayal, JAV –
993 Banteng, YKT -Yakut, HNW – Hanwoo, LUX – Luxi, HOL – Holstein, KHO - Kholmogory, YNB
994 – Yanbian, JXR – JaxianRed, NNY – Nanyang, NDM – nDama, ANK – Ankole, DNZ – Dianzhong,
995 KEN – Kenana, OGD – Ogaden, BRH – Brahman, BRN – Boran, JIA - Ji'an, LEI - Leiqiong, WNN
996 – Wannan. B) Residuals of the maximum-likelihood phylogenetic tree. As implemented in TreeMix,
997 the residual covariance between each pair of populations i and j is divided by the average standard
998 error across all pairs. This scaled residual is then plotted in each cell (i,j) . Colours are described in
999 the palette on the right. Residuals above zero represent populations that are more closely related to
1000 each other in the data than in the best-fit tree and thus are candidates for admixture events (SE –
1001 standard error).

1002

1003 **Figure 3.** Forty eight missense SNPs with high frequency alleles present in Yakut cattle ($\geq 70\%$) but
1004 with low frequency of the same allele in Holstein, Kholmogory, and Hanwoo ($\leq 10\%$); shades of red
1005 represent allele frequency and conservation score for Yakut cattle allele encoded amino acids more
1006 often found in orthologous positions of other mammals, while shades of green indicate those Yakut
1007 cattle allele encoded amino acids that are found rarely in other species; TAU – taurine; IND –
1008 indicine.

1009

1010 **Figure 4.** High frequency SNPs in Yakut cattle but low frequency/absence of the same allele in
1011 Holstein, Kholmogory, and Hanwoo (total of 13,257 SNP). Each circle from the periphery to the
1012 centre shows the following datasets: bovine autosomes, indicine-like genome intervals from the
1013 RFMix analysis in different shades of grey (light to dark indicating indicine-like interval frequencies
1014 in the Yakut population ranging from 50% to 100%), Yakut alleles found in bovine species in red,
1015 in African and Chinese taurine breeds in blue, and exclusively in Indian, African and Chinese indicine
1016 breeds in green.

1017

1018 **Figure 5.** GO analysis of genes overlapping ‘indicine-like’ genome intervals ($\geq 60\%$ of the total
1019 gene length) as identified by RFMix.

1020

1021 **Figure 6.** Phylogenetic tree of 114 mammals from nine phylogenetic orders (100). Carriers of the
1022 common histidine (H) amino acid are in blue while the carriers of the glutamine (Q) variant are in red
1023 (cattle has both blue and red indicating the Yakut cattle specific H100Q mutation); the outer circle in
1024 orange highlights species that either hibernate or enter torpor (Supplementary Information 17).
1025 Images are representative of the species that have the glutamine variant, Yakut cattle
1026 (glutamine/histidine) and Kholmogory cattle (histidine).

1027

1028 **Figure 7.** SNPs in the *NRAP* gene. A) Haplotype diversity in the *NRAP* gene for Hanwoo (HNW),
1029 Holstein (HOL), Kholmogory (KHO), and Yakut (YKT) breeds. B) F_{ST} analysis of SNPs from the
1030 1000 Bull Genomes Run 7 in the Yakut population against all other individuals; C) Sensitivity of the
1031 H100Q change modelled in the NRAP protein; D) Extract of local alignment of the exon 4 of the
1032 *NRAP* gene showing the H100Q mutation and adjacent amino acids in the Yakut cattle (Yakut),
1033 reference Hereford genome (Cattle) and seven mammals.

1034

1035 **Figure 8.** Summary of copy number variant regions (CNVRs) in Yakut, Kholmogory, Holstein, and
 1036 Hanwoo breeds. A) intersection plot of CNVRs for the four breeds (HOL – Holstein, HNW –
 1037 Hanwoo, KHO – Kholmogory, YKT – Yakut); colours indicate: black, breed-specific CNVRs; red,
 1038 intersection between two breeds; green intersection among three breeds; blue intersection among all
 1039 four breeds. B) GO analysis of genes in breed-specific CNVRs in Yakut (YKT) and Kholmogory
 1040 (KHO) (FDR<0.05); only unique GO terms present in either Yakut or Kholmogory breeds were
 1041 reported.

1042

1043 **Figure 9.** Reactome pathway enrichment analysis for Yakut cattle genes identified to be under
 1044 selection by the dN/dS analysis (FDR<0.05).

1045 **Tables**

1046 **Table 1.** Summary of samples and genome sequencing statistics (hard filtered biallelic SNP).

	Holstein	Kholmogory	Hanwoo	Yakut	Total
Samples	20	20	20	20	80
Coverage	10.60±5.1	11.68±1.2	10.21±3.13	11.18±0.76	
SNPs	14,448,782	15,426,893	18,914,454	17,025,030	27,487,145
ts/tv	2.13	2.20	2.23	2.23	2.15
Synonymous*	56,249	49,976	56,568	75,798	109,774
Missense*	32,216	31,089	35,740	35,919	64,221
CDS	53,523	63,551	70,590	92,202	137,744
3' UTR	33,537	36,100	45,514	46,886	74,332
5' UTR	6,345	7,419	8,473	8,557	13,814
Intron	2,927,686	3,100,776	3,898,489	3,457,265	5,676,380
Intergenic	11,489,562	12,280,607	14,965,760	13,485,886	21,685,165
MamCNEs	82,533	85,465	108,525	89,593	167,632
CetCNEs	42,643	43,811	56,327	45,934	84,970
RumCNEs	74,576	76,886	100,8716	83,653	153,684

1047 *Synonymous and missense mutations were annotated by dNdScv. Abbreviations are as follows:
1048 CNEs (conserved non-coding elements); MamCNEs (Mammalian CNEs); CetCNEs (Cetartiodactyl
1049 CNEs); RumCNEs (Ruminant CNEs).

Table 2. High-frequency derived alleles in Yakut cattle genomes

SNPs	Statistics
High-frequency allele in Yakut cattle genomes*	13,257
Yakut cattle-specific	263
Low frequency in European taurine genomes**	4,866
Absent from European taurine genomes	8,128
Found in indicine genomes	7,546
Found in bovine genomes	5,760
Found <i>only</i> in African, Chinese, and Asian taurine genomes	2,018
Found <i>only</i> in Hanwoo genomes	184
Found <i>only</i> in indicine genomes	166
Missense alleles	48
Found in Chinese taurine genomes	46
Found in African taurine genomes	42
Found in indicine genomes	39
Found in Hanwoo genomes	37
Absent from European taurine genomes	37
Found in bovine genomes	30
Evolutionary conserved (Yakut cattle allele)	19
Evolutionary conserved (European taurine allele)	10
Yakut cattle-specific	1

*Allele frequency $\geq 70\%$ in Yakut cattle genomes and $< 10\%$ in Hanwoo, Holstein and Kholmogory cattle genomes; ** $< 10\%$ in Holstein, and Kholmogory cattle genomes.

1054 **Table 3.** Association between allelic variants at NRAP residue 100 and habitat across 113
1055 mammalian species.

1056

Residue	Amino acid	No. species		R ²	P value*
		Torpor/hibernation	No torpor/hibernation		
100	<i>His</i>	2**	95	0.77	2.72 x 10 ⁻⁰⁶
	<i>Gln</i>	12	4		

1057 * phylogenetic logistic regression test.

1058 **tenrec – hibernates when it is hot rather than cold and the alpine squirrel. One species, the golden
1059 hamster, has not been included in the analysis as the only hibernating species which shows arginine
1060 (Arg), a descendant of glutamine (Gln) at the NRAP 100 position.

Fig. 1

Fig. 2

Fig. 4

GO: Yakut indicine-like genes (frq>=0.6)

Fig. 5

Fig. 6

Fig. 7

Fig. 8

Reactome pathway analysis

Fig. 9