nature portfolio

https://doi.org/10.1038/s41586-024-08112-6

Supplementary information

The genomic natural history of the aurochs

In the format provided by the authors and unedited

The genomic natural history of the aurochs

Supplementary Material Guide

This file contains the following six sections: Section 1: Population genetic analysis of Bos genomes; Section 2: Uniparental markers of *Bos*; Section 3: Exploring *Bos* population structure; Section 4: MSMC2 analyses of Ne history and divergence times; Section 5: Imputation and ROH Analysis; Section 6: Domestic cattle. The file also includes Supplementary Figs. 1–29, Supplementary Tables 1-5 and additional references.

Other Supplementary Data are available in a separate data file. Supplementary Data 1 includes information on the 50 new samples presented here. All specimens have been sampled with written consent of their respective steward institute. Specimens BZL1, BZL2, NVL1, NVL3, KOL9, KOL10, ROS001 and ROS002 are sampled under the ASIAPAST project (ERC Action No 772957), Christian-Albrechts-Universität, Kiel, Germany, and are subject to material transfer agreement complying with the Nagoya Protocol. Data from these specimens are published for non-commercial use only. Utilisation for purposes other than non-commercial scientific research may infringe the conditions under which the genetic resources were originally accessed, and should not be undertaken without contacting and/or seeking permission from the original provider of the genetic material.

Supplementary Information Section 1: Population genetic analysis of *Bos* genomes

Identity-By-State Tree

A Neighbour-Joining tree computed using an IBS matrix of 180 samples revealed several groupings of *Bos* based on geographical, temporal and domestic/wild identity (Figure S1). A clade of North Asian aurochsen emerges despite their sampling from a large geographical range; from Tula in Western Russia (Tula1; 41.5kya) to Lake Baikal in Siberia (Baikal1; 12.2kya), and Gyumri in the Caucasus (Gyu1; 13.9kya). This clade is also temporally diverse, although Holocene North Asian aurochs outgroup the core Pleistocene clade in a stepping stone manner. Notably, the sole Holocene Caucasian aurochs (Gyu2; 7.1kya) does not join this clade, instead branching closer to the domestic clades. European aurochsen also cluster together over a wide temporal range; from 47+ kya (Rhi1) to 4.7kya (YoA). The sole African aurochs sampled here (Th7; 8.8kya) forms a clade with a single Bronze Age Levantine domesticate (Tmq3; 2.9kya).

Subsequent clades comprise geographically-congruent groups of domestic cattle. The exception to this is a single Anatolian aurochs (Ch22, 7.1kya), which clusters with ancient Southwest Asian domesticates. Southern European taurine breeds also group away from other European domesticates, likely due to the influence of minor indicine introgression. A diverse group of hybrid (taurine x indicine) animals then fall along the branch leading to indicine animals. Finally, two indicine clades emerge based on geographical origin (South and East Asian). The tree was rooted by the outgroup Water Buffalo genome.

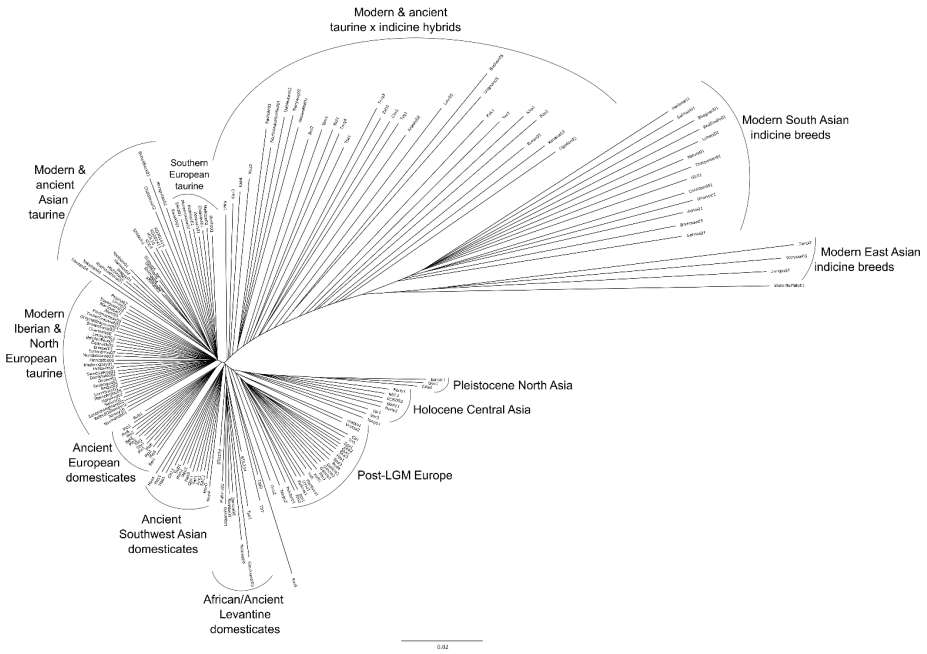

Figure S1. Neighbour-Joining Tree of Identity-By-State matrix computed with modern and ancient *Bos* genomes.

Principal Components Analysis

Principal components analyses (PCA) were computed using PCAngsd. The first PCA was computed on 151 modern domestic *Bos* genomes using 2,951,363 SNPs. A maximum of two animals per breed were included except in the case of hybrids. PC1 differentiated taurine and indicine ancestry; South Asian and East Asian indicine breeds fell to the extreme right of this PC while European taurine breeds fell to the left. In roughly the centre of PC1 are breeds that may be described as taurine x indicine hybrids including Southwest Asian, East Asian and West African breeds. Several taurine breeds from East Asia, West Africa and southern Europe drifted from the main taurine cluster toward the indicine side of the PC1. PC2 separated East Asian and South Asian indicine.

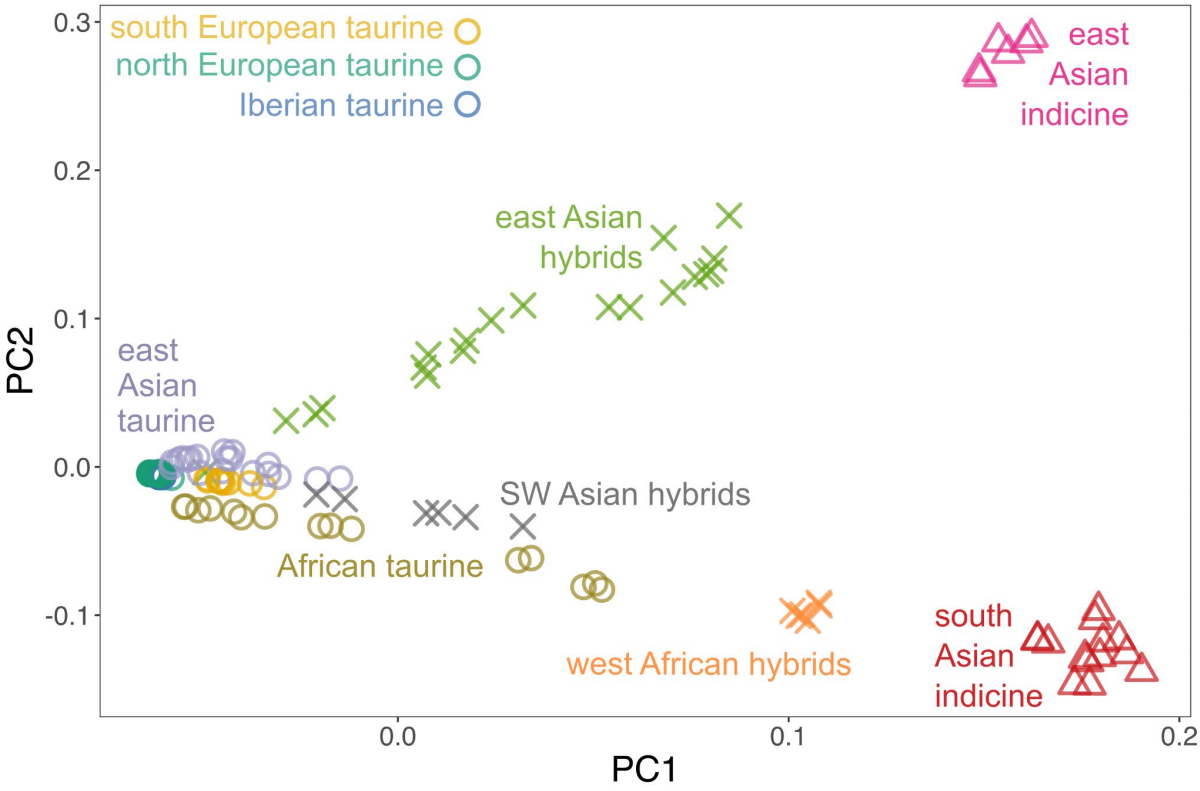

Figure S2. Principal components analysis on 2,951,363 SNPs of modern individuals of *Bos*.

A second PCA was computed on 231 ancient and modern *Bos* genomes using 2,877,051 SNPs. PC1 and PC2 recapitulated the patterns of Figure S2; PC1 differentiated taurine versus indicine ancestry and PC2 differentiated South Asian indicine versus East Asian indicine ancestry. In PC3, Pleistocene and Holocene European aurochsen and Pleistocene North Asian aurochsen fell to the left of this PC while domestic taurine genomes fell to the left. Within the domestic cluster, ancient and modern domestic cattle from Europe, Africa, and Asia group together geographically. Modern and ancient taurine x indicine hybrid domesticates pulled away from the taurine cluster toward the indicine pole. Several Asian domesticates drifted slightly away from this domestic cluster toward the wild pole.

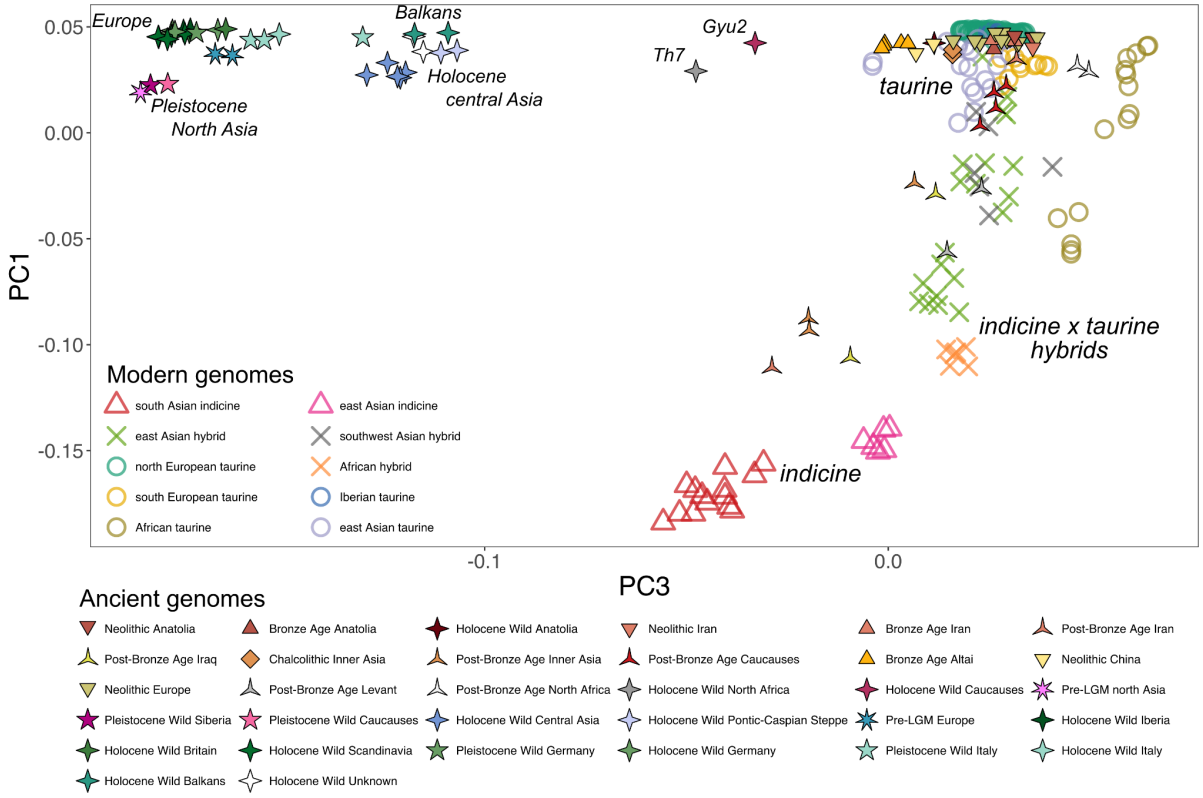

Figure S3. Principal components analysis on 2,877,051 SNPs of modern and ancient individuals of *Bos*. A maximum of two animals per breed were included except in the case of hybrids.

A third PCA of 156 non-indicine *Bos* genome variation was computed using 1,622,606 SNPs. PC1 is controlled by North Asian variation while PC2 is driven by wild European ancestry. Overall, this depicts three poles of *Bos primigenius* diversity: European, Southwest Asian/Near Eastern taurine and North Asian. A Holocene Caucusian aurochs (Gyu2; 7.2kya) and a Holocene North African aurochs (Th7; 8.9kya) drift away from the domestic cluster while a wild bull from Neolithic Anatolia (Ch22; kya) fell comparatively closer toward the domestic cluster. Holocene Central Asia aurochsen fell within the centre of the PCA space. More easterly aurochsen from Borly, Roschinskoe and Novoilinka VI are pulled closer to this Eastern cluster compared to aurochsen from Varfolomeevka. An aurochs recovered from a region in the Former Soviet Union (Tango1) also clusters with samples from Varfolomeevka (Var1, Var2).

Structure within the domestic pole emerged in greater focus with ancient and modern clustering by geography. Ancient Southwest Asian domestics fall in the centre of the diversity. Ancient and modern European domesticates clustered together, being pulled slightly toward the wild European pole. Ancient and modern Asian domesticates similarly drifted toward the North Asian cluster; although there is a cline of affinity with the ancient domesticates from the Altai region displaying the largest affinity with North Asia.

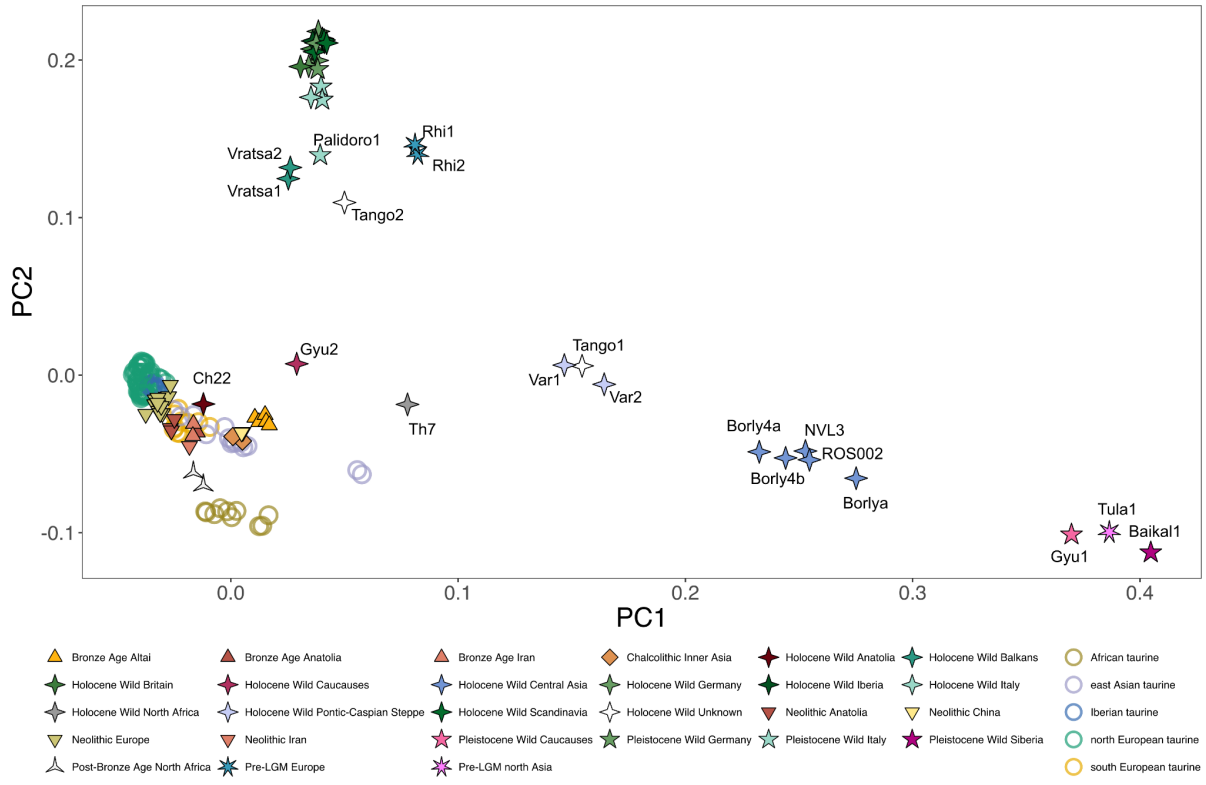

Figure S4. Principal components analysis on 1,622,606 SNPs of ancient and modern individuals of *Bos*, excluding indicine breeds and their hybrids.

A degree of genetic flux was observed within the European pole with clusters corresponding to the southern European refugial peninsula (Iberia, Italy, Balkans). To further investigate this, a final PCA comprising European wild and Anatolian genomes was computed with 1,321,543 SNPs. Here, PC1 separates European wild and Anatolian ancestry. The European clusters were

principally controlled by the extent of influence from Southwest Asia, with early Holocene Iberian animals displaying the least affinity toward Southwest Asia in PCA space. Early Scandinavian, German and British aurochsen also clustered with these, which strongly implicates Iberia as the major source of northwestern European recolonisation. More recent animals of (YoA) drift from this Iberian cluster, possibly influenced by domestic introgression. Domestic introgression into the wild is explicitly evident in the case of the British aurochs YoA, which harbours a domestic haplotype (T3) imported to Europe by early farmers.

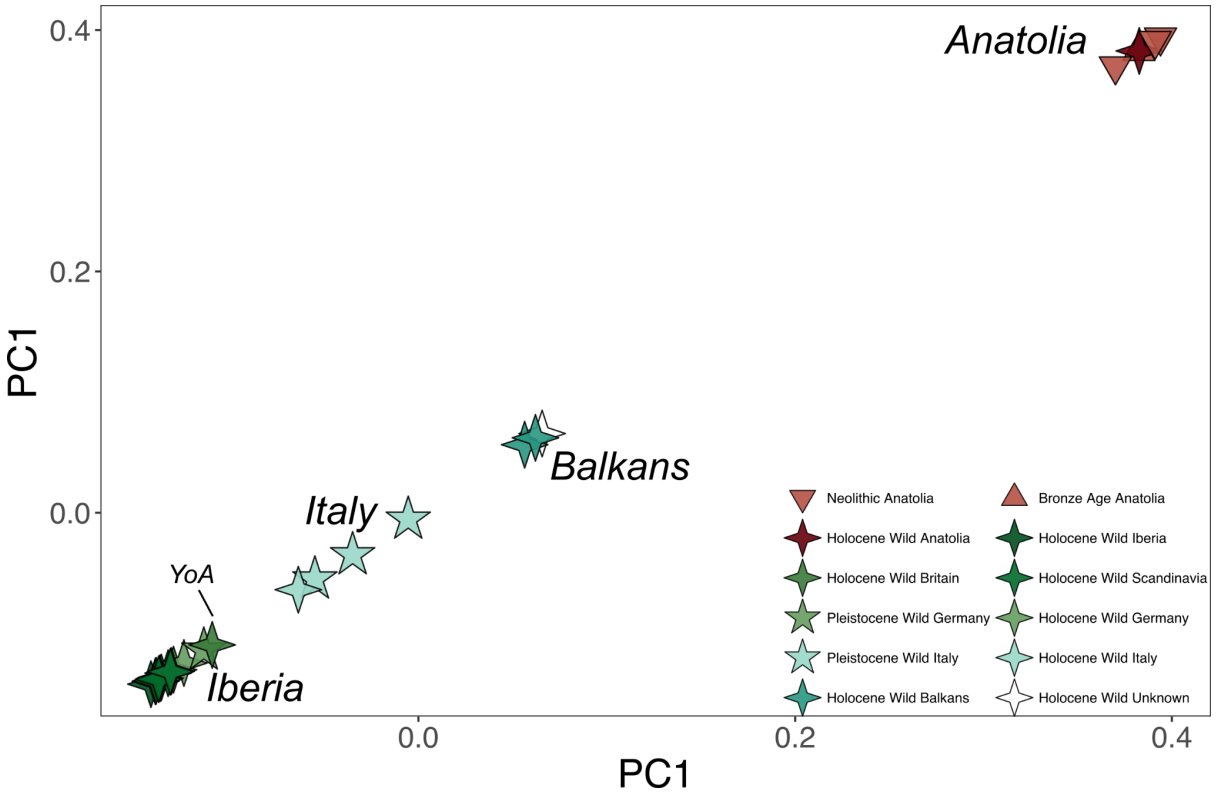

Figure S5. Principal components analysis on 1,321,543 SNPs of ancient wild and domestic European and Anatolian *Bos* genomes.

Unsupervised clustering of ancestral components

Unsupervised model-based clustering methods are commonly used in population genomics to assign an individual a proportion of ancestry to *K* number of genetically distinct source populations. The value of *K* is sometimes chosen based on hitherto known aspects of the studied populations. For example, *K=3* is often used in cattle studies based on BovineHD

Genotyping BeadChip SNPs as this *K* value is well-characterised in distinguishing between indicine, European taurine and African taurine ancestry [1]. It is more common in exploratory work to trial a range of potential *K* values to evaluate the most appropriate using a number of likelihood methods. NGSadmix was run using a range of K values (1-10) and the likelihood value from each run of NGSadmix was extracted and written to a logfile. The results of K2-7 are displayed in Figure S6-7, separated into modern and ancient for clarity. This process was repeated 50 times and this file was used with Clumpak [2] to calculate the most likely *K* based on methods detailed in [3]. The most likely *K* was determined to be K=3, followed by K=4.

- *K=2* described two ancestral components that best describe *Bos primigenius* versus *Bos namadicus* ancestry. The exception is that Pleistocene North Asian aurochs ancestry is modelled as ~5% *Bos namadicus*, ~95% *Bos primigenius*.
- *K=3* described three ancestral components; first, European aurochs, European taurine and Asian taurine ancestry. Second, North Asian aurochs and African taurine ancestry. Third, indicine ancestry.
- *K=4* described four ancestral components that have been previously discussed in PCA analysis: European aurochs, North Asian aurochs, Near Eastern domestic taurine and domestic indicine variation.
- *K=5* describes European aurochs, North Asian aurochs and indicine components as in *K=4,* but now splits domestic taurine variation into two groups: modern European/Asian taurine and African taurine breeds. The taurine component of ancient Southwest Asian and European domesticates and modern Southwest Asian hybrid breeds are modelled as a composite of these components. Similarly, aurochsen of Southwest Asia and North Africa were modelled as composites of all non-indicine ancestries.

- *K=6* described variation as above but split modern European/Asian taurine ancestries. Some ancient Southwest Asian domesticates and Southwest Asian breeds were modelled as a composite of these three taurine components. Neolithic Anatolian breeds were modelled as the same ancestral component as Asian taurine breeds.
- *K=7* described variation as above but split indicine ancestry into East Asian and South Asian components.

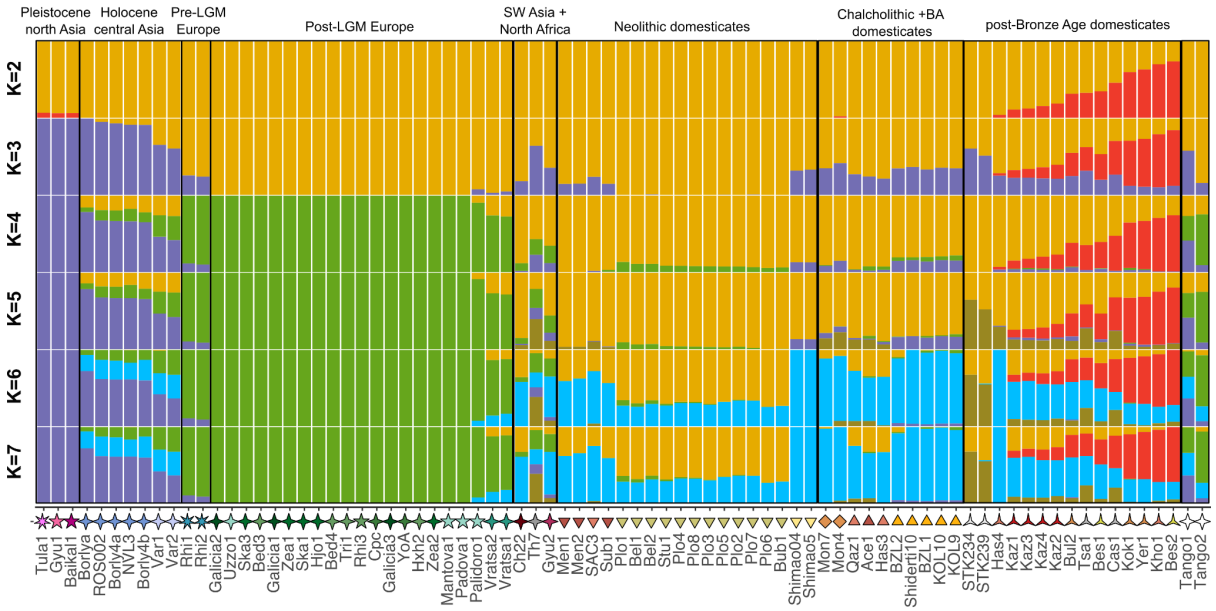

Figure S6. Results from unsupervised NGSadmix analysis of ancient *Bos* with K=2-7 ancestral components.

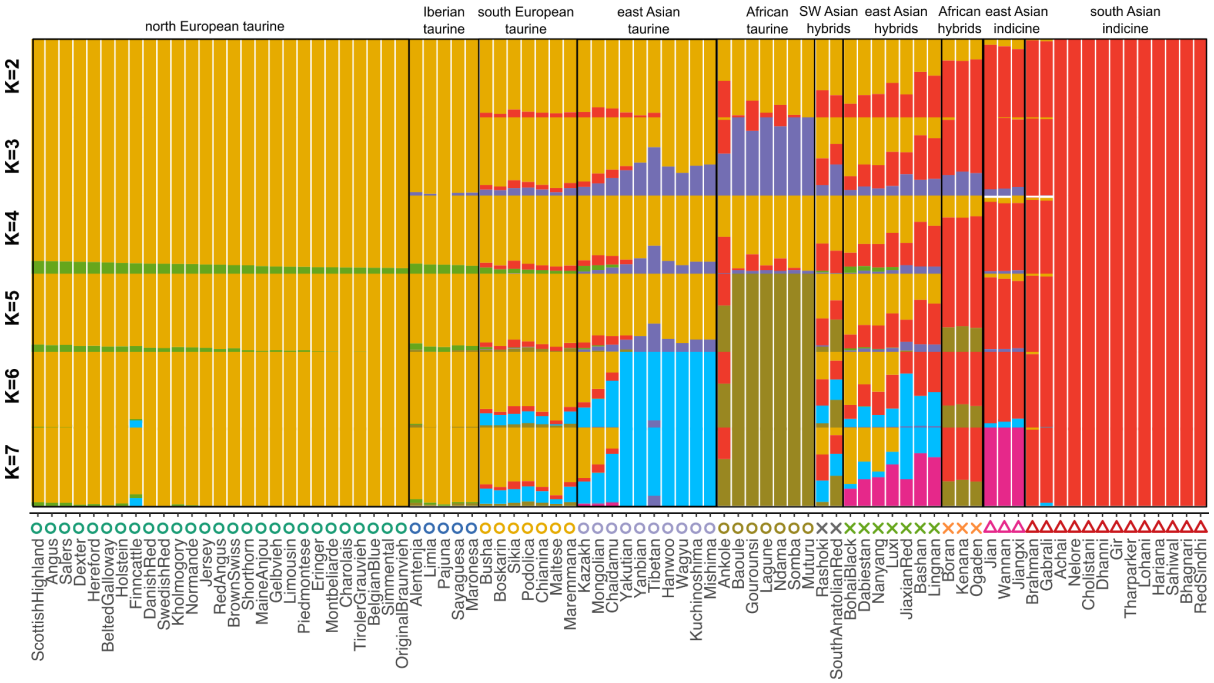

Figure S7. Results from unsupervised NGSadmix analysis of modern *Bos* with K=2-7 ancestral components.

Two aurochsen of unknown origin from the former Soviet Union harbouring a domestic T3 allele are included in this dataset, Tango1 and Tango2. In *K=4*, Tango1 displays an ancestral profile that is remarkably similar to that of Varfolomeevka, implying Tango1 may come from the Pontic-Caspian Steppe. Tango2 derives most of its ancestry from European aurochsen based on PCA clustering and ADMIXTURE modelling, suggesting it is of more western origin.

Modelling ancestry in Holocene Asian aurochs

Holocene Asian aurochs are depicted as hybrid animals in NGSadmix K=4 analysis, harbouring European, Southwest Asian and North Asian ancestral components. qpAdm was employed to explicitly test if Holocene Asian aurochs can be modelled using three or less of these implied sources. Models were accepted if the P value > 0.05. If multiple models passed this significance threshold, the model with the lowest number of source populations was accepted. We tested if Holocene Asian aurochs could be modelled with three left/source

populations (western Europe, Neolithic Anatolia and North Asia). Pre-LGM Rhine samples (Rhi1 and Rhi2; Rhine), *Bes2*, *Th7*, samples from Monjukli Depe (Mon4 & Mon7; Monjukli), and Water Buffalo were used as right populations. Only one model of ancestry (three sources) was accepted in all Holocene Asian aurochs while all two and one source models were rejected. Both two and three source models were accepted for Ch22, a Neolithic Anatolian aurochs.

Supplementary Table 1. Accepted ancestry models for Holocene Asian aurochs (N=9) using qpAdm with three sources (western Europe (N=12), Neolithic Anatolia (N=3) and North Asia (N=3)).

Sample	Accepted model	P Value
Borlya	Three sources	8.97E-01
Borly4a	Three sources	3.58E-01
Borly4b	Three sources	2.81E-01
NVL3	Three sources	3.12E-01
ROS002	Three sources	1.16E-01
Var2	Three sources	3.60E-01
Var1	Three sources	5.23E-01
Gyu2	Three sources	8.53E-02
Ch22	Two ancestries (Neolithic Anatolia and western Europe)	4.94E-01

Bronze Age Central Asian domesticate animals were also modelled in NGSadmix (K=4) with three ancestral components. To test if Central Asian aurochs could be modelled using two sources (North Asia and Central Asian domestic) we performed the same test as above, substituting Central Asian domesticates (Altai) for Neolithic Anatolia. In this test, all Central Asian aurochs were modelled with three sources (western Europe wild, North Asian wild and Central Asian domesticate) while all models were rejected for aurochs from Southwest Asia and the Caucasus, suggesting that admixed Central Asian domesticates were inappropriate as a source. Nevertheless, it provides evidence that Central Asian aurochs are hybrids deriving their ancestry from three different populations.

Supplementary Table 2. Accepted ancestry models for Holocene Asian aurochs (N=9) using qpAdm with three sources (western Europe (N=12), ancient Central Asian domesticate (N=5) and North Asia (N=3)).

Sample	Accepted model	P Value
Borlya	Three sources	8.53E-02
Borly4a	Three sources	4.94E-01
Borly4b	Three sources	7.87E-02
NVL3	Three sources	9.20E-01
ROS002	Three sources	8.66E-01
Var2	Models rejected	N/A
Var1	Models rejected	N/A
Gyu2	Models rejected	N/A
Ch22	Models rejected	N/A

Supplementary Information Section 2: Uniparental markers of *Bos*

Bovini mitochondrial tree

A Maximum Likelihood phylogeny was constructed using modern and ancient *Bovini* sequences (Supplementary Data 4). Extant domestic taurine and indicine cattle and extinct aurochs haplogroups populate the *Bos primigenius* clade. The mitochondrial sequences from pre-LGM European aurochsen (the Rhine samples) outgroup all extant taurine and indicine animals but remain within *Bos primigenius* variation. These sequences are denoted by a star.

The discordant mitochondrial placement of *Bison* emerges in this phylogenetic tree. European wisent (*Bison bonasus*) are placed closer to the *Bos primigenius* clade than to the extinct steppe bison (*Bison priscus*) or to the extant American bison (*Bison bison)*. The wisent clade features a deep split between the two major haplogroups, *Bb1* and *Bb2.* All American bison samples form a sub-clade within the larger clade of Caucasian/American bison. The other *Bos* species (i.e. gaur, gayal and banteng) displayed poor species sorting, as has been previously reported

[4]. Some Gaur samples form clades with both banteng and gayal despite deep divergences for other haplotypes.

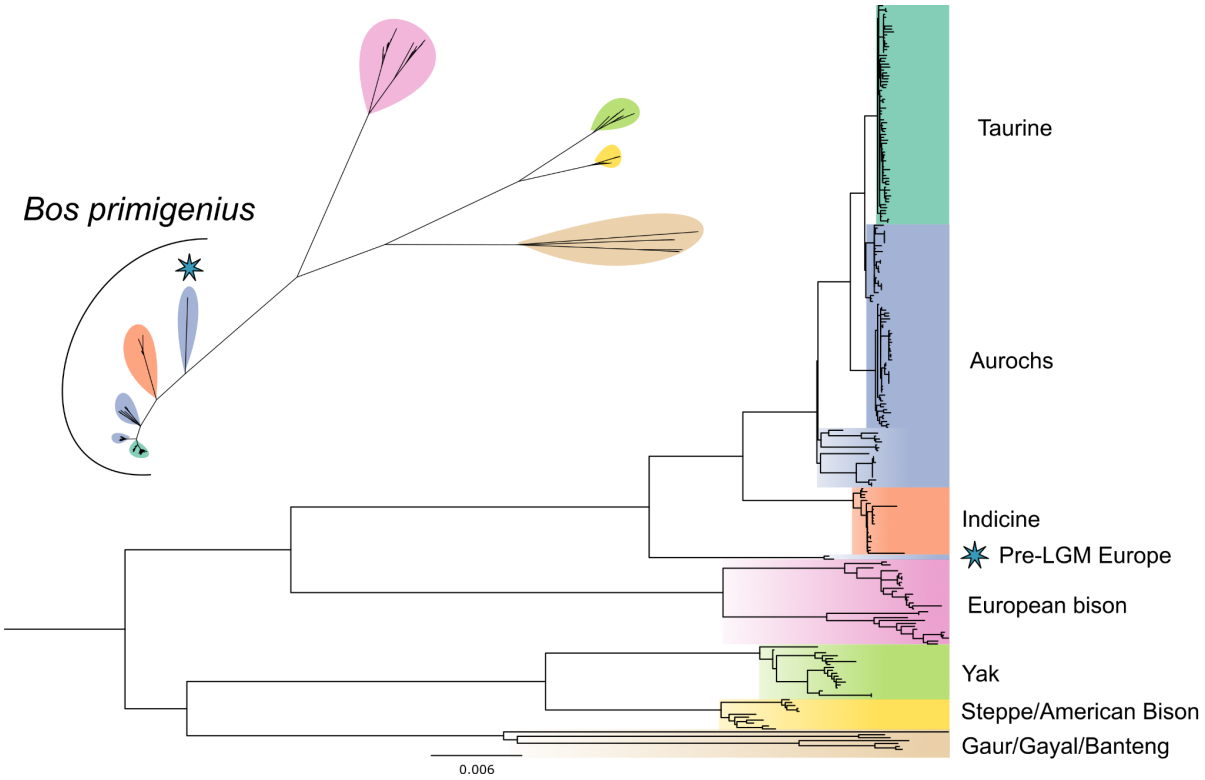

Figure S8. A Maximum Likelihood phylogeny constructed from mitochondrial sequences of ancient and modern Bovini samples.

Bayesian analysis of *Bos primigenius* mitochondria

All BEAST runs converged after 250 million chains, with all parameters reaching an Effective Sampling Size of >300. The maximum clade credibility tree is displayed in Figure S9. The oldest radiocarbon dated animals (Rhi1; 47+ kya, Rhi2; 46.1 kya) harbour newly described haplotypes (G) that outgroup all known taurine and indicine mtDNA variation. The divergence of the G haplogroup (median node age = 336,319 BP, HPD 95% = 389,610-286,028 BP) occurred during MIS8 and preceded TMRCA of *Bos primigenius* and *Bos namadicus* based on mitochondrial and nuclear sequence divergence dating. Despite this, Y chromosome phylogenies and autosomal variation of the Rhine samples firmly place them within the *Bos primigenius* clade and not as an outgroup species.

The next deepest divergence is the well-characterised split of *Bos primigenius* and *Bos namadicus* occurring during MIS6 (median node age = 183,998, HPD 95% = 217,317-154,849 BP). After the taurine/indicine split, a rapid emergence of five haplogroups occurs during MIS5, in what may be described as a pentafurcation. K, C, R, E, and PQT all diverged between 96,133 and 63,165 HPD 95% years ago. All divergences are well-supported for these with a prior <0.9, except for RE|PQT (prior=0.28). The P haplotype subsequently diverged from QT prior to the LGM (median node age = 46,982 BP, HPD 95% = 55,397-38,779 BP).

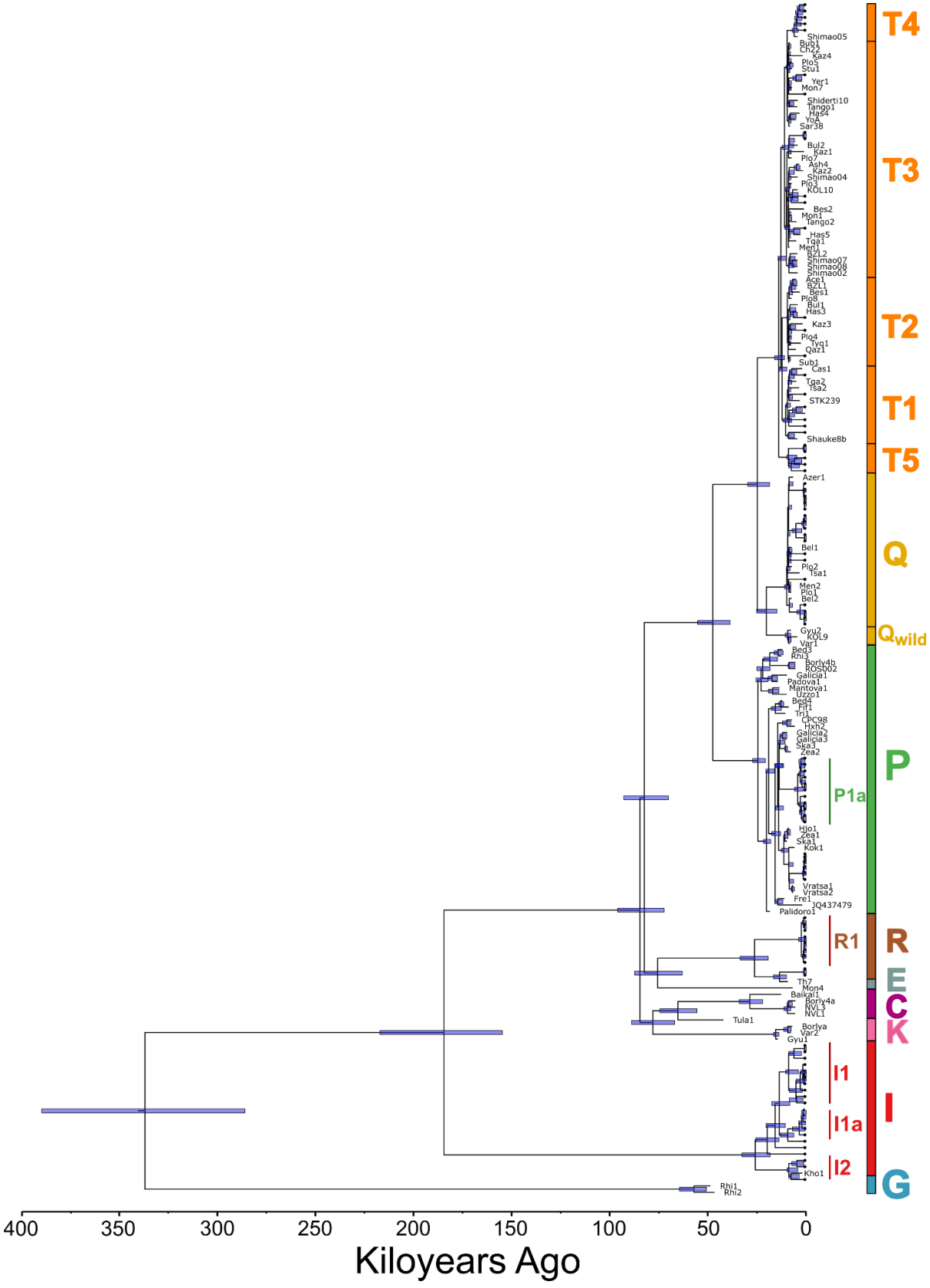

Figure S9. A maximum clade credibility Bayesian phylogeny constructed from mitochondrial sequences of ancient and modern *Bos* samples.

Supplementary Table 3. Node ages of haplogroup divergences.

Tree Node	Median node age (BP)	Node age 95% HPD (BP)	Node Prior
G|IKCREPQT	336,319	389,610-286,028	1
I|KCREPQT	183,998	217,317-154,849	0.9998
KC|REPQT	84,060	96,133-72,395	1
RE|PQT	81,864	93,040-70,246	0.2752
K|C	77,509	89,034-67,102	0.9528
R|E	75,052	87,515-63,165	0.9954
P|QT	46,982	55,397-38,779	1
Q|T	24,242	29,875-18,704	1

Y chromosomal phylogeny

Genetic surveys of modern cattle have revealed a deep divergence between taurine breeds (haplogroups Y1 and Y2) and indicine breeds (Y3) [5–7]. Geographical structure has been found within modern taurine populations; the Y1 haplogroup is found mostly in northwestern European breeds while the Y2 haplotype is found predominantly in southern European and Anatolian breeds. Increased resolution in the Y2 clade also revealed two geographical sub-haplogroups: a Y2b clade manifested mostly in East Asian taurines, while a Y2a clade is found mostly in southern European taurines[8]. Y3 was also partitioned into Y3a found in Chinese indicine breeds and Y3b in South Asian indicine breeds.

The multifasta of 206 males with <15% missingness was loaded into SeaView Version 5.0.1 [9] and a Maximum Likelihood tree was constructed with PhyML with 100 bootstraps [10] (Extended Data Figure 7). The deepest node within our ML described the divergence of *Bos namadicus* (Y3; represented by their descendants, indicine cattle) versus all other aurochsen (Y1, Y2, and the newly described Y4). A Mediaeval Iraqi hybrid (Bes2) fell into the Y3b clade, while an Iron Age Uzbekistani sample (Kok1) had a placement basal to all Y3 haplotypes in this phylogeny. The deepest node within *Bos primigenius* variation described a divergence of North Asian aurochsen (Y4) versus European and Southwest Asian aurochsen and all taurines (Y1 and Y2). Within Y4, the pre-LGM North Asian aurochsen (Tula1; 41.5kya) clustered with post-LGM Siberian aurochs (Baikal1; 12.2kya) while a Pleistocene Caucasus aurochs (Gyu1; 13.9kya) clustered with a Holocene aurochs from Pontic-Caspian Steppe (Var1; 7.5kya). While only two aurochs bulls (Var1 and Var2) were genotyped from Holocene Central Asia, each had a distinct genotype (Y4, Y2a).

The Y chromosomes of the two pre-LGM European aurochsen (Rhi2; 46.1kya, Rhi1; 47kya+) outgroup all extant taurine diversity. Along the taurine branch, Northwest European breeds cluster tightly in the Y1 clade with a southern Iberian *Pajuna* breed outgrouping the main cluster. All modern diversity was outgrouped by a Neolithic Bulgarian aurochs (Vratsa1; 6.1kya). Structure was observed within the Y2 clade; Y2a consisted mostly of southern European taurine breeds and was outgrouped by a smaller clade of West African taurine cattle. The Y2a branch also expanded with several ancient clades that outgroup all modern Y2a diversity. The first clade comprised geographically diverse Eurasian Holocene aurochs ranging from Iberia (Galicia2) to the Pontic-Caspian Steppe (Var2). The second clade comprised Holocene Southwest Asian aurochsen (Ch22 and Gyu2) and a single Neolithic European domesticate (Plo6). Finally, a Bronze Age Iranian domesticate (Has3) outgrouped all Y2a

diversity. The Y2b clade similarly displayed previously undescribed structure. While Y2b was represented by a diverse clade of modern Eurasian breeds, a second clade of ancient domesticates also emerged, comprising two Neolithic bulls from Anatolia (Sub1, Men1) and two Neolithic European bulls (Plo3, Plo4). Interestingly, this Neolithic haplogroup was only sampled in a single modern breed; an African *Ankole* bull. Finally, a Y2c clade is described here for the first time with 11 aurochsen dispersed across Europe; from Iberia, Italy, Britain, Germany and Scandinavia.

Supplementary Information Section 3: Exploration of *Bos* population structure

Exploration of post-glacial populations

To begin exploration of the data, we produced TreeMix graphs with four populations representing putative postglacial populations of aurochs: Europe, North Asia, Near Eastern (represented by Neolithic Anatolia) and South Asian (represented by Indicine breeds). European populations were tested using post-LGM western European samples. Water buffalo was used as an outgroup (Figure S10).

With 0 migration edges, the graph schematic mirrors the mitochondrial and Y chromosomal phylogeny of the populations. However, large residuals deviating from zero indicate that a simple bifurcating tree is an incomplete model. Incorporating 1 migration edge results in an admixture edge of the Neolithic Anatolian branch to South Asian indicine. Previous analyses have similarly shown a minor contribution of Southwest Asian taurine ancestry within modern indicine breeds [11,12]. Whether this is due to modern breed formation, ancient taurine admixture during indicine domestication, or gene flow between pre-Neolithic populations of aurochs remains equivocal. To further explore the relationships of these populations, a graph with 2

migration edges was also computed. This reduced all residuals to 0 and introduced an admixture edge from an unknown outgroup population into western Europe.

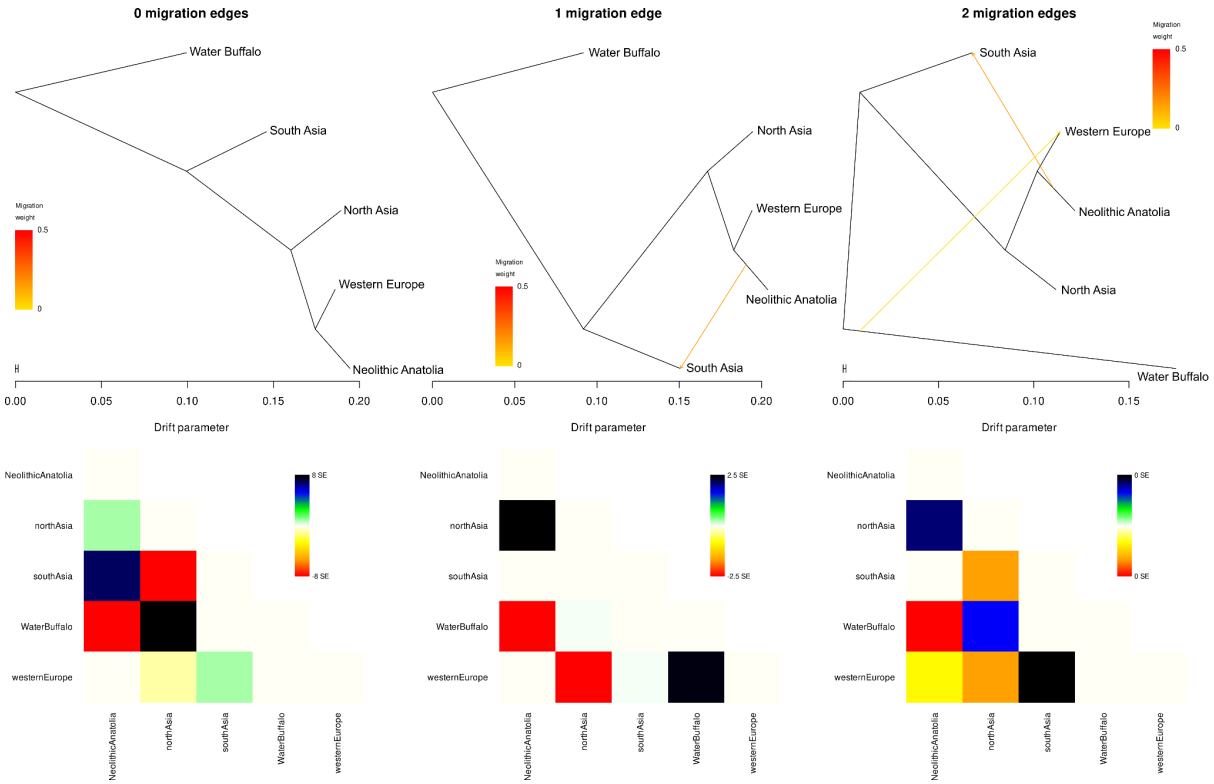

Figure S10. TreeMix graphs and associated residuals of the four putative postglacial populations of aurochs, varying in migrations edges from 0-2.

To further explore models of aurochs population history, ADMIXTOOLS2 [13] was used to find admixture graph models which best fit the observed SNP data. Using the same populations as above, f_2 values were extracted for pairs of populations using extract_f2(maxmiss = 1, auto_only=F). To determine the optimal number of migration events, find_graphs() was run 100 times for m=0-8 with no constraints, setting water buffalo as the outgroup. The worst residual and the log-likelihood of each best-fit model were recorded and the distribution of both statistics was assessed for each m to determine the least migration events which may explain the relationships of the populations (Figure S11). At m=2, a small proportion of models

were observed with their largest outlier |Z| scores falling < 3 while log-likelihood scores appeared unimodal. Therefore, *m*=2 was selected as the number of migration events.

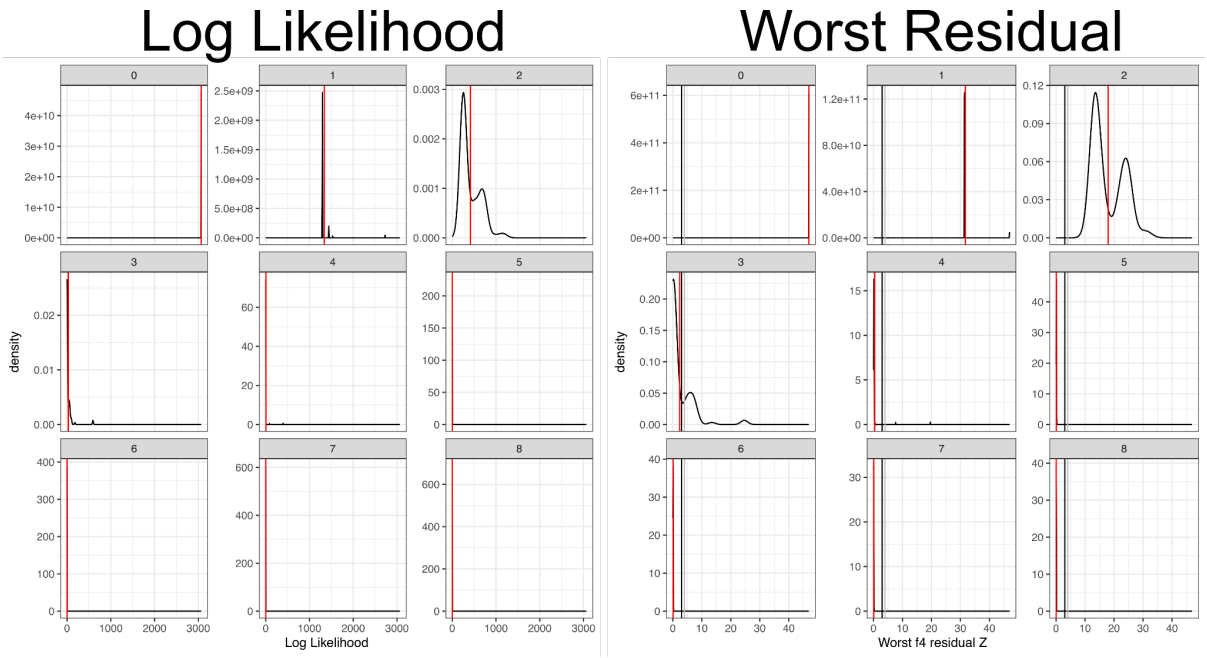

Figure S11. The distribution of the Log Likelihood and Worst Residual values of find_graphs() over 100 iterations for m=0-8.

This level of complexity (m=2) was then explored using find_graphs() for 250 iterations each. The final filtered graphs were then manually assessed for features common among the filtered graphs. Of the 250 graphs explored, only one was excluded due to a residual >3. The remaining 249 graphs were visually inspected and 11 distinct graphical configurations were observed. The key features of the 249 graphs are reported (Supplementary Table 4). The two most common features of all graphs were the overall graph schematic mirroring the observed mtDNA and Y chromosomal phylogeny, and a minor input into the European wild population from an unsampled outgroup population (82.7% of graphs). The majority (163/206) of these graphs model a ghost introgression in Europe from an unsampled population which outgroups all populations including South Asian indicine. The next most common feature was a minor

input of Neolithic Anatolia ancestry into indicine cattle (72.7% of graphs). An additional 12 graphs (4.8%) modelled a minor input of Indicine ancestry contributing to Neolithic Anatolia. Finally, a majority of graphs (71.5%) modelled the North Asian population as unadmixed.

Supplementary Table 4. Most common features of Admixture Graph exploration.

Feature	Number of graphs
Neolithic Anatolia has a minor input into South Asia	181
Indicine has a minor input into Neolithic Anatolia	12
European Wild has input from an unsampled outgroup population	206
North Asian population is modelled as an unadmixed population	178
North Asian population is modelled as an admixed population	71
Graph mirrors mtDNA/Y chromosome tree	206

Two graph schematics satisfied the following features: Graph mirrors mtDNA/Y chromosome tree, Neolithic Anatolia has a minor input into South Asia, European Wild has input from an unsampled outgroup population, and North Asian population is modelled as an unadmixed population. The first (Figure S12A) (log-likelihood score: 0.025) modelled a 35% contribution from an unsampled outgroup *Bos primigenius* population while the second (log-likelihood score: 0.015; the best log-likelihood score of all graphs) modelled a 1% contribution from an unsampled outgroup *Bos* population (Figure S12B). Using qpgraph_resample_multi and compare_fits functions, the likelihood scores of the graphs were compared with 100 bootstraps resampling SNP block sets. The graphs were statistically indistinguishable from each other.

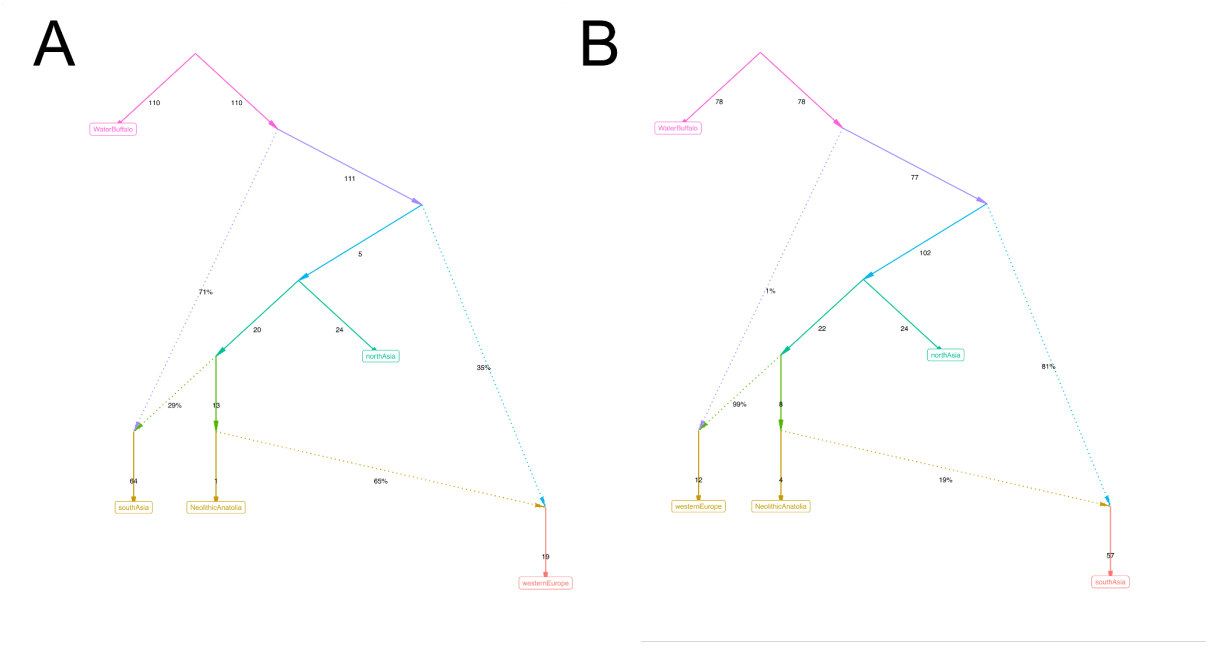

Figure S12. Two graph schematics satisfying the most common features found in AdmixtureGraph exploration.

Modelling Pre-LGM Europe

Next, we incorporated a population comprising Pre-LGM European aurochsen from the Upper Rhine Valley (Rhine). While these aurochsen group with post-LGM Europe aurochsen in clustering analyses, they also harbour a divergent, outgroup mitochondrial haplotype. Using find_graphs() exploration as before, *m=3* was the lowest number of admixture edges required to produce graphs with all residuals <3. This level of complexity (m=3) was then explored using find_graphs() for 250 iterations each. Of the 250 graphs explored, 90 were excluded due to one or more residuals >3. The remaining graphs were visually inspected and a graph schematic (Figure S13) was accepted based on the constraints of the previous graph exploration.

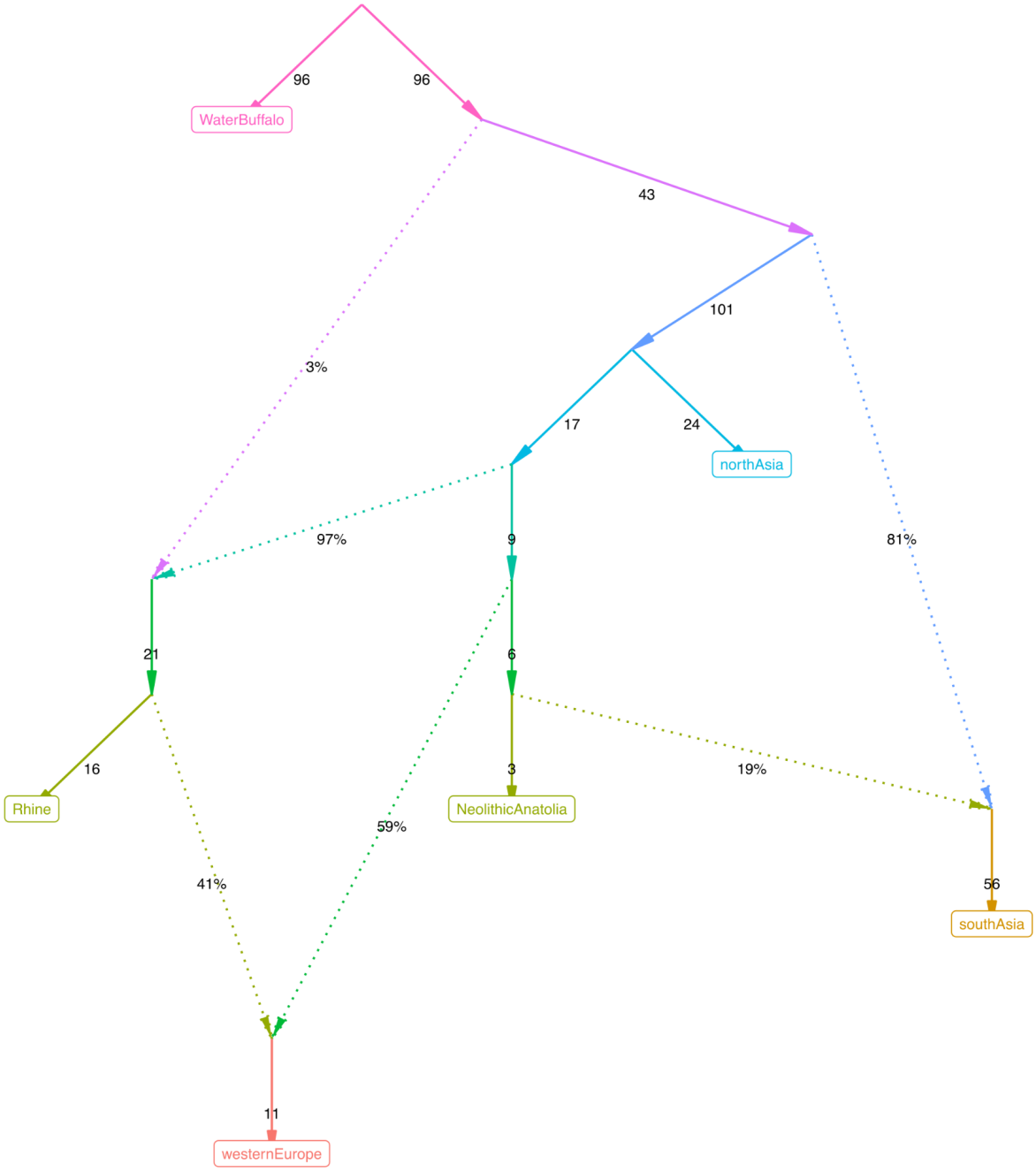

Figure S13. AdmixtureGraph (log-likelihood score: 0.265) schematic modelling aurochs populations, *m=3*.

In this graph schematic, the outgroup ghost introgression increased to 3% contributing to a hypothetical ancestral population of the Rhine Valley and Western Europe. Notably, the Rhine aurochsen also harbour an outgroup mitochondrial haplotype despite a clearly taurine Y chromosome that ingroups from indicine variation and genomic affinity to European genomes.

While this non-monophyly may be explained by an incomplete lineage sorting framework, we argue that several lines of evidence support introgressive hybridisation. First, the presence of the outgroup haplotype. Second, this autosomal modelling of ghost introgression. Third, a palaeontological record of European aurochsen for up to ~650kya despite a more recent (~200kya) coalescence of all sampled aurochsen (Figure 3). Introgressive hybridisation is a relatively common phenomenon during species range expansions and often proceeds asymmetrically from the local population into the colonising population [14]. During warmer periods of the Pleistocene the ranges of some warm-adapted species expanded into hybrid zones with cold-adapted sister species, facilitating introgressive hybridisation. For example, extant brown bears harbour polar bear mitochondria from a hybridisation event when brown bears expanded into polar bear ranges during warmer periods of the Pleistocene [15]. Such a framework is congruent with the observed data for late Pleistocene aurochsen arriving in Europe.

Within this graph schematic, the Western European population was modelled as a hybrid population of a population related to the Rhine Valley and an ancestral population of Neolithic Anatolia. The greater affinity to Southwest Asia was explicitly tested by D statistics, (*D* (water buffalo, *Neolithic Anatolia*; *Rhine Valley, post-LGM Europe*). This revealed allele-sharing with Southwest Asian aurochs populations greatly increased in post-LGM European aurochsen relative to Pleistocene Rhine Valley aurochsen.

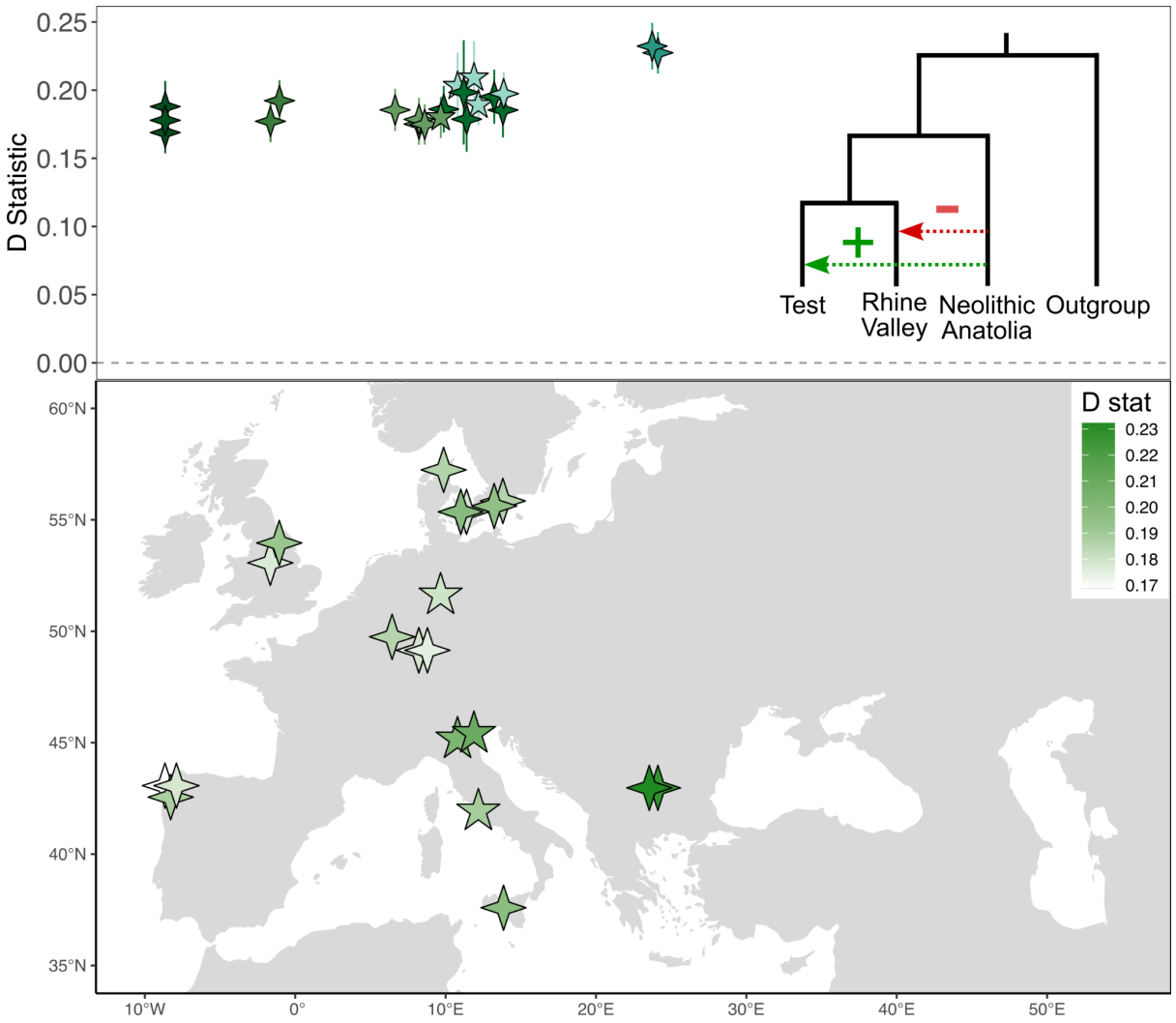

Figure S14. Allele-sharing of Neolithic Anatolian domesticates with European aurochsen relative to the Pre-LGM Rhine Valley population.

Modelling Europe

In NGSadmix analysis (*K=4*), Post-LGM European aurochsen are modelled as deriving most ancestry from one ancestral component. Italian genomes derive all ancestry from this component except for one Italian aurochs (Palidoro1, 18kya), which has ~9.2% of its genome modelled with a Southwest Asian component (Figure S6). Similarly, a PCA of European genomes depicts a cline of affinity to Anatolian ancestry which is correlated to the distance from Southwest Asia, peaking in the Balkans and steadily decreasing toward Iberia and Britain (Figure S5).

We first attempted to fine-scale model Europe by geography (Iberia, Britain [excluding YoA due to domestic introgression], Scandinavia, Germany, Italy, Balkans) using Neolithic Anatolia as a proxy for wild Southwest Asia. Using find_graphs() as before, *m=2* was determined as the lowest number of admixture edges required to produce graphs with all residuals <3. This level of complexity (m=2) was then explored using find_graphs() for 250 iterations each, assessing each graph relative to the highest log-likelihood model using the qpgraph_resample_multi(nboot=100), as above. Of the 250 graphs explored, 147 were excluded due to a residual >3. All remaining graphs depicted the schematic representation presented in Extended Data Figure 1.

Within this schematic, Neolithic Anatolia outgroups all European populations, reflecting previous population genetic analyses. Balkan and Italian populations receive input from a hypothetical population related to Neolithic Anatolia in differing proportions (12 and 41%) while western and northern Europeans do not. This once again reflects previous analyses which suggested a greater affinity of the Balkans and Italy to Southwest Asia, mirroring geography. Interestingly, the Balkans is modelled as a hybrid population with western Europeans while Italy derives most ancestry from a more hypothetical basal European population. In this schematic, all western and northern European populations derive from the same source as Iberia, strongly implicating Iberia as the refugial source of postglacial northern European recolonisation.

As Palidoro1 (18K) presents a different ancestral profile compared to the three other Italian genomes, we split these populations into two (Palidoro1 and Italy) and tested if the previous schematic can be replicated. Using find_graphs() as before, *m=3* was determined as the lowest number of admixture edges required to produce graphs with all residuals <3. Here, the same

schematic was achieved by modelling Palidoro1 and Italy deriving different levels of Southwest Asian ancestry and Palidoro1 deriving its European ancestry from a more basal source (Figure S15).

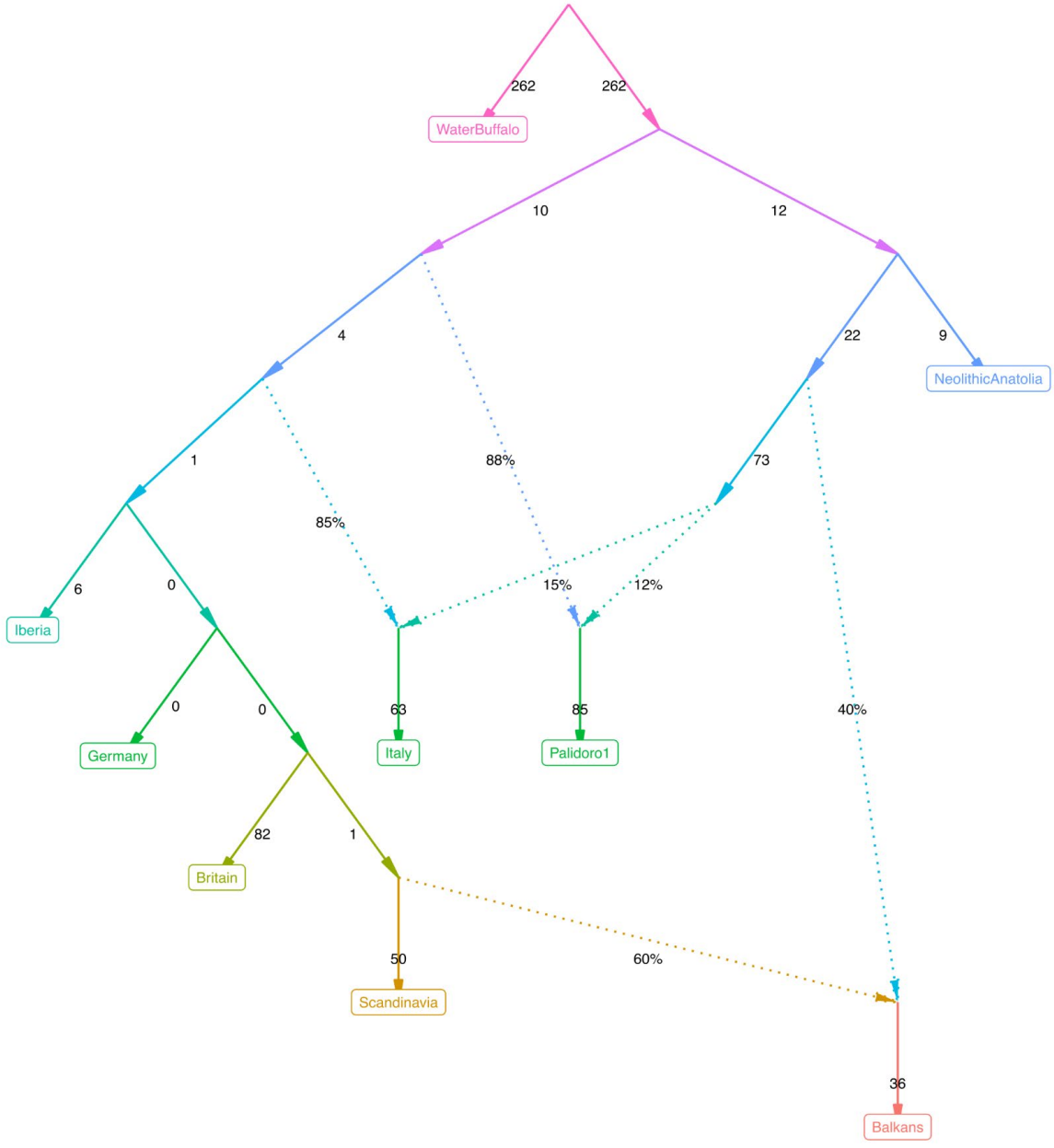

Figure S15. AdmixtureGraph (log-likelihood score: 35.199) schematic modelling aurochs populations, *m=3*.

We then explored models of population history by incorporating the Italian population (including Palidoro1) with our previous population sets. Using find_graphs() as before, *m=4* was determined as the lowest number of admixture edges required to produce graphs with all residuals <3. This level of complexity (m=4) was then explored using find_graphs() for 250 iterations each, assessing each graph relative to the highest log-likelihood model using the qpgraph_resample_multi(nboot=100), as above. Of the 250 graphs explored, 14 were excluded due to a residual >3. The remaining graphs were visually inspected and a graph schematic (Figure S16) was accepted based on the constraints of the previous graph exploration.

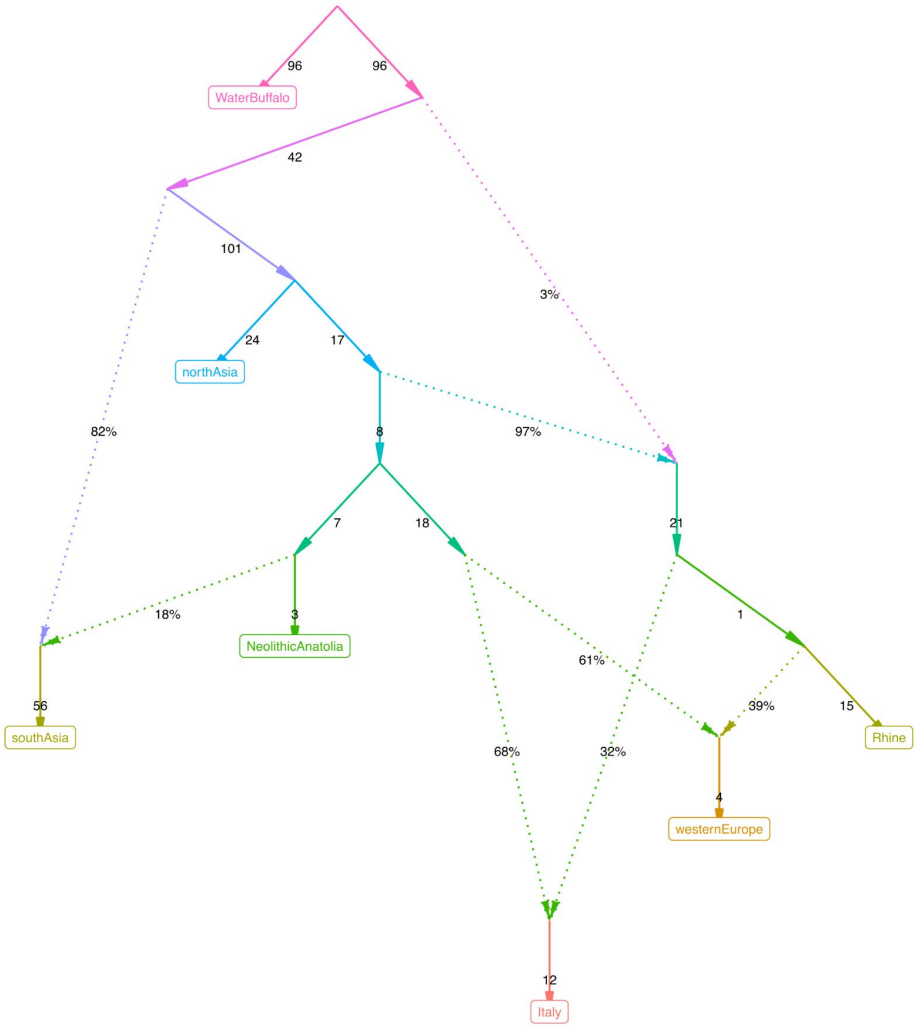

Figure S16. AdmixtureGraph (log-likelihood score: 9.164) schematic modelling aurochs populations, *m=4*.

In this graph schematic, the Western European population was once again modelled as a hybrid population of populations related to the Rhine Valley and Neolithic Anatolia. Italy is similarly modelled as such but with a greater input from the population related to Neolithic Anatolia. The relative affinity to Southwest Asia in European genomes was explicitly tested by D statistics, (*D* (water buffalo, *Neolithic Anatolia*; *Bed3, test Europe*). Bed3 was selected as H2 due to its age (11.6kya), provenance (Germany) and high depth of coverage. This recapitulated previous analyses that found increased affinity of post-LGM Balkan/Italian genomes to Southwest Asia. The sole postglacial western European aurochs with significant allele-sharing with Southwest Asia is a Neolithic British aurochs, YoA.

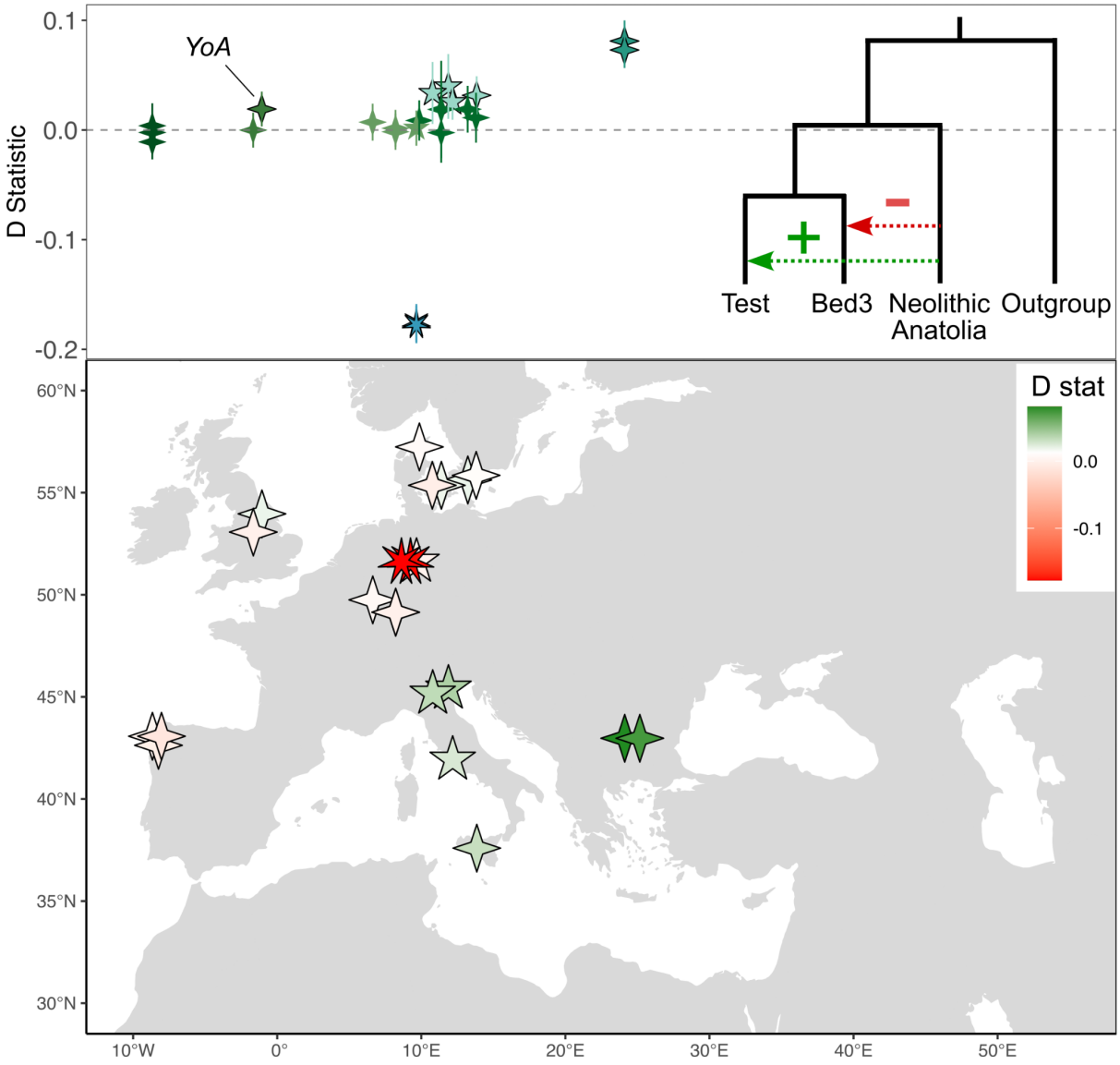

Figure S17. Allele-sharing of Neolithic Anatolian domesticates with European aurochsen relative to the Bed3.

Supplementary Information Section 4: MSMC2 analyses of *Ne* history and divergence times

PSMC'

MSMC2 was employed to infer changes in the past effective population size (*Ne*) of different ancient *Bos* genomes [16]. First, MSMC2 was run on individual, unphased genomes to compute

changes in *Ne* in the two haplotypes present in the genome (referred to as PSMC'). MSMC2 was run on each genome with 20 bootstraps and the time segment pattern: --timeSegmentPattern 25*1+1*2+1*3. Curves were scaled with a mutation rate of 1.26e-8 and generation time of 6, as in [8]. The PSMC' curves of 9 high-coverage ancient genomes are displayed in Figure S18.

The PSMC' curve trajectories of *Bos* genomes depicted differing patterns of demographic history. The PSMC' of the Mediaeval Iraqi taurine x indicine domesticate Bes2 diverged from all other genomes first, reflecting the deep divergence of *Bos primigenius* and *Bos namadicus* (discussed further below). The *Ne* of all *Bos primigenius* populations mirrored one another until around ~100kya when North Asian lineages (Baikal1, Gyu1) continued to decrease in *Ne* while European and Southwest Asian (Sub1) lineages began to stabilise. The *Ne* of all *Bos primigenius* genomes began to climb around ~50kya, with the North Asians experiencing the highest incline. All *Bos primigenius* lineages depicted a sharp collapse in *Ne,* starting around ~35kya just prior to the LGM.

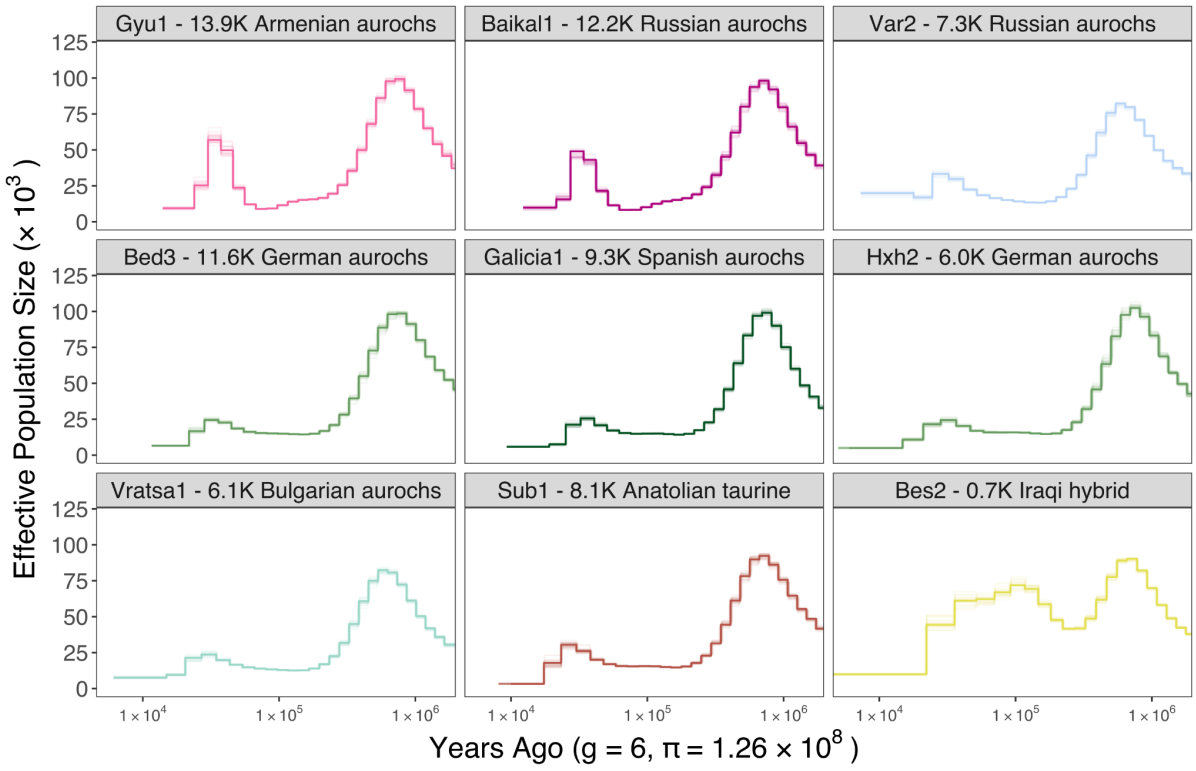

Figure S18. PSMC' curves of 9 unphased, ancient genomes.

PSMC' was also computed for a range of modern genomes including European, East Asian and African taurine, East Asian and South Asian indicine, and taurine x indicine hybrids of Africa and East Asia (Figure S19). The patterns of the taurine genomes were similar to each other and to that of the Neolithic Anatolian, further suggesting a shared ancestral demographic history of all taurine animals. South Asian indicines diverge in their demographic from taurines around ~200-300 kya; in contrast to the taurines (and the European and North Asian aurochsen), South Asian genomes displayed an emphatic increase in *Ne* at this time. This elevated *Ne* is also seen in the East Asian indicines, except here the increase in *Ne* is followed by secondary elevation of Ne before collapsing. The indicine x taurine hybrids display intermediate aspects of these secondary spikes of *Ne*.

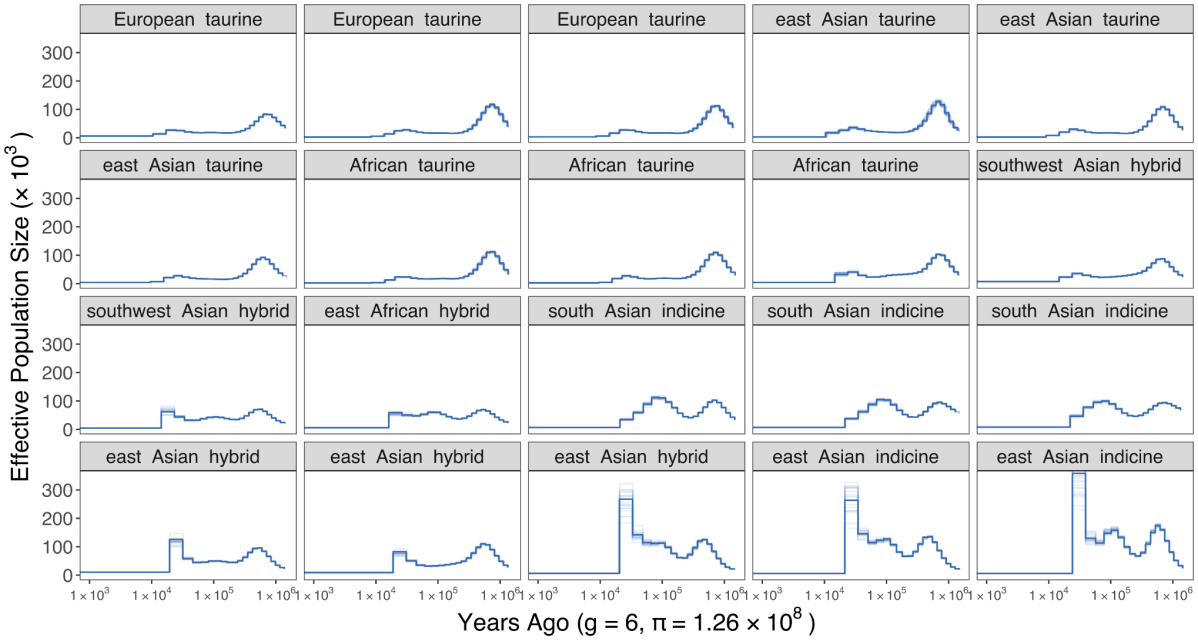

Figure S19. PSMC' curves of 20 unphased, modern genomes.

Introgression from *Bos javancius* (banteng) has previously been detected in East Asian indicine breeds [8,17]. This increased diversity of introgressed haplotypes may have inflated *Ne* in East Asian indicine compared to South Asian indicine. Prior to ~ 70kya, all South Asian and East Asian indicine depict an elevated *Ne* compared to all *Bos primigenius* genomes (North Asian, European and taurine). This elevated *Ne* may reflect a more suitable habitat for *Bos namadicus* but also may be due to some ancient admixture with other Bovini lineages. We explicitly tested allele-sharing of *Bos* populations with wild Eurasian Bovini relative to western European aurochs using *D* (water buffalo, Bovini sp.; western Europe, *Bos* test) using the SNP set called on Bovini outgroups. In this test we consistently observed significant excess allele sharing between indicine and outgroup Bovini species compared to western European populations, with East Asian *Bos indicus* showing most elevated values (Extended Data Figure 5). All other *Bos primigenius/Bos taurus* populations suggest clade integrity, except Pleistocene North Asia which displayed marginally significant allele sharing with Bovini outgroups - with *D* values an order of magnitude lower than that shown by *Bos indicus*.

Pseudo-diploid X chromosome PSMC' analysis

To complement the results of our MSMC2 cross-population analysis, we also used MSMC2 on pairs of haploid X chromosomes from two different male genomes to estimate divergence times between populations. Divergence time between populations can be investigated by combining two male haploid X chromosomes and computing a PSMC' on the "pseudo-diploid" chromosome. At the point of divergence, the *Ne* of the pseudo-diploid X chromosome is expected to become extremely large due to the lack of more recent coalescence events [19]. Male X chromosomes are haploid and thus also considered to be perfectly phased chromosomes. The mutation rate was scaled down by 25% due to the lower mutation of the X chromosome compared to the autosome (after [20]). Additionally, we shifted the beginning of the curve to the midpoint of males of different ages.

Generally, the point of divergence is determined heuristically based on visual inspection of the PSMC curves. We arbitrarily defined the beginning of isolation as the midpoint of the time segment which had a 10% higher *Ne* than that of the previous time segment eventually leading to a $Ne > 200 \times 10^3$. Visual inspection of the curves reiterated two aspects of the cross-population analysis. First, a deep divergence between all *Bos primigenius* populations (i.e. European, Southwest Asian and North Asian) vs *Bos indicus* (proxy for *Bos namadicus)* occurring prior to 200kya. Second, the split between North Asian and European/Southwest Asian populations occurred during the MIS5, as inferred by the sharp incline of the PSMC curve among all population pairs. The divergence between European and southwest Asian populations was determined to have occurred just prior to the LGM by MSMC2 cross-population analysis. While we observe an increase in *Ne* between these populations at this time, the increase in *Ne* is comparatively modest. This may be interpreted as a less definitive isolation between European and southwest Asian populations through the LGM.

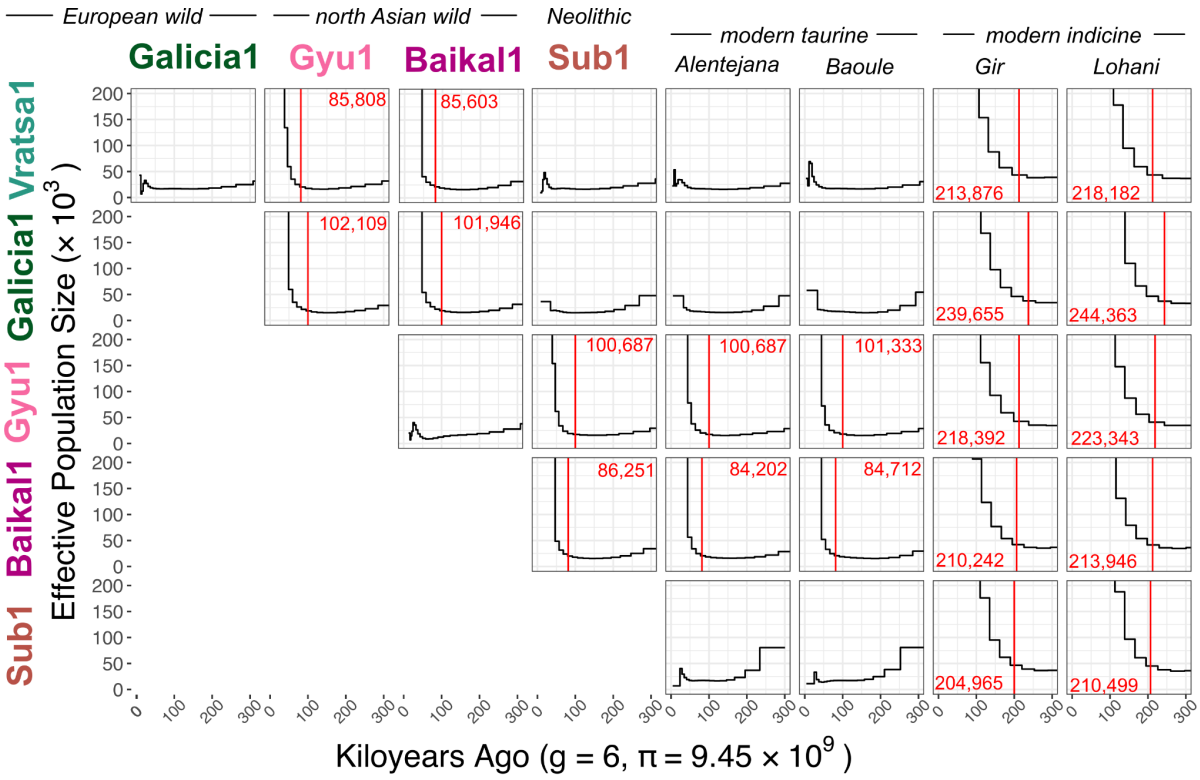

Figure S20. Pseudo-diploid X chromosome PSMC analysis curves depicting population splits.

Supplementary Information 5: Imputation and ROH Analysis

Imputation accuracy of ancient samples

Imputation is capable of calculating accurate, phased diploid genotypes from low coverage data in ancient cattle and aurochsen[21]. We used the phased dataset of 21,672,432 SNPs and 240 samples to impute ancient samples and investigate runs of homozygosity (ROH). First, we tested the imputation accuracy of aurochs ancestry with this dataset by downsampling the high-coverage ancient samples to an average genome-wide coverage of 0.5X and 1X. The high-coverage sample was then removed from the full, phased dataset and the downsampled sample was imputed with this reference panel (239 samples). The accuracy of the imputed calls was tested against the high quality genotype calls using Picard's GenotypeConcordance tool. The imputation accuracy was high for the ancient Neolithic Anatolian and European aurochsen as

previously reported[21]. The imputation accuracy of the North Asian samples was comparatively lower, possibly due to rare alleles not present in the reference dataset.

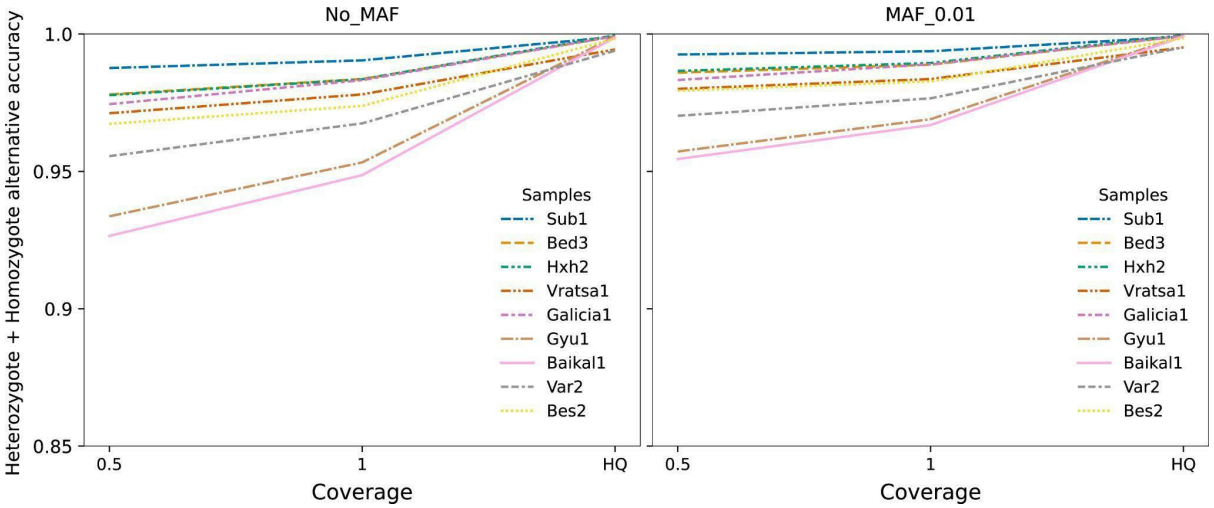

Figure S21. Accuracy of imputation of heterozygous and homozygous alternative sites at 0 and 0.01 MAF bins and different downsampled coverages (0.5X, 1X, HQ).

As the imputed calls were intended for ROH analysis, we also compared ROH profiles of our imputed, downsample calls to the high-coverage calls. The high-quality and imputed 0.5X downsample calls of the European aurochs, Holocene Asian aurochsen and domesticates produced very similar patterns of ROH. In contrast, the imputed 0.5X downsample calls of the Pleistocene North Asian samples produced ROH patterns that overestimated short ROH and underestimated longer ROH compared to the HQ profiles (Figure S22), possibly due to rarer haplotypes that were difficult to impute with the dataset of mostly modern domestic animals. When we imputed the Pleistocene North Asian aurochsen downsampled to 1X, a more accurate ROH profile was produced. We proceeded with including all aurochs-types since the only Pleistocene North Asian aurochs we imputed was 3.45X (Tula1).

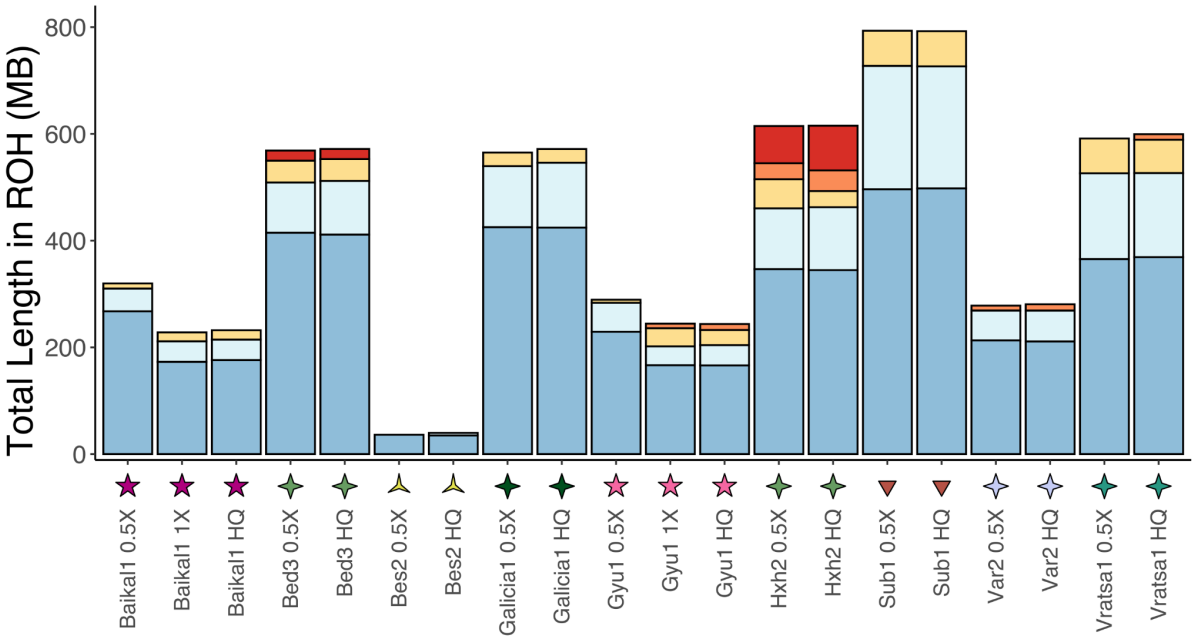

Figure S22. ROH profile comparison between the downsampled imputed aurochsen and the HQ aurochs.

ROH analysis

The patterns of ROH revealed several aspects of *Bos* demography. High levels of short ROH are present in European, Southwest Asian and North Asian aurochsen. This contrasts with indicine breeds and may reflect a bottleneck event shared by all sampled *Bos primigenius,* also seen in PSMC'. In Europe, there is a marked increase in total ROH load between pre-LGM and Post-LGM population, possibly caused by the reduced and fractured range of European aurochs due to glaciation. In contrast, all Pleistocene North Asian aurochsen share similar profiles of ROH, with a modest increase in medium ROH in post-LGM specimens. Holocene Asian aurochs display the lowest total load of ROH of all aurochs, likely due to their diverse ancestries. Two aurochsen from island populations (Sicily and North Jutland) display extreme levels of medium to long ROH due to recent consanguineous mating associated with island populations [22].

The imprints of domestication are also evident in the patterns of ROH. The earliest domesticates from Anatolia and Europe display elevated medium-sized ROH compared to the aurochs of Europe and Southwest Asia which is consistent with a reduced population size of domestic animals but with inbreeding avoidance behaviour. Later domesticates of Bronze Age Iran and Altai Mountain region display a reduction in ROH load, possibly due to hybridisation with local wild populations. Despite clear evidence of wild introgression, a Chinese Neolithic domesticate (Shimao05) displays elevated medium-sized ROH that are comparable to that of Neolithic Anatolia and Europe and may reflect a further bottleneck event as taurine cattle were dispersed into East Asia around 5-4kya [23]. Taurine x indicine hybrids display the lowest ROH load of all ancient domesticates, reflecting the diversity present in indicine cattle. Long ROH dominate the profiles of modern taurine cattle breeds reflecting the consanguinity of modern breed formation. These patterns are most extreme in European taurine breeds where up to ⅓ of the genome comprises ROH > 1MB.

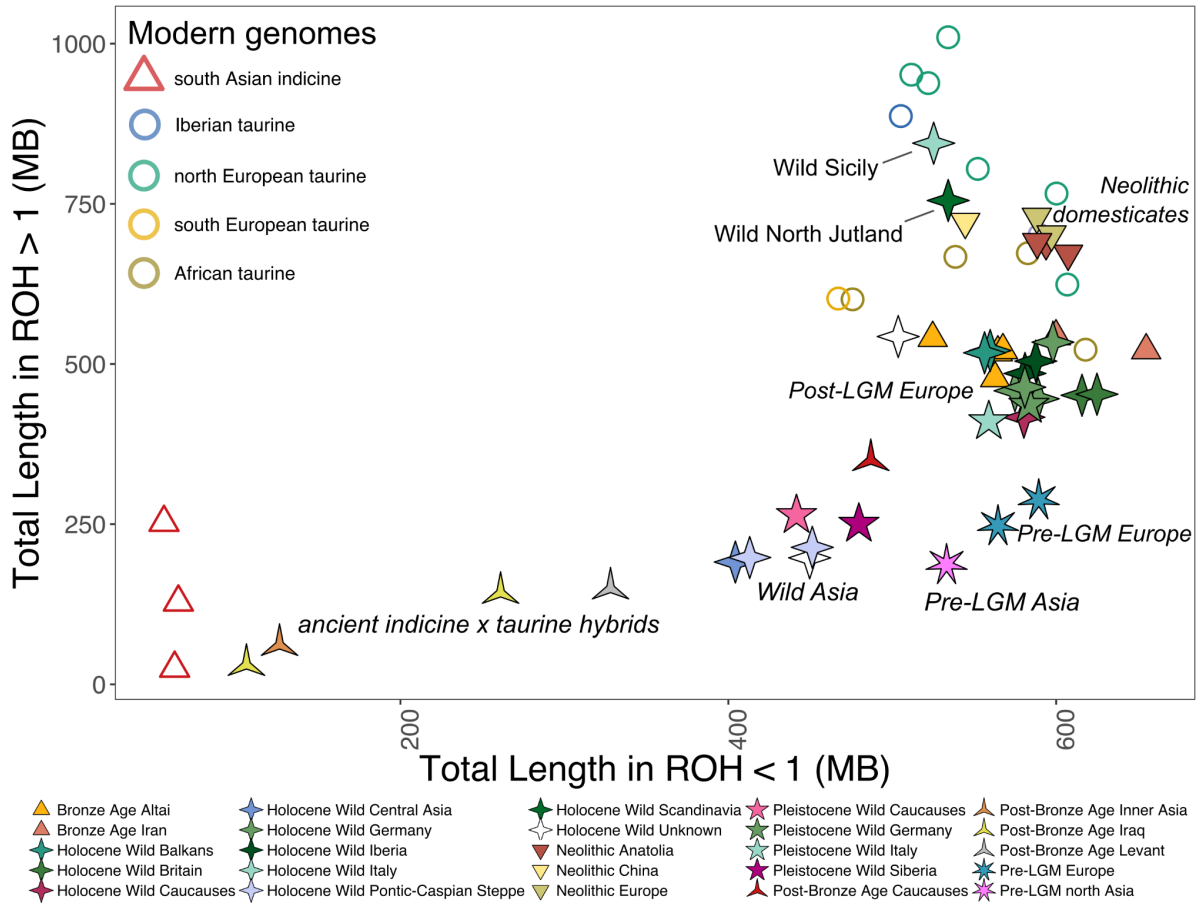

Figure S23. Scatterplot of the total length of ROH >2MB and ROH <1MB in ancient and modern genomes.

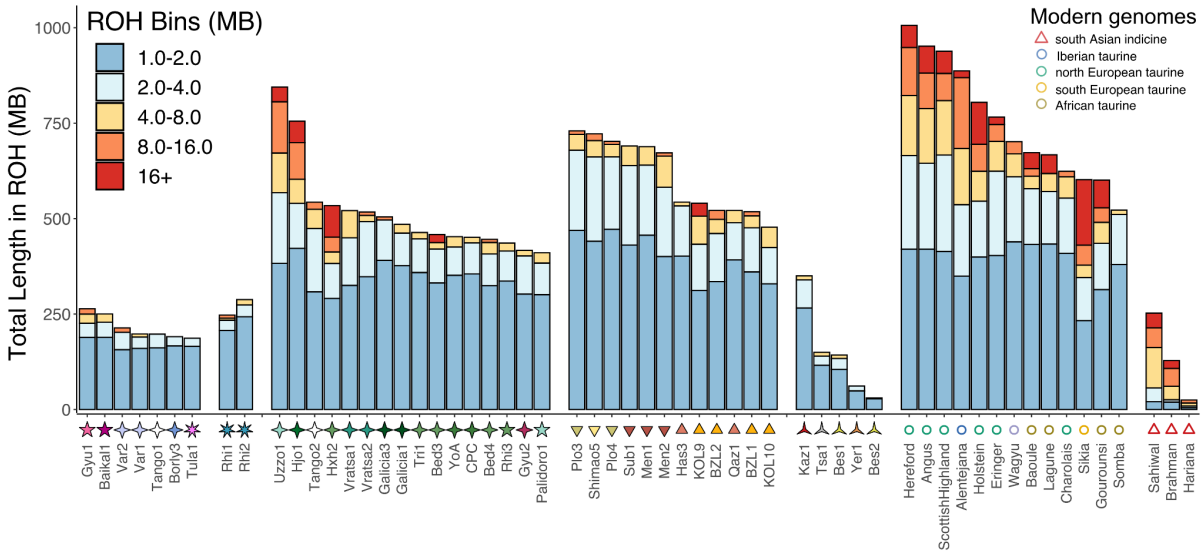

Figure S24. ROH profiles in ancient and modern genomes.

Supplementary Information Section 6: Domestic cattle

Earliest local wild introgression in domestic cattle

PCA and NGSadmix analysis depicted a persistence of Southwest Asian aurochs ancestry as taurine cattle migrated with early farmers into Europe and Asia, although with some degree of genetic flux. NGSadmix and D statistics revealed widespread admixture of indicine with taurine domesticates most prominently first appearing in Near Eastern domesticates around ~4kya, shortly after the 4.2k dry climatic event (Figure S25;[12]) which may confound *D statistics* to detect wild introgression due to the underlying treeness of *Bos*. For this reason, cattle with significant and high levels of indicine introgression were not plotted on Figures S28-30. Cattle from Chalcolithic Monjukli Depe (Turkmenistan; 6.4kya) were modelled with a minority North Asian aurochs component in NGSadmix K=4 analysis but, interestingly, these cattle also have significant allele sharing with indicine cattle (Figure S25). Modern indicine cattle are modelled with a minority contribution from a population related to Neolithic Anatolia. The increased allele-sharing between the cattle of Monjukli Depe may have been due to ancient admixture from an Inner Asian taurine population during the indicine domestication. This imbalance in allele sharing may also be due to an earlier diffusion of indicine ancestry from the Indus Valley.

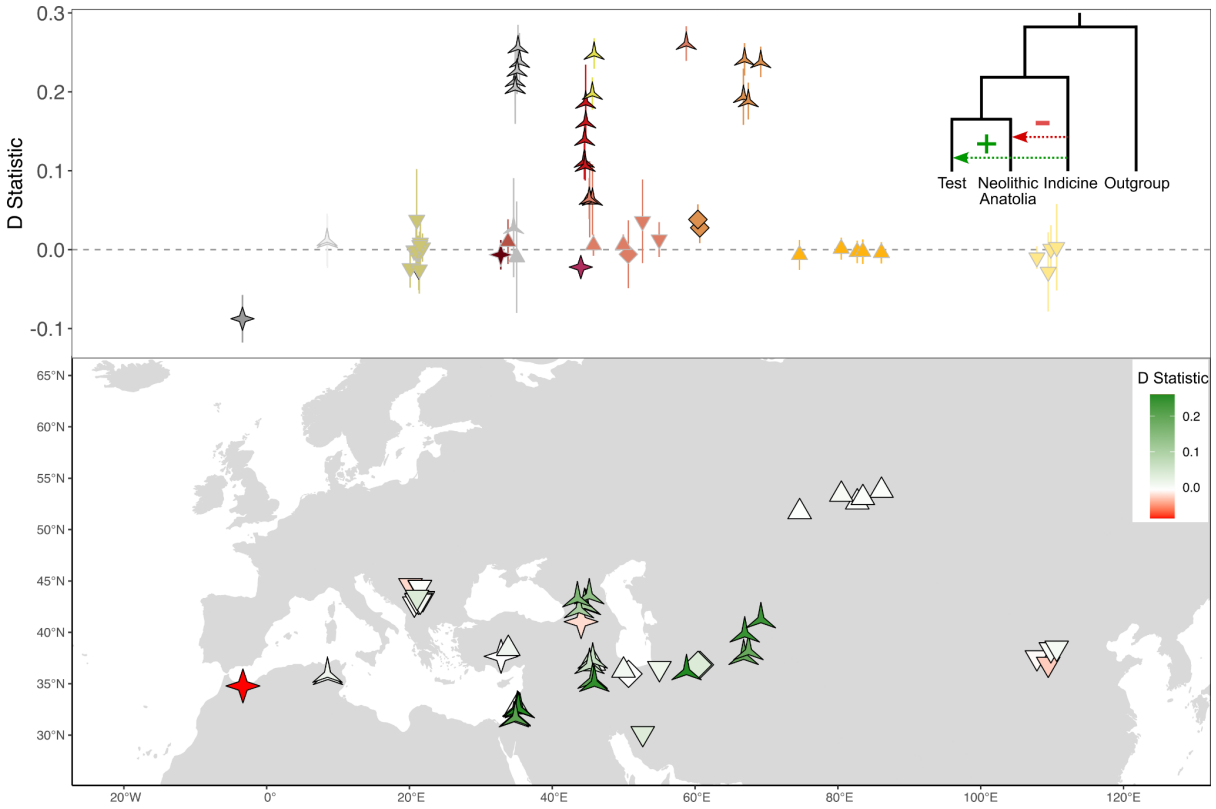

Figure S25. Allele-sharing of South Asian indicine with ancient domesticates relative to Neolithic Anatolia.

NGSadmix analysis and *D* statistics revealed early introgression of local wild variation into the imported taurine in Europe and Africa, as has previously been reported [12]. Here, we show for the first time that admixture with local wild aurochsen also occurred in early Asian cattle as they were dispersed into Central Asia and East Asia (Figures S3; S26-28). This local introgression appears embedded in modern breeds (Supplementary Data 5) with European wild ancestry peaking in British and Irish breeds (Scottish Highland, Salers, Belted Galloway, Dexter, Angus), as previously described [24]. Iberian breeds also display a consistent European wild aurochs signal, as previously detected [25]. North Asian aurochs ancestry persists in almost all modern Central and East Asian cattle in NGSadmix *K=4*, with some breeds (Mishima, Kuchinoshima) also showing excess allele sharing of wild Eastern ancestry. Significant excess wild North African allele sharing is only detected in one modern breed (Muturu), likely due to

the substantial imprint of indicine introgression on the African continent, as previously described [26,27].

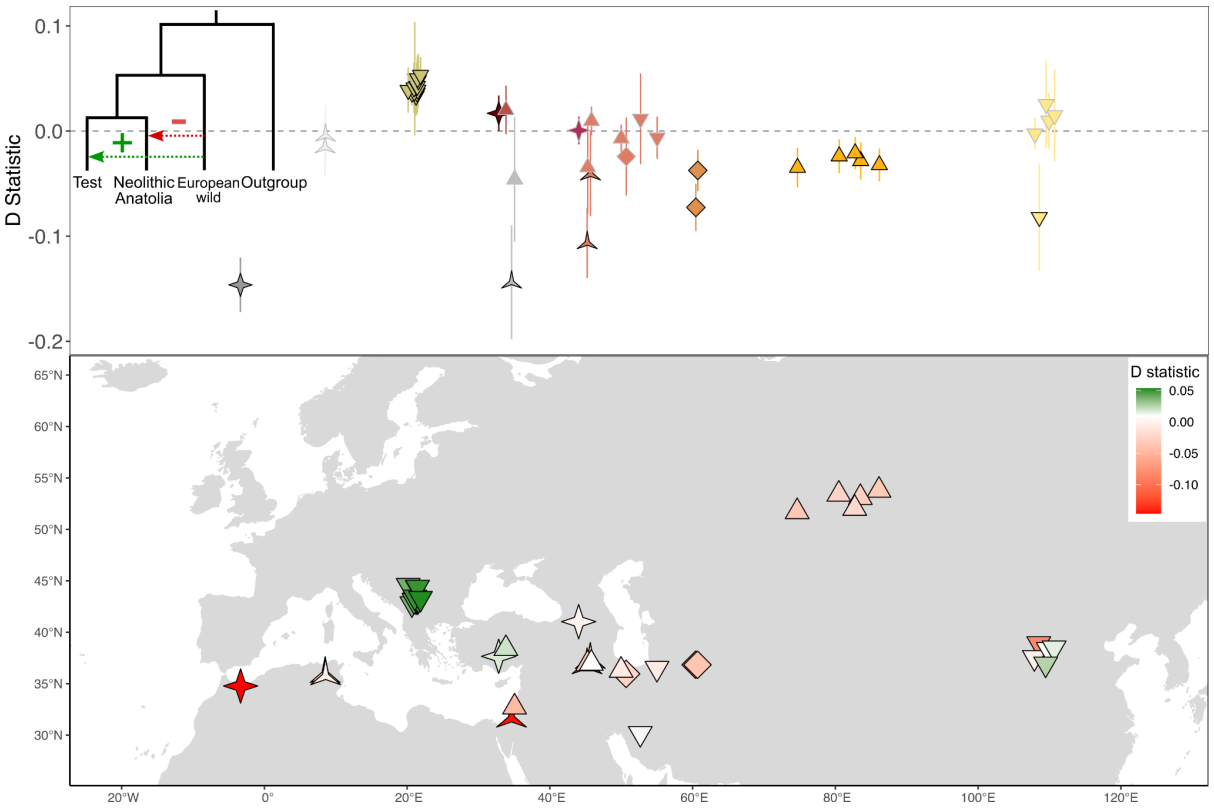

Figure S26. Allele-sharing of European aurochsen with ancient domesticates relative to Neolithic Anatolia.

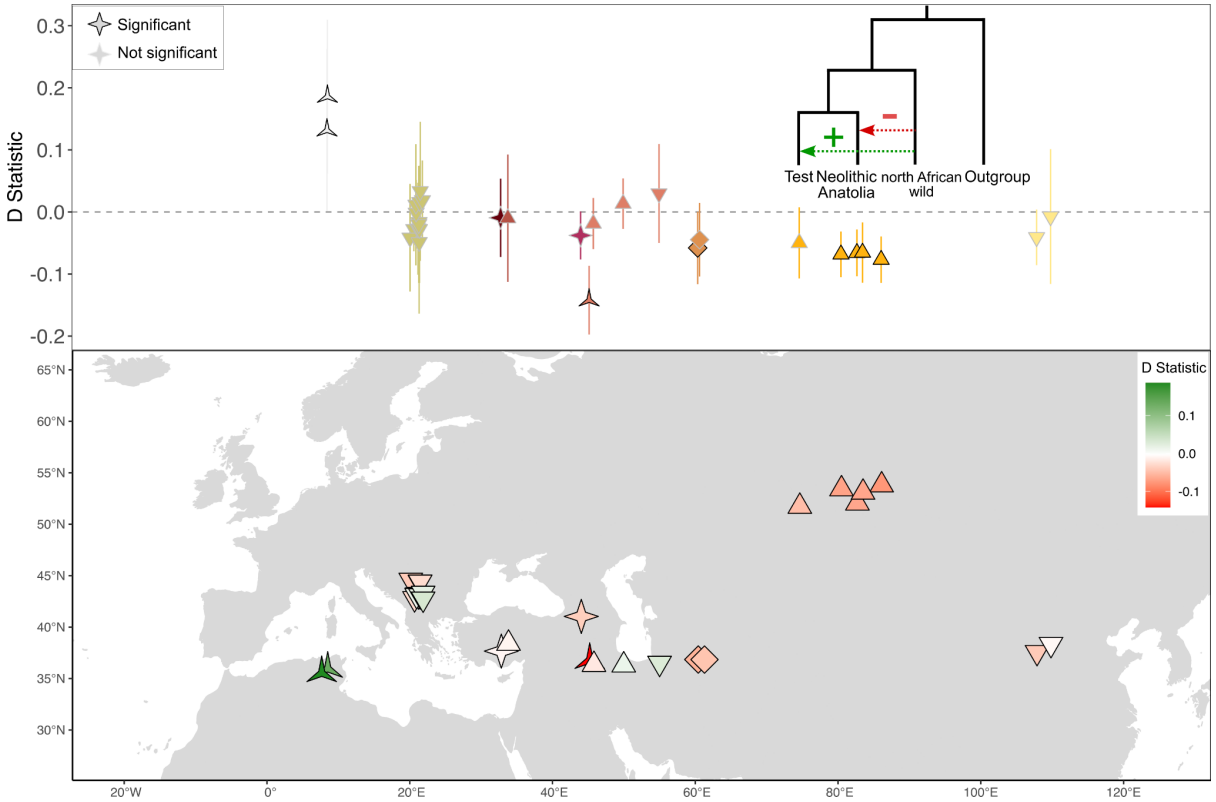

Figure S27. Allele-sharing of North African aurochsen with ancient domesticates relative to Neolithic Anatolia.

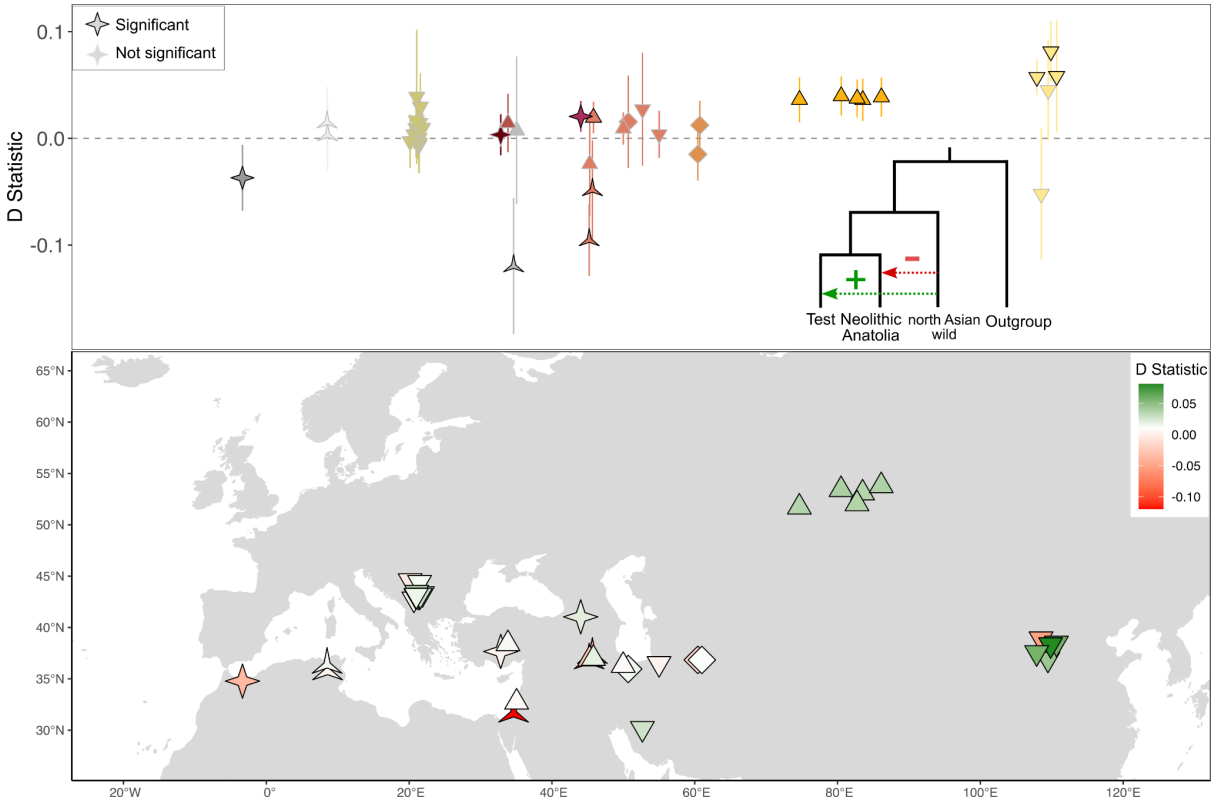

Figure S28. Allele-sharing of North Asian aurochsen with ancient domesticates relative to Neolithic Anatolia.

Maternal aurochs capture

Toothcomb-like topologies were observed in the domestic mitochondrial phylogenies which indicated proliferation of animals after a domestication event of a small number of breeding females. While the T mega-haplogroup dominates modern taurine cattle, the exact process of domestication of the T sub-haplotypes is equivocal [28–3031–33], although the founder population of aurochs cows contributing to domestic cattle is thought to have been limited [34,35]. Based on the Bayesian analysis of C14 time-stamped mitochondrial sequences, TMRCA of the T mega-haplogroup procedes the proposed *Bos primigenius* domestication time (~10.5kya) in Southwest Asia (median node age = 13,373 BP, HPD 95% = 16,123-11,042 BP). In contrast, TMRCA of the T1-4 haplogroups (or T1-T2 and T3 haplogroups separately) overlap with the proposed timing of domestication, subsequently diversifying into sub-haplotypes. The divergence time of T/Q was prior to the domestication of taurine cattle, perhaps during the LGM (median node age = 21,129 yrBP, HPD 95% = 26,930 - 15,798 yrBP), confirming that they were not domesticated from a common maternal ancestor. The timing of TMRCA of Q supports a recruitment timing very early in cattle domestication, perhaps contemporaneously to the T aurochs capture (median node age = 9,266 yrBP, HPD 95% = 11,062-8,333 yrBP). Overall, this strongly implies that the original maternal capture of taurine capture was of very few aurochs cows. The separate domestication of indicine cattle occurred later than taurine cattle in the Indus Valley, around 8 kya [36]. The three domestic indicine haplogroups (I1, I1a, I2) sampled here coalesce to around ~8.5kya, similarly suggesting very few aurochs cow captures.

While the P haplotype represents the major extinct European aurochs haplotype, it is also unexpectedly harboured in East Asian cattle breeds [37]. Interestingly, the expansion of this P1a haplogroup occurred approximately 3.7kya (median node age = 3,709 yrBP, HPD 95% = 5,997-1,747 yrBP), corresponding to a time when taurine cattle are believed to have spread from Central Asia to East Asia between 5-4kya [23]. In ancient Kazakhstan, maternal lineages of P are found in two ~5kya wild samples (ROS002 and Bor4b). While this provides explicit evidence of the existence of the P haplogroup in ancient Central Asia, these ancient samples are not maternal ancestors to the modern domestic P haplotype (P1a). In addition to P1a, P is harboured in an Iron Age Ubezekistani hybrid domesticate (Kok1) and modern Austrian taurine animals, suggestive of multiple captures of the P haplotype. A post-domestication maternal capture of the R haplotype is also inferred from the internal node coalescences. The only sampled wild R haplotype is harboured in a Moroccan aurochs (Th7, 8.8kya) while the domestic R haplotypes have been sampled rarely in modern southern Italian breeds and Iron Age Tunisian cattle [38].

In addition to recruitment of wild haplotypes into domestic stock, there are several instances of clear domestic introgression into the wild. Explicit evidence is found in three aurochs that harbour domestic T3 haplotypes despite being sampled outside of the assumed natural range of the T aurochs (Tango1; 4.0kya, Tango2; 4.4kya and YoA; 4.7kya). YoA also has significantly higher allele-sharing with Neolithic Anatolian cattle compared to other western European aurochs (Figure S17). The existence of female domestic haplotypes in the wild emphasises the existence of reciprocal gene flow between wild and domestic cattle in ancient farming communities.

Supplementary Table 5. The Most Recent Common Ancestor of major haplogroups.

Haplotype	TMRCA median node age (yrBP)	TMRCA HPD 95% node age (yrBP)
Q	9,266	11,062-8,333
T	13,373	16,123-11,042
T1	9,863	11,718-8,446
T2	8,931	9,851-8,241
T3	10,419	12,317-8,877
T4	5,486	7,226-4,081
T5	8,378	9,991-4,843
T1'T2'T3'	12,234	14,435-10,311
I	25,308	32,787-18,447
I1	8,285	10,466-4,006
I1a	8,718	13,289-6,341
I2	8,231	9,821-4,316
R1	1,862	3,880-545
P	23,930	27,372-20,869
P1a	3,709	5,997-1,747

Paternal aurochs capture

It has been speculated that local wild restocking of cattle may have been a purposeful strategy of early farmers, possibly to acquire new advantageous traits that would adapt the imported Anatolian cattle to their new surroundings [24,39]. Wild hybridisation of domesticates was pervasive in early cattle husbandry and likely male-biased (Extended Data Figure 7).

The earliest sampled domestic bulls of central and northwestern Anatolian (Sub1 and Men2) and Europe (Plo3 and Plo6) cluster together in a distinct, newly described Y2b clade. This particular sub-clade of Y2b is not sampled in any modern European or Southwest Asian bulls, only appearing in a single African *Ankole* animal. Another Neolithic European domesticate (Plo6) clusters with two Southwest Asian aurochs (Ch22; 7kya, Gyu2 7.9kya) in a newly described Y2a clade which was not sampled in any modern sample. These Neolithic clades may represent the ancestors of the original recruitment of aurochs in the Middle and Upper Euphrates Valley and would suggest a near-complete turnover of male taurine ancestry has occurred since the Neolithic. This contrasts with the maternal lineages where most taurine cattle retain the T/Q mitochondrial haplotype (the likely original domestic maternal lineages), reflecting the strength of founder effects in Y chromosomes compared to mitochondria in breed formation [40]. A sole Bronze Age Iranian domesticate (Has3; 4.1kya) also outgroups Y2a, suggesting that, along with the modern haplogroups, there were at least 7 male contributions within *Bos taurus*.

An initial hypothesis to explain the geographical structure posited that Y2 was the original Anatolian haplotype domesticated in the Fertile Crescent while Y1 was a wild European haplotype that was introgressed into domestic herds of northern European cattle[5]. Later studies using aDNA refuted an Anatolian origin for Y2, as the Y2 haplotype was genotyped in several Neolithic European aurochsen carrying P mitochondria [41]. This suggested either widespread male-domesticate introgression occurred into Neolithic aurochsen, or more parsimoniously, that Y1 does not define the European aurochs Y chromosome. The earliest occurrence of the Y1 haplogroup is harboured in Neolithic Bulgarian aurochs (Vratsa1; 6.1kya), which outgroups all modern variation and supports a domestic Y1 origin via European wild introgression. However, with only one wild occurrence, it is not possible to ascertain if its presence in Vratsa1

is because this is the natural range of the haplogroup or if it is due to introgression into the wild population by domesticates.

Selection

A total of 62 outlier regions between modern taurine breeds and aurochs were detected in the sliding windows *Fst* analysis (Supplementary Data 6). Within these outlier regions, 47 genes directly overlapped with the outlier regions (16 unannotated). When we expanded the gene search to the 100kb flanking regions, we identified 163 genes (57 unannotated).

Eight genes were associated with the carcass/body weight and growth (STEAP3 [42], THADA [43], SHISA2 [44], AMMECR1L [45], GPD2 [46], SREBF2 [47], ACVR1 [48], MYO9A [49]). A further three genes were associated with milk production (PCCA [50], PEX14 [51], ADIPOQ [52]) and hormonal/reproductive pathways (ESR1 [53], CYP19A1 [54], SPATA17 [55]). Other genes of interest included those related to heat-shock (HSP90AA1 [56]), brain development (NLGN1 [57]) (MUC2, MUC6 [58]) mucins in the gastrointestinal tract.

Ancient genomes have revealed putative selected traits in early domesticates of other species through selection scans. For example, selection of aesthetic traits like coat colour has been detected in domestic pigs and goats [59,60]. This is the first time a selection scan has been attempted in wild versus domestic cattle populations. We identified some groups of genes that may be associated with the domestication of cattle, such as genes associated with body size and milk production. However, using modern breeds as the domestic population may be problematic in this way and these signals may be representative of more recent genetic effects of cattle breeding. Further, the use of European aurochsen is not a perfect representative of the Near Eastern wild population from which taurine cattle were domesticated.

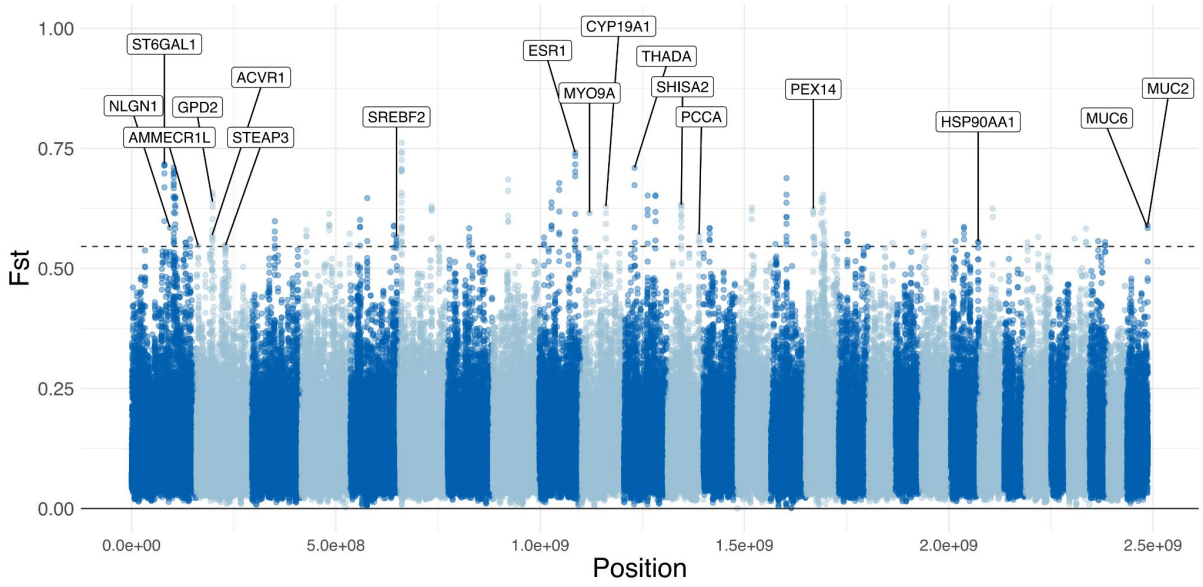

Figure S29. Sliding *Fst* calculated between modern taurine breeds and aurochsen to investigate regions in the genome that are strongly differentiated between wild and domestic populations. Genes of interest are annotated.

Bos exploitation in ancient Central Asia

We also sampled *Bos* remains from Eneolithic sites of the Botai/Tersek culture of northern Central Asia (Novoilinka VI, Borly 4, Roshchinskoe). The Botai people are best known for experimenting with horse husbandry [61], and were genetically distinct from the Yamnaya and Elunino populations despite their geographic and temporal proximity [62,63]. Botai-Tersek culture *Bos* samples were modelled primarily as North Asian wild with minor components of European wild and Near Eastern ancestries (Extended Data Figure 4), and are most similar to an earlier Central Asian Neolithic sample (Borlya; 6.6kya). This strongly suggests that the Botai communities were exploiting local herds of aurochs.

Domesticated cattle, sheep and likely goats were first husbanded in the Altai region by Afanasievo communities by the second half of the fourth millennium BCE. These communities shared ceramic styles and mortuary practices with the Yamnaya herders of the Pontic-Caspian

Steppe and human aDNA studies provided explicit genetic links between these populations [62,63]. Ancient DNA analysis of sheep from Afanasievo settlements has also revealed the presence of Near Eastern haplotypes, indicating that they were herding domestic sheep rather than the local wild population [64]. Here, we sampled cattle from the Early Bronze Age Elunino culture of the Altai Krai region (Berezovaya Luka & Kolyvanskoe); Elunino cultural formation appears to have emerged out of indigenous forest-steppe Neolithic Bolshemys hunter-gatherer-fisher traditions influenced by the arrival of domesticated animals to the region. These Bronze Age cattle are modelled as mostly Near Eastern and harboured the domestic mitochondrial haplotype T3, indicating that domestic cattle arrived to the region by at least 4kya. Interestingly, a cow from Kolyvanskoe also harboured a Qwild haplotype, a haplogroup sampled in aurochs of the Caucuses and Pontic-Caspian Steppe (Gyu2; 7.1kya and Var1; 7.5kya). The Bronze Age Elunino culture exploited imported domestic taurine cattle, unlike the Botai/Tersek culture. It is interesting to note that the domestic cattle of Elunino culture are also modelled as having the highest proportion of wild components of all sampled domesticates (approximately 20%). The proportions of the wild components Elunino cattle reflect the ancestral proportions of the Central Asian aurochs, suggesting that wild hybridisation occurred locally.

Supplementary References

1. Bovine HapMap Consortium *et al.* Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–532 (2009).
2. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
3. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals

using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).

4. Sinding, M.-H. S. *et al.* Kouprey (Bos sauveli) genomes unveil polytomic origin of wild Asian Bos. *iScience* **24**, 103226 (2021).
5. Götherström, A. *et al.* Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proc. Biol. Sci.* **272**, 2345–2350 (2005).
6. Ginja, C., Telo da Gama, L. & Penedo, M. C. T. Y chromosome haplotype analysis in Portuguese cattle breeds using SNPs and STRs. *J. Hered.* **100**, 148–157 (2009).
7. Edwards, C. J. *et al.* Dual Origins of Dairy Cattle Farming – Evidence from a Comprehensive Survey of European Y-Chromosomal Variation. *PLoS ONE* vol. 6 e15922 Preprint at https://doi.org/10.1371/journal.pone.0015922 (2011).
8. Chen, N. *et al.* Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* **9**, 2337 (2018).
9. Gouy, M., Tannier, E., Comte, N. & Parsons, D. P. Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation. *Methods Mol. Biol.* **2231**, 241–260 (2021).
10. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
11. Kumar, P. *et al.* Admixture analysis of South Asian cattle. *Heredity* **91**, 43–50 (2003).
12. Verdugo, M. P. *et al.* Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science* **365**, 173–176 (2019).
13. Maier, R. *et al.* On the limits of fitting complex models of population history to f-statistics. *Elife* **12**, (2023).
14. Currat, M., Ruedi, M., Petit, R. J. & Excoffier, L. The hidden side of invasions: massive introgression by local genes. *Evolution* **62**, 1908–1920 (2008).
15. Cahill, J. A. *et al.* Genomic evidence of geographically widespread effect of gene flow

from polar bears into brown bears. *Mol. Ecol.* **24**, 1205 (2015).

16. Schiffels, S. & Wang, K. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. *Methods Mol. Biol.* **2090**, 147–166 (2020).
17. Decker, J. E. *et al.* Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* **10**, e1004254 (2014).
18. Gou, X., Wang, Y., Yang, S., Deng, W. & Mao, H. Genetic diversity and origin of Gayal and cattle in Yunnan revealed by mtDNA control region and SRY gene sequence variation. *J. Anim. Breed. Genet.* **127**, 154–160 (2010).
19. Palkopoulou, E. *et al.* Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* **25**, 1395–1400 (2015).
20. Bergström, A. *et al.* Grey wolf genomic history reveals a dual ancestry of dogs. *Nature* **607**, 313–320 (2022).
21. Erven, J. A. M. *et al.* A high coverage Mesolithic aurochs genome and effective leveraging of ancient cattle genomes using whole genome imputation. *bioRxiv* 2024.01.23.576850 (2024) doi:10.1101/2024.01.23.576850.
22. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* **19**, 220–234 (2018).
23. Peng, L., Brunson, K., Jing, Y. & Zhipeng, L. Zooarchaeological and Genetic Evidence for the Origins of Domestic Cattle in Ancient China. *Asian Perspectives* **56**, 92–120 (2017).
24. Park, S. D. E. *et al.* Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biol.* **16**, 234 (2015).
25. Upadhyay, M. R. *et al.* Genetic origin, admixture and population history of aurochs (Bos

primigenius) and primitive European cattle. *Heredity* **119**, 469 (2017).

26. Freeman, A. R. *et al.* Admixture and diversity in West African cattle populations. *Mol. Ecol.* **13**, 3477–3487 (2004).
27. Mwai, O., Hanotte, O., Kwon, Y.-J. & Cho, S. African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. *Asian-australas. J. Anim. Sci.* **28**, 911–921 (2015).
28. Achilli, A. *et al.* Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr. Biol.* **18**, R157–8 (2008).
29. Achilli, A. *et al.* The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PLoS One* **4**, e5753 (2009).
30. Bonfiglio, S. *et al.* Origin and spread of Bos taurus: new clues from mitochondrial genomes belonging to haplogroup T1. *PLoS One* **7**, e38601 (2012).
31. Bradley, D. G., MacHugh, D. E., Cunningham, P. & Loftus, R. T. Mitochondrial diversity and the origins of African and European cattle. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5131–5135 (1996).
32. Bradley, D. G., Loftus, R. T., Cunningham, P. & MacHugh, D. E. Genetics and domestic cattle origins. *Evolutionary Anthropology: Issues, News, and Reviews* vol. 6 79–86 Preprint at https://doi.org/10.1002/(sici)1520-6505(1998)6:3<79::aid-evan2>3.0.co;2-r (1998).
33. Mannen, H. *et al.* Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. *Mol. Phylogenet. Evol.* **32**, 539–544 (2004).
34. Bollongino, R. *et al.* Modern taurine cattle descended from small number of near-eastern founders. *Mol. Biol. Evol.* **29**, 2101–2104 (2012).
35. Scheu, A. *et al.* The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genetics* vol. 16 Preprint at https://doi.org/10.1186/s12863-

015-0203-2 (2015).

36. Chen, S. *et al.* Zebu cattle are an exclusive legacy of the South Asia neolithic. *Mol. Biol. Evol.* **27**, 1–6 (2010).
37. Mannen, H. *et al.* Cattle mitogenome variation reveals a post-glacial expansion of haplogroup P and an early incorporation into northeast Asian domestic herds. *Scientific Reports* vol. 10 Preprint at https://doi.org/10.1038/s41598-020-78040-8 (2020).
38. Ginja, C. *et al.* Iron age genomic data from Althiburos – Tunisia renew the debate on the origins of African taurine cattle. *iScience* **26**, 107196 (2023).
39. Orlando, L. The first aurochs genome reveals the breeding history of British and European cattle. *Genome biology* vol. 16 225 (2015).
40. Carolino, N. & Gama, L. T. Indicators of genetic erosion in an endangered population: The Alentejana cattle breed in Portugal1. *J. Anim. Sci.* **86**, 47–56 (2008).
41. Bollongino, R., Elsner, J., Vigne, J.-D. & Burger, J. Y-SNPs do not indicate hybridisation between European aurochs and domestic cattle. *PLoS One* **3**, e3418 (2008).
42. Edea, Z. *et al.* Signatures of positive selection underlying beef production traits in Korean cattle breeds. *Hanguk Tongmul Chawon Kwahakhoe Chi* **62**, 293–305 (2020).
43. Boitard, S. & Rocha, D. Detection of signatures of selective sweeps in the Blonde d'Aquitaine cattle breed. *Anim. Genet.* **44**, 579–583 (2013).
44. Du, L. *et al.* Genome-Wide Association Study Based on Random Regression Model Reveals Candidate Genes Associated with Longitudinal Data in Chinese Simmental Beef Cattle. *Animals (Basel)* **11**, (2021).
45. Kenny, D., Sleator, R. D., Murphy, C. P., Evans, R. D. & Berry, D. P. Detection of Genomic Imprinting for Carcass Traits in Cattle Using Imputed High-Density Genotype Data. *Front. Genet.* **13**, 951087 (2022).

46. Corrêa, M. S. L. *et al.* A network-based approach to understanding gene-biological processes affecting economically important traits of Nelore cattle. *Anim. Genet.* **55**, 55–65 (2024).

47. Gamarra, D., Aldai, N., Arakawa, A., Pancorbo, M. & Taniguchi, M. Association of a polymorphism in SREBP1 with fatty acid composition and lipogenic gene expression in beef cattle breeds. (2019) doi:10.21203/rs.2.15242/v1.

48. Cheng, J. *et al.* The ACVR1 gene is significantly associated with growth traits in Chinese beef cattle. *Livest. Sci.* **229**, 210–215 (2019).

49. Mohammadabadi, M., Bordbar, F., Jensen, J., Du, M. & Guo, W. Key Genes Regulating Skeletal Muscle Development and Growth in Farm Animals. *Animals (Basel)* **11**, (2021).

50. Weikard, R., Goldammer, T., Brunner, R. M. & Kuehn, C. Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk production in cows. *Physiol. Genomics* **44**, 728–739 (2012).

51. Gutiérrez-Gil, B., Arranz, J. J. & Wiener, P. An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. *Front. Genet.* **6**, 167 (2015).

52. Lecchi, C. *et al.* Characterisation of adiponectin and its receptors in the bovine mammary gland and in milk. *Vet. J.* **203**, 296–301 (2015).

53. Jing, R., Gu, L., Li, J. & Gong, Y. A transcriptomic comparison of theca and granulosa cells in chicken and cattle follicles reveals ESR2 as a potential regulator of CYP19A1 expression in the theca cells of chicken follicles. *Comp. Biochem. Physiol. Part D Genomics Proteomics* **27**, 40–53 (2018).

54. Vega, W. H. O. *et al.* Variants in the CYP19A1 gene can affect in vitro embryo production traits in cattle. *J. Assist. Reprod. Genet.* **35**, 2233–2241 (2018).

55. Qanbari, S. *et al.* Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* **12**, 318 (2011).

56. Hu, L. *et al.* The HSP90AA1 Gene Is Involved in Heat Stress Responses and Its Functional Genetic Polymorphism Are Associated with Heat Tolerance in Holstein Cows. *J. Dairy Sci.* (2024) doi:10.3168/jds.2023-24007.

57. Liu, Y. *et al.* Discovery of Genomic Characteristics and Selection Signatures in Southern Chinese Local Cattle. *Front. Genet.* **11**, 533052 (2020).

58. Hoorens, P. R. *et al.* Genome wide analysis of the bovine mucin genes and their gastrointestinal transcription profile. *BMC Genomics* **12**, 140 (2011).

59. Frantz, L. A. F. *et al.* Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. *Proceedings of the National Academy of Sciences* **116**, 17231–17238 (2019).

60. Daly, K. G. *et al.* Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* **361**, 85–88 (2018).

61. Gaunitz, C. *et al.* Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* **360**, 111–114 (2018).

62. Damgaard, P. de B. *et al.* 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).

63. Anthony, D. W. The horse, the wheel, and language. *The Horse, the Wheel, and Language* (2010).

64. Hermes, T. R. *et al.* Mitochondrial DNA of domesticated sheep confirms pastoralist component of Afanasievo subsistence economy in the Altai Mountains (3300–2900 cal BC). *Archaeological Research in Asia* **24**, 100232 (2020).