UPPSALA
UNIVERSITET

The genetics behind early cattle keeping

Frida Johnson

Degree project in biology, Master of science (2 years), 2010
Examensarbete i biologi 30 hp till masterexamen, 2010
Biology Education Centre, Uppsala University and Biology Education Centre and Subdepartment of Evolutionary Biology
Supervisors: Anders Götherström and Emma Svensson

Abstract

Ancient DNA can help us to infer evolutionary changes in selection processes, which can no longer be detected among modern populations. Cattle breeding has been an important process in human history, especially for the milk drinking populations in Europe, but when and where selection for traits of interest started in domesticated cattle is still not known. In this project I aim to investigate selection signatures in milk traits. By using SNP typing of coding nuclear genes in the casein region on chromosome 6, a region that harbours genetic variants with a large effect on both protein content and milk yield, it is possible to measure linkage disequilibrium, allele frequencies and infer haplotypes. Applying this technique to ancient cattle material from Hungary, no sign of selection was found neither in allele frequencies nor in haplotypes between aurochs and domesticates. A divergent LD map between the populations is the best indication that some selection occurred in the casein cluster during this time, although not detectable in my allele frequency data. Thus, SNP typing could be used as means to explore selective traces in domestication processes, but further studies with more animals is needed to test the hypothesis and provide proper support for or against it.

Introduction

The first genetic material recovered from an extinct species was published in 1984 on a member of the horse family (Higuchi *et al.* 1984). The field of ancient DNA (aDNA) has since then increased rapidly. In only a little more than two decades research has gone from retrieval of small fragments of mitochondria DNA to large-scale population genetic studies (Shapiro *et al.* 2004), phenotypically important loci (Svensson *et al.* 2007) and whole mitochondria sequences of extinct species (Cooper *et al.* 2001, Haddrath *et al.* 2001, Gilbert *et al*. 2007, Gilbert *et al.* 2008). In the very near future it is actually probable that complete ancient genomes will appear. Including ancient samples in population genetic studies allow us to investigate events from which signals are no longer present in modern day populations and thus help us to make better inferences of the past. Ancient work on European brown bears, for example, demonstrate that population genetic samples over large temporal scales are essential in order to understand past population dynamics and provide tools for reconstructing the evolutionary process (Valdiosera *et al.* 2007). In the future the field of ancient DNA wills most likely impact our understanding of important evolutionary processes. Indeed, the understanding of demographic processes has already been affected by aDNA studies, e.g. in a study by Dalén *et al.* (2007) aDNA reveals lack of postglacial habitat tracking in the arctic fox providing new insights into how species respond to climate change. This study was only made possible by using ancient mitochondrial DNA. The development of new techniques, better suited for the recovery of aDNA, enables the analysis of nuclear markers, thus providing the

possibility to address questions about early breeding and selection in domestic animals (Jaenicke-Després *et al.* 2003, Svensson *et al.* 2007, Svensson *et al.* 2008, Ludwig *et al.* 2009).

Dairy is one of the most prominent agricultural sectors in Europe, particularly in the northern and western parts. Despite the fact that milk is produced in every European country there are strong geographical northwest / southeast clines in milk production and consumption. It can be postulated that this cline is an echo of two processes that occurred in European prehistory, the introduction of domestic animals from their centre of domestication in the Near East and selection for the ability to digest lactose as an adult. These two processes may have been intimately linked in time and space, as selection of the ability to digest milk co-evolved with increased production of fresh milk from dairy animals (Fig. 1).

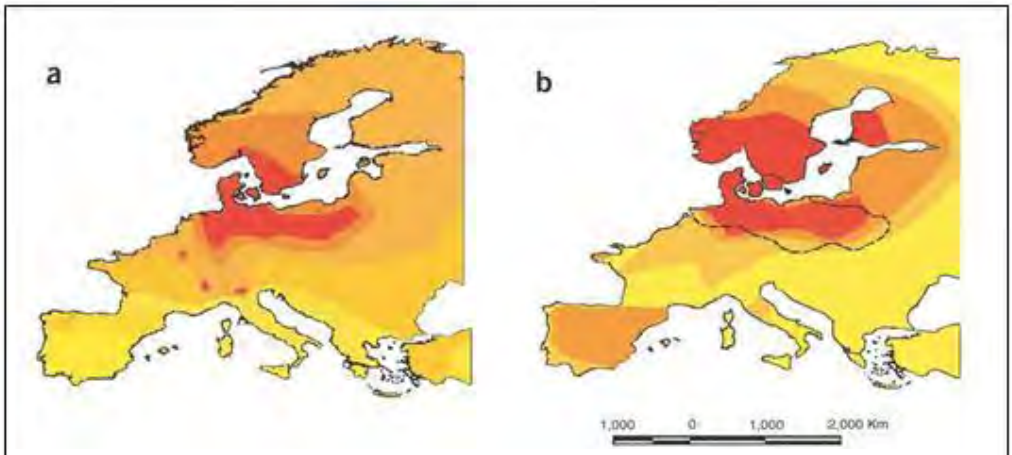

Figure 1: Geographical coincidence between milk gene diversity in cattle (a) and lactase persistence in humans (b). Locations of Neolithic cattle farming sites dotted line in (b). (With permission from Beja-Pereira *et al.* 2003).

As domestic animals, capable of providing milk, were only introduced into Europe relatively recently, i.e. from 10,000 years ago, the rise of dairy farming and milk consumption must have conferred a massive selective advantage to explain the current high frequency of lactose persistence found in some areas of Europe (Fig. 1a). Some populations, particularly those of northwest European descent (Fig. 1b) can drink raw milk throughout life, being lactose tolerant. A variant of the lactase (LCT) gene, the T allele (actually a mutation upstream the gene), correlates with the known distribution of lactose tolerance in Europe, and is either under strong positive selection or linked to a mutation under strong positive selection, probably the single most advantageous genetic trait that has arisen in humans in the last 30,000 years. (Enattah *et al.* 2002, Bersaglieri *et al.* 2004). This selective advantage became

significant after the introduction of domesticated animals in Europe but when and where is still unknown.

If the place and time of dairying in the Neolithic economy could be understood in greater depth and detail we may be closer to understanding the wider significance of cattle to humans throughout history and thus the actual foundation of our lifestyle that the domestication brought with it. The most likely hypothesis, the Secondary product hypothesis, states that lactose tolerance became advantageous later in the Neolithic or Early Bronze Age when pastoral farming was more firmly established (Sherratt *et al.* 1981). However the hypothesis has been questioned, for example a study on organic residues preserved in archaeological pottery have provided direct evidence for the use of milk as early as the 7th millennium BC (Evershed *et al.* 2008), and Spangenberg *et al.* (2008) showed the existence of dairy farms in Europe by the 4th millennium BC. Other studies have shown selection for milk traits already during medieval time in Swedish cattle (Svensson *et al.* 2007) but the situation in central Europe is so far not studied at all. It is claimed that milk has been a valuable and widely used food source in Europe for over 8000 years (Beja-Pereira *et al.* 2004). Much more recently domesticated cattle have become highly specialized for milk production following strong artificial selection increasing the frequency of favorable alleles at the loci affecting milk production traits (Goddard *et al.* 2009, Choi *et al.* 2009).

The casein region on chromosome 6 in bovines has been shown to harbor genetic variants with a large effect on both protein content and milk yield, producing approximately 80 percent of the protein content of milk in cattle (Farrell *et al.* 2004). The region which is often picked up by scans for selection in modern cattle (e.g. Hayes *et al.* 2008, Flori *et al.* 2009) is composed of four genes; αs1-, β-, αs2-, k-casein (CSN1S1,CSN2, CSN1S2 and C2N3, respectively) (Fig. 2). Nilsen *et al.* (2009) constructed a SNP map for the region identifying haplotype structures and single SNPs associated with milk traits. The β-gene (represented by the following markers in this study, CSN2-BMC_9215 and CSN_67) is the most likely loci harbouring the underlying causative variation, with especially two alleles (CSN2-BMC_9215 G allele and CSN_67 C allele) associated with high protein and milk yields in Norwegian Red cattle (Nilsen *et al*. 2009).

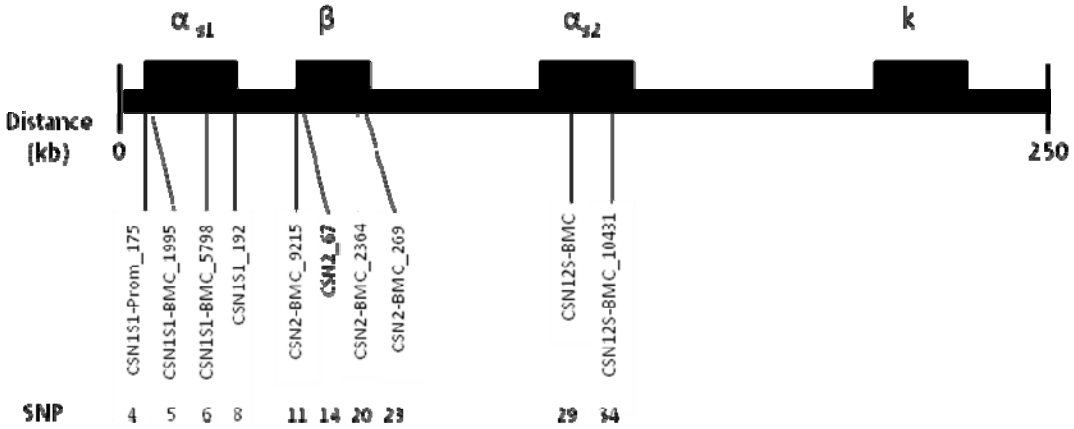

Figure 2. The bovine casein locus with positions of SNPs analyzed in this study marked. The SNPs are numbered according to Nilsen *et al.* 2009.

The aim of this project is to investigate if there are any signs of early selective breeding for milk traits in Europe early after domestication by looking at aurochs and domesticates from a Bronze Age site, Százhalombatta, in Hungary. By looking at haplotypes and allele frequencies from single nucleotide polymorphism (SNP) markers in prehistoric aurochs and cattle, selection signatures may be identified. This is done by targeting 10 SNPs in the casein region on chromosome 6. To my knowledge, this is the first time a number of closely linked SNPs are gathered from a functionally important region in ancient material, and it will, for the first time, yield estimated levels of LD and inferred haplotypes in an ancient population.

Material and Method

(a) *Samples and site information*

The Hungarian material (Százhalombatta) consists of four aurochs and five domestic specimens dated to 3207- 3548 BP (Table 1). All aurochs were identified as male and all domesticates as female with molecular tools (as in Svensson *et al.* 2008, E Svensson unpublished data) as well as morphologically (M.Vretemark unpublished data).

Százhalombatta-Földvár is situated on the west bank of the Danube, thirty kilometers south of Budapest. The site consists of a large tell evolved during the Early and Middle Bronze Age (2000 – 1500 BC). The region represents the main east-west communication line during Bronze Age, forming a complex settlement system with specialized functions in production and trade. Thus the material has potential to show influences transmitted between the Black Sea and Scandinavia (http://www.eoec.org/webpages/szazhalombatta/background.htm Date visited 11 Jan 2010).

One aurochs, also identified as male (Svensson unpublished data), from the south coast of England was also included. For primer optimization post-mediaeval historic material from the harbor in Marstrand contextually dated to the 18th century was used. This material has previously been shown to be very well preserved (Malmström *et al.* 2007). In order to determine ancestral alleles 5 bison (*Bison bison*) and 8 zebu (*Bos indicus*) were genotyped.

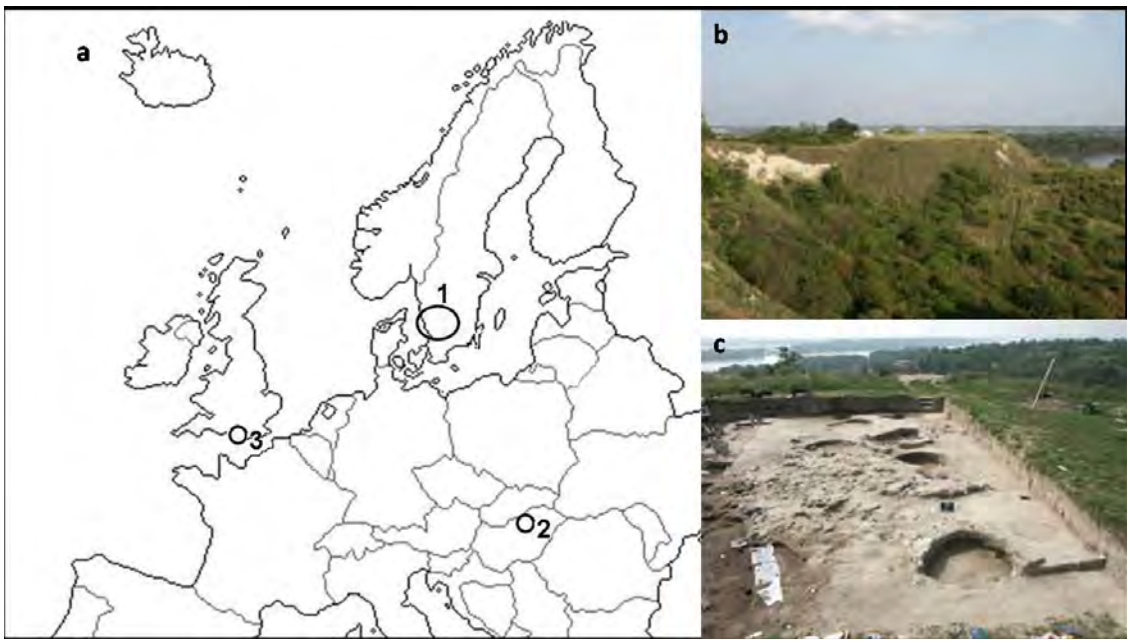

Figure 3. a) Site locations of where historic and prehistoric material was sampled. (1) Marstrand, (2) Szazhalombatta, (3) English coast, Southampton. b) The tell in Százhalombatta and c) The excavation site. Figure b and c taken from http://www.eoec.org/webpages/szazhalombatta/background.htm

(b) ***Genetic analysis***

Primers were designed for 15 SNPs covering approximately 120 kb in and around the casein region (nucleotide position 87903788-88470917 in Btau 4.0) for primer sequences see supplementary table S1. To increase amplification success very short amplicons (53 -88 bp) were designed.

(c) ***Sample extraction and amplification***

All pre-PCR work on ancient samples was done in an isolated lab facility following strict routines, including daily cleaning with bleech and UV-irradiation in fume hoods at nights. Special lab clothing, including full bodysuits, face mask and gloves were used. Contamination was monitored with negative controls in the form of no-template blanks, both PCR and DNA

extraction blanks, which were given the same treatment and processed alongside the other samples. All extracts were screened for human contamination with a 70bp primer, designed to amplify both authentic ancient cattle mtDNA and contaminating human mtDNA. Five out of fifteen SNPs amplified in human DNA even after optimization and were thus withdrawn from further analysis. Positive PCR products were sequenced on a PSQTM 96MA (Biotage) using pyrosequencing TM technology and 25 µl of PCR product. The SNPs were recognized using PSQTM 96MA and edited by the PSQTM 96MA SNP software. Two datasets were created from the obtained genotype data, one collecting all successfully typed samples and one with more strict routines. In the latter all samples in a PCR were withdrawn if a blank control was found contaminated. Only genotypes with a perfect hit with pyrosequencing technique, correct reference peaks and nucleotide levels were kept in this sample set. All statistical and data analyses were performed with the strict dataset. Accepted guidelines for aDNA research were followed (Gilbert & Willerslev 2006). For details on DNA extraction, PCR conditions and SNP typing, see the supplementary material.

Table 1. Description of ancient specimen included in the study. With ^{14}C and ^{13}C values as analyzed by Göran Possnert at Ångströmslaboratoriet, Uppsala

Sample ID	Location	Classification	$\Delta^{13}C$‰pdb	^{14}C age BP
n2	Százhalombatta	Domestic	-20,5	3296 ± 33
n3	Százhalombatta	Domestic		*
N4	Százhalombatta	Domestic	-20,3	3513 ± 35
N5	Százhalombatta	Domestic		*
N6	Százhalombatta	Domestic	-20	3368 ± 36
U1	Százhalombatta	Auroch	-22.5	3241 ± 34
U7	Százhalombatta	Auroch	-22,2	3365 ± 31
U8	Százhalombatta	Auroch		*
Uj	England	Auroch	-23,2	3325 ± 35

*Samples come from the same context as the 14C dated specimens.

(d) Data analysis and statistics

Allelic dropout was calculated for all markers as in Gagneaux *et al.* (1997). Logistic regression as implemented in STATISTICA 9 was used to test for correlation between fragment length and success rate. The allele frequency and heterozygosity was calculated by hand for all samples. Haploview (Barrett *et al.* 2005) was used for computation of linkage disequilibrium (LD) measured by r^2 in domesticates and aurochs.

(e) Haplotype reconstruction

At autosomal loci, the genetic material carried by a diploid individual can be thought of as being composed of two haplotypes, each containing the genetic information from one of the two homologous chromosomes. The common techniques for SNP typing do not provide the allele information separately from each of the two chromosomes, they just give genotype information. Given a set of genotypes, the problem of finding the corresponding two haplotypes for each genotype is called *phasing* the genotypes (Stephens *et al.* 2001). Knowledge about haplotypes carried by sampled individuals is very useful when mapping LD and analyzing population evolutionary history, since genetic inheritance operates through the transmission of chromosomal segments. Here we used the software PHASE (version 2.1.1) based on a Bayesian statistical model for haplotype reconstruction, using default settings and repeating the reconstruction 3 times with same results.

Results & Discussion

Five of the nine Hungarian bones were C14 dated to 3207 - 3548 BP, along with the English aurochs (Table 1). The amplification success rate of ancient DNA was generally high (44%), with short fragments (53-78bp) having a higher success rate (56%) than longer ones (84-88bp) (35%), $p = 0.02$ (Fig. 4). Small marker specific deviations may be due to primers with different quality scores. All samples were successfully typed more than once for every marker and all contained high proportion (on average 90, 5 % - 100%) of cattle mtDNA compared to human mtDNA, which is a strong support for authentic results (Table S3).
The risk of allelic dropout, a heterozygote not detected, in five PCR replicates was minimal ($p < 0,05$) for all markers. On average each genotype was confirmed by 7 replicates (Table S3). Even though the probability for a specific sample to suffer dropout in a specific locus is low, there is a risk of 65% that at least one of all the observations is suffering from dropout.

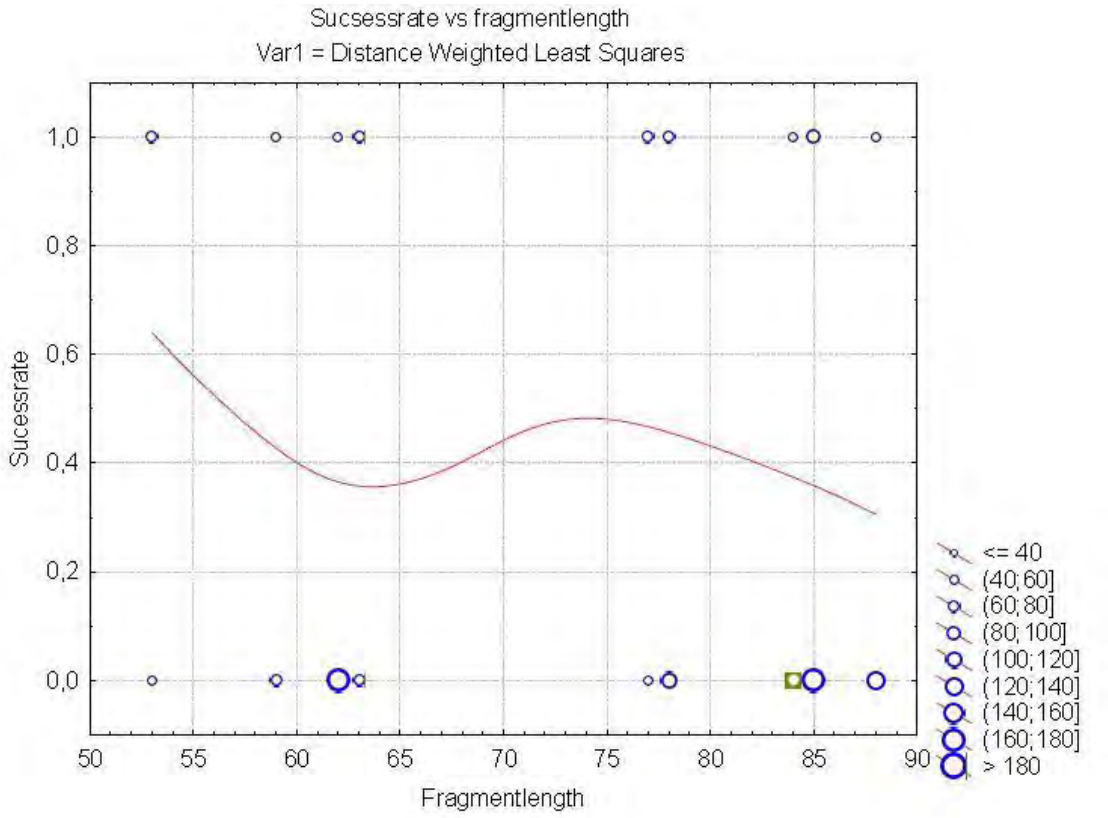

Figure 4. Success rate over fragment size. Overall there was a significant correlation between the length of the PCR product and the genotyping success.

a) Allele frequencies and casein haplotypes

The observed heterozygosity (H_o) was overall low, especially for the aurochs population. A bottleneck or genetic drift are possible scenarios but the small sample size (n=3) may also affect H_o. Allele frequencies for each SNP were compared between the three groups, aurochs, domesticates and Norwegian Red cattle (modern). There was no difference in allele frequencies for SNP 4, 5, 8, 11, 20, 23 or 34 (Fig. 5). Both aurochs and domesticate allele frequencies for SNP 6 (CSN1S1-BMC_5798), 14 (CSN2_67), 20 (CSN2-BMC_2364) and 29 (CSN1S2-BMC) differed slightly from modern Norwegian cattle population.

The results give no indication that selective breeding for milk had started during Bronze Age, but more animals must be typed to verify this. The bison ancestral allele could be determined for 6 of the SNPs, in the other 4 SNPs the bison was polymorphic.

Table 2. Ancestral allele, observed heterozygosity (Ho) and allele frequencies among samples.

SNP	Ancestral allele	Aurochs Hungary		Domesticate Hungary		Cattle Norway[a]	Bison	Zebu
		H_0	Allele frequency	H_0	Allele frequency	Allele frequency	Allele frequency	Allele frequency
CSN1S1-Prom_175	G	0	1,0 (A)	0,15	0,8 (A)	0,876(A)	0(A)	0,714(A)
CSN1S1-BMC_1995	-	0	1,0 (G)	0,17	0,9 (G)	0,969(G)	0,8(G)	0,714(G)
CSN1S1-BMC_5798	C	0	0,375 (T)	0,2	0,3 (T)	0,589(T)	0(T)	0,143 (T)
CSN1S1_192	G	0	1,0 (A)	0,09	0,8 (A)	0,873(A)	0(A)	0,429(A)
CSN2-BMC_9215	-	0,11	0,625 (T)	0,15	0,4 (T)	0,599(T)	0,5(T)	0,857(T)
CSN2_67	C	0,07	0,5 (C)	0,07	0,4 (C)	0,6(C)	1,0(C)	0,857 (C)
CSN2-BMC_2364	G	0,19	0,375 (A)	0,2	0,4 (A)	0,552(A)	0(A)	0,071(A)
CSN2-BMC_269	-	0,15	0,875 (T)	0,03	0,9 (T)	0,919(T)	-	0,750(T)
CSN1S2-BMC	-	0,09	0,5 (G)	0,03	0,4 (G)	0,634(G)	0,1(G)	0,8(G)
CSN1S2-BMC_10431	A	0	1,0 (A)	0,05	0,9 (A)	0,955(A)	1,0 (A)	0,786 (A)

[a] Published data (Nilsen *et al.* 2009)

In total, 6 different haplotypes were constructed in the aurochs and domesticate population by the software PHASE(2.1.1). Two were unique and could not be found among the modern haplotypes. An overall decrease in haplotype frequency from (0.81) in ancient to (0.74) in modern population could be observed. The loss of unique haplotypes and alleles could be due to breeding, since a higher diversity is expected in a region closer to the domesticating centre. Of the remaining 4 haplotypes, one was aurochs specific and three domesticate specific in my Hungarian dataset (Table 2). The probability for each haplotype was in general high (1.0) for n2, u1 and uj. (0.938) for n3, (0.981) for n4, (0.913) for n5, (0.854) for n8. But lower for haplotype n6 (0.156) and u7 (0.4156). Since only 10 of 38 SNPs used to construct the haplotypes for modern cattle in the region were typed and analyzed, comparison with modern haplotypes must be speculative. Also analysis on modern specimen for these 10 SNPs would be needed to exclude more unique haplotypes and assure difference between domesticates and aurochs. No particular divergence could be found only looking at haplotypes between the populations since both cluster into same haplogroups associated with both reduced and increased milk and protein content.

If looking at individual haplotypes a major decrease in frequency could be seen for haplotype 1 from ancient to modern population. Looking at the populations separately demonstrate that the aurochs are mainly responsible for this high frequency in Bronze Age population, with a frequency of (0.5) compared to domesticates (0.36). This is interesting since the study on modern breeds detected four haplotypes that associate with reduced milk and protein

production; all of these contain, as in haplotype 1, the A-allele of CSN2_67 in addition to the G-allele of CSN2-BMC_9215 (Nilsen et al. 2009). Looking at the other SNPs show an increase in modern cattle for haplotype 3,4 and 6, all containing the milk yielding CSN2-BMC_9215 T- allele in addition to the CSN2-67 C - allele. Also this time the aurochs population is responsible for the pattern, having lower haplotype frequencies than domesticates.

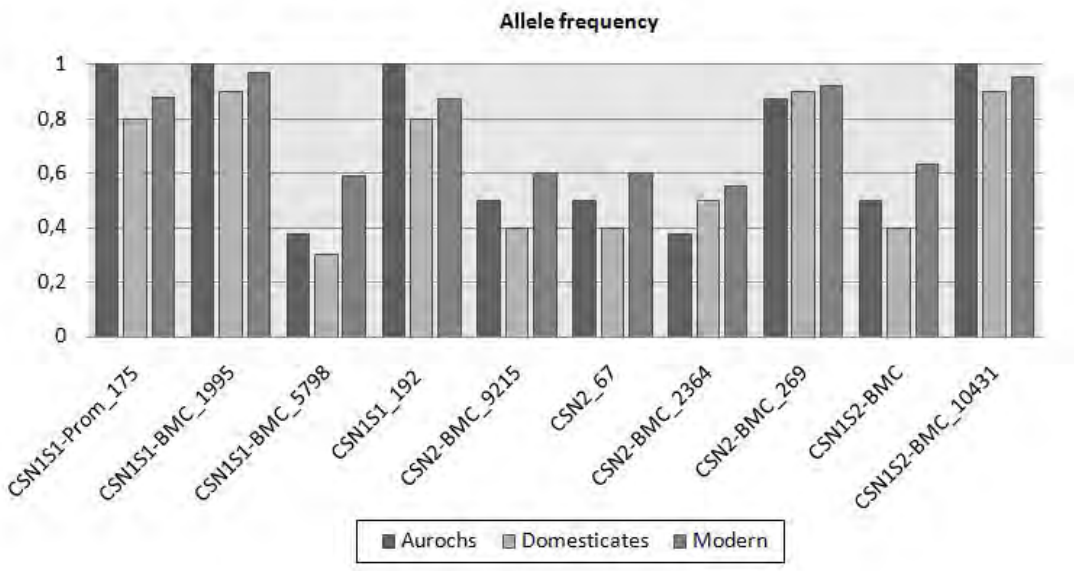

Figure 5. Variation in allele frequencies (major allele) in the 10 SNPS among ancient domesticates, aurochs and modern specimen (20th century). Data for the modern cattlefrom Nilsen *et al.* 2009.

Table 3. Casein haplotypes in five domestic cattle (n2-n6) and four aurochsen (u1, u7, u8, uj) constructed by software PHASE.

			Marker										
Haplotype	Specimen	Freq (%)	4	5	6	8	11	14	20	23	29	34	Haplotype Nilsen[a]
1	n2, n3, n5, n6 u1, uj	0,445	A	G	C	A	G	A	G	T	A	A	3/6 (0,16)
2	n3	0,052	A	G	T	A	T	C	A	T	A	A	-
3	n4, u7, u8	0,178	A	G	T	A	T	C	A	T	G	A	1/5 (0,39)
4	n4,n5	0,104	G	G	C	G	T	C	A	T	G	A	4 (0,13)
5	n6	0,009	A	C	T	A	G	A	G	C	G	G	-
6	u7	0,021	A	G	C	A	T	C	G	C	G	A	8/12 (0,04)

b) Extent of LD

Linkage disequilibrium (LD) was calculated with phased haplotypes between pair of loci in the two populations using Haploview (Barrett *et al.* 2005). LD between pairs of loci varied from complete disequilibrium to no disequilibrium. The aurochs population generates a different LD pattern than domesticates with LD between different loci and in a limited region (Fig 6). The domesticated population gives rise to LD in three blocks with possible recombination between them (Fig.7). Both populations show complete LD between the milk associated markers 11 (CSN2-BMC_9215) and 14 (CSN2_67) which is also confirmed by the study on modern breeds (Nilsen *et al.* 2009). The disparate pattern between the populations could suggest that there have been selection for breeding in domesticated specimens during this time. But the small sample size (n=8 chromosomes for aurochs and n= 10 chromosomes for domesticates) could also be responsible for the pattern seen. An LD map from Zebu specimen (*Bos indicus*) was generated showing overall low LD values, apart from SNP 11 and 14 were LD was complete (Fig S1). This is true for all datasets and indicates the importance of these two markers in the casein region.

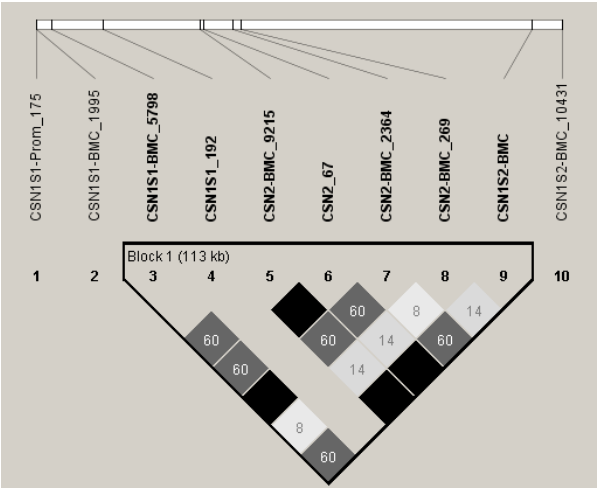

Figure 6. LD across the casein segment in aurochs. Each square contains the level of LD measured by r2 between the markers specified; dark tones correspond to increasing levels of r2; triangles indicate division by loci.

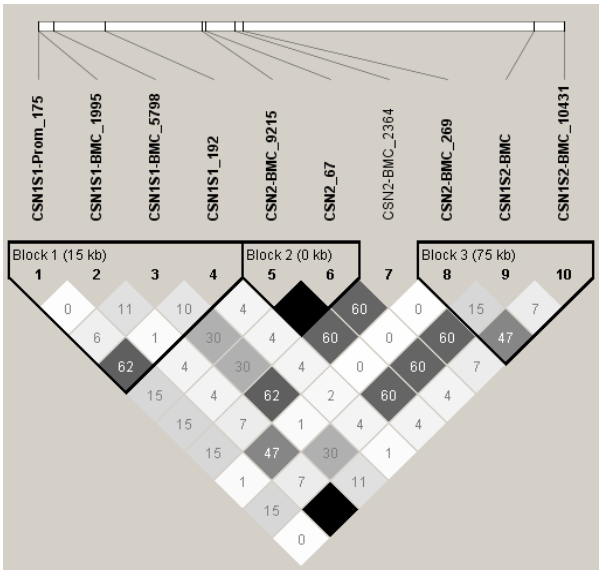

Figure 7. LD across the casein segment in domesticates. Each square contains the level of LD measured by r2 between the markers specified; dark tones correspond to increasing levels of r2; triangles indicate division by loci.

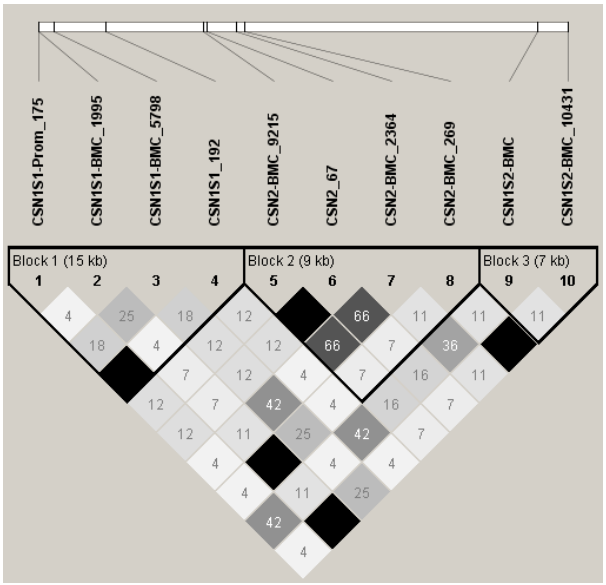

Figure 8. LD across the casein segment in all ancient animals. Each square contains the level of LD measured by r2 between the markers specified; dark tones correspond to increasing levels of r2; triangles indicate division by loci.

(c) Conclusions and future perspectives

The results show no differences in allele frequencies at individual SNPs or haplogroups associated with milk traits between aurochs and domesticates. However, change in haplotype frequencies between ancient and modern population confirm an increase among domesticates in haplotypes containing the milk producing alleles C and T at marker 11 and 14, with aurochs having a higher frequency of haplotypes containing the milk reducing alleles A and G at the same markers. The small sample size may be unable to detect changes in allele frequencies and haplotype structure, but the result could also just reflect the small sample size in itself. Perhaps is haplotype data more informative than allele data in studies of population differences with ancient DNA data of small samples sizes.

The most notable pattern in the study can be seen within the LD map, with a divergent pattern of LD between pair of loci among aurochs and domesticates. An indication of differences in selection pattern among domesticates and aurochs during Bronze Age. However, the data set is small and this conclusion should be viewed as a hypothesis in need of further testing with more individuals, more typing to confirm genotypes in some SNPs and test with a neutral marker to out rule other demographic events

Future studies will involve coalescent simulation by the software COMPASS (Jakobsson 2009) in order to investigate possible different demographic histories, caused by selection, for some alleles associated with milk traits of interest. We will also be able to estimate effective population size using LD. On the whole my study demonstrates that it is possible to amplify and analyze ancient nuclear DNA with both authenticity and high success rate, indicating what a tremendous potential aDNA studies have in the future for studying selection signatures through time.

Acknowledgements

The material from Hungary and Marstrand was kindly provided by Dr. Maria Vretemark. The English specimen was provided by Dr. Jaco Weinstock. The samples from Clad Hallan were extracted by Simon McGrory. Thanks to Emma Svensson and Anders Götherström for their encouragement, advice and helpful comments and of course, making this study possible.

References

Barrett J.C, Fry B, Maller J, Daly M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Oxford University Press* **2:**263 265. Beja-Pereira A, Luikart G. 2003. Gene-culture coevolution between cattle milk protein genes and human lactase genes *Nature Genetics* **35:** 311 – 312.

Beja-Pereira A, Luikart G, England PR, Bradley DG, Jann OC, Bertorelle G, Chamberlain AT, Nunes TP, Metodiev S, Ferrand N, Erhardt G. 2003. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat Genet* **35:**311-313.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* **74:** 1111-1120.

Choi J, Curtis P.V.T, Gill G.A. 2009. Grefenstette J.J, Matukumalli L.K, Villa-angulo R. High-resolution haplotype block structure in the cattle genome. *BMC Genetics* **10:**19

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409:** 704-707.

Dalén L, Nyström V, Valdiosera C, Germonpré, Sablin M, Turner E, Angerbjörn A, Arsuaga J.L, Götherström A. 2007. Ancient DNA reveals lack of postglacial habitat tracking in the artic fox. *PNAS* **104,** 6726-6729.

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Järvelä I. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30,** 233-237

Evershed RP, Payne S, Sherratt AG, Copley MS, Coolidge J, Urem-Kotsu D, Kotsakis K, Ozdoğan M, Ozdoğan AE, Nieuwenhuyse O, Akkermans PM, Bailey D, Andeescu RR, Campbell S, Farid S, Hodder I, Yalman N, Ozbaşaran M, Biçakci E, Garfinkel Y, Levy T, Burton MM. 2008. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature*. **25,** 528-31.

Farrell H, Jimenez-Flores R, Bleck G, Brown E, Butler J, Creamer L, Hicks C, Hollar C, Ng-Kwai-Hang K, Swaisgood H. 2004. Nomenclature of the proteins of cow´s milk – sixth revision. *J Dairy Sci.* **87,** 1641-1674.

Flori L, Fritz S, Jaffréziv F, Boussaha M, Gut I, Heath S, Foulley JL, Gautier M. 2009. The genome response to artificial selection: a case study in dairy cattle. *PLoS One* **12:**4

Gagneux P, Boesch C, Woodruff DS. 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Mol Ecol* **6:**861-8.

Gilbert MT, Willerslev E. 2006. Authenticity in ancient DNA studies. *Med Secoli*. **18:**701-23.

Gilbert MT, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, Tikhonov A, Dalén L, Kuznetsova T, Kosintsev P, Campos PF, Higham T, Collins MJ, Wilson AS, Shidlovskiy F, Buigues B, Ericson PG, Germonpré M, Götherström A, Iacumin P, Nikolaev V, Nowak-Kemp M, Willerslev E, Knight JR, Irzyk GP, Perbost CS, Fredrikson KM, Harkins TT, Sheridan S, Miller W, Schuster SC. 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **28:**1927-30.

Gilbert MT, Drautz DI, Lesk AM, Ho SY, Qi J, Ratan A, Hsu CH, Sher A, Dalén L, Götherström A, Tomsho LP, Rendulic S, Packard M, Campos PF, Kuznetsova TV, Shidlovskiy F, Tikhonov A, Willerslev E, Iacumin P, Buigues B, Ericson PG, Germonpré M, Kosintsev P, Nikolaev V, Nowak-Kemp M, Knight JR, Irzyk GP, Perbost CS, Fredrikson KM, Harkins TT, Sheridan S, Miller W, Schuster SC. (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proc Natl Acad Sci U S A.* **17:** 8327-32.

Goddard M, McEwan J, Hayes B, MacEachern S. (2009) An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (Bos taurus) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. *BMC Genomics.* **10:**181

Hayes BJ, Lien S, Nilson H, Olsen HG, Berg P, Maceachern S, Potter S, Meuwissen TH. 2008. The origin of selection signatures on bovine chromosome 6. *Anim Genet.* **39:**105-11.

Haddrath O, Baker A. J. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesus. *Proc. R. Soc. Lond. B* **268:** 939-945.

Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312:** 282-284.

Janicke-Després V, Buckler ES, Smith BD, Gilbert MT, Cooper A, Doebley J, Pääbo S. 2003. Early allelic selection in maize as revealed by ancient DNA. *Science* **14:**1206-1208.

Jakobsson M. 2009. COMPASS: a program for generating serial samples under an infinite sites model. *Oxford University Press* **21:** 2845-2847.

Ludwig A, Pruvost M, Reissmann M, Benecke N, Brockmann GA, Castaños P, Cieslak M, Lippold S, Llorente L, Malaspinas AS, Slatkin M, Hofreiter M. (2009) Coat color variation at the beginning of horse domestication. Science 24:485.

Nilsen H, Olsen HG, Hayes B, Sehested E, Svendsen M, Nome T, Meuwissen T, Lien S. 2009. Casein haplotypes and their association with milk production traits in Norwegian Red cattle. *Genet Sel Evol.* **20:**41-24.

Malmström H, Svensson E, Gilbert M.T.P, Willerslev E, Götherström A, Holmlund G. 2007. More on contamination: The use of asymmetric molecular behavior to identify authentic ancient human DNA. *Mol. Biol. Evol*. **24:** 998 – 1004.

Ronaghi M, Shokralla S, Gharizadeh B. 2007 Pyrosequencing for discovery and analysis of DNA sequence variations. *Pharmacogenomics* **10:**1437-1441. Shapiro B, Drummond AJ, Rambaut A, et al. (27 co authors) 2004. Rise and fall of the Beringian steppe bison. *Science* **306:** 1561-1565.

Sherratt, A. 1981 Plough and pastoralism: aspects of the secondary products revolution, pp 261–305. In Pattern of the Past: Studies in honour of David Clarke, edited by I Hodder, G Isaac and N Hammond. Cambridge University Press: Cambridge

Spangenberg JE, Matuschik I, Jacomet S, Schibler J. 2008. Direct evidence for the existence of dairying farms in prehistoric Central Europe (4th millennium BC). *Isotopes Environ Health Stud* **44:** 189-200.

Stephens, M., Smith, N.J., Donnelly, P. 2001 A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68:** 978-989.

Svensson E.M, Anderung C, Baubliene J, Persson P, Malmström H, Smith C, Vretmark M, Daugnora L, Götherström A. 2007. Tracing genetic change over time using nuclear SNPs in ancient and modern cattle. *Animal Genetics* **38:** 378 – 383.

Svensson E, Götherström A. 2008. Temporal fluctuations of Y-chromosomal variation in Bos taurus. *Biol. Lett.* **4:** 752-754.

Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR. 1998. Technical note:improved DNA extraction from ancient bones using silica-based spin columns. *AM J Phys Anthropol,* **105:** 539-543.

Valdiosera C, Garcia N, Anderung C, Dalén L, Crégut-Bonnoure E, Kahlke R.D, Stiller M, Brandström M, Thomas M.G., Arsuaga J.L, Götherström A, Barnes I. 2007 Staying out in the cold: glacial refugia and mitochondrial DNA phylogeny in ancient European brown bears. *Molecular Ecology* **16:** 5140-5148.

SUPPLEMENTAL MATERIAL & METHODS

DNA extraction

Bones from each specimen was decontaminated with 1 J/m2 irradiation in a UV-crosslinker (Techtum Lab) before extraction. An automated drill (Dremmel) was used to generate approximately 100mg bone powder from the specimen. The powder were subsequently incubated at 38° C in 1 ml buffer containing 0.5 M EDTA, 1 M urea and 100 µg/ml proteinase K over the night. The sample was then centrifuged at 2000 rpm for 5 min and the supernatant transferred to an Amicon Ultra-15 centrifugal filter (Millipore, MA, USA) which was then spun at 4000g for 10 to 15 minutes until it was concentrated down to 100ul. The extract was then purified using QIAquickTM silicabased spin columns (Qiagen, Hilden, Germany) based on the method by Yang and others (1997). The extract was added to a spin column together with 500µl of PB buffer, following centrifugation at 13 000 rpm for 1 minute. Then 750 µl PE buffer was added and the column was centrifuged twice, until all liquid was gone. DNA was then eluded by adding 50 µl EB buffer to the column membrane and centrifuging for 1 minute at 13 000 rpm. 5µl sample was used in an 55µl PCR with a 70bp primer amplifying authentic ancient cattle DNA. Samples from Cladh Hallan (Scotland) were also genotyped but due to lack of material and problems with instrument not enough data for conclusive results were obtained.

DNA Amplification and genotyping

DNA was amplified with a 70bp primer, to screen for proportion authentic ancient cattle mtDNA in samples, as well as primers targeting 10 different SNPs on the casein region on chromosome 6. The PCR was in 30 µl volumes with 1x reaction buffer (Naxo, Estonia), 2,5 mM MgCl2, 0,2 mM of each dNTP, 0,18 µM of each primer, 1U Smart-Taq DNA polymerase (Naxo), 1-2 µl extract and ddH2O to a final volume of 30 µl. Rat serum albumin (RSA) was used (10g/L) to improve amplification success. Contamination was monitored with blank controls for every 8 reactions. Reactions were run on a PTC-225 DNA engine tetrad (MJ Research, Waltham, USA) with an initial 7 minute activation step at 94° C, then 37 -48 cycles of 94° C for 1 min, 50 / 52 / 55 or 58° C for 1 minute and 72° C for 1 minute. The program also included a final extension step at 72° C for 7 minutes. Positive PCR products were genotyped on a PSQTM 96MA (Biotage) using pyrosequencing technology (Ronaghi et al. 2009) and 30 – 50ul of PCR product.

Table S1. PCR and sequencing primers for the 10 markers used in study and their optimal annealing temperature.

Gene	Name	Sequence 5´ - 3	Annealing temp (°C)
CSN1S1-Prom_175	CSN1S1_175F	Ccattccatttcctgtataatgag[a]	50
	CSN1S1_175Rbio	tctaaggagagtttacaacaaaga	
	pyroCSN1S1_175	tccatttcctgtataatgag	
CSN1S1-BMC_1995	CSN1S1-BMC_1995Fbio	taaactctccttagaatttcttgg	55
	CSN1S1-BMC_1995R	cacataggaaatcaatgttctgtt	
	pyroCSN1S1-BMC_1995	aaatcaatgttctgttcag	
CSN1S1-BMC_5798	CSN1S1 BMC 5798F-bio	ttcgctgaggtgtaatatcttcaa	55
	CSN1S1 BMC 5798R	agtccaaatgccctcttctctaa	
	pyroCSN1S1 5798 BMC	tctaaaaaggtttgtgaga	
CSN1S1_192	CSN1S1 F-bio	ccccatcattctctgacatcc	55
	CSN1S1 R	ttaccaccacagtggcatagtagt	
	CSN1S1 pyro	acagtggcatagtagtcttt	
CSN2-BMC_9215	CSN2_9215Fbio	cacaattatttcaccacatgactc	55
	CSN2_9215R	tccctcatatgctcatatatcaaa	
	pyroCSN2_9215	atgctcatatatcaaaaaca	
CSN2_67	CSN2 8101 F-bio	ctttgcccagacacagtctcta	55
	CSN2 8101 R	ggtttgagtaagaggagggatgtt	
	pyro CSN2 8101	tgtgggaggctgtta	
CSN2-BMC_2364	CSN2_2364F-bio	gagctagcatatttagtcaagaga	52
	CSN2_2364R	cctttcaaatagcacatgtatcc	
	pyroCSN2_2364	gcacatgtatcctaacaac	
CSN2-BMC_269	CSN2_269F-bio	tcaaaattggtgagagacagtcat	52
	CSN2_269R	acatttccctttctcaggaagat	
	pyroCSN2_269	gatgctttacatattgtgc	
CSN1S2-BMC_3312	CSN1S2 3312F	ggtcgtgactagaagaggttctgt[a]	55
	CSN1S2 3312R-bio	agctatgaggctcagcataagg	
	pyroCSN1S2 3312	ctagaagaggttctgtggt	
CSN1S2-BMC_10431	CSN1S2_10431F	acgggtctgtctggaatgga[a]	58
	CSN1S2_10431R-bio	ggcaaactgaatccagcatatcct	
	pyroCSN1S2_10431	agccaatggcctcct	

[a]Biotinylated in the 5´end

SUPPLEMENTAL RESULTS

Allelic dropout formula

$$P\,(false\ homozygote) = k\left(\frac{k}{2}\right)^{\frac{n-1}{k}} = \frac{observed\ nr\ of\ allelic\ dropouts}{all\ observations\ from\ heterozygous\ individuals}$$

n = antalet typningar

Table S2. Allelic dropout for different markers. A minimum of four successful SNP identifications, five for marker 8, for each historic sample were considered sufficient for inclusion in the statistical analysis.

Marker	4	5	6	8	11	14	20	23	29	34
P (n= 4)	0,0078	0,0042	0,043	0,025 (n=5)	0,023	0,025	0,017	0,035	0,025	0,040

Table S3. Proportion of cattle fragments compared to human fragments in all samples with the 70bp primer system.

Sample	Locality	%70bp	Average %
MS1	Marstrand	100, 97,8	98,9
MS2	Marstrand	96,7, 100	98,4
MS5	Marstrand	97,5, 98,1, 96,5	97,4
MS7	Marstrand	100, 98,1	99
MS9	Marstrand	96,3	96,3
MS10	Marstrand	97,3, 97,8	97,6
ML1	Marstrand	89,5, 97,1	93,3
ML2	Marstrand	93, 97,6	95,3
ML3	Marstrand	96,9, 97,7, 100	97,3
ML5	Marstrand	100, 97,1, 100	100
ML8	Marstrand	100, 97,8	99
ML10	Marstrand	96,4, 97,2	98,9
ML11	Marstrand	94,8	94,8
Domestic			
Nöt 2	Hungary	84,98,95,98,96,2,80,7,100,93,5,98,5,97,3	94
Nöt 3	Hungary	100, 96,6, 97, 97, 97, 96, 96,8, 100, 100, 97,4, 100, 100	98,2
Nöt 4	Hungary	96, 100, 93, 94, 95,9, 96,5, 100, 100, 100, 100, 100, 100	98
Nöt 5	Hungary	98, 97, 95, 95,8, 98,6, 97, 100, 92,5, 98,2, 96,5, 98,7	97
Nöt 6	Hungary	96,5, 97,8, 86, 95,8, 97, 100, 100, 98,1, 97,7, 92,5	96
Aurochs			
U1	Hungary	100, 96,3, 98,4, 90,6, 98,5, 95,6	96,6
UJ	Hungary	100, 96,8	98,4
U7	Hungary	60,8, 100, 69,9, 98,9, 98,6, 97,1, 94,6	88,8
U8	Hungary	97,7, 100, 100, 100, 97,1	99
U9	Hungary	90, 91	90,5

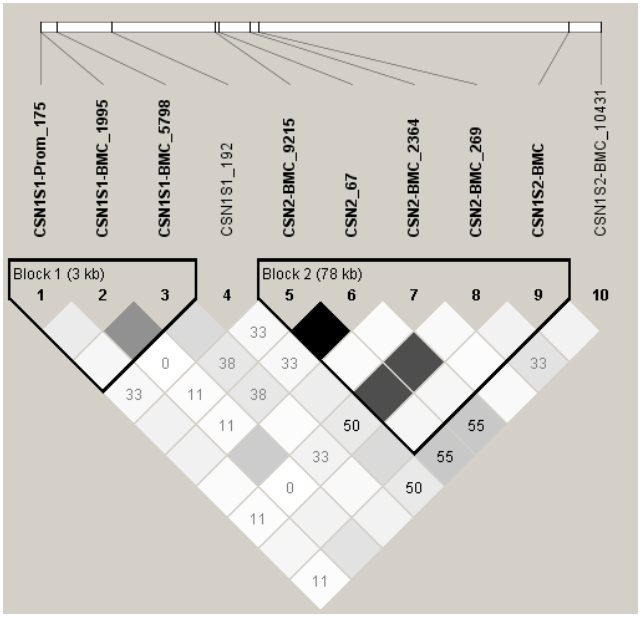

Figure S1. LD across the casein segment in Zebu. Each square contains the level of LD measured by r^2 between the markers specified; dark tones correspond to increasing levels of r^2; triangles indicate division by loci.

Table S4. Ancient samples replication and genotype. All replicates are listed together with the genotype used in the analysis. The genotypes are produced with strict routines excluding possible contaminated samples or those vaguely indicated by pyrosequencing technique.

Sample ID	4			5			6			8			11			14			20			23			9			34
	AA	AG	GG	GG	GC	CC	CC	CT	TT	AA	AG	GG	GG	GT	TT	AA	AC	CC	AA	AG	GG	TT	TC	CC	GG	GA	AA	AA	AG	GG
Nöt 2	VIIII			AA	VII		GG	VII		CC	V		AA	VI		GG	VI		AA		IIII	GG	VI		TT		V	AA	V		AA
Nöt 3	IIII			AA	VIII		GG	II	I	III	CT	XVI	I		AA	I	V	V	GT	I	II	III	AC	II	II	I	AG	VII		TT		VI	AA	IIII		AA
Nöt 4	VI	I	III	AG	VII		GG	I	II	II	CT	III	I	II	AG			V	TT			IIII	CC	IIII			AA	VII		TT	VI		III	GG	IIII		AA
Nöt 5	IIII	IIII	I	AG	VI		GG	VII			CC	III	I	IIII	AG	III	I	VII	GT	III		V	AC	I	I	V	AG	VI		TT	II		III	AG	V		AA
Nöt 6	XIII	II		AA	III	V	GC	III	III	III	CT	VI	I	I	AA	VIII			GG	IIII			AA		II	III	GG	V	I	I	TC	VIII	I	II	AG	II	I	I	AG
U1	VIII			AA	IIII		GG	IIII			CC	V			AA	X			GG	IIII			AA		I	V	GG	III			TT	I		III	AA	IIII			AA
U7	VIII			AA	VI		GG	IIII		IIII	CT	VI			AA		II	VI	TT			X	CC	II	IIII	III	AG	IIII	II		TC	VIII	I	III	AG	IIII			AA
U8	VIII			AA	V		GG			V	TT	IIII			AA	I	I	III	TT		I	III	CC	V			AA				-	VIII	I		GG	IIII			AA
U9	II			AA	III		GG	I		I	CT	I			AA		I	I	GT	I			AA				-				-				-				-
UJ	IIII			AA	VII		GG	V			CC	X			AA	X			GG	VIIII		I	AA			V	GG	IIII			TT	III	I	V	AA	VI			AA

Table S5. Genotypes for modern samples, Bison and Zebu.

Sample ID	Specimen	4	5	6	8	11	14	20	23	29	34
						Genotype					
B1	Bison	GG	GG	CC	GG		CC	GG		AA	AA
B2	Bison		GG	CC	GG		CC	GG		AA	AA
B3	Bison	GG	GC	CC	GG		CC	GG		AA	AA
B4	Bison	GG	GG	CC	GG		CC	GG		AG	AA
B5	Bison	GG	GC	CC	GG	TG	CC	GG		AA	AA
KZ14	Zebu	AG	GC	CC	AG	TT	CC		TT		AA
KV24	Zebu							GG	TT	GG	AA
KV27	Zebu	AA	GG	CC	AG	TT	CA	GG	TT	AG	AA
KV30	Zebu	AA	GC	TC	AG	TG	CA	GG	TC	GG	AG
KV3	Zebu	AG	GC	TC	AG	TG	CA	GG	TC	GG	AG
KV4	Zebu	AA	GC	CC	AG	TT	CC	GA			AA
RZ8	Zebu	AA	GG	CC	GG	TT	CC	GG		AG	AG
RZ9	Zebu	GG	GG	CC	AG	TT	CC	GG	TT		