bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Complex admixture preceded and followed the extinction of wisent in the wild

Karolina Węcek[1,2], Stefanie Hartmann[2], Johanna L. A. Paijmans[2], Ulrike Taron[2], Georgios Xenikoudakis[2], James A. Cahill[3], Peter D. Heintzman[3], Beth Shapiro[3,4], Gennady Baryshnikov[5], Aleksei N. Bunevich[6], Jennifer J. Crees[7], Roland Dobosz[8,9], Ninna Manaserian[10], Henryk Okarma[11], Małgorzata Tokarska[12], Samuel T. Turvey[7], Jan M. Wójcik[12], Waldemar Żyła[8], Jacek M. Szymura[1], Michael Hofreiter[2], Axel Barlow[2]

[1] Department of Comparative Anatomy, Institute of Zoology, Jagiellonian University, Kraków, Poland

[2] Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany

[3] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, 95064, USA

[4] University of California Santa Cruz Genomics Institute, University of California, Santa Cruz, CA, 95064, USA

[5] Gennady Baryshnikov, Laboratory of Theriology, Zoological Institute of the Russian Academy of Sciences, Russia

[6] State National Park, Belovezhskaya Pushcha, Kamieniuki, Belarus

[7] Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

[8] Upper Silesian Museum, Bytom, Poland

[9] Department of Zoology, Faculty of Biology and Environmental Protection, University of Silesia, Katowice, Poland

[10] Institute of Zoology, Armenian National Academy of Sciences, Yerevan, Armenia

[11] Institute of Nature Conservation, Polish Academy of Sciences, Kraków, Poland

[12] Mammal Research Institute, Polish Academy of Sciences, Białowieża, Poland

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Keywords: Admixture, ancient DNA, European bison, *Bison bonasus*, domestic cattle, aurochs, hybridisation

Corresponding authors:

Karolina Węcek

Department of Comparative Anatomy, Institute of Zoology, Jagiellonian University, Gronostajowa 9, 30-387 Kraków, Poland

Fax: (+48) 12 664 51, email: k.wecek@gmail.com

Axel Barlow

Institute for Biochemistry and Biology. University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam OT Golm, Germany

Fax: (+49) 33 197 72 921, email: axel.barlow.ab@gmail.com

Running title: Genomic evidence of wisent admixture.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Retracing complex population processes that precede extreme bottlenecks may be impossible using data from living individuals. The wisent (*Bison bonasus*), Europe's largest terrestrial mammal, exemplifies such a population history, having gone extinct in the wild but subsequently restored by captive breeding efforts. Using low coverage genomic data from modern and historical individuals, we investigate population processes occurring before and after this extinction. Analysis of aligned genomes supports the division of wisent into two previously recognised subspecies, but almost half of the genomic alignment contradicts this population history as a result of incomplete lineage sorting and admixture. Admixture between subspecies populations occurred prior to extinction and subsequently during the captive breeding program. Admixture with the *Bos* cattle lineage is also widespread but results from ancient events rather than recent hybridisation with domestics. Our study demonstrates the huge potential of historical genomes for both studying evolutionary histories and for guiding conservation strategies.

Introduction

The last known wild wisent, or European bison (*Bison bonasus*), was shot and killed in 1927, marking the extinction of this species in the wild (Pucek, 1991). As a result of an intensive captive breeding program and a series of re-establishments, today the species again occupies part of its former range in Central and Eastern Europe. The total population of free-ranging wisent now stands at 5,553 (Raczyński, 2014), and the International Union for Conservation of Nature no longer considers the wisent as an endangered species (IUCN 2008).

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The decline of the original wisent population was protracted, and their subsequent restitution has been complex. In historical times, wisent ranged extensively across semi-open habitats and broadleaved, mixed and coniferous forests in Western Europe, from what is today France in the west to the Volga River and the Caucasus in the east, with the northernmost range limits around 60° north (Fig. 1; Kuemmerle et al., 2011; Kerley et al., 2012; Bocherens et al., 2015). However, ongoing habitat fragmentation and overhunting eradicated most populations. By the end of the 19th century, there were only two populations of wisent left in the wild that were assigned to separate subspecies: in Białowieża Forest (Lowland wisent, *B. b. bonasus*) and in the western Caucasus Mountains (Caucasian wisent, *B. b. caucasicus*). Finally, even these populations collapsed; the last wild Lowland wisent was shot in Poland in 1919 followed by the last Caucasian animal in 1927 (Pucek, 1991).

In 1924, the captive population consisted of only 54 individuals (29 males and 25 females). However, detailed analysis of pedigrees (Olech 2009, Slatis 1960) has shown that some individuals in this population were the direct descendents of others, and that some individuals left no modern descendents. Thus, the true founding population of wisent was considerably smaller, and is thought to comprise of just 12 individuals (Slatis 1960). All but one of these 12 founders were Lowland wisent, almost half of which came from a population established in 1865 in Pless (now Pszczyna, Poland). The remaining founder was a Caucasian wisent bull named M100 KAUKASUS, that represented the last surviving pure Caucasian wisent in captivity. The modern herds that are derived from this founding population are managed as two separate genetic lines. The Lowland line (L) derives from 7 Lowland founders (4 males and 3 females), and is thus considered to represent a pure Lowland wisent lineage. The Lowland-Caucasian (LC) line originates from all 12 founders (5 males, 7 females), which

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

included the last remaining Caucasian wisent bull (Slatis, 1960). Descendants of the LC line thus represent a mixture of Lowland and Caucasian wisent ancestry.

Although the wisent restitution undoubtedly represents a tremendous conservation success, several factors may limit the long-term viability of the species, many of which are applicable to ex-situ conservation strategies in general. A factor that has received particular attention is that of reduced genetic variability, which may be correlated with a lowered resistance to disease and parasites in wisent (e.g.: Krasińska & Krasiński, 2007 and references therein; Adaszek et al., 2014; Karbowiak et al., 2014a; 2014b and the references therein; Krzysiak et al., 2014; Majewska et al., 2014; Moskwa et al., 2014; Oleński et al., 2015; Panasiewicz et al., 2015), and also seriously impacts conservation programs for other threatened species (Altizer et al., 2007). Although genetic variability and geneflow among living wisent herds has been investigated using limited sets of genetic markers (e.g. Gralak et al., 2004; Luenser et al., 2005; Radwan et al., 2007; Wójcik et al., 2009; Babik et al., 2012; Tokarska et al., 2009a, 2009b, 2015), to date no published study has investigated the genetic composition of extant wisent at the level of the complete genome. Similarly, while preserved specimens of some founding individuals that may retain genetic information still exist within museum collections, the genetic variability among these founders and their subsequent contribution to living herds remains unquantified.

A second threat for wisent is potential hybridisation with domestic cattle (*Bos taurus*). Although F1 hybrid bulls are sterile, hybrid female offspring are not (Basrur, 1968), and would therefore have the potential to reintegrate back into the wild population. It is known that domestic cattle were grazed in the Białowieża Forest of Poland and Belarus for many years at the same time as wisent were present there, though no births of hybrids in natural

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conditions were recorded (Krasińska & Krasiński 2007). Hybridisation with domestics in general is of wider concern for threatened species (Wolf et al., 2001). Well known examples of this kind of hybridisation include Przewalski's horse and domestic horse (Sarkissian et al., 2015), mouflon and domestic sheep (Lühken et al., 2009), wolf/coyote and dog (Bohling & Waits, 2011), wildcat and domestic cat (Steyer et al., 2016), and game birds and domestic fowl (Arrieta et al., 2013). However, in the case of wisent, the extent of cattle admixture into living herds remains undetermined, as does the admixture that may have occurred prior to the establishment of the captive breeding program. Such information is critical to assess the magnitude of the threat that cattle admixture represents to the long term viability and integrity of the species.

Here we present low-coverage whole genome sequencing data from both modern genetic lines, as well as from four historical samples representing two of the original founding individuals and two individuals of the now extinct Caucasian wisent subspecies. Using these data, we compare the relative contributions of the founding individuals to the genomes of living wisent. We also determine the magnitude and distribution of Caucasian wisent ancestry in modern and founding individuals. Furthermore, we uncover evidence of admixture with the cattle/aurochs lineage that has resulted in a significant component of mixed ancestry in the genomes of both founding and modern wisent. Our results have important implications for both understanding the evolutionary history and for the future conservation management of wisent. Finally, we demonstrate the huge potential of genomic approaches, in particular applied to historical samples, for conservation management of endangered species.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS

Sequencing of wisent genomes

We conducted shotgun sequencing of wisent genomes using Illumina technology and mapped the resulting sequence reads to a reference genome assembly of the Asian water buffalo, *Bubalus bubalis* (GenBank accession no. GCA_000471725.1), which represents an outgroup to the wisent/cattle clade (Fig. S1; Hassanin et al., 2013). In total we sequenced seven individuals that we divide into three categories. Detailed sample information, including provenance, is provided in Table 1 and sample localities are shown in Figure 1.

Modern wisent – three modern individuals representing both genetic lines. These comprise two individuals from the L line (MdL1, mean read depth 1.59x; and MdL2, mean read depth 1.63x; from the Polish and Belarussian parts of the Białowieża Forest, respectively) and one individual from the LC line (MdLC, from Dydiowa in the Bieszczady Mountains, mean read depth 1.49x).

Founding wisent – two individuals assignable to the Lowland wisent subspecies *B. b. bonasus* from the initial breeding population originating from Pszczyna, both of which contributed to the establishment of both the L and the LC genetic lines: foundress F42 PLANTA (1904-1931, mean read depth 1.82x) and her male offspring, M158 PLATEN (1926-1933, mean read depth 1.36x), who was fathered by another founder M45 PLEBEJER (1917-1937).

Caucasian wisent – two individuals from the early 1900's representing the now extinct Caucasian wisent subspecies *B. b. caucasicus* (Cc1, from Kubańskaja Obłast, mean read depth 1.17x; and Cc2, from North Ossetia-Alania, mean read depth 0.92x).

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For each individual, we collapsed mapped reads into a single pseudo-haploid genome sequence by randomly selecting a single high quality nucleotide from the read stack at each position of the reference genome, following the procedure described by Cahill et al. (2013; 2015). This procedure disregards heterozygous positions, where only one allele will be sampled, but should not introduce any biases in allele sampling. Ancient DNA fragments frequently contain miscoding lesions resulting from postmortem DNA degradation, the most common of which involves the deamination of cytosine to uracil, which causes C to T substitutions in the resulting data (e.g. Dabney et al., 2013b). This pattern is present at varying levels in sequence data from our historical samples (Fig. S2-S5), and so we restricted all subsequent analyses to transversion sites only to avoid any confounding effects of DNA damage.

Genomic divergence

We investigated patterns of nuclear genomic divergence among wisent by conducting pairwise comparisons of the number of transversion differences occurring along a sliding window of 1Mb, producing a distribution of genomic divergence for each wisent pair. The resulting probability densities showed that nuclear genomic divergence is broadly similar among all modern and founding wisent (Fig. 2a). The two founding individuals, PLANTA and PLATEN, are somewhat less diverged from one another than either is from all modern wisent (Fig. 2a), reflecting their mother-son relationship. Slightly increased divergence is observed between the modern LC line individual (MdLC) and all modern and founding wisent (Fig. 2a), which may reflect the increased component of Caucasian wisent ancestry in this individual resulting from the captive breeding program.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genomic divergence between Caucasian and both modern and founding wisent greatly exceeds that occurring between the latter two groups (Fig. 2a-c). Substantial divergence is also found between the two Caucasian wisent individuals. One of these Caucasian wisent (Cc1) was found to be less diverged from modern and founding wisent than other Caucasian wisent individual (Cc2), suggesting the presence of not only substantial genetic diversity but also substantial population structure in the extinct Caucasian wisent subspecies.

We also investigated mitochondrial genome variability among all individuals subjected to nuclear genome sequencing, in addition to eight other modern wisent (Tab. S1). Sequence analysis revealed that all investigated modern wisent, both founding wisent, and a single historical Caucasian wisent (Cc1), share a single haplotype. The haplotype occurring in the second historical Caucasian wisent (Cc2) differed from this widely shared haplotype by a single transition site. These results hint at a major loss of mitochondrial haplotype diversity prior to the extinction of wisent in the wild. This inference is supported by additional haplotypes that we recovered from three ancient middle Holocene wisent from Austria and one ancient middle Holocene Caucasian wisent from Armenia (Tab. 1). All these ancient haplotypes are substantially divergent from those found in modern and historical wisent, suggesting a substantial loss of haplotype diversity, potentially within the last ~1500 years.

Neighbour-joining phylogenetic analysis of total nuclear genomic divergence supports paraphyly of Caucasian wisent (Fig. 2d), as does analysis of mitochondrial haplotypes (Fig. S1). We further investigated the population history of wisent by dividing aligned nuclear genome sequences into non-overlapping 1MB blocks and subjecting each block to maximum-likelihood phylogenetic analysis. For this analysis, we included both Caucasian wisent and both modern L line wisent as representatives of the Lowland wisent

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

subspecies, with water buffalo as outgroup. Founding wisent and the modern LC line wisent were not included to avoid any confounding effects of direct ancestor-descendent relationships and documented Caucasian wisent introgression (Slatis, 1960) on phylogenetic interpretation, respectively. We found that 57% of the investigated genomic blocks support reciprocal monophyly of Caucasian and Lowland wisent (Fig. 3). We therefore conclude that this most likely represents the true population history. All alternative topologies occur, individually, at a much lower frequency. Nevertheless, almost half of the genome sequence alignment of these individuals contradicts the true population history.

In order to interpret wisent genomic divergence in the context of total species genetic diversity, we obtained data from the NCBI Short Read Archive for seven domestic cattle and seven yak (*Bos grunniens*; Tab. S4) and subjected them to the same analysis pipeline. Genomic divergence among modern wisent was found to be similar to that found among domestic cattle, and exceeded that found among yak (Fig. 4). We conducted equivalent comparisons for modern wisent and these other bovid species but with the inclusion of transition as well as transversion sites. Interestingly, the distribution of genomic divergence for pairs of wisent was bimodal in all cases (Fig. 4), but the relative levels of genomic divergence between species were similar to that measured using only transversion sites.

Wisent geneflow and admixture

We investigated patterns of admixture among wisent using the *D* statistic test for admixture (Green et al., 2010). This test identifies any imbalance in the number of derived alleles that either of two closely related individuals share with a candidate introgressor. A significant excess of derived alleles shared between one individual and the introgressor provides

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

evidence of admixture between them (Durand et al., 2011). For all *D* statistic tests, we used water buffalo (*Bubalus bubalis*) as outgroup for allele polarisation.

We first investigated patterns of derived allele sharing among modern wisent, and found no statistically significant signal of admixture between the modern LC line individual and either modern L line individual (Tab. S5). Between modern and founding wisent, we found that modern L line wisent share a significantly greater number of derived alleles with founding wisent than the modern LC line individual does (Fig. 5a), indicating a greater contribution of the two founding wisent investigated here to the L line, relative to the LC line, which is consistent with pedigrees (Slatis, 1960).

We then investigated admixture involving Caucasian wisent. We found a significant excess of derived allele sharing between one founding wisent, PLANTA, and one Caucasian wisent, Cc2, relative to the other founding wisent, her son PLATEN (Fig. 5b). This indicates that a proportion of the genome of PLANTA can be attributed to admixture with Caucasian wisent. Furthermore, we can deduce that PLEBEJER, the father of PLATEN, must have possessed a lower level of Caucasian wisent admixture than PLANTA, and that PLATEN himself was likely admixed to some degree through inheritance from PLANTA. The detection of admixture involving one Caucasian wisent (Cc2) but not the other (Cc1) further supports the existence of genetic structure in Caucasian wisent inferred from estimates of genomic divergence (Figs. 2b & 2c).

Next, we investigated evidence of Caucasian wisent admixture among modern wisent. Consistent with expectations, we found that the modern LC line individual (MdLC) shares an excess of derived alleles with one of the Caucasian wisent (Cc1) relative to modern L line

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

individuals (Fig. 5c). We did not, however, detect such an excess between the modern LC line individual and the second Caucasian wisent (Cc2), relative to the modern L line individuals. We can therefore infer that the last surviving Caucasian wisent, KAUKASUS, whose living descendants comprise the modern LC line, was more closely related to Caucasian wisent individual Cc1 that to individual Cc2.

We further investigated Caucasian ancestry in the genome of the modern LC line individual by performing phylogenetic analysis of non-overlapping 1MB genomic blocks. This analysis involved the modern LC line wisent (MdLC), the founding wisent (PLATEN) that was found to be least admixed with Caucasian wisent, Caucasian wisent (Cc1) and domestic cattle, with water buffalo as outgroup. These datasets were mapped to the reference genome of the zebu, *Bos indicus* (Canavez et al., 2012), to benefit from the chromosome-level assembly of that genome. Of the investigated genomic blocks, we find that 22% return the modern LC line and Caucasian wisent as monophyletic (Fig. 6), and may therefore represent introgressed segments of Caucasian wisent ancestry in this modern LC line individual. Around 8% of these blocks are likely to result from incomplete lineage sorting, based on the frequency of occurrence of the opposing topology (Fig. 6), producing an overall estimate of 14% of the genome of the modern LC individual that results from Caucasian wisent admixture, most likely inherited from the bull KAUKASUS. In addition to providing an estimate of admixture proportions, our method is also able to accurately map admixed segments of the genome (Fig. 6a). Many of these segments span multiple megabase blocks. For example, a contiguous 22MB admixed block is observed on chromosome 4, which may span as much as 33MB under the assumption that intervening blocks with missing data are linked to adjacent ones. Relative admixture proportions also vary among chromosomes in this individual. For

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

example, chromosome 27 almost entirely lacks Caucasian wisent ancestry whereas around 50% of chromosomes 4 and 15 are likely derived from admixture.

Finally, we investigated evidence of Caucasian wisent ancestry in the modern L line. We found that both modern L line wisent share a significant excess of derived alleles with Caucasian wisent (Fig. 5b), relative to founding wisent. Thus, modern L line individuals appear more admixed with Caucasian wisent than the two founding wisent investigated here. This admixture signal could result from either variable admixture proportions among founding individuals, or recent geneflow between the L and LC lines, although *D* statistic comparison of modern individuals failed to detect the latter (see above). We further investigated these alternative hypotheses by comparing the sizes of putatively admixed genomic blocks. Recent geneflow results in large contiguous genomic blocks derived from admixture in the genomes of the recipient population, which are broken up over time as a result of recombination. Pedigree information provides an approximate date for geneflow from Caucasian wisent into the modern LC line around 90 years ago (15-22 generations), the result of which are many intact multi-megabase genomic blocks derived from Caucasian wisent in the modern LC line individual (Fig. 6). We compared the sizes of these blocks with those of putative Caucasian wisent ancestry in a modern L line individual, and found the abundance of large blocks to be considerably lower in the latter (Fig. 7). This rejects recent admixture and instead supports variable admixture proportions among the founding herd in explaining the observed signal of Caucasian wisent admixture in this modern L line individual.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Admixture with the cattle/aurochs lineage

We investigated potential admixture between wisent and the cattle/aurochs lineage using pseudo-haploid sequences generated from short read data of two domestic Holstein cows and an ancient aurochs (*Bos primigenius*; Park et al., 2015), the extinct species from which cattle were domesticated and that lived sympatrically with wisent up until its extinction around 400 years ago (van Vuure, 2005). First, we looked for significant differences in derived allele sharing between cattle and wisent, relative to the aurochs. We found that all investigated wisent share a significant excess of derived alleles with cattle relative to aurochs (Fig. 5d). This suggests either admixture between wisent and domestic cattle, or alternatively, admixture with aurochs, if aurochs populations were highly structured and the admixing individuals were from a population more closely related to domestic cattle than the British aurochs used in this analysis.

We also compared derived allele sharing with cattle among the individual wisent. We found that all modern wisent investigated here share a significant excess of derived alleles with cattle relative to any founding or Caucasian wisent (Fig. 5e). Variable admixture was also observed among founding wisent. Specifically, PLANTA shares more derived alleles with domestic cattle than either PLATEN or the Caucasian wisent do. Finally, we estimated the genomic fraction in the modern wisent that could be attributed to cattle/aurochs admixture, above that occurring in the founding wisent, using the $\hat{f}$ statistic (Durand et al., 2011). This analysis showed that 2.4% to 3.2% of the sampled modern wisent genomes can be attributed to admixture with the cattle/aurochs lineage, respectively, that is not the result of direct inheritance from the sampled founding wisent (Tab. S6).

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The increased admixture signal observed in all modern wisent relative to founding wisent could result from either variable cattle/aurochs admixture proportions among the founding herd from which all modern wisent are derived, or alternatively, from recent admixture with modern cattle. However, the fact that we do not find evidence of any complete genomic 1MB blocks resulting from cattle admixture in the modern LC line individual (Fig. 6a) argues strongly against recent cattle admixture, and instead supports variable cattle/aurochs admixture among the founding herd in explaining the excess of derived alleles shared among domestic cattle and modern wisent.

DISCUSSION

Retracing complex population histories can be challenging. In particular, admixture involving populations or species that are now extinct may be impossible based solely on data from living individuals (Hofreiter et al., 2015). Through the use of low-coverage genomic data from modern and historical wisent, including from the now extinct Caucasian wisent subspecies, we have revealed the complexity of wisent evolution. This complex history involved not only admixture resulting directly from the captive breeding program, but also older processes occurring prior to their extinction in the wild, which included admixture with another bovid lineage (Fig. 8).

Wisent evolution and admixture

The accepted view of wisent evolution is of two distinct subspecies, Lowland wisent and Caucasian wisent, that both underwent dramatic population declines, with the last few surviving individuals serving as founders of the modern L (Lowland only) and LC (Lowland and Caucasian) lines (Pucek et al., 2004). Our results show that this model is an oversimplification. We find evidence of at least two highly differentiated populations within

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Caucasian wisent, with one of these showing greater pairwise similarity with Lowland wisent than with the second Caucasian population, at the level of the complete genome (Fig. 2). However, analysis of aligned nuclear genomic blocks from four individuals returns Caucasian and Lowland wisent as reciprocally monophyletic across slightly more than half of the genomic alignment (Fig. 3), providing support that this topology reflects the true history of population divergence. Thus, among any two Caucasian wisent and any two modern (or founding Lowland) wisent, we may expect that any single locus has only around 50% probability of reflecting the true history of population divergence. Moreover, increased sampling of individuals is likely to further reduce this proportion, potentially to such an extent that, at a given sampling level, the true evolutionary history cannot be untangled from the effects of random drift and more recent admixture. This result reinforces the notion that phylogeny-based interpretation may be inappropriate at the level of the complete genome, and that alternative, more flexible models will be required to keep pace with our ability to generate such data (Hofreiter et al., 2015).

A further implied assumption of the traditional view of wisent evolution is that, with the exception of the Caucasian bull KAUKASUS, all founding wisent represented “pure” Lowland wisent (Pucek et al., 2004; Tokarska et al., 2015). On this basis, the modern L line that is derived only from the latter can also be considered as pure Lowland wisent, referable to the subspecies *B. b. bonasus*. Our results also demonstrate that the notion of wisent subspecies purity is flawed in the sense that founding Lowland individuals were in fact admixed with Caucasian wisent to varying degrees. We demonstrate this both directly for the founder PLANTA in comparison to another founder, her son PLATEN, and also indirectly, by the elevated signal of Caucasian wisent admixture in modern L line wisent, relative to these founders, most likely a result of inheritance from other founding individuals not included in

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

this analysis (Fig. 7). The notion of subspecies purity therefore disregards the fact that admixture between Caucasian and Lowland wisent almost certainly occurred prior to the extinction of wisent in the wild, and such admixture could therefore be regarded as part of the normal population processes and dynamics of this species.

The notion of subspecies purity has driven efforts to ensure that free-living L and LC herds do not come into contact (Pucek et al., 2004), and also motivated genetic investigations of living populations that may have been recipients of geneflow from the opposing genetic line (Tokarska et al., 2015). The latter study investigated the modern L line population that currently inhabits the eastern, Belarussian part of the Białowieża Forest. Some individuals were found to possess a microsatellite allele that was common among Caucasian wisent but absent in all studied Lowland wisent, which these authors interpreted as evidence of recent admixture with the modern LC line. Although the individual from this population (modern L line, MdL2) that we sequenced does not possess this putative Caucasian wisent allele, we nevertheless detected evidence of Caucasian wisent admixture above that occurring in founding wisent for this individual (Fig. 5). However, the small size of admixed blocks, in addition to non-significant *D* comparisons of modern lines, supports variable Caucasian wisent admixture among founding wisent in explaining this result, which may also account for the occurrence of putative Caucasian wisent alleles in other individuals from this population. Future studies of such individuals using the methodology applied here would provide a robust test of these alternative hypotheses.

Wisent conservation and de-extinction

The issue of low genetic variability among living wisent is considered as cause for concern, and has been demonstrated by several population-level studies using various molecular and

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

biochemical markers, such as blood-group systems and blood serum proteins (Sipko et al., 1995; Gębczyński & Tomaszewska-Guszkiewicz 1987; Hartl & Pucek, 1994; Sipko et al., 1997), mtDNA (this study; Tiedemann et al., 1998; Burzyńska et al., 1999; Wójcik et al., 2009; Hassanin et al., 2012), nuclear gene sequences (Sipko et al., 1994; Udina et al., 1994; Kamiński & Zabolewicz 1997; Udina & Shakhaev 1998; Burzyńska & Topczewski, 1999; Radwan et al., 2007; Babik et al., 2012; Hassanin et al., 2013), microsatellites (Gralak et al., 2004; Luenser et al., 2005; Tokarska et al., 2009a; 2009b) and SNPs (Tokarska et al., 2009b, 2015). In apparent contrast to these results, we find relatively high levels of genomic divergence among both modern and founding wisent, being approximately equal to or exceeding that found between pairs of cattle or yak, respectively. Pairwise divergence between modern and Caucasian wisent was found to be even higher, indicating a major loss of genetic diversity from the species as a whole following the extinction of the latter population.

The discrepancy between estimates of genetic variation obtained from population-level studies and from pairwise nuclear genomic divergences, as reported here, may result from several factors. First, application of microsatellites or SNP markers developed for cattle to wisent will likely lead to ascertainment bias (Albrechtsen et al., 2010), particularly if diversity estimates for wisent are then compared back to those obtained for cattle using the same marker set. Admixture with the cattle/aurochs lineage detected in this study adds a further layer of complexity in interpreting data from molecular markers developed for cattle. Second, although genetic diversity and pairwise genetic divergence are both measures of genetic variation in the broad sense, these measures are different and not necessarily interchangeable. Thus, wisent may lack allelic diversity, but the alleles that do exist may be highly divergent from one another. Indeed, the bimodal distributions observed for pairwise comparisons based on transitions and transversions are compatible with high levels of relatedness -- likely

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

inbreeding -- among modern wisent, and low mitochondrial haplotype diversity among modern and historical individuals in comparison to that found among middle Holocene wisent (Fig. S1) provides further evidence of a loss of allelic variation. Under this scenario of low allelic diversity but high nuclear allelic divergence, given continued population expansion and sufficient time, nuclear genomic diversity may potentially be restored as recombination dissects and shuffles divergent chromosomes over successive generations.

The Caucasian wisent is extinct in the wild, but a fraction of its genepool survives in the genomes of modern wisent. Our results provide not only a direct measure of this admixture in a modern LC line individual, but also allow us to map with relative accuracy chromosomal segments that are inherited from Caucasian wisent. Although sometimes controversial, the concept of de-extinction has generated considerable interest (Sherkow & Greely, 2013), and attempts are currently underway to generate animals that, at least superficially, resemble the quagga (*Equus quagga*; Heywood, 2013), an extinct subspecies of plains zebra, and also the aurochs (van Vuure, 2005), by careful selective breeding of their living relatives. In both of these cases, selective breeding and the ultimate success of the project are based solely on morphological criteria. Our study demonstrates that, at least in principle, by generating chromosomal admixture maps for multiple living representatives of the LC line, it would be possible to selectively breed an animal that is, at the genomic level, highly similar to a Caucasian wisent.

Admixture with the cattle/aurochs lineage

Hybridisation of wild species with their domesticated close relatives is a subject of considerable discussion and concern for conservation management (Ellstrand et al., 2010). Frequently, such events are deemed to be detrimental to the recipient wild species, as is the

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

case for American bison (*Bison bison*) herds found to be admixed with cattle (Halbert & Derr, 2007). However, the introgression of alleles from domestics into wild populations may also provide the basis for adaptation (Fuelner et al., 2013; Anderson et al., 2009), and allow populations to take advantage of new ecological niches (Monzón et al., 2014). In any case, the identification of admixture is the essential first step to guide conservation policy, and measuring admixture proportions among individuals is likely the second essential step for implementing conservation measures. Our results achieve both of these objectives.

We detect a significant proportion of admixture with cattle relative to aurochs in the genomes of all wisent except founder PLATEN. However, since D statistic is a relative test, and PLATEN is the offspring of admixed founder PLANTA, it is reasonable to infer the PLATEN is also admixed to some extent. Although our comparisons consistently found an excess of derived alleles shared between cattle and wisent, relative to aurochs and wisent, it is difficult to completely exclude admixture with aurochs as an alternative explanation, given the lack of knowledge of population structure in aurochs. Testing these two alternatives would require data from additional aurochs populations from within the core distribution of wisent, and would be a valuable direction for future research.

The timing of admixture also has implications for conservation management. Specifically, the removal of individuals resulting from very recent hybridisation may be deemed appropriate (Halbert & Derr, 2007). The small size of cattle admixed blocks in modern wisent (at least undetectable at a 1MB scale) clearly rejects recent cattle admixture for the individuals investigated here. Instead, admixture must have occurred prior to the establishment of the captive breeding program, and the admixture signal detected in modern wisent results from inheritance from the founders that were admixed with cattle to varying degrees. Thus, based

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

on the current evidence, cattle introgression appears of low concern for wisent conservation for the following reasons: 1. admixture does not appear to have occurred since the establishment of the captive breeding program, although screening of additional individuals may be desirable to further support this generalisation; 2. the number of intervening generations separating living wisent from the F1 hybrids is likely sufficient that all living wisent are admixed to some extent (Chang, 1999); and 3. our results may in fact reflect admixture with aurochs, rather than domestic cattle, although this hypothesis requires further investigation.

Conclusions

The ability to detect admixture is of key importance for both evolutionary and applied conservation studies. However, interpretation of a significant signal of admixture, in terms of both evolutionary inference and the formulation of management strategies, may require information its timing. Using seven low-coverage wisent genomes from both modern and historical wisent we have revealed multiple instances of admixture, but moreover, because the approximate age of introgression of Caucasian wisent into the modern LC line is known, through comparisons of the sizes of likely admixed genomic segments we have inferred the relative ages of other admixture events. This unique historical information, coupled with the ability to recover genomic data from historical samples, establish wisent as an exemplary taxon for the study of admixture in wild populations. As new analytical methods for studying admixture are developed, wisent can serve as a valuable empirical test of both their performance and utility.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

MATERIALS AND METHODS

Complete details of all samples and specimens used in this study are shown in Table 1.

Laboratory methods, modern samples

DNA was extracted from tissue samples of three modern wisent using either a DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions (sample MdL1) or by phenol/chloroform extraction (Sambrook & Russell, 2001). We mechanically sheared the DNA of the modern samples using a Covaris S220 sonicator to an average fragment length of 500 bp and prepared indexed Illumina libraries from 500 ng of each modern DNA extract using a published double-stranded protocol (Meyer & Kircher, 2010) with modifications (Fortes & Paijmans, 2015). Library molecules from 450 bp to 1000 bp were then selected using a Pippin Prep Instrument (Sage Science).

Laboratory methods, historical samples

DNA extraction from four museum specimens as well as sequencing library preparation steps preceding amplification were performed in a dedicated ancient DNA laboratory (Evolutionary Adaptive Genomics Group, Potsdam University, Germany). DNA extracts were prepared from horn and bone powder obtained by grinding in a mixer mill (MM 400, RETSCH). DNA extraction followed the protocol of Dabney et al. (2013a), except for horn samples where we used a different digestion buffer containing 10mM Tris buffer (pH 8.0), 10 mM NaCl, 5 mM $CaCl_2$, 2.5 mM EDTA (pH 8.0), 2% SDS (Shapiro & Hofreiter, 2012). The museum samples were already fragmented due to degradation, so were not sonicated. We used 25 µl of each DNA extract to construct single-stranded indexed Illumina libraries according to the protocol of Gansauge & Meyer (2013).

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sequencing

Final library concentrations and the distribution of insert sizes were determined using a 2200 TapeStation (Agilent Technologies) and Qubit HS-assay (Thermo Fisher Scientific), respectively. Each library was then sequenced using an Illumina NextSeq 500 instrument. For modern libraries we used a High Output Kit (75 bp paired-end sequencing), for libraries obtained from historical horn samples we used High Output Kits (75 bp single-end and 150 bp paired-end) and each library built from historical bone samples was sequenced separately with High Output Kits (75 bp single-end and paired-end). Full details of sequencing results are provided in Supporting Tables S2 and S3.

Data processing, mapping and pseudo-haploidisation

For paired-end data, we trimmed adapter sequences and merged overlapping read pairs using SeqPrep (https://github.com/jstjohn/SeqPrep), requiring a minimum read length of 30 bp (-L 30), minimum overlap of 15 bp (-o 15), and a minimum merge quality of 13 (-q 13). Adapters occurring at the 3' ends of single-end reads were trimmed using cutadapt (Martin, 2011), also requiring a minimum length of reads of 30 (-m 30). We then mapped the resulting data to the zebu (*Bos indicus*; GenBank accession no. GCA_000247795.2) and water buffalo (*Bubalus bubalis*; GenBank accession no. GCA_000471725.1) nuclear genomes and wisent mitochondrial genome (KW, unpublished) using BWA aln version 0.7.8 (Li & Durbin, 2009) with default 0.04 mismatch value. We removed duplicate reads likely resulting from PCR amplification using samtools rmdup (Li et al., 2009). Detailed descriptions of the mapping results are provided in Supporting Tables S2 and S3. We then generated pseudo-haploid sequences as described by Cahill et al. (2015) and used these for further analysis.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Pairwise genomic divergence

Pairwise genomic divergence was calculated by dividing genomic alignments into non-overlapping 1MB blocks and calculating the proportion of transversions, or transitions plus transversions (comparisons of modern individuals only), for each pair of individuals, accounting for the presence of missing data. Blocks with > 75% missing data were disregarded. Probability densities were generated by kernel density estimation in R (R Core Team, 2014) using default parameters. Full details of comparative data generated for domestic cattle and yak (data from the NCBI Short Read Archive) are provided in Supporting Table S4.

Mitochondrial genome analysis

In addition to the mitochondrial genomes generated from the seven modern and four historical specimens, we generated four mitogenomes from ancient middle Holocene wisent individuals (Tab. 1). These ancient samples were radiocarbon dated at either the Oxford University Radiocarbon Accelerator Unit (Oxford, UK) using ultrafiltered collagen and accelerator mass spectrometry (Ramsey et al. 2004a, 2004b) or the VERA-Laboratorium Institut für Isotopenforschung und Kernphysik (Vienna, Austria). We calibrated radiocarbon dates using the IntCal13 calibration curve (Reimer et al., 2013) in OxCal v4.2 (https://c14.arch.ox.ac.uk/oxcal/OxCal.html).

DNA was extracted from the Holocene *B. bonasus* samples (Bb1-Bb3) at the Henry Wellcome Ancient Biomolecules Centre (Oxford University, UK), following Shapiro et al. (2004). We extracted the Holocene *B. b. caucasicus* sample (Cc3) in the specialist Paleogenomics facility at UC Santa Cruz, following Rohland et al. (2010). DNA library construction, mitochondrial target enrichment, sequencing, and sequence data processing protocols for the four Holocene samples followed approach four in Heintzman et al. (2016),

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

except that the whole mitochondrial genome consensus sequence was retained. The mean read depth of these Holocene consensus sequences ranged from 14.1 to 165.7x. The consensus sequences for the four Holocene and one historic Caucasian bison have been deposited in GenBank with accession numbers XXXXXXXXX-XXXXXXXX. The mitogenomic sequences from the remaining modern and historic samples were identical to either haplotypes previously published or the historic Caucasian bison haplotype (Tab. S1).

We assessed phylogenetic relationships among wisent mitochondrial haplotypes, as well as their placement within the wider bovin (tribe Bovini) tree. Wisent sequences were aligned with those of 12 other bovin taxa, including the extinct steppe bison (*Bison priscus*) and aurochs (*Bos primigenius*) (Tab. S1). We excluded two previously published wisent mitochondrial genomes from the phylogenetic analysis (GenBank accessions: HQ223450, HM045017/NC_014044), as these sequences were considered problematic. Specifically, HQ223450 has multiple insertions totaling 9 bp in the ND4 coding region, and HM045017/NC_014044 has multiple indels and point mutations concentrated in the large rRNA and ND3 coding regions. Sequence alignment, partitioning, model testing, and phylogenetic and associated statistical support methods followed the ordinal-level analyses of Heintzman et al. (2015), except that we used the *B. bison* reference mitochondrial genome (NC_012346) for partitioning. We selected the following models of molecular evolution for the six partitions: GTR+I+G (CP1, 3803 bp; rRNAs, 2541 bp), GTR+G (CP3, 3803 bp), HKY+I+G (CP2, 3803 bp; tRNAs, 1526 bp; control region, 927 bp). We used the saola (Pseudoryx nghetinhensis; Tab. S1) as outgroup in both the maximum likelihood and Bayesian analyses, following Bibi (2013).

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

D statistic tests

The *D* statistic involves 4 genomes: a genome from each of two sister populations (P1 and P2), a genome from a third population as the potential source of introgression (P3), and an outgroup genome (O) to identify the ancestral state (identified as the A allele). We identified variable positions at which P3 possessed the derived allele (B) and presence of the derived allele is variable among P1 and P2, leading to two possible patterns: either ABBA or BABA. Under the scenario of incomplete lineage sorting without geneflow these patterns should occur with equal frequency and the expected *D* value will be zero. An excess of ABBA or BABA patterns is interpreted as evidence of admixture. However, it might also arise from nonrandom mating in the ancestral population due to population structure (Eriksson & Manica, 2012). To determine the ancestral state we used the water buffalo genome. In all tests involving data mapped to the zebu genome, we took into consideration the autosomes only. We performed a total of 105 comparisons considering all possible combinations of wisent, all wisent with either domestic cattle or aurochs as candidate admixer, and domestic cattle and aurochs with all wisent as candidate admixer. These results are reported in Supporting Table S4. $\hat{f}$ test (Green et al., 2010; Durand et al., 2011) was used to estimate the proportion of the genome derived from admixture. This test requires two individuals of the candidate introgressor species that are not themselves admixed. For our datasets this was possible only for admixture involving the cattle/aurochs lineage. For both *D* statistic and $\hat{f}$ test, significance was assessed using a weighted block jackknife using 1Mb blocks (Green et al., 2010; Durand et al., 2011). The weighted block jackknife tests if admixture signals are uniform across the whole genome and therefore reflect the same population history. By removing one at a time blocks of adjacent sites (larger than the extent of linkage disequilibrium) and computing the variance of the *D* statistic or $\hat{f}$ values over the entire genome *M* times leaving each block of the genome in turn, and then multiplying by *M* and taking the square root we generated the

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

standard error. The number of standard errors by which D or $\hat{f}$ differs from zero is the Z score. The results with Z-scores greater than 3 in absolute value were qualified as statistically significant (Green et al., 2010).

Nuclear genome phylogenetic tests

The aligned pseudohaploid sequences were divided into non-overlapping blocks of 1MB. If each of the five taxa contained no more than 50% gaps within a window, the sequence data were recoded into binary characters to only score transversions (Rs: 0, Ys: 1), otherwise the window was recorded as having insufficient data. A Maximum Likelihood phylogeny under the BINGAMMA model and with the water buffalo as outgroup was then computed for each alignment with sufficient data using RaxML (Stamatakis, 2014). The topology of each phylogeny was evaluated using a custom Perl script that made use of the ETE3 software (Huerta-Cepas et al., 2016). The lengths of admixed genomic regions was estimated by counting the number of consecutive 1MB blocks returning the respective tree topology. Due to the presence of blocks with insufficient data, these measurements are likely to be underestimates. Evaluation of the lengths of genomic regions was conducted using the empirical cumulative distribution function in R, with default parameters.

ACKNOWLEDGEMENTS

We would like to thank Andrzej Imiołczyk and his colleagues from the Specimen Preparation Laboratory of the Upper Silesian Museum in Bytom. We thank Andrew Fields and Joshua Kapp for technical assistance. The study has been funded by National Science Center (NCN) grant 2011/03/N/NZ8/00086 to KW.

AUTHOR CONTRIBUTIONS

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Conceived the study - KW, JMS, AB, MH

Designed laboratory experiments - AB, KW, JLAP, PDH

Performed lab work - KW, UT, GX, PDH, BS

Coordinated data analysis - KW, SH, AB, JAC, JLAP, PDH

Performed data analysis - KW, SH, AB, PDH

Coordinated writing of the manuscript - KW, AB, MH

Obtained funding - KW, JMS, MH

Provided samples - GB, ANB, JJC, RD, NM, HO, MT, STT, JMW, WŻ

All authors read, gave comments and helped revise the final version of the manuscript.

COMPETING INTERESTS

The authors declare that they have no competing financial interests.

REFERENCES

Adaszek L, Dzięgiel B, Krzysiak M, Skrzypczak M, Adaszek M, Staniec M, Wyłupek D, Winiarczyk S. 2014. Detection of *Borrelia burgdorferi* sensu lato DNA in the blood of wild bison from Białowieza primeval forest in eastern Poland. *Pol J Vet Sci* **17**:713-715.

Albrechtsen A, Nielsen FC, Nielsen R. 2010. Ascertainment biases in SNP chips affect measures of population divergence. *Mol Biol Evol* **27**:2534-2547. doi: 10.1093/molbev/msq148.

Altizer S, Nunn CL, Lindenfors P. 2007. Do threatened hosts have fewer parasites? A comparative study in primates. J Anim Ecol **76**:304-314. doi: 10.1111/j.1365-2656.2007.01214.x

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Anderson LA, Christianson GB, Linden JF. 2009. Stimulus-specific adaptation occurs in the auditory thalamus. *J. Neurosci* **29**:7359–7363.doi: 10.1523/JNEUROSCI.0793-09.2009.

Arrieta RS, Lijtmaer DA, Tubaro PL. 2013. Evolution of postzygotic reproductive isolation in galliform birds: analysis of first and second hybrid generations and backcrosses. *Biol J Linnean Soc* **110**:528–542. doi: 10.1111/bij.12153.

Babik W, Kawalko A, Wójcik JM, Radwan J. 2012. Low major histocompatibility complex class I (MHC I) variation in the European bison (*Bison bonasus*). *J Hered* **103**:349-359. doi: 10.1093/jhered/ess005.

Basrur PK. 1968. Hybrid Sterility. In: Benirschke K, editor. *Comparative Mammalian Cytogenetics*. Springer. 107-131.

Benecke, N. 2005. The Holocene distribution of European bison – the archaeozoological record. *Munibe* **57**:421-28.

Bibi F. 2013. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evol Biol* **13**:166. doi: 10.1186/1471-2148-13-166.

Bocherens H, Hofman-Kamińska E, Drucker DG, Schmölcke U, Kowalczyk R. 2015. European bison as a refugee species? Evidence from isotopic data on Early Holocene bison and other large herbivores in northern Europe. *PLoS One* **10**:e0115090. doi: 10.1371/journal.pone.0115090.

Bohling JH, Waits LP. 2011. Assessing the prevalence of hybridization between sympatric Canis species surrounding the red wolf (*Canis rufus*) recovery area in North Carolina. *Mol Ecol* **20**:2142-2156. doi: 10.1111/j.1365-294X.2011.05084.x.

Burzyńska B, Olech W, Topczewski J. 1999. Phylogeny and genetic variation of the European bison Bison bonasus based on mitochondrial DNA D-loop sequences. *Acta Theriol* **44**:253-262.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Burzyńska B, Topczewski J. 1995. Genotyping of *Bison bonasus* kappa-casein gene following DNA sequence amplification. *Anim Genet* **26**:335-336.

Cahill JA, Green RE, Fulton TL, Stiller M, Jay F, Ovsyanikov N, Salamzade R, St John J, Stirling I, Slatkin M, Shapiro B. 2013. Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. *PLoS Genet* **9**:e1003345. doi: 10.1371/journal.pgen.1003345.

Cahill JA, Stirling I, Kistler L, Salamzade R, Ersmark E, Fulton TL, Stiller M, Green RE, Shapiro B. 2015. Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol Ecol* **24**:1205–1217. doi: 10.1111/mec.13038.

Canavez FC, Luche DD, Stothard P, Leite KR, Sousa-Canavez JM, Plastow G, Meidanis J, Souza MA, Feijao P, Moore SS, Camara-Lopes LH. 2012. Genome sequence and assembly of *Bos indicus*. *J Hered* **103**:342-8. doi: 10.1093/jhered/esr153.

Chang, JT. 1999. Recent common ancestors of all present-day individuals. *Adv. in Appl. Probab* 1002-1026.

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, Garcia N, Paabo S, Arsuaga JL, Meyer M. 2013a. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A* **110**:15758–15763. doi: 10.1073/pnas.1314445110.

Dabney J, Meyer M, Pääbo S. 2013b. Ancient DNA damage. *Cold Spring Harb Perspect Biol* **5**: a012567. doi: 10.1101/cshperspect.a012567.

Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, Jónsson H, Bar-Gal GK, Albrechtsen A, Vieira FG, Petersen B, Ginolhac A, Seguin-Orlando A, Magnussen K, Fages A, Gamba C, Lorente-Galdos B, Polani S, Steiner C, Neuditschko M, Jagannathan V, Feh C, Greenblatt CL, Ludwig A, Abramson NI, Zimmermann W, Schafberg R, Tikhonov A, Sicheritz-Ponten T, Willerslev E, Marques-Bonet T, Ryder OA,

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

McCue M, Rieder S, Leeb T, Slatkin M, Orlando L. 2015. Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Curr Biol* **25**:2577-2583. doi: 10.1016/j.cub.2015.08.032.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol* **28**:2239–2252. doi: 10.1093/molbev/msr048.

Ellstrand NC, Heredia SM, Leak-Garcia JA, Heraty JM, Burger JC, Yao L, Nohzadeh-Malakshah S, Ridley CE. 2010. Crops gone wild: evolution of weeds and invasives from domesticated ancestors. *Evol Appl* **3**:494-504. doi: 10.1111/j.1752-4571.2010.00140.x.

Eriksson A, Manica A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc Natl Acad Sci U S A* **109**:13956–13960. doi: 10.1073/pnas.1200567109.

Feulner PG, Chain FJ, Panchal M, Eizaguirre C, Kalbe M, Lenz TL, Mundry M, Samonte IE, Stoll M, Milinski M, Reusch TB, Bornberg-Bauer E. 2013. Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Mol Ecol* **22**:635-649. doi: 10.1111/j.1365-294X.2012.05680.x.

Fortes GG, Paijmans JLA. 2015. Analysis of whole mitogenomes from ancient samples. In Whole Genome Amplification. Editor T. Kroneis, Humana Press, USA. Available on ArXiv.

Gansauge MT, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc* **8**:737–748. doi: 10.1038/nprot.2013.038.

Gębczyński M, Tomaszewska-Guszkiewicz K. 1987. Genetic variability in the European bison. *Biochem Syst Ecol* **15**:285-288.

Gralak B, Krasińska M, Niemczewski C, Krasiński ZA, Żurkowski M. 2004. Polymorphism of bovine microsatellite DNA sequences in the lowland European bison. *Acta Theriol* **49**:449-456.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. 2010. A draft sequence of the Neandertal genome. *Science* **328**:710–722. doi: 10.1126/science.1188021.

Halbert ND, Derr JN. 2007. A comprehensive evaluation of cattle introgression into US federal bison herds. *J Hered* **98**:1-12.

Hartl GB, Pucek Z. 1994. Genetic depletion in the European bison *Bison bonasus* and the significance of electrophoretic heterozygosity for conservation. *Conserv Biol* **8:**167-174.

Hassanin A, An J, Ropiquet A, Nguyen TT, Couloux A. 2013. Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of Laurasiatherian mammals: Application to the tribe Bovini (Cetartiodactyla, Bovidae). *Mol Phylogenet Evol* **66**:766–75. doi: 10.1016/j.ympev.2012.11.003.

Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, Ruiz-Garcia M, Catzeflis F, Areskoug V, Nguyen TT, Couloux A. 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C R Biol* **335**:32-50. doi: 10.1016/j.crvi.2011.11.002.

Hassanin A, Ropiquet A. 2007. Resolving a zoological mystery: the kouprey is a real species. *Proc Biol Sci* **274**:2849–55. doi: 10.1098/rspb.2007.0830

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Heintzman PD, Froese D, Ives JW, Soares AE, Zazula GD, Letts B, Andrews TD, Driver JC, Hall E, Hare PG, Jass CN, MacKay G, Southon JR, Stiller M, Woywitka R, Suchard MA, Shapiro B. 2016. Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. *PNAS USA*. doi: 10.1073/pnas.1601077113.

Heintzman PD, Zazula GD, Cahill JA, Reyes AV, MacPhee RD, Shapiro B. 2015. Genomic data from extinct North American Camelops revise camel evolutionary history. *Mol Biol Evol* **32**:2433–2440. doi: 10.1093/molbev/msv128.

Heywood P. 2013. The quagga and science: what does the future hold for this extinct zebra? *Perspect Biol Med* **56**:53-64. doi: 10.1353/pbm.2013.0008.

Hofreiter M, Paijmans JLA, Goodchild H, Speller CF, Barlow A, Fortes GG, Thomas JA, Ludwig A, Collins MJ. 2015. The future of ancient DNA: Technical advances and conceptual shifts. *Bioessays* **37**:284–293. doi: 10.1002/bies.201400160.

Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, analysis and visualization of phylogenomic data. *Mol Biol Evol* **33**:1635-1638. doi: 10.1093/molbev/msw046.

Kamiński S, Zabolewicz T. 1997. DDE I restriction fragment length polymorphism in the *Bison bonasus* utranslated region of kappa-casein gene. *Acta Academiae AC Technicae Olstenensis, Zootechnica* **47**:174-153.

Karbowiak G, Demiaszkiewicz AW, Pyziel AM, Wita I, Moskwa B, Werszko J, Bień J, Goździk K, Lachowicz J, Cabaj W. 2014a. The parasitic fauna of the European bison (*Bison bonasus*) (Linnaeus, 1758) and their impact on the conservation. Part 1. The summarising list of parasites noted. *Acta Parasitol* **59**:363-371. doi:10.2478/s11686-014-0252-0.

Karbowiak G, Demiaszkiewicz AW, Pyziel AM, Wita I, Moskwa B, Werszko J, Bień J, Goździk K, Lachowicz J, Cabaj W. 2014b. The parasitic fauna of the European bison

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(*Bison bonasus*) (Linnaeus, 1758) and their impact on the conservation. Part 2. The structure and changes over time. *Acta Parasitol* **59**:372-9. doi:10.2478/s11686-014-0253-z.

Kerley GIH, Kowalczyk R, Cromsigt JPGM. 2012. Conservation implications of the refugee species concept and the European bison: king of the forest or refugee in a marginal habitat? *Ecography* **35**:519–529. doi: 10.1111/j.1600-0587.2011.07146.x

Krasińska M, Krasiński ZA. 2007. European bison. The Nature Monograph. Mammal Research Institute Polish Academy of Science, Białowieża.

Krzysiak MK, Dudek K, Krajewska M, Bednarek D, Szulowski K. 2014. Serological studies to determine the occurrence of Johne's disease and mycoplasma infection in the Northern-East Polish population of European bison (*Bison bonasus*). *Pol J Vet Sci* **17**:721-723. doi: 10.2478/pjvs-2014-0105.

Kuemmerle T, Hickler T, Olofsson J, Schurgers G, Radeloff VC. 2011. Reconstructing range dynamics and range fragmentation of European bison for the last 8000 years. *Diversity Distrib* **18**:47–59. doi: 10.1111/j.1472-4642.2011.00849.x.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. doi: 10.1093/bioinformatics/btp324.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi: 10.1093/bioinformatics/btp352.

Luenser K, Fickel J, Lehnen A, Speck S, Ludwig A. 2005. Low level of genetic variability in European bisons *Bison bonasus* from the Bialowieza National Park in Poland. *Eur. J Wildl Res* **51**:84-87. doi: 10.1007/s10344-005-0081-4.

Lühken G, Wagner HW, Seichter D, Hecht W, Erhardt G. 2009. Genetic Characterization of a Sheep-Dwarf Goat Hybrid. *Cytogenet Genome Res* **125**:158–161. doi:10.1159/000227841.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Majewska AC, Werner A, Cabaj W, Moskwa B. 2014. The first report of Toxoplasma gondii antibodies in free-living European bison (*Bison bonasus bonasus* Linnaeus). *Folia Parasitol* **61**:18-20.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**:10-12. doi: 10.14806/ej.17.1.200.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc **6**:pdb.prot5448. doi: 10.1101/pdb.prot5448.

Monzón J. 2014. First regional evaluation of nuclear genetic diversity and population structure in northeastern coyotes (*Canis latrans*). *F1000Res* **3**:66. doi: 10.12688/f1000research.3567.1.

Moskwa B, Bień J,Goździk K, Cabaj W. 2014. The usefulness of DNA derived from third stage larvae in the detection of *Ashworthius sidemi* infection in European bison, by a simple polymerase chain reaction. *Parasit Vectors* **7**:215. doi:10.1186/1756-3305-7-215.

Oleński K, Tokarska M, Hering DM, Puckowska P, Ruść A, Pertoldi C, Kamiński S. 2015. Genome-wide association study for posthitis in the free-living population of European bison (*Bison bonasus*). *Biol Direct*: **10**:2. doi: 10.1186/s13062-014-0033-6.

Panasiewicz G, Zamojska A, Bieniek M, Gizejewski Z, Szafranska B. 2015. Persistent Müllerian duct syndrome (PMDS) in the Polish free-ranged bull populations of the European bison (*Bison bonasus* L.). 2014. *Anim Reprod Sci* **152**:123-136. doi:10.1016/j.anireprosci.2014.11.012.

Park SD, Magee DA, McGettigan PA, Teasdale MD, Edwards CJ, Lohan AJ, Murphy A, Braud M, Donoghue MT, Liu Y, Chamberlain AT, Rue-Albrecht K, Schroeder S, Spillane C, Tai S, Bradley DG, Sonstegard TS, Loftus BJ, MacHugh DE. 2015. Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

phylogeography and evolution of cattle. *Genome Biol* **16**:234. doi: 10.1186/s13059-015-0790-2.

Pucek Z, Belousova IP, Krasińska M, Krasiński ZA, Olech W. 2004. European bison. Status Survey and Conservation Action Plan. IUCN/SSC Bison Specialist Group, IUCN. Gland, Switzerland and Cambridge, ix+54.

Pucek Z. 1991. History of the European bison and problems of its protection and management. In: Bobek B, Perzanowski K, Regelin WL, editors. *Global trends in wildlife management*. Kraków-Warszawa (Poland). Świat Press.

Raczyński J. 2014. Księga Rodowodowa Żubrów. Białowieski Park Narodowy, Białowieża.

Radwan J, Kawałko A, Wójcik JM, Babik W. 2007. MHC-DRB3 variation in a free-living population of the European bison, *Bison bonasus*. *Mol Ecol* **16**:531-540.

Ramsey CB, Gigham T, Leach P. 2004a. Towards high-precision AMS: progress and limitations. *Radiocarbon* **46**:17-24.

Ramsey CB, Higham T, Bowles A, Hedgeset R. 2004b. Improvements to the pretreatment of bone at Oxford. *Radiocarbon* **46**: 155-63.

Rohland N, Siedel H, Hofreiter M. 2010. A rapid column-based ancient DNA extraction method for increased sample throughput. *Mol Ecol Resour* **10**:677–683. doi: 10.1111/j.1755-0998.2009.02824.x.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Ramsey CB, Buck CE, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Haflidason H, Hajdas I, Hatté C, Heaton TJ, Hoffmann DL, Hogg AG, Hughen KA, Kaiser KF, Kromer B, Manning SW, Niu M, Reimer RW, Richards DA, Scott EM, Southon JR, Staff RA, Turney CSM, van der

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Plicht J. 2013. IntCal13 and Marine13 radiocarbon age calibration curves 0-50,000 years cal BP. *Radiocarbon* **55**:1869–1887.

Sambrook J, Russell DW. 2001. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press.

Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**:e34131. doi: 10.1371/journal.pone.0034131.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MT, Barnes I, Binladen J, Willerslev E, Hansen AJ, Baryshnikov GF, Burns JA, Davydov S, Driver JC, Froese DG, Harington CR, Keddie G, Kosintsev P, Kunz ML, Martin LD, Stephenson RO, Storer J, Tedford R, Zimov S, Cooper A. 2004. Rise and fall of the Beringian steppe bison. *Science* **306**:1561–1565.

Shapiro B, Hofreiter M. 2012. Ancient DNA. Methods and protocols. New York, Humana Press.

Sherkow JS, Greely HT. 2013. Genomics. What if extinction is not forever? *Science* **340**:32-33. doi: 10.1126/science.1236965.

Sipko TP, Rautian GS, Udina IG, Rakitskaya, TA. 1995. Investigation of blood group polymorphism in European bison (*Bison bonasus*). *Russ J Genet* **32**:356-351.

Sipko TP, Rautian GS, Udina IG, Strelchenko, NS. 1997. Conservation of genetic material from endangered and economically important species in the establishment of cryobanks. In: Physiology and general biology reviews. TM Tupaev and NK Kolcov, eds, 35-99.

Sipko TP, Udina IG, Badagueva YN. Sulimowa GE. 1994. Polymorphism of kappa-casein gene in the family Bovidae: comparative analysis. *Russ J Genet* **30**:203-207.

Slatis HM. 1960. An analysis of inbreeding in the European bison. *Genetics* **45**:275–287. PMID: 17247924.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312-3. doi: 10.1093/bioinformatics/btu033.

Steyer K, Kraus RHS, Mölich T, Anders O, Cocchiararo B, Frosch C, Geib A, Götz M, Herrmann M, Hupe K, Kohnen A, Krüger M, Müller F, Pir JB, Reiners TE, Roch S, Schade U, Schiefenhövel P, Siemund M, Simon O, Steeb S, Streif S, Streit B, Thein J, Tiesmeyer A, Trinzen M, Vogel B, Nowak C. 2016. Large-scale genetic census of an elusive carnivore, the European wildcat (*Felis s. silvestris*). *Conserv Genet* 1-17. doi:10.1007/s10592-016-0853-2.

Tiedemann R, Nadlinger K, Pucek Z. 1998. Mitochondrial DNA-RFLP analysis reveals low levels of genetic variation in European bison *Bison bonasus*. *Acta Theriol Suppl* **5**:83-87.

Tokarska M, Bunevich AN, Demontis D, Sipko T, Perzanowski K, Baryshnikov G, Kowalczyk R, Voitukhovskaya Y, Wójcik JM, Marczuk B, Ruczyńska I, Pertoldi C. 2015. Genes of the extinct Caucasian bison still roam the Białowieża Forest and are the source of genetic discrepancies between Polish and Belarusian populations of the European bison, *Bison bonasus*. *Biol J Linnean Soc* **114**:752–763. doi:10.1111/bij.12470.

Tokarska M, Kawalko A, Wójcik JM, Pertoldi C. 2009a. Genetic variability in the European bison (*Bison bonasus*) population from Białowieża forest over 50 years. Biol J Linn Soc **97**:801-809. doi: 10.1111/j.1095-8312.2009.01203.x

Tokarska M, Marshall T, Kowalczyk R, Kawałko A, Wójcik JM, Pertoldi C, Kristensen TN, Loeschcke V, Gregersen VR, Bendixen C. 2009b. Effectiveness of microsatellite and SNP loci (BovineSNP50 BeadChip, Illumina®) in parentage assignment and identity analysis in species with low genetic diversity: the case of European bison. *Heredity* **103**:326-332. doi:10.1038/hdy.2009.73.

Udina IG, Shakhaev GO.1998. Restriction fragment length polymorphism (RFLP) of exon 2 of the MHC Bibo-DRB3 gene in European bison *Bison bonasus*. *Acta Theriol* **43**:75-82.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Udina IG, Sokolova SS, Sipko TP, Sulimova GE. 1994. Comparative characteristics of DNA polymorphism for major histocompatibility complex loci DQB and DRB in representatives of the Bovidae family. *Genetika* **30**:356-360.

van Vuure C. 2005. Retracing the Aurochs: History, Morphology and Ecology of an extinct wild Ox. Pensoft Publishers. Sofia-Moscow. ISBN 954-642-235-5.

Wolf DE, Takebayashi N, Riesberg LH, 2001. Predicting the risk of extinction through hybridization. Conserv Biol **15**:1039–1053.

Wójcik JM, Kawałko A, Tokarska M, Jaarola M, Vallenback P, Pertoldi C. 2009. Post-bottleneck mtDNA diversity in a free-living population of European bison: implications for conservation. *J Zool* **277**:81-87. doi: 10.1111/j.1469-7998.2008.00515.x

FIGURE AND TABLE LEGENDS

Figure 1. Map of Western Europe showing the putative historical range of Lowland wisent (shaded green) and Caucasian wisent (shaded grey) based on bone remains and written records (according to Benecke et al., 2005; Kuemmerle et al., 2011; Tokarska et al., 2011; Bocherens et al., 2015) and sample locations. Black circles indicate contemporary free-ranging modern L line herds and white circles indicate modern LC line herds. Purple and peach circles denote the locations of investigated modern L (MdL1, MdL2) and LC (MdLC) line wisent, respectively, orange squares show the location of the Holocene wisent (Bb1-3) and blue and yellow triangles indicate historical founding wisent from the Pszczyna population (PLANTA and PLATEN) and the extinct Caucasian wisent (Cc1-3), respectively.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2. Pairwise genomic divergence among wisent. A, B, C, show probability densities for pairwise transversion divergence (x axes) along a sliding window of 1Mb. Individual plots show all pairwise comparisons among modern and founding individuals (A); comparisons of modern and founding individuals and Caucasian wisent Cc1 (B); and comparisons of Caucasian wisent Cc2 and all other individuals (C). Specific comparisons discussed in the text are identified by colours, according to the key at the top right of each plot. Schematic neighbour-joining phylogeny based on whole genome distances (D).

Figure 3. Population history of Lowland and Caucasian wisent, estimated using two representatives of each. The pie chart shows the percentage of 1Mb genomic blocks supporting each alternatively rooted tree topology. The indicates fraction of genome blocks returning both wisent subspecies as reciprocally monophyletic. Dark and light blue colours show the next most frequently encountered topologies in which Caucasian wisent are paraphyletic (dark blue: Cc2 most divergent, and light blue: Cc1 most divergent, the first of which is compatible with estimates pairwise genomic divergence (see Fig 2d).

Figure 4. Comparison of pairwise genomic divergence within three bovid species: wisent (*Bison bonasus*), domestic cattle (*Bos taurus*), and yak (*Bos grunniens*). Probability densities were calculated along a sliding window of 1Mb from transversions only (A), and from transitions and transversions (B). For wisent, only modern samples included.

Figure 5. Results of *D* statistic analysis. Red and grey points show significant and nonsignificant *D* values (x axis), respectively, and show: the genetic contribution of the founders (F) to the modern individuals (A); Caucasian wisent (Cc) admixture with modern L and LC herds (MdL and MdLC) and one founder, PLANTA, relative to the least Caucasian

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

admixed wisent, founder PLATEN (B); Caucasian wisent admixture among modern wisent (C); apparent cattle (DC) admixture with all investigated wisent (W) relative to aurochs (Aur) (D); variance in cattle/aurochs admixture among wisent (W) compared to PLATEN (E). Detailed *D* statistic results are provided in Supporting Table S5.

Figure 6. Genomic admixture map (A) of Caucasian bison ancestry in the modern LC line individual (MdLC). Colored blocks indicate 1Mb genomic blocks returning alternative tree topologies, blue blocks are compatible with the species tree; yellow blocks return the monophly of the modern LC line and Caucasian wisent, and likely result from admixture and to a lesser extent incomplete lineage sorting; red blocks return the monophyly of PLATEN and Caucasian wisent and likely result from incomplete lineage sorting. “X” shows blocks with missing data. The pie chart (B) shows the percentage of 1Mb genomic supporting each tree topology identified by colours, according to the key presented above.

Figure 7. Variation in the sizes of genomic blocks in modern L (blue) and LC (red) likely resulting from Caucasian wisent admixture. Plots show cumulative probability densities calculated at a scale of 1MB. Genomic blocks in the LC line wisent (red) result from admixture occurring around 90 years ago; the lower abundance of larger admixed blocks in the modern L line wisent support that this admixture event preceded the former. The plots have been truncated to aid visualisation, and single blocks of 18MB and 22MB in the LC line individual are not shown. The largest block size detected in the modern L line individual was 8MB.

Figure 8. Schematic diagram showing inferred admixture among wisent, and among wisent and the cattle/aurochs lineage. Arrows indicate the direction of the geneflow. Black lines

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

indicate admixture between wisent and cattle/aurochs, yellow lines/arrow - between Caucasian and founding or modern wisent respectively, and the blue arrow - from founders to modern wisent.

Table 1. Basic information about sampled individuals. The last column is the average coverage for each sample after mapping it to either the water buffalo (*Bubalus bubalis*) nuclear reference genome (GenBank accession no. GCA_000471725.1) or the American bison (*Bison bison*) mitochondrial reference genome (GenBank accession no. NC_012346.1).

SUPPORTING INFORMATION

Figure S1. A phylogeny of wisent (A) and the Bovini (B), inferred from a partitioned maximum likelihood (ML) analysis of whole mitochondrial genomes. (A) is an expansion of the region in the blue box in (B). The green box in (A) highlights the two haplotypes found in sampled historic and modern individuals: one in Caucasian wisent (Cc2) and the other one identical to haplotype published already (JN632602; Hassanin et al., 2012) in all remaining historical and modern wisent (see Table S1). Tips are coloured based on whether the haplotype occurs in modern individuals (black), or only in historical or ancient individuals, and therefore likely extinct (red). The outgroup, Pseudoryx nghetinhensis, is not shown. Branch support is indicated by bootstrap percentages based on 500 ML bootstrap replicates (above branches) and Bayesian posterior probabilities (below branches).

Figure S2. DNA fragmentation (upper four plots) and deamination (lower two plots) patterns of founding wisent, PLANTA, sequencing reads. For deamination plots, red lines show the frequency of C to T substitutions (Y axes) in the sequenced historical DNA fragments relative

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

to the reference genome at the 5' (left plot) and 3' (right plot) fragment ends. X axes show sequenced positions moving internally from the 5' (positive values) and 3' (negative values) fragment ends. Elevated rate of C to T substitutions at fragment ends are indicative of DNA damage.

Figure S3. DNA fragmentation and deamination patterns of founding wisent, PLATEN, sequencing reads.

Figure S4. DNA fragmentation and deamination patterns of Caucasian wisent, Cc1, sequencing reads.

Figure S5. DNA fragmentation and deamination patterns of Caucasian wisent, Cc2, sequencing reads.

Table S1. Species and accession data for sequences included in the mitochondrial genome phylogenetic analysis.

Table S2. Full details of sequencing and mapping results for modern wisent samples. SR – merged reads, PE – unmerged reads.

Table S3. Full details of sequencing and mapping results for archival wisent samples. SR – merged reads, PE – unmerged reads.

Table S4. Full details of mapping results for data downloaded from the NCBI Short Read Archive. SR – merged reads, PE – unmerged reads.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S5. Detailed *D* statistic results. In red significant results are shown. Results highlighted in grey are presented also in Figure 5.

Table S6. $\hat{f}$ results. In the last column percentage of the genome resulted from hybridisation is given.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

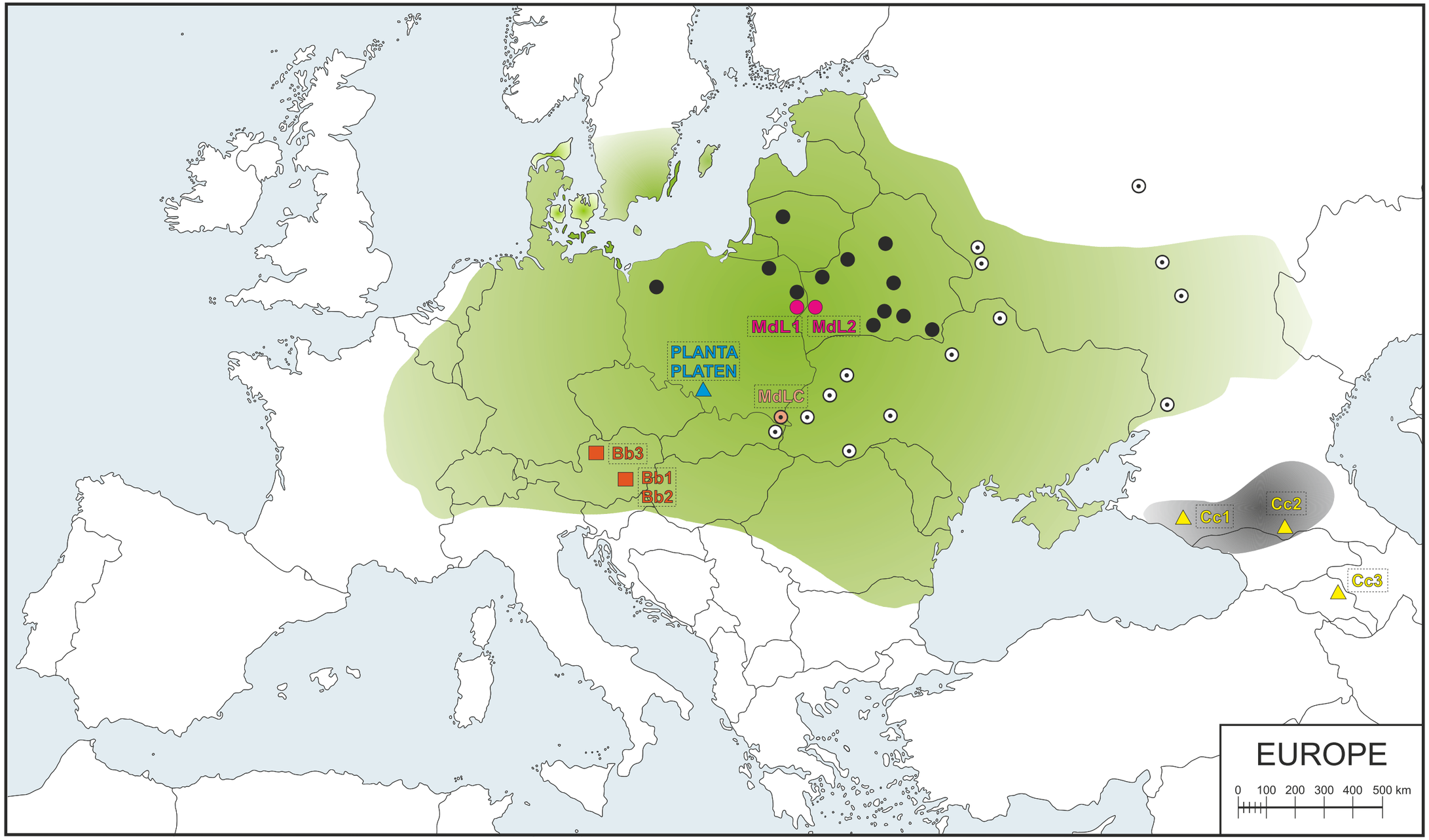

Figure 1. Map of Western Europe showing the putative historical range of Lowland wisent (shaded green) and Caucasian wisent (shaded grey) based on bone remains and written records (according to Benecke et al., 2005; Kuemmerle et al., 2011; Tokarska et al., 2011; Bocherens et al., 2015) and sample locations. Black circles indicate contemporary free-ranging modern L line herds and white circles indicate modern LC line herds. Purple and peach circles denote the locations of investigated modern L (MdL1, MdL2) and LC (MdLC) line wisent, respectively, orange squares show the location of the Holocene wisent (Bb1-3) and blue and yellow triangles indicate historical founding wisent from the Pszczyna population (PLANTA and PLATEN) and the extinct Caucasian wisent (Cc1-3), respectively.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

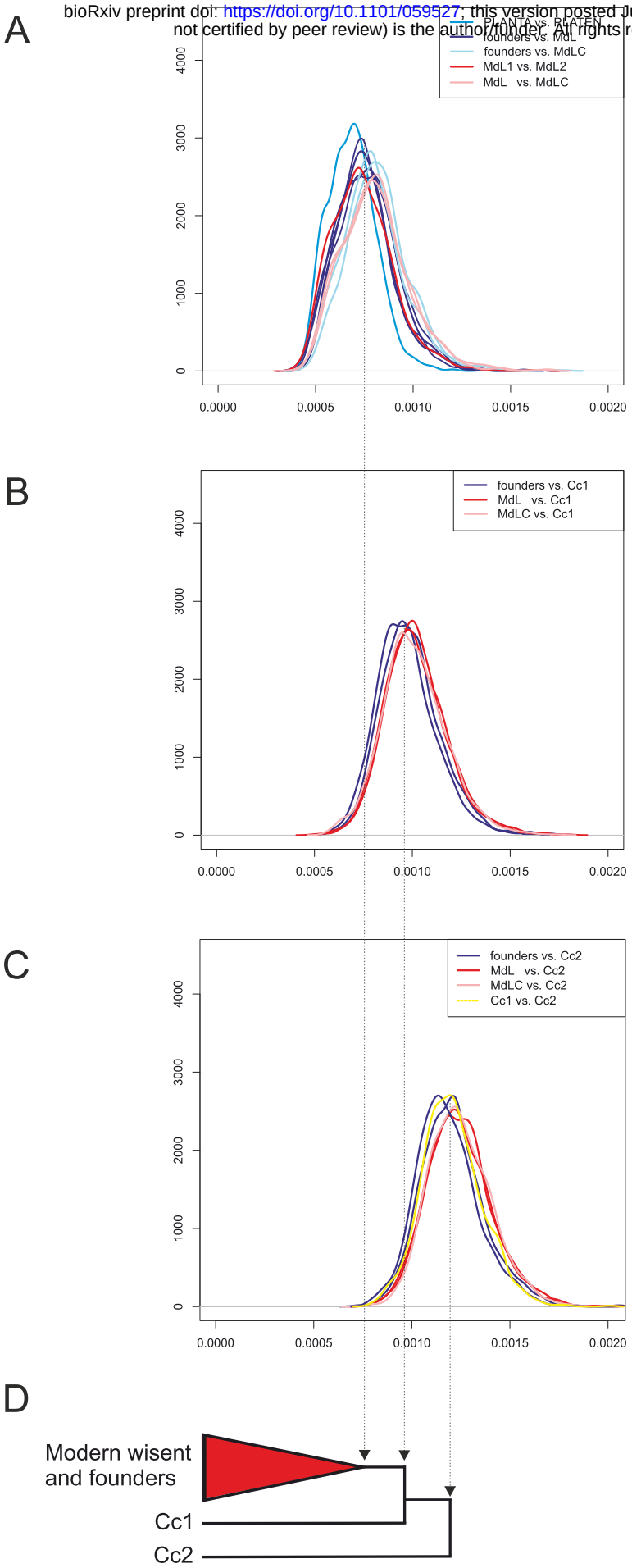

Figure 2. Pairwise genomic divergence among wisent. A, B, C, show probability densities for pairwise transversion divergence (x axes) along a sliding window of 1Mb. Individual plots show all pairwise comparisons among modern and founding individuals (A); comparisons of modern and founding individuals and Caucasian wisent Cc1 (B); and comparisons of Caucasian wisent Cc2 and all other individuals (C). Specific comparisons discussed in the text are identified by colours, according to the key at the top right of each plot. Schematic neighbour-joining phylogeny based on whole genome distances (D).

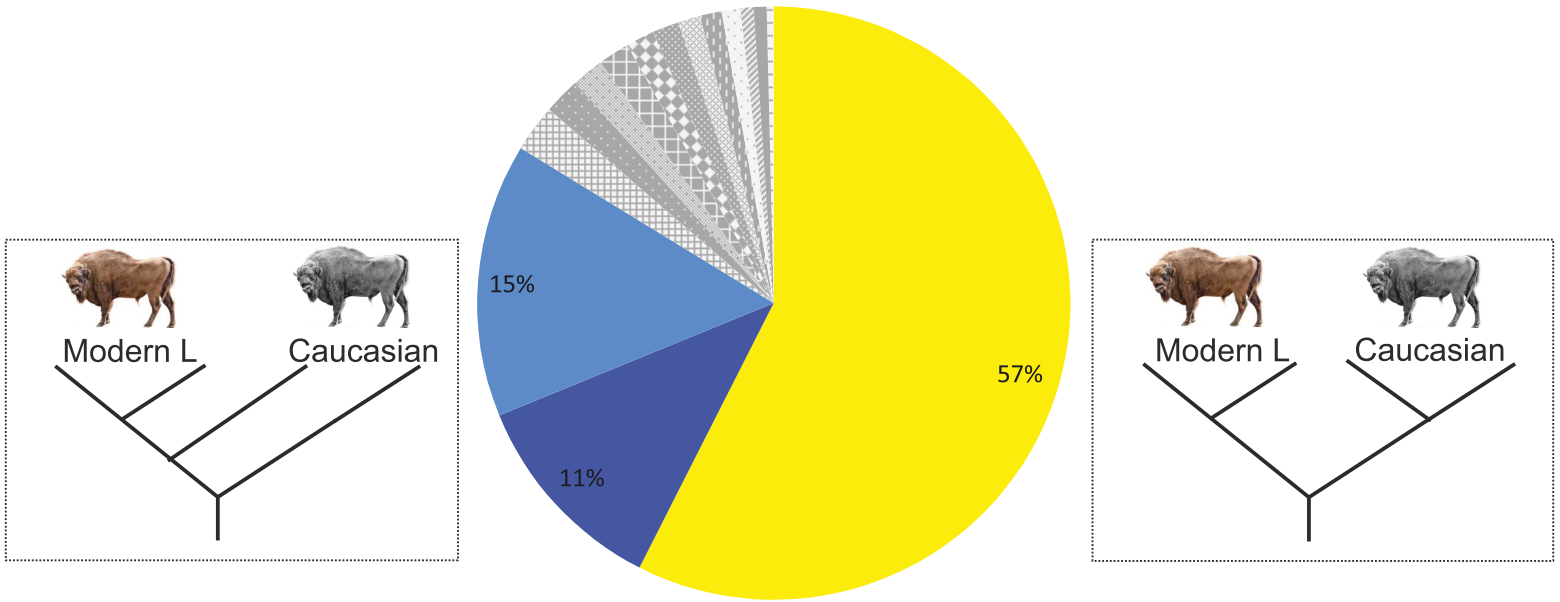

Figure 3. Population history of Lowland and Caucasian wisent, estimated using two representatives of each. The pie chart shows the percentage of 1Mb genomic blocks supporting each alternatively rooted tree topology. The indicates fraction of genome blocks returning both wisent subspecies as reciprocally monophyletic. Dark and light blue colours show the next most frequently encountered topologies in which Caucasian wisent are paraphyletic (dark blue: Cc2 most divergent, and light blue: Cc1 most divergent, the first of which is compatible with estimates pairwise genomic divergence (see Fig 2d).

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

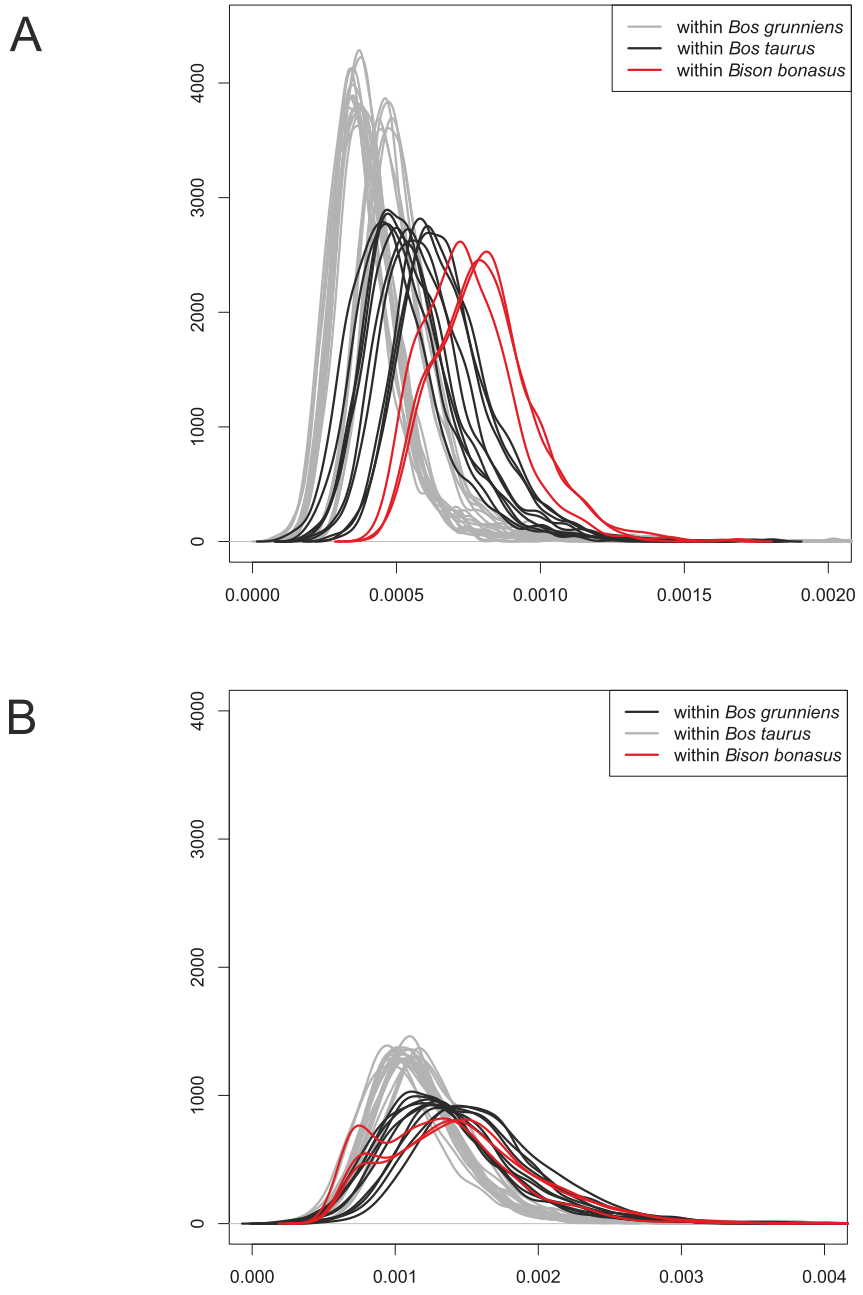

Figure 4. Comparison of pairwise genomic divergence within three bovid species: wisent (*Bison bonasus*), domestic cattle (*Bos taurus*), and yak (*Bos grunniens*). Probability densities were calculated along a sliding window of 1Mb from transversions only (A), and from transitions and transversions (B). For wisent, only modern samples included.

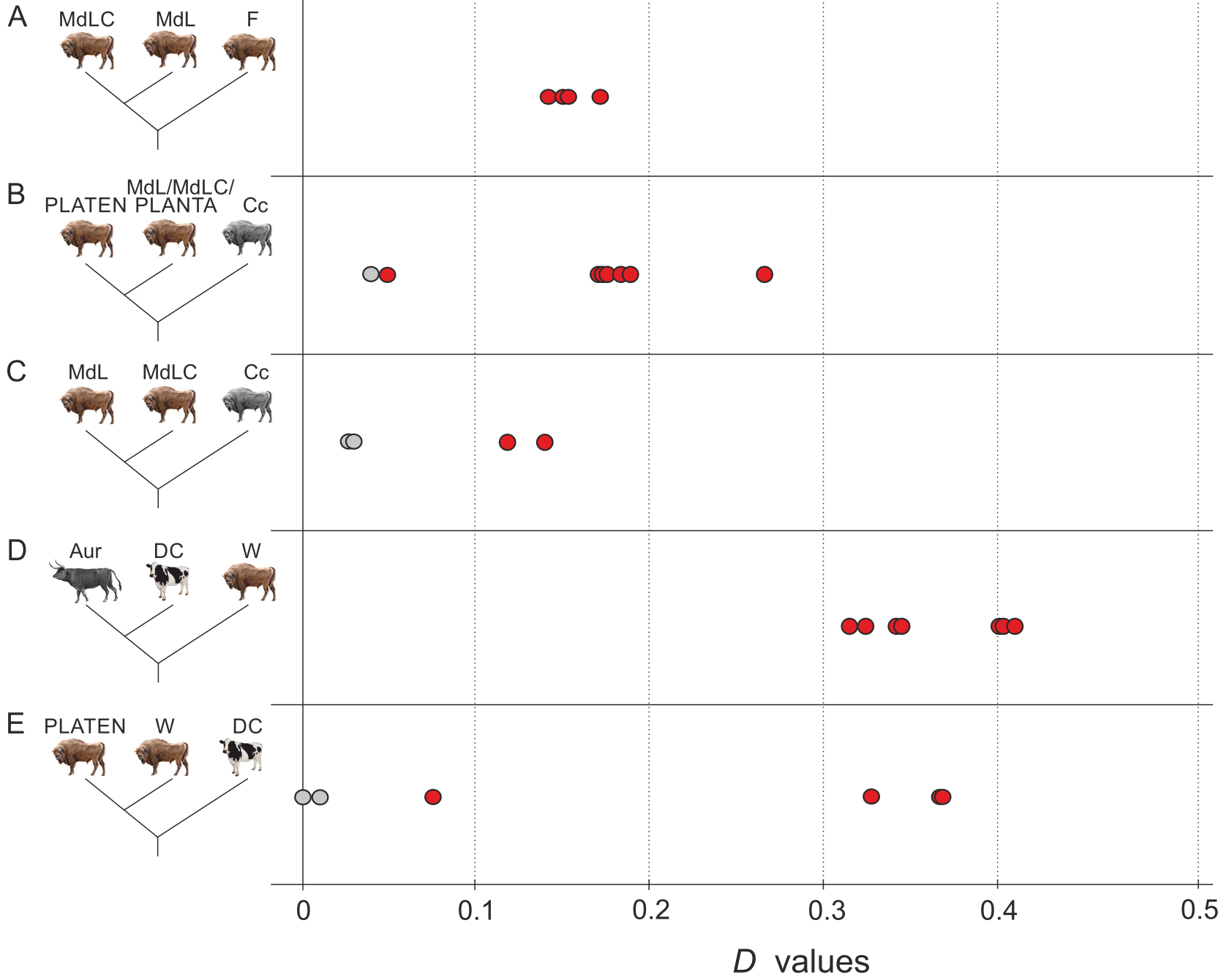

Figure 5. Results of D statistic analysis. Red and grey points show significant and nonsignificant D values (x axis), respectively, and show: the genetic contribution of the founders (F) to the modern individuals (A); Caucasian wisent (Cc) admixture with modern L and LC herds (MdL and MdLC) and one founder, PLANTA, relative to the least Caucasian admixed wisent, founder PLATEN (B); Caucasian wisent admixture among modern wisent (C); apparent cattle (DC) admixture with all investigated wisent (W) relative to aurochs (Aur) (D); variance in cattle/aurochs admixture among wisent (W) compared to PLATEN (E). Detailed D statistic results are provided in Supporting Table S5.

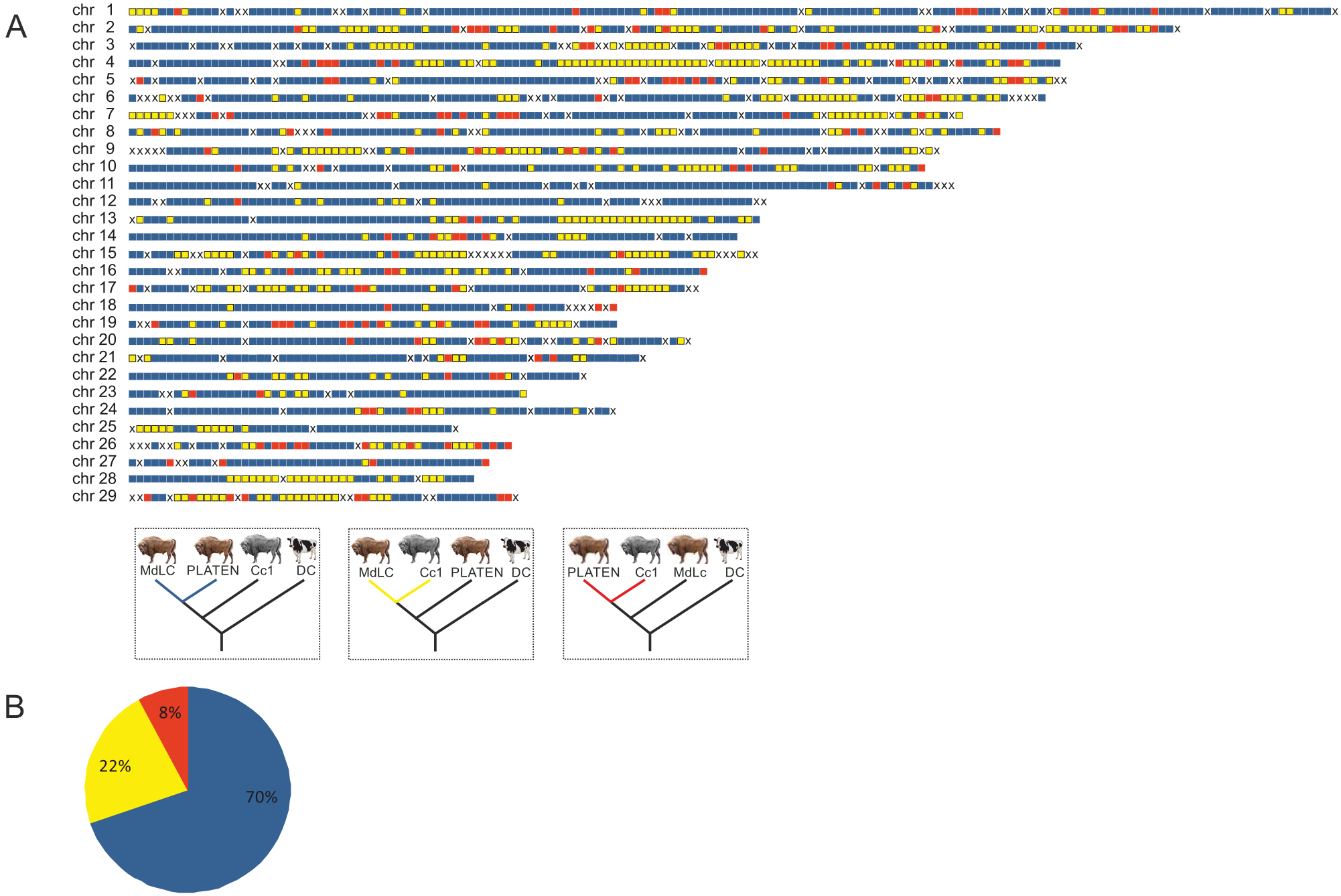

Figure 6. Genomic admixture map (A) of Caucasian bison ancestry in the modern LC line individual (MdLC). Colored blocks indicate 1Mb genomic blocks returning alternative tree topologies, blue blocks are compatible with the species tree; yellow blocks return the monophly of the modern LC line and Caucasian wisent, and likely result from admixture and to a lesser extent incomplete lineage sorting; red blocks return the monophyly of PLATEN and Caucasian wisent and likely result from incomplete lineage sorting. “X” shows blocks with missing data. The pie chart (B) shows the percentage of 1Mb genomic supporting each tree topology identified by colours, according to the key presented above.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

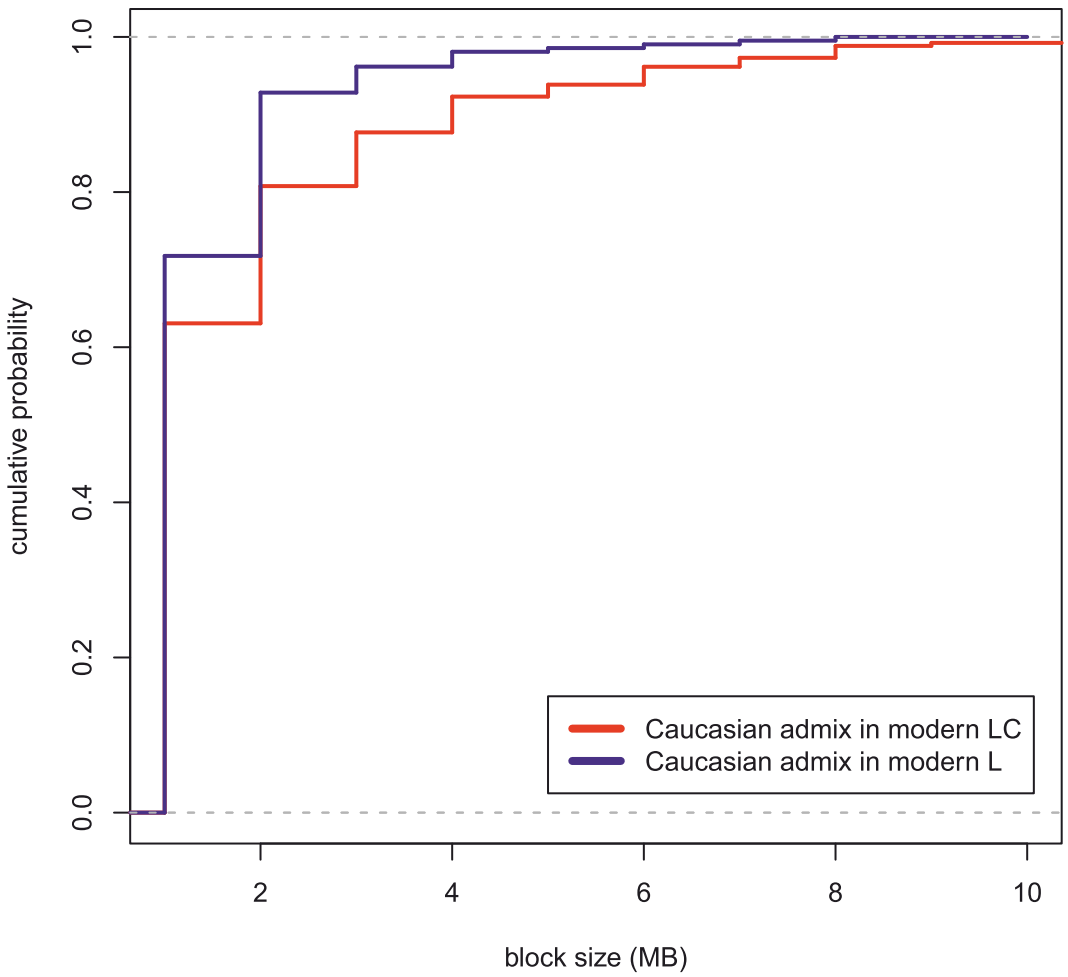

Figure 7. Variation in the sizes of genomic blocks in modern L (blue) and LC (red) likely resulting from Caucasian wisent admixture. Plots show cumulative probability densities calculated at a scale of 1MB. Genomic blocks in the LC line wisent (red) result from admixture occurring around 90 years ago; the lower abundance of larger admixed blocks in the modern L line wisent support that this admixture event preceded the former. The plots have been truncated to aid visualisation, and single blocks of 18MB and 22MB in the LC line individual are not shown. The largest block size detected in the modern L line individual was 8MB.

bioRxiv preprint doi: https://doi.org/10.1101/059527; this version posted July 15, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

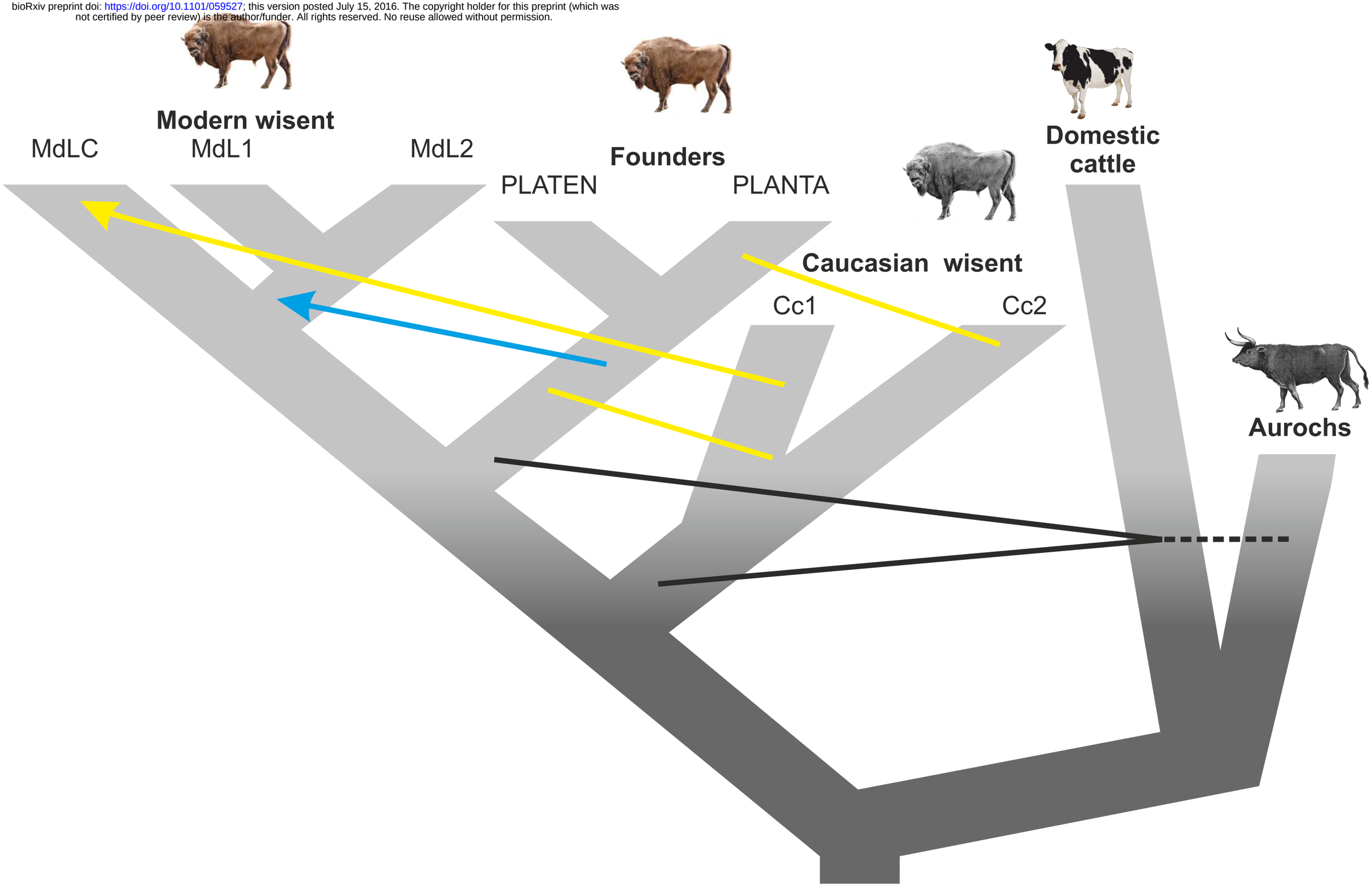

Figure 8. Schematic diagram showing inferred admixture among wisent, and among wisent and the cattle/aurochs lineage. Arrows indicate the direction of the geneflow. Black lines indicate admixture between wisent and cattle/aurochs, yellow lines/arrow - between Caucasian and founding or modern wisent respectively, and the blue arrow - from founders to modern wisent.

Table 1. Basic information about sampled individuals. The last column is the average coverage for each sample after mapping it to either the water buffalo (*Bubalus bubalis*) nuclear reference genome (GenBank accession no. GCA_000471725.1) or the American bison (*Bison bison*) mitochondrial reference genome (GenBank accession no. NC_012346.1).

Sample code	Sample name	Taxon	Group	Sample location	Sample origin	Tissue type	Year of death*/ radiocarbon age	Water buffalo nDNA mapping coverage/ mtDNA coverage
MdL1	Z3	*Bison bonasus*	Modern Lowland line (L)	Białowieża Forest, Polish part, Poland		kidney	2006	1.59 (n)
MdL2	868			Białowieża Forest, Belarussian part, Belaruss		muscles	2009	1.63 (n)
MdLC	Dy		Modern Lowland-Caucasian line (LC)	Dydiowa, Bieszczady Mountains, Poland		muscles	2012	1.49 (n)
PLANTA	F42 PLANTA	*Bison b. bonasus*	Founders	Pszczyna, Poland	Upper Silesian Museum, Bytom, Poland	horns	1931	1.50 (n)
PLATEN	M158 PLATEN						1933	1.39 (n)
Cc1	8853	*Bison b. caucasicus*	Caucasian wisent	Kubańskaja Obłast	Zoological Institute RAS, Sankt Petersburg, Russia	skull bones	1911*	1.07 (n)

Cc2	22533			North Ossetia-Alania	State Darwin Museum, Moscow, Russia		1949*	0.92 (n)
Cc3	AF005		Holocene Caucasian wisent	Sevan Lake region, Armenia		horn core	^{14}C: 4972±29 OxA-31935 5724-5657	58.10 (mt)
Bb1	BS587	*Bison b. bonasus*	Holocene wisent	Styria, Austria	VMNH H-1981-28-6	tibia	^{14}C: 1480±70 VERA 0145 1511-1302	115.56 (mt)
Bb2	BS589				VMNH H-1977-49-1	femur	^{14}C: 1980±45 VERA 0142 1987-1886	165.75 (mt)
Bb3	BS607			Upper Austria, Austria	VMNH H-1979-48-1	femur	^{14}C: 1370±50 VERA 0143 1338-1265	14.10 (mt)

* for Caucasian individuals (Cc1, Cc2) year of collecting the sample is given. VNHM, Vienna Museum of Natural History. n; nuclear, mt; mitochondrial. Radiocarbon dates include the ^{14}C age (years before present; yr BP), ^{14}C accession (where known), and calibrated age BP (1σ). Calibrated dates follow the IntCal13 calibration curve.