Article

Discoveries

Title

Deciphering the wisent demographic and adaptive histories from individual whole-genome sequences

Authors and Affiliations

Mathieu GAUTIER[1,2§], Katayoun MOAZAMI-GOUDARZI[3], Hubert LEVEZIEL[4], Hugues PARINELLO[5], Cécile GROHS[3], Stéphanie RIALLE[5], Rafał KOWALCZYK[6], Laurence FLORI[3,7§]

[1] CBGP, INRA, CIRAD, IRD, Supagro, 34988 Montferrier-sur-Lez, France

[2] IBC, Institut de Biologie Computationnelle, 34095 Montpellier, France

[3] GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

[4] UGMA, INRA, Université de Limoges, 87060 Limoges, France

[5] MGX-Montpellier GenomiX, c/o Institut de Génomique Fonctionnelle, Montpellier, France

[6] Mammal Research Institute, Polish Academy of Sciences, 17-230 Białowieża, Poland

[7] INTERTRYP, CIRAD, IRD, 34398 Montpellier, France

[§]Corresponding authors

Emails:

mathieu.gautier@supagro.inra.fr

laurence.flori@jouy.inra.fr

Abstract

As the largest European herbivore, the wisent (*Bison bonasus*) is emblematic of the continent wildlife but has unclear origins. Here, we infer its demographic and adaptive histories from two individual whole genome sequences via a detailed comparative analysis with bovine genomes. We estimate that the wisent and bovine species diverged from $1.7x10^6$ to 850,000 YBP through a speciation process involving an extended period of limited gene flow. Our data further support the occurrence of more recent secondary contacts, posterior to the *Bos taurus* and *Bos indicus* divergence (*ca.* 150,000 YBP), between the wisent and (European) taurine cattle lineages. Although the wisent and bovine population sizes experienced a similar sharp decline since the Last Glacial Maximum, we find that the wisent demography remained more fluctuating during the Pleistocene. This is in agreement with a scenario in which wisents responded to successive glaciations by habitat fragmentation rather than southward and eastward migration as for the bovine ancestors.

We finally detect 423 genes under positive selection between the wisent and bovine lineages, which shed a new light on the genome response to different living conditions (temperature, available food resource and pathogen exposure) and on the key gene functions altered by the domestication process.

Introduction

The European bison, *Bison bonasus* (BBO), also known as the wisent, is an emblematic species of the European wildlife. The oldest fossil remains were found in Eastern Europe and trace back to the Late Pleistocene, approximately 12,000 Years Before Present (YBP) (Benecke 2005). The wisent is closely related to the North American bison (*Bison bison*), both species assumed to be derived from the extinct long-horned steppe bison (*Bison priscus*) which was widespread across the Northern hemisphere in the mid and upper Pleistocene (i.e., 781,000 to 11,700 YBP). However, several morphological and behavioral traits distinguish the wisent from its American relative. In particular, adults have generally a smaller body length (Krasińska and Krasiński 2002), display a less hairy body and tend to be mixed feeders that combine browsing and grazing (Kowalczyk, et al. 2011; Merceron, et al. 2014) while American bisons are essentially grazer. These various characteristics might be viewed as adaptation to living constraints associated to forest habitat that has been considered as a refuge (Kerley, et al. 2012; Bocherens, et al. 2015). During the mid and late Holocene, the wisent was distributed in central and eastern Europe with a longest survival in north-eastern Europe (Benecke 2005). Nevertheless, the progressive replacement of the open steppe by forest cover in early Holocene and the human population growth associated with the spread of farming during the middle Holocene lead to a reduction of its habitat (Kuemmerle, et al. 2012; Bocherens, et al. 2015). Hence, by the Middle Age, the wisent was only present in a few natural forests of North-Eastern Europe and finally faced extinction in the wild at the beginning of the 20th century (Pucek 2004). A species restoration program relying on a few tens of specimens kept in European zoos and private breeding centers was early initiated in the 1920's (Pucek 2004; Tokarska, et al. 2011). Thanks to these efforts, the wisent world population now consists of over 5,000 animals (including about 3,500 free-living individuals) that are mostly maintained in forests of eastern Europe (Kerley, et al. 2012; Raczyński 2015).

The wisent is among the few large mammalian terrestrial species that survived the massive megafauna extinction of the last glacial/interglacial transition period, ca. 25,000-10,000 YBP (Lister and A.J. 2008; Lorenzen, et al. 2011). Characterizing the specific conditions of its survival may thus help understanding how organisms evolve in response to environmental change. In this study, we explore the wisent demographic and adaptive histories using individual whole-genome sequences from two males of the wisent lowland line (Tokarska, et al. 2011) living in the Białowieża Forest. To that end, we perform a detailed comparative analysis of these newly generated wisent genome sequences with individual genome

sequences representative of different European cattle breeds. Domestic cattle, the wild ancestor of which, the aurochs (*Bos primigenius*) went extinct in the 17th century, is indeed among the closest wisent relatives (Verkaar, et al. 2004) and benefits from abundant genomic resources (Liu, et al. 2009). Our genome-wide comparison provides an accurate characterization of the divergence between the bovine and wisent lineages and an estimation of their population size histories during Pleistocene and Holocene. In addition, the identification and functional annotation of genes under positive selection between both lineages shed light into the biological functions and the key adaptive traits that were affected by environmental constraints.

Results and discussion

Comparison of wisent and cattle genomes reveals low amount of nucleotide divergence.

High-throughput sequencing data from two wisent bulls, namely BBO_3569 and BBO_3574, were each aligned onto the UMD3.1 cattle (BTA) genome assembly (Liu, et al. 2009). The read mapping statistics showed overall good performances (Table S1) with noticeably 97.4% (respectively 96.6%) of the reads being properly paired for the BBO_3569 (respectively BBO_3574). Conversely, following the same read mapping procedure but considering as a reference the *OviAr3* genome assembly (Jiang, et al. 2014) from the more distantly related domestic sheep (OAR) strongly altered the statistics (Table S1). Altogether, these mapping results provided support to consider the cattle genome as closely related enough to provide a good reference to assemble our wisent reads. For autosomes, an approximately 10X-coverage was achieved for both individuals (9.81X and 11.6X for BBO_3569 and BBO_3574 respectively) with more than 95% of the (bovine) reference sequence covered by at least one read (Table 1). As expected for males, the bovine chromosome X was only slightly more than half covered (ca. 56.8% and 55.8% of the autosome coverage for BBO_3569 and BBO_3574 respectively) due to the inclusion of pseudo-autosomal regions (PAR) in the estimation. Conversely, the (bovine) mitochondrial genome was found highly covered (>300X) for both individuals.

Interestingly, the estimated autosomal nucleotide divergence between the bovine genome and each wisent individual genome was found equal to 0.870% for BBO_3569 and 0.882% for BBO_3574 (Table 1) supporting a low amount of genome divergence between the two species. The estimated nucleotide divergence was about 1.37 times smaller for the

chromosome X than for autosomes, as already observed in the Chimpanzee vs. Human genome sequence comparison (Mikkelsen, et al. 2005). This is likely related to a higher mutation rate in the male compared to the female germ line in mammalian species (Li, et al. 2002; Mikkelsen, et al. 2005). Also, the nucleotide divergence increased in distal chromosomal regions in both autosomes and X-chromosome comparisons while read coverage remained mostly uniform (Figure S1). Such a regional pattern, also reported in the aforementioned Chimpanzee vs. Human genome sequence comparison (Mikkelsen, et al. 2005), might similarly result from the physical properties of the corresponding distal regions (i.e. high local recombination rate, high gene density and higher GC content). Indirectly, observing such a pattern for wisent sequences mapped onto the cattle genome assembly supports a low amount of chromosome rearrangements between the two species. Finally, the estimated nucleotide divergence for the mitochondrial genomes were about 4 times higher than for autosomes. This suggests an absence of recent maternal bovine introgression at least within the lineages of the two wisents we considered.

The wisent demography has been more fluctuating than the bovine one during the whole Pleistocene.

We further sought to characterize the genetic variability of the wisent population that has long been considered as a threatened species and is still classified as "vulnerable" by the International Union for Conservation of Nature (IUCN) red list (http://www.iucnredlist.org/). The estimated average (autosomal) heterozygosities, as approximated by the population mutation rate ($\theta = 4N_e\mu_g$ where μ_g represents the mutation rate per site and per generation and N_e the effective population size) measured on whole individual genomes, were consistent when considering either the BBO_3569 ($\hat{\theta} = 1.37 \times 10^{-3}$) or the BBO_3574 ($\hat{\theta} = 1.45 \times 10^{-3}$) individual. Interestingly, both estimates were similar to the ones obtained on four individual cattle genomes representative of the Angus (AAN_0037), Jersey (JER_0009), Holstein (HOL_0101) and Simmental (SIM_0043) European cattle breeds which ranged from $\hat{\theta} = 1.27 \times 10^{-3}$ to $\hat{\theta} = 1.97 \times 10^{-3}$ (Figure S2 and Table S2). The small heterozygosity observed for wisent might be explained by the sharp decline of its effective population size during the last 20,000 years (see below), but also by its more recent history since wisent experienced a strong bottleneck at the end of the First World War (Wójcik, et al. 2009; Tokarska, et al. 2011). However, we noted that our wisent estimates remained within the range observed for the four different European cattle breeds. This suggests that the

constitution of the Białowieża wisent herd from a small number of presumably poorly related founders and subsequent management (Pucek 2004) allowed to recover a reasonable amount of variability for the current population.

More generally, the patterns of variability observed at a local individual genome scale are informative about the demographic history of the population (Li and Durbin 2011). From our two individual genomes, we estimated the wisent past N_e using the Pairwise Sequentially Markovian Coalescent model (PSMC) introduced by Li and Durbin (2011), and compare it to that estimated for the bovine species based on each of the four cattle individual genomes aforementioned. Estimates of the scaled effective population sizes (in units of $4N_e\mu_g$) backward in time measured in scaled units of $2\mu_y T$ (where T is in YBP, and μ_y represents the mutation rate per site and per year) are plotted in Figure S3 for each of the six individual genome analyses (see Figure S4 for bootstrap confidence intervals of each history). Overall, a similar trend was observed within each species, irrespective of the individual considered (although the AAN_0037 profile was more dissimilar than the three other cattle ones probably as a result of an overall lower coverage for this individual). By contrast, marked differences were found when comparing the inferred wisent and bovine population size histories. In particular, from backward time (in scaled units) $t=10^{-3}$ to $t=5.10^{-5}$ (Figure S3), wisent displayed more pronounced oscillations and from twice to three times higher effective population sizes. Because the PSMC model relies on local density of heterozygosites, using the bovine genome assembly as a reference to derive the analyzed individual wisent genomes (see Material and Methods) could have lead to some inferential biases (Nevado, et al. 2014). Nevertheless, the inferred wisent population size histories remained similar when considering a BBO consensus genome as a reference to derive the individual genome sequences (Figure S5). Similarly, given the close relatedness of the bovine and bison species, chromosomal rearrangements might remain negligible and should not have substantial effect on the overall inferred histories.

To facilitate biological interpretation of the observed demographic signals, we transformed the population size history profiles assuming $\mu_y = 2.10 \times 10^{-9}$ (Liu, et al. 2006) to translate the time scale in YBP. Further, to translate the effective population sizes (N_e) in real units, we assumed a generation time of g=6 years (Keightley and Eyre-Walker 2000; Gautier, et al. 2007), leading to $\mu_g = \mathrm{g} \times \mu_y = 1.26 \times 10^{-8}$. The resulting backward in time estimates of N_e are plotted in Figure 1. For both the wisent and bovine species, a sharp decline in N_e started at the end of the last glacial period, predating cattle domestication approximately 10,000 YBP as

already suggested for the latter species by MacLeod *et al.* (2013). This most presumably originates from a growing impact of human activities via hunting intensification and land anthropization especially after the development of agriculture (Soares, et al. 2010). Interestingly, the estimated wisent population sizes clearly oscillated according to the succession of glacial Pleistocene periods to *ca.* 500,000 YBP (viewed backward in time) with interglacial periods coinciding with lower population sizes. Conversely, from the Last Glacial Maximum (LGM) approximately 20,000 YBP (Clark, et al. 2009) to 150,000 YBP, the estimated bovine population sizes only slightly decreased, remaining always between 10,000 to 20,000 individuals (more than two to three times lower than wisents, as mentioned above). Hence, even if the fluctuating pattern observed for wisent might correspond to actual changes in the population sizes, the comparison with the bovine species might lead to an alternative scenario of population fragmentation. In such a scenario, the alternation of population isolation events due to fragmentation of the habitats during glacial periods leads to an increase of the overall estimated N_e while a more continuously interbreeding between populations during interglacial periods leads to a decrease in the overall estimated N_e. Indeed, as recently evidenced by Mazet *et al.* (2015) when considering structured populations, changes in the gene flow patterns result in changes in the estimated N_e (even if the actual N_e remained constant) as PSMC assumes a panmictic model. As a result, one might expect an increase (respectively a decrease) in the estimated N_e when gene flow and thus the amount of connectivity between (sub)populations decreases (respectively increases). It is not possible from the PSMC analyses to distinguish between a scenario involving habitat fragmentation and a scenario of actual population size changes. However, it is tempting to speculate from the divergent patterns in the inferred population size across the wisent and bovine species that the ancient aurochs distribution range, possibly restrained to southern regions with more stable habitats, has only been partially overlapping the wisent one (hence the lower N_e). In agreement with the "refugium theory" (Hofreiter and Stewart 2009), both species might have followed different strategies to survive during Pleistocene glaciations, the wisent remaining in small disconnected refugia while the aurochs migrated southward and possibly eastward (Mona, et al. 2010). Conversely, during interglacial periods, wisent may have retrieved a larger distribution area with a reconnection of isolated populations while the aurochs may have recolonized northern and central areas. Accordingly, the divergence of the wisent and the American bison which took place *ca.* 230,000 YBP (Hassanin, et al. 2013), might result from an absence of reconnection between populations separated in Eurasian and North-American refugia. Note also that the most recent common ancestor of all the Beringian bison

lived from 164,000 to 111,000 YBP (Shapiro, et al. 2004), a period that lies within the second peak observed in our estimated wisent population size history (Figure 1).

Finally, in agreement with MacLeod et al. (2013) that used a different inference model (and a different time calibration procedure), we also observed a steep increase (viewed backward in time) starting ca. 150,000 YBP in the inferred bovine population sizes that might actually be explained by the divergence of the taurine (BTA) and zebu (BIN) lineages (see e.g., Figure S5 in Li and Durbin, 2011). Indeed, this divergence event is dated between 117,000 and 275,000 YBP based on diversity of cattle mitochondrial sequences (Bradley, et al. 1996) and at 147,000 YBP (with a 95% Highest Posterior Density interval ranging from 84,000 to 219,000 YBP) using modern and radiocarbon-dated ancient mitochondrial DNA (Ho, et al. 2008). A similar increase starting about 1 MYA could also be observed for the ancestral wisent population size which coincides with the timing of BTA/BBO speciation (see below).

The wisent and bovine species diverged in the early Pleistocene and experienced more recent secondary contacts

To characterize the divergence between the BBO and BTA lineages, we further relied on the coalescent Hidden Markov Models (CoalHMM) framework (Hobolth, et al. 2007). Two models were considered corresponding to *i)* an isolation (I) model which assumes that the ancestral population split into two populations at time τ_D into the past (Mailund, et al. 2011); and *ii)* an isolation-with-migration (IM) model which assumes that the ancestral population started to split into two populations at time $\tau_2=\tau_1+\tau_M$ and that these two populations exchanged gene with a migration rate M during a period τ_M until a later time τ_1 when gene flow stopped (Mailund, et al. 2012). Under the I-model and assuming as above $\mu_y = 2.10 \times 10^{-9}$, the median estimate of τ_D (over 10 Mbp segments) was found equal to 1.20×10^6 YBP (ranging from 1.02×10^6 to 1.36×10^6) and 1.23×10^6 YBP (ranging from 1.01×10^6 to 1.51×10^6) when considering the alignments onto the UMD 3.1 bovine genome assembly of the BBO_3569 and BBO_3574 whole genome sequences respectively (Figure 2A). As expected, the τ_D estimates were always intermediate between the estimated start (τ_2) and end (τ_1) times of the divergence period obtained under the IM-model (Figure 2B). Hence, when considering the BBO_3569 (respectively BBO_3574) sequence the median estimates of τ_1 and τ_2 allowed to define a period ranging from 0.846×10^6 YBP to 1.70×10^6 YBP (respectively from 0.864×10^6 YBP to 1.72×10^6 YBP) for the divergence of the BBO and BTA lineages. During the corresponding divergence periods, the estimated number of migrants (per

generation) was found limited with median values of 0.56 and 0.61 for the BBO_3569 and BBO_3574 analyses respectively (Figure 2C). It should also be noticed that the ancestral population sizes estimated under both the I-model (median values of 62,120 and 65,740 for the BBO_3569 and BBO_3574 analyses respectively) and the IM-model (median values of 50,750 and 53,430 for the BBO_3569 and BBO_3574 analyses respectively) were in agreement with those obtained above with the PSMC for the corresponding time period (Figure 1). Finally, as shown in Figure 2D, the comparison of the I- and IM-models of speciation provided a clear support in favor of the IM-model. From these CoalHMM analyses, we thus conclude that the BBO and BTA lineages diverged in the early Pleistocene (from 1.7×10^6 to 850,000 YBP), in agreement with previous studies that relied on alternative time calibrations (Bradley, et al. 1996; Troy, et al. 2001). Moreover, the divergence between these two lineages involved an extended period of limited gene flow similar to the speciation process previously reported in great apes (Mailund, et al. 2012).

However, both the I- and IM models were not designed to capture signals from more recent gene flow events associated to secondary contacts between the bovine and wisent species. To that end, we compared the relative abundance of ABBA and BABA site patterns (Green, et al. 2010; Durand, et al. 2011) defined across the following four taxons: *i)* BBO, *ii)* BTA, *iii) Bos indicus* (BIN), and *iv)* OAR. Assuming a (((BTA;BIN);BBO);OAR) phylogeny (e.g., (Buntjer, et al. 2002; Ho, et al. 2008; Jiang, et al. 2014)), ABBA (respectively BABA) sites are those at which the derived allele ("B") is shared between the nonsister BIN and BBO (respectively BTA and BBO) lineages whereas BTA (respectively BIN) carries the ancestral allele ("A"), as defined by the OAR outgroup. In the absence of (recent) gene flow, both patterns may only result from incomplete lineage sorting and should be equally abundant in the genome while gene flow events between the BTA and BBO lineages would lead to a significant excess of BABA over ABBA sites (Durand, et al. 2011). Whether considering the BBO_3569 or the BBO_3574 sequencing data to define the BBO reference (see Material and Methods), we found a slight but significant excess of "BABA" (n=49,337 and n=58,032 for BBO_3569 and BBO_3574 respectively) over "ABBA" (n=48,004 and n=56,357 for BBO_3569 and BBO_3574 respectively). Accordingly, the *D-statistic* defined as the normalized difference in the counts of ABBA and BABA sites (Green, et al. 2010; Durand, et al. 2011) was found significantly negative for both the BBO_3569 and BBO_3574 derived consensus sequences with values equal to -1.25% (s.d.=0.032% ; Z=39.0) and -1.37% (s.d.=0.031% ; Z=44.4) respectively. The corresponding proportion of BBO ancestry into (European) cattle was found equal to $\hat{f} = 0.143\%$ (s.d.=0.004%) which is actually a

conservative minimum (Green, et al. 2010). Assuming a constant population size, the bias is equal to t_a/t_s=0.28/0.85=0.33 in the worst scenario since t_a represents the timing of the BBO and BTA admixture (<275,000 YBP, the lower bound of the estimated divergence between the taurine and zebu lineages, see above) and t_s the timing of the bovine/wisent speciation (>850,000 YBP, the upper bound found above under the IM-model). Even if the resulting corrected proportion ($\widehat{f^*} \approx 0.4\%$) remains very small, the observed (significant) footprints of BBO ancestry into the genomes of European cattle supports the occurrence of secondary contacts between these two lineages that are posterior to the *Bos taurus* and *Bos indicus* divergence (*ca.* 150,000 YBP, see above). Analyzing non-European (e.g., African) taurine genomes might allow to refine the timing of such secondary contacts by assessing whether or not they still occurred after (taurine) cattle domestication (*ca.*, 10,000 YBP). To that respect also, sequencing individuals (or fossil remains) from extant (*Bison bison*, most particularly) or extinct (e.g., *aurochs* and *Bison priscus*) sister species would also be very informative. Conversely, it should be noticed that genomic data from *Bison* species would also allow to quantify the proportion of bovine ancestry (if any) in the BBO genome.

Identification and functional annotation of genes under positive selection

To identify genes under positive selection between the wisent and bovine lineages, we computed the *Ka/Ks* ratio of nonsynonymous (*Ka*) and synonymous (*Ks*) substitution rates (Kimura 1983) for 17,073 protein sequence alignments between the BBO and BTA genomes. As expected from the effect of purifying selection (Hurst 2002), the average *Ka/Ks* was found equal to 0.273 (ranging from 0.00 to 17.2) with a distribution highly shifted towards 0 (Figure S6). Transcripts with *Ka/Ks*>1 were further considered under positive selection leading to the annotation of 425 genes ready to functional and gene network analyses (see Material and Methods, Table S3). Overall, the most significant functions underlying these genes are related to i) nervous system; ii) immune and inflammatory responses; iii) embryonic and organ development; iv) cellular morphology and organization and; v) skeletal and muscular disorders (Table S4). Figure 3 illustrates the connection of the selected genes that were annotated with their key underlying functions and Table 2 gives a more detailed list of functions and sub-functions. Interestingly, many genes are related to several functions (Table S5), suggesting a pleiotropic role and more strikingly, 85% of the genes participated to a global gene network which is defined by six significant networks connected by up to 11 common molecules (Table S6 and Figure S7). Note that very similar results are obtained

when restricting the functional analysis to the 359 selected genes that are one to one orthologs between cattle and sheep (see Table 2 and Table S4). Thanks to the method we used to define BBO/BTA gene sequence alignments, this suggests that our annotation remains somewhat robust to extra variation that could have been identified when mapping BBO reads in genic regions belonging to large gene families or due to Copy Number Variants in the sequenced BBO individuals.

Among the genes under selection, some underlie obvious distinctive features between wisent and cattle providing in turn insights into the wisent adaptive history (Table 2, Figure 3 and Table S6). First, KRT74 and DCS3 that are involved in wooly hair development and hypotrichosis, respectively (Ayub, et al. 2009; Shimomura, et al. 2010) and GPR50 that plays a role in thermogenesis (Bechtold, et al. 2012) might be directly related to wisent adaptation to colder climatic conditions (Table 2, Table S4). Second, genes encoding olfactory and taste receptors are probably footprints of feeding behavior modifications, resulting *e.g.*, from food resource differences in forest *versus* steppe habitat (trees-shrubs *versus* grass). Third, the many genes related to immune and inflammatory responses (through their key role in activation, migration, binding, expansion and modulation of a broad range of immune cells, Table S5) may sign adaptation of wisent and bovine to different pathogen exposures. Fourth, several genes were found involved in lipid metabolism or mammary gland development (Table S5), likely the result of selection in cattle for improved milk production performances.

More strikingly, the functional analysis of the genes under selection highlighted physiological functions associated to the domestication process (Figure 3, Table 2, Table S6), as expected from the close relatedness of wisent and domestic cattle. These functions underlie both key processes of nervous system (*e.g.*, neurogenesis, neurulation, remodeling of dendrites, differentiation of some neural cells) and by-product phenotypes of domestication (*e.g.*, skeletal and muscular disorders, hair and skin properties, vision, hearing, reproduction). Our analysis thus gives an empirical support to the unified explanation of the "domestication syndrome" in mammals (*i.e.*, the general combination of observed traits in domestic mammals) formulated by Wilkins and collaborators (Wilkins, et al. 2014) following the pioneering work by D.K. Belyaev and L. Trut (Belyaev 1979; Trut, et al. 2009). This explanation accords a central role to the neural crest through its developmental reduction as a result of the primary domestication pressure (*i.e.*, the taming of animals). Furthermore, because neural crest cells are also cellular precursors of many different cells (*e.g.*, osteocytes, chondrocytes, odontocytes and melanocytes), this reduction indirectly produces various secondary phenotypic changes (*e.g.*, skeletal and craniofacial morphological modifications or

coat-colour changes). For instance, such an indirect relationship is well supported by studies on coat coloration in wild and domesticated animals (Cieslak, et al. 2011) and the description coat-colour associated mutations with pleiotropic effects (Reissmann and Ludwig 2013). Also, the numerous genes detected under selection and involved in immune response (Table 2 and Figure 3) might be indirectly related to the domestication process. During this process and afterwards, herding conditions may have incidentally lead to an increased pathogen exposure of domestic cattle due to proximity with individuals from the same or different species (Freeman, et al. 2008).

Conclusions

To characterize the wisent demographic and adaptive histories, we carried out whole genome sequencing of two males from the Białowieża lowland line. Although still considered as a vulnerable species, our results show that the conservation plan and subsequent management practices have been efficient to recover a reasonable amount of genetic variability that now compares to that observed in commercial cattle breeds. We further confirmed at the nucleotide level, the close relatedness of the wisent and cattle species without any evidence for recent gene flow (i.e., during historical times) between these two species. We estimated that the divergence between the bovine and wisent lineages occurred in the early Pleistocene through a speciation process involving limited gene flow, lasting from $1.7x10^{-6}$ to 850,000 YBP. We also found evidence for more recent secondary contacts, posterior to the *Bos taurus* and *Bos indicus* divergence (*ca.* 150,000 YBP), between the wisent and (European) taurine cattle lineages. Interestingly, whole individual genome based demographic inference highlighted contrasting patterns in both species that might be reminiscent of different adaptive strategies (habitat fragmentation versus migration) to survive Pleistocene glaciations. Our results are indeed in agreement with a scenario in which wisent survive in refugee pockets during glaciations (leading to habitat fragmentation) while aurochs migrate southward (and possibly eastward) where climate remained more temperate (Sommer and Nadachowski 2006). It is tempting to speculate that these two alternative strategies might have contributed to the survival of both species to the large mammals Quaternary extinction.Our results also show that wisent and aurochs display a similar trend towards extinction from the Last Glacial Maximum, 20,000 years ago that is concomitant to human population growth (spread of farming and increased hunting pressure), to climate warming and to vegetation changes (e.g., replacement of open-steppe by forests). For the wisent, in particular, the Holocene period was

characterized by a tendency to shift towards forested habitats imposing a significant diet change (Kerley, et al. 2012; Bocherens, et al. 2015). These new constraints left footprints at the genomic level as illustrated by the genes that we found under selection when comparing the wisent and bovine lineages, i.e., genes involved in feeding behavior and in adaptation to temperature conditions and to pathogen exposure. Conversely, wisent being the closest extant wild relative species of domestic cattle, several of the genes under selection could be related to the adaptive response to cattle domestication via their implication in nervous system development and in the expression of by-product domestication phenotypes. Strikingly, this result is in line with unified explanation of the domestication syndrome in mammals formulated by Wilkins *et al.* (2014) giving a shared developmental connection of these diverse traits via neural crest cells.

Materials and Methods

Sample origin.

Genomic DNA of two male wisents, namely BBO_3569 and BBO_3574, was extracted from blood samples collected in 1991 in the Białowieża forest. More precisely, BBO_3569 and BB0_3574 belong to the genetically isolated pure lowland line which originates from seven founders kept in zoo and private breeding centers and used for the species restoration program at the beginning of the 1920's (Tokarska, et al. 2009). In 1952, 40 of their descendants that were born in captivity were re-introduced to the wild in the Polish part of the Białowieża forest, the so-called lowland line now including more than 2,000 individuals (i.e., about half of the world wisent population).

High-throughput sequencing of two wisents.

The Illumina TruSeq DNA sample preparation kit (FC-121-2001, Illumina Inc., San Diego,USA) was used according to the manufacturer's protocol. Libraries were then validated on a DNA1000 chip on a Bioanalyzer (Agilent) to determine size and quantified by qPCR using the Kapa library quantification kit (KAPA) to determine concentration.

The cluster generation process was performed on cBot (Illumina Inc.) by using the Illumina Paired-End DNA sample preparation kit (FC-102-1001, Illumina Inc.). Both BBO individuals were further paired-end sequenced on the HiSeq 2000 (Illumina Inc.) using the SBS (Sequence By Synthesis) technique. Base calling was achieved by using the RTA software (Illumina Inc.). Reads of 100 bp from both sides of the fragments were thus obtained after this step. Quality control of the sequences was checked using the FastQC software.

Mapping wisent sequencing reads onto the bovine genome.

In total, 331,975,598 and 407,585,788 reads paired in sequencing were available for BBO_3569 and BBO_3574 respectively after Illumina Quality Check. After removal of sequencing adapters, these reads were mapped onto the UMD3.1 cattle (*Bos taurus*) reference genome assembly (Liu, et al. 2009) using default options of *aln* and *sampe* programs from the *bwa* (version 0.6.2) software package (Li and Durbin 2009). In total, 92.7% (respectively 92.2%) of the reads from the BBO_3569 (respectively BBO_3574) library were successfully mapped onto the bovine assembly, 94.5% (respectively 93.6%) of which being properly paired. Read alignments with a mapping quality Phred-score MAPQ<20 and PCR duplicates were further removed using the *view* (option *–q 20*) and *rmdup* programs from the *samtools* (version 0.1.19) software package (Li and Durbin 2009). The resulting bam files are available for download from the Sequence Read Archive repository (http://www.ncbi.nlm.nih.gov/sra) under the accession number SRP070526. It should be noted that mapping the wisent reads onto the preliminary Y-chromosome sequence from the BosTau7 assembly (available at http://www.genome.ucsc.edu/) lead to a high proportion of reads improperly paired and only about 12% of the BTAY sequence was covered with an unexpected high read coverage (>60X). We thus decided not to consider the chromosome Y in further analyses.

Construction of individual wisent consensus genome sequences.

For the analyses of BBO/BTA cattle divergence, consensus genome sequences were built for each individual by first generating a *mpileup* file considering each European bison *bam* alignment file separately and using the *mpileup* program from the *samtools* (version 0.1.19) software package (Li and Durbin 2009) run with –C 50 and –d 5000 options. For each BBO individual and at each position (in the bovine reference assembly), the retained consensus base was randomly sampled among the aligned bases after discarding those showing a Base Alignment Quality BAQ>30 (allowing to account for uncertainty resulting from small indels). Such a procedure was aimed at limiting biases towards the bovine reference base at BBO heterozygous sites. Finally, positions with a read depth DP<3 and DP>30 (i.e., three times the average individual read coverage), or for which the retained base was supported by less than 2 reads (to limit sequencing error biases) were treated as N (not called) in the consensus sequences.

Whole genome sequence of four individual bovine genomes.

Four bovine individual whole genome sequences were obtained from the 1000 bull genome projects data (Daetwyler, et al. 2014) stored in the NCBI *sra* archive website

(http://www.ncbi.nlm.nih.gov/sra). More precisely, the four males selected were AAN_0037, HOL_0101, JER_0009 and SIM_0043 and belonged to the Angus, the Holstein, the Jersey and the Simmental European taurine breeds respectively and were sequenced at a roughly similar coverage (8.2X, 9.6X, 11X and 10X respectively) than the European bisons. As for the European bison sequencing data, reads with a MAQ<20 and PCR-duplicates were filtered out from the downloaded *bam* files using the *view* (option *–q 20*) and the *rmdup* programs from the *samtools* (version 0.1.19) software package (Li and Durbin 2009).

Testing for recent gene flow events between the BBO and BTA lineages using the ABBA-BABA statistics.

To count the number of sites with ABBA and BABA patterns across the BTA, BIN, BBO and OAR lineages, we first identified sites displaying a "BA" pattern by comparing the consensus sequences dervived from each individual BBO to the OAR reference genome assembly. To that end and following the same procedure and program options as the ones described above for the mapping of reads onto the UMD3.1 bovine genome assembly (see Results section and Table S1), we mapped the sequencing reads originating from the BBO_3569 and BBO_3574 European bisons onto the *OviAr3* OAR genome assembly (Jiang, et al. 2014). Based on the resulting *mpileup* alignment files, we then defined for each BBO individual a consensus base as described above (section "Construction of individual consensus genome sequences") except that the upper read depth threshold was set to 20 (to account for the lower coverage when mapping reads onto the ovine genome as summarized in see Table S1). The "BA" sites were those at which the consensus base differed from the OAR reference. To further identify whether the "A" or "B" allele was present in the BTA and BIN genomes, the OAR sequences surrounding each of these sites were extracted (60 nt upstream and 60 nt downstream) and aligned onto the BTA UMD3.1 (Liu, et al. 2009) and BIN genome (Canavez, et al. 2012) assemblies using the *blat* (version v35x1) software with default options except for the minimum sequence identity that was set to 95% (Kent 2002). For a given comparison, sequences aligning to more than 10 positions were discarded from further analyses (otherwise, the alignment displaying the highest score was retained). Similarly, sites for which a base different from the "A" and "B" alleles was present in the BIN or BTA sequence or the 2 upstream and downstream flanking nucleotides did not perfectly match were although discarded. In total, 10,035,538 (respectively 11,533,956) "BA" sites could be called in both the BIN and BTA assemblies based on the BBO_3564 (respectively BBO_3574) genomic data. From the observed numbers of "ABBA" and "BABA", we then computed the *D*-

statistics and both their associated standard error and their corresponding Z-score (to assess whether *D* significantly differs from zero) using a Weighted Block Jacknife with 5 Mb blocks (Green et al., 2010; Durand et al., 2011). To further estimate the proportion of BBO ancestry in the BTA genome, we relied on the estimator proposed by Durand et al. (2011) here defined as $\hat{f} = \frac{S(ZEB,BTA,BBO\_3574,OAR)}{S(ZEB,BBO\_3574,BBO\_3569,OAR)}$, where S(P1,P2,P3,P4) represents the difference of the "ABBA" and "BABA" sites numbers assuming a ((P1;P2);P3);P4) phylogeny and estimated following a procedure similar to the one described above. As for the *D-statistic* estimates, the standard deviation of $\hat{f}$ was computed using a Weighted Block Jacknife with 5 Mb blocks.

Estimation of genetic heterozygosity from individual genomes.

Genetic heterozygosities were estimated from the individual genome alignments onto the UMD 3.1 cattle assembly (i.e., based on the *mpileup* files described above) using *mlrho* version 2.8 (Haubold, et al. 2010). Note that *mlrho* actually implements a maximum likelihood estimator of the population mutation rate ($\theta = 4N_e\mu_g$), which fairly approximates heterozygosity under an infinite sites model (and providing θ is small), while simultaneously estimating sequencing error rates and accounting for binomial sampling of parental alleles (Lynch 2008). For a given individual genome alignment, only sites covered by 3 to 30 reads (after discarding bases with a BAQ<25) were retained in the computation (Table S2).

Population size history inference using the psmc.

Effective population sizes in units of $4N_e\mu_g$ were estimated backward in time (in units of $2\mu_y T$) based on individual whole genome sequences under the PSMC model using the program *psmc* (version 0.6.4) program (Li and Durbin 2011). Briefly, the analyzed individual whole genome sequences were obtained from the above described *mpileup* files (aligned onto the UMD3.1 cattle assembly) that were processed (individually) with *bcftools* (version 0.1.19) (Li and Durbin 2009) using the *–c* option. The resulting *vcf* files were further converted into *fastq* files using the *vcf2fq* utility of the *vcfutils.pl* program available in the *samtools* (version 0.1.19) suite (Li and Durbin 2009) discarding sites located less than 5 nucleotides from an indel (*–l 5* option) and covered by less than 3 or more than 30 reads (*-d 3 -D 30* options). As originally described, the *psmc* program was finally run with default options except for the pattern of atomic time intervals (*-p* option) that were set to "4+25*2+4+6" on a reduced version of the (autosomal) genomes (*psmcfa* files), with consecutive sites grouped into consecutive bins of 100 nucleotides marked as "K" (at least one heterozygote), "T" (no heterozygote sites) or "N" (less than 90 called sites).

Characterization of the divergence between BBO and BTA lineages under the coalHMM framework.

For each BBO individual consensus genome sequence, genome alignment onto the UMD 3.1 bovine reference assembly (see the section "Construction of individual consensus genome sequences" above) was divided into 10 Mbp non overlapping segments. Segments mapping to the (bovine) X-chromosome or with more than 500,000 (5 %) missing information (i.e., an uncalled base in either the BTA or BBO sequence) were further discarded leading to a total of 123 and 177 available segments for the BBO_3569 and BBO_3574 alignments respectively. Maximum likelihood inference was then carried out independently for each segment using the scripts *isolation-model.py* (version 1.1) for the I-model, and the *initial-migration-model.py* (version 1.2) for the IM-model, from the *IMCoalHMM* software package (https://github.com/mailund/IMCoalHMM). More precisely, the parameters underlying the I-model include the split time τ_D (in units of $\mu_y T$) and also the ancestral effective population size (in scaled units of $\theta = 4N_e\mu_g$) and the recombination rate ρ (in scaled units of μ_g) that both parameterize the underlying coalescent process. The parameters underlying the IM-model include the times τ_1 (completion of the split) and the IM period length τ_M that both are units of $\mu_y T$, the migration rate M (in number of migration per substitution) and θ and ρ (as for the I-model). Note that the product $M \times (\theta/2)$ corresponds to the estimated number of migrant per generation (in the usual units of $2N_e m$ where m is the net migration rate) over the IM period. Finally, for model comparison purposes, the Akaike Information Criterion (AIC) was computed as $AIC_I=2x(3-l)$ for the I-model and $AIC_{IM}=2x(5-l)$ for the IM-model where l represents the estimated log-likelihood of the corresponding model.

Detection of genes under selection.

Based on the alignment of the BBO_3569 and BBO_3574 individual sequences onto the bovine UMD 3.1 reference genome assembly (*mpileup* file described above), a consensus BBO sequence was derived for all the n=22,091 (bovine) ENSEMBL protein-coding sequences as defined in the UCSC genome database (http://genome.ucsc.edu/). At a given position, the retained consensus base corresponded to the most represented one among all the aligned reads from both the BBO_3569 and BBO_3574 individuals and after discarding bases with a BAQ<30. Positions with a read depth DP<3 and DP>60 (three times the average coverage of BTA genome assembly by the combined BBO_3569 and BBO_3574 read sequencing data) were treated as N (not called) in the consensus sequences. BBO consensus protein sequences with more than 10% "N" were discarded from further analysis leading to a

total of n=19,372 remaining BBO/BTA alignments (with 1.11% "N" on average). For 1,306 (6.74%) of these, no nucleotide differences could be observed between the obtained BBO consensus sequence and the BTA one. Nonsynonymous (Ka) and synonymous (Ks) substitution rates were then computed using *KaKs_calculator v1.2* (Zhang, et al. 2006) assuming a MLWL model to estimate the number of synonymous and nonsynonymous sites (Zhang, et al. 2006) allowing the estimation of the *Ka/Ks* ratio for a total of 17,073 protein coding sequences.

Functional annotation of candidate genes under positive selection and gene network analysis.

Among the 17,073 Ensembl transcripts ID corresponding to the protein coding sequences for which a *KaKs* ratio was estimated, 873 transcripts with a *KaKs* ratio above one were considered under positive selection.

Functional annotation of these transcripts and networks analyses were carried out with the Ingenuity Pathway Analysis software (IPA, Ingenuity®Systems, www.ingenuity.com). Among the 873 transcripts with a *KaKs* ratio above one, 481 transcript ID were mapped to the Ingenuity Pathway Knowledge Base (IPKB) and were representative of 425 individual genes ready for analysis (Table S3). A total of 405 of these transcripts were further identified as one to one orthologs between cattle and sheep, using Ensembl orthology information (http://www.ensembl.org/biomart/). The top significant functions and diseases (p-value<0.05) were obtained by comparing functions associated with the 425 (among which 359 were one to one orthologs between cattle and sheep) genes against functions associated with all genes in our reference set (14,465 transcripts mapped in IPKB from the 17,073 transcripts ID that were analyzed) using the right-tailed Fisher's exact test. A total of 262 genes (among which 151 were one to one orthologs between cattle and sheep) participated to the significant functions thus determined (Table S5).

Among the 425 genes, 367 genes (among which 324 were one to one orthologs between cattle and sheep) were included in network analyses. For each network that contains at most 140 molecules, a score S was computed based on a right-tailed Fisher exact test for the overrepresentation of the genes with a *KaKs* ratio>1 (S=-log(p-value)). A network was considered as significant when S>3.

Acknowledgements and funding information

573 The authors wish to thank Anna Madeyska-Lewandowska (Private Veterinary School,
574 Warsaw, Poland) that provided the *Bison bonasus* samples in the early 1990's as part of a
575 collaborative research project on the characterization of casein variants. We are also grateful
576 to the four anonymous reviewers for their very helpful and constructive comments. This work
577 was supported by an INRA (Institut National de La Recherche Agronomique)-Animal
578 Genetics Department Grant (AFROSEQ project).

579 Sequencing data (bam files) are available for download from the Sequence Read Archive
580 repository (http://www.ncbi.nlm.nih.gov/sra) under the accession number SRP070526.

581

TABLES

Table 1: Read mapping statistics from the alignment of each BBO_3569 and BBO_3574 European bison genome sequences onto the UMD 3.1 cattle genome assembly (Liu, et al. 2009).

Chromosome Type (BTA)	Size (in bp)	Coverage (% of sequence covered)*		Average nucleotide divergence in % (number of sites compared)**	
		BBO_3569	BBO_3574	BBO_3569	BBO_3574
Autosomes	2,512,082,506	9.81 (95.5)	11.6 (95.7)	0.870 ($2.27x10^9$)	0.882 ($2.29x10^9$)
X-chromosome (including PAR)	148,823,899	5.57 (90.9)	6.48 (91.2)	0.636 ($1.11x10^8$)	0.642 ($1.17x10^8$)
Mitochondria	16,338	397 (97.7)	302 (97.1)	3.58 ($1.18x10^4$)	3.41 ($1.20x10^4$)

*For each individual, the average read coverages (after alignment) over autosomes, the X-chromosome and the mitochondria are given together with the overall percentages of sites from the corresponding reference sequences covered with at least one read.

** For each type of chromosomes, the average nucleotide divergence between the cattle reference genome and each European bison individual consensus sequence is given (see Material and Methods).

596 **Table 2: Main genes under selection listed by key functions and by presumed**

597 **adaptation.**

Key Function	Color in Figure 3	Sub-function	Genes	Presumed adaptation
Hair development	grey	Growth of hair follicules	CDKN2A	Adaptation to temperature conditions (cold/temperate)
		Wooly hair	KRT74*	
		Hypotrichosis	DSC3*	
Thermogenesis	grey		GPR50*	
Gustation	white		PKD2L1*, TAS1R1*, TAS2R16*, TAS2R46*	Adaptation to available food resources and vegetation diversity (Forest/steppe habitat)
Nervous system	purple	Olfaction	ADH7, CNGB1*, Olfr1178*, Olfr1179, Olfr1231*, Olfr1280, Olfr1353, Olfr1358*, Olfr1535, Olfr424*, Olfr49, Olfr541, Olfr600*, Olfr606, Olfr610*, Olfr711, Olfr867*, Olfr905, Olfr922*, Olfr963*, OR10AG1, OR10J5*, OR10V1*, OR12D3*, OR13F1, OR1E2*, OR1J1*, OR1M1*, OR2AG2*, OR2D2*, OR2M5, OR4C12, OR4C46*, OR4F15*, OR4F6, OR51A7*, OR52E8*, OR52H1, OR52R1, OR5C1, OR5M11*, OR6K2*, OR6N2*, OR6Y1*, OR7G3, OR8A1, OR9K2	
		Neurogenesis, neurulation	CD44*, CD9*, CDKN2A, GZMB*, NEIL3*, OR8A, SIX1*, SP4*, TTLL8*, VASP*	Wildlife/Domestication
Hearing	blue		CLIC5*, FOXL1, SIX1*, SYNE4*	Wildlife/Domestication
Vision	blue		CNGB1*, GPNMB, LCA5*, MYOC*, OCA2*, TMEM5	Wildlife/Domestication
Pigmentation	blue		OCA2*	Wildlife/Domestication
Sketetal and muscular development	blue		CDKN2A, CKAP2L*, FAM111B*, FOXI1*, KLF5, NCOA1*, NDUFS6, PIGV*, PLG*, RNF135*, SLC17A3*, WISP3*	Wildlife/Domestication
Reproduction	orange		ACR, AGER*, CD44*, CD48, CD55*, CD9*, CDKN2A, FETUB*, FOXI1*, NCOA1*, PER2*, PLG*, SFTPC*, SPAM1*, ZP2*	Wildlife/Domestication
Immune and inflammatory responses	pink		ADAM8, AGER*, ANGPTL3*, APOBEC3B*, CCL16*, CCL24*, CCL5*, CD180*, CD1E*, CD244*, CD4*, CD44*, CD48, CD55*, CD72*, CD9*, CDKN2A, CFH, CKAP2L*, CXCL16*, EVI2A*, FCAR*, FCRL1*, FCRL3*, HBD*, HLA-B, HLA-DMB*, ICAM1*, IFI44*, IFNA16, IFNAR1*, KLRC1, KLRD1*, Klrk1*, LAG3*, MAVS*, OR12D3*, PGLYRP2*, PIGV*, PLG*, PPP1R15A*, PRSS16*, RHBDD3*, RTP4, SFTPD, SIRPA, SLC17A3*, SLC39A4*, SPN*, TF*, THPO*, TMPRSS11D*, TNFSF9*, TRIM40*, ULBP3, VASP*, XAF1*	Resistance/Tolerance to pathogens (linked to wildlidfe/domestication)
Metabolism	yellow	Lipid metabolism	ACADL*, ANGPTL3*, APOF*, BCO2*, CCL5*, CD4*, CD9*, CIDEC*, CYP4B1*, GHRL*, GPLD1*, HSD17B3*, MOGAT3, PLIN2*, RDH16*	Adaptation to diet Artificial selection on dairy traits
Embryonic and organ development	green	Mammary gland development	CD44*, CDN2A PLG*,	Artificial selection on dairy traits
		Other	ACR, AGER*, CD44*, CD9*, CDKN2A, CKAP2L*, CLDN15*, FCAR*, FETUB*, FOXI1*, ICAM1*, IFNAR1*, KLF5, MAVS*, NCOA1*, ODC1, PLG*, POU2F1*, RNF135*, SFTPD, SIX1*, SPAM1*, TF*, THPO*, VASP*, VGLL1*, ZP2*	

598 *One to one orthologs between cattle and sheep

FIGURE LEGENDS

Figure 1. Population size histories inferred from the two wisent and the four bovine genomes under the PSMC model.

Backward in time (in YBP) estimates of the effective population sizes derived from the *psmc* analyses of the BBO_3569 and BBO_3574 individual wisent genomes and the AAN_0037, HOL_0101, JER_0009 and SIM_0043 bull bovine genomes and assuming $\mu_y = 2.10 \times 10^{-9}$ (Liu, et al. 2006) and g=6 (Keightley and Eyre-Walker 2000; Gautier, et al. 2007). At the bottom of the figure, the timing of cattle domestication and the zebu (BIN) and taurine (BTA) divergence (Ho, et al. 2008) are indicated by a dark red triangle while the timing interval of the bovine and wisent divergence (see the main text) is indicated by dark red arrows. Similarly, European ice ages (Würm, Riss, Mindel and Günz glacial stages) according to the Penck and Bruckner work (Penck and Brückner 1901-1909) cited in (Elias 2013) are indicated by blue intervals at the top of the figure. Finally, the Last Glacial Maximum (LGM), ca. 20,000 YBP (Clark, et al. 2009), is indicated by a vertical blue dotted line.

Figure 2. Characterization of the divergence between the BBO and BTA lineages under the isolation and the isolation-with-migration models.

A. Estimation of the time τ_D of divergence between the BBO and BTA lineages under the Isolation model (that assumes a clean split). The two violin plots show the distribution of τ_D estimates obtained for each 10 Mbp non-overlapping segments from the alignments of the BBO_3569 (n=123 segments) and BBO_3574 (n=177 segments) whole genome sequence onto the UMD 3.1 bovine genome assembly. Time scale was translated into YBP assuming $\mu_y = 2.10 \times 10^{-9}$ (Liu, et al. 2006).

B. Estimation of starting (τ_2) and ending (τ_1) time estimates of the divergence between the BBO and BTA lineages under the Isolation-with-Migration model (that assumes the two ancestral lineages exchange migrants between τ_2 and τ_1). The violin plots show the distribution of τ_2 and τ_1 time estimates obtained for each 10 Mbp non-overlapping segments from the alignments of the BBO_3569 (n=123 segments) and BBO_3574 (n=177 segments) whole genome sequence onto the UMD 3.1 bovine genome assembly. Time scale was translated into YBP assuming $\mu_y = 2.10 \times 10^{-9}$ (Liu, et al. 2006).

C. Estimation of the average number of migrant per generation ($2N_e m$) during the divergence of the BBO and BTA lineages under the Isolation-with-Migration model. The two violin plots

show the distribution of the migration rate estimates obtained for each 10 Mbp non-overlapping segments from the alignments of the BBO_3569 (n=123 segments) and BBO_3574 (n=177 segments) whole genome sequence onto the UMD 3.1 bovine genome assembly.

D. Model comparisons between the isolation and the Isolation-with-Migration models. The two violin plots show the distribution of the difference between the Akaike Information Criterion for the isolation (AIC_I) and for the isolation-with-migration (AIC_{IM}) obtained for each 10 Mbp non-overlapping segments from the alignments of the BBO_3569 (n=123 segments) and BBO_3574 (n=177 segments) whole genome sequence onto the UMD 3.1 bovine genome assembly. Because the model with the smallest *AIC* should be preferred, the distributions provide strong support in favor of the IM-model (AIC_{IM} is always lower than AIC_I).

Figure 3. Representation of the genes with a Ka/Ks >1 connected to their key functions. Each global function and its links to the corresponding key genes are differently colored i.e. in purple for nervous system, blue for by-products of domestication, green for functions related to embryonic, organ development, organismal abnormality and hereditary disorders, pink for immune and inflammatory responses, orange for reproduction, grey for hair development and thermogenesis, yellow for lipid metabolism and white for gustation. Global functions contain genes belonging to several Ingenuity Pathway Analysis functional categories.
Gene symbols are colored in red and color intensity is correlated to Ka/Ks value.

SUPPLEMENTARY TABLE LEGENDS

Table S1: Mapping statistics of the European bison high throughput sequencing reads onto the UMD3.1 bovine reference genome and *OviAr3* ovine reference genome assemblies.

Table S2: Maximum likelihood estimates of the (autosomal) heterozygosities ($\hat{\theta}$) and sequencing error rates ($\hat{\varepsilon}$) for the BBO_3569 and BBO_3574 individual European bison genomes and the AAN_0037, JER_0009, HOL_0101 and SIM_0043 individual genomes belonging to the Angus, Jersey, Holstein and Simmental European cattle breeds respectively. The table also gives the number of sites available in the estimation procedure (i.e., covered by 10 to 100 read nucleotides after discarding bases with a BAQ<25)

Table S3: List of the 481 Ensembl transcript ID with a Ka/Ks >1 (from 1.003 to 8.289) mapped in the Ingenuity Pathway Knowledge Base and of their corresponding gene symbol and Entrez gene name. One to one orthologs between cattle and sheep are indicated.

Table S4: List of the top five significant diseases and biological functions in each main functional categories obtained using Ingenuity Pathway Analysis for the 481 (405) transcripts with a Ka/Ks >1 (one to one orthologs between bovine and sheep).

Table S5: List of the significant functional categories with their corresponding functional annotation, p-value and annotated molecules.

Table S6: List of the significant networks including genes with a Ka/Ks >1, obtained with the Ingenuity Pathway Analysis software.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Regional nucleotide divergence between the cattle genome and the BBO_3569 (A) and BBO_3574 (B) European bison genome sequences over contiguous 10-Mb segments covering the 29 bovine autosomes and the X-chromosome.

Figure S2: Distribution of the heterozygosity estimates across the 29 bovine autosomes based on the BBO_3569 and BBO_3574 individual European bison genomes and for the AAN_0037, JER_0009, HOL_0101 and SIM_0043 individual genomes belonging to the Angus, Jersey, Holstein and Simmental European cattle breeds respectively.

Figure S3: Backward in time (in units of $2\mu_y T$) estimates of the (scaled) effective population sizes (in units of $4N_e\mu_g$) from the analyses of the BBO_3569 and BBO_3574 individual wisent genomes and the AAN_0037, HOL_0101, JER_0009 and SIM_0043 bull bovine genomes

Figure S4: Boostrap confidence intervals of the population size histories inferred from the two wisent and the four bovine genomes under the PSMC model. For each individual genome, the dark red curve represents the direct estimation (as in Figure 1A)

while the 100 orange curves represent estimated history obtained on bootstrap genome samples (Li and Durbin 2011).

Figure S5: Comparison of the population histories inferred by the PSMC algorithm from the two wisent genomes derived by either aligning the underlying sequencing reads onto the UMD 3.1 cattle genome assembly (in black) or the consensus BBO reference assembly (in red).

Figure S6: Distribution of the Ka/Ks estimates obtained for all the BBO/BTA protein sequence alignments.

Figure S7. List of genes with a Ka/Ks ratio>1 participating to the six overlapping significant networks.

Gene symbols are colored in red and color intensity is correlated to Ka/Ks value.

References

Ayub M, Basit S, Jelani M, Ur Rehman F, Iqbal M, Yasinzai M, Ahmad W. 2009. A homozygous nonsense mutation in the human desmocollin-3 (DSC3) gene underlies hereditary hypotrichosis and recurrent skin vesicles. *Am J Hum Genet* 85:515-520.

Bechtold DA, Sidibe A, Saer BR, Li J, Hand LE, Ivanova EA, Darras VM, Dam J, Jockers R, Luckman SM, et al. 2012. A role for the melatonin-related receptor GPR50 in leptin signaling, adaptive thermogenesis, and torpor. *Curr Biol* 22:70-77.

Belyaev DK. 1979. The Wilhelmine E. Key 1978 invitational lecture. Destabilizing selection as a factor in domestication. *J Hered* 70:301-308.

Benecke N. 2005. The Holocene distribution of European bison-the archaeozoological record. *Munibe (Antropologia-Arkeologia)* 57:421-428.

Bocherens H, Hofman-Kamińska E, Drucker DG, Schmolcke 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.

Bradley DG, MacHugh DE, Cunningham P, Loftus RT. 1996. Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci U S A* 93:5131-5135.

Buntjer JB, Otsen M, Nijman IJ, Kuiper MT, Lenstra JA. 2002. Phylogeny of bovine species based on AFLP fingerprinting. *Heredity (Edinb)* 88:46-51.

731 Canavez FC, Luche DD, Stothard P, Leite KR, Sousa-Canavez JM, Plastow G, Meidanis J, Souza
732 MA, Feijao P, Moore SS, et al. 2012. Genome sequence and assembly of Bos indicus. *J Hered*
733 103:342-348.
734 Cieslak M, Reissmann M, Hofreiter M, Ludwig A. 2011. Colours of domestication. *Biol Rev Camb*
735 *Philos Soc* 86:885-899.
736 Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW,
737 McCabe AM. 2009. The Last Glacial Maximum. *Science* 325:710-714.
738 Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brondum RF, Liao X, Djari A,
739 Rodriguez SC, Grohs C, et al. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of
740 monogenic and complex traits in cattle. *Nat Genet* 46:858-865.
741 Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely
742 related populations. *Mol Biol Evol* 28:2239-2252.
743 Elias SA. 2013. History of Quaternary science. . In: Elias SA, editor. Encyclopedia of Quartenary
744 Science, second edition. Amsterdam: Elsevier. p. 10-18.
745 Freeman AR, Lynn DJ, Murray C, Bradley DG. 2008. Detecting the effects of selection at the
746 population level in six bovine immune genes. *BMC Genet* 9:62.
747 Gautier M, Faraut T, Moazami-Goudarzi K, Navratil V, Foglio M, Grohs C, Boland A, Garnier JG,
748 Boichard D, Lathrop GM, et al. 2007. Genetic and haplotypic structure in 14 European and African
749 cattle breeds. *Genetics* 177:1059-1070.
750 Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz
751 MH, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:710-722.
752 Hassanin A, An J, Ropiquet A, Nguyen TT, Couloux A. 2013. Combining multiple autosomal introns
753 for studying shallow phylogeny and taxonomy of Laurasiatherian mammals: Application to the tribe
754 Bovini (Cetartiodactyla, Bovidae). *Mol Phylogenet Evol* 66:766-775.
755 Haubold B, Pfaffelhuber P, Lynch M. 2010. mlRho - a program for estimating the population mutation
756 and recombination rates from shotgun-sequenced diploid genomes. *Mol Ecol* 19 Suppl 1:277-284.
757 Ho SY, Larson G, Edwards CJ, Heupink TH, Lakin KE, Holland PW, Shapiro B. 2008. Correlating
758 Bayesian date estimates with climatic events and domestication using a bovine case study. *Biol Lett*
759 4:370-374.
760 Hobolth A, Christensen OF, Mailund T, Schierup MH. 2007. Genomic relationships and speciation
761 times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS*
762 *Genet* 3:e7.
763 Hofreiter M, Stewart J. 2009. Ecological change, range fluctuations and population dynamics during
764 the Pleistocene. *Curr Biol* 19:R584-594.
765 Hurst LD. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* 18:486.

766 Jakob W, Schröder HD, Rudolph M, Krasiński ZA, Krasińska M, Wolf O, Lange A, J.E. C, Frölich K.
767 2000. Necrobacillosis in free-living male European Bison in Poland. *Journal of Wildlife Diseases*
768 36:248-256.
769 Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W, et al.
770 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344:1168-
771 1173.
772 Keightley PD, Eyre-Walker A. 2000. Deleterious mutations and the evolution of sex. *Science* 290:331-
773 333.
774 Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome Res* 12:656-664.
775 Kerley GIH, Kowalczyk R, Cromsigt PGM. 2012. Conservation implications of the refugee species
776 concept and the European bison: king of the forest or refugee in a marginal habitat? *Ecography*
777 35:519-529.
778 Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.
779 Kowalczyk R, Taberlet P, Coissac E, Valentini A, Miquel C, Kamiński T, Wojcik JM. 2011. Influence
780 of management practices on large herbivore diet-Case of European bison in Białowieża Primeval
781 Forest (Poland). *Forest Ecology and Management* 261:821-828.
782 Krasińska M, Krasiński ZA. 2002. Body mass and measurements of the European bison during
783 postnatal development. *Acta Theriologica* 47:85-106.
784 Kuemmerle T, Hickler T, Olofsson J, Schurgers G, Radeloff V. 2012. Reconstructing range dynamics
785 and range fragmentation of European bison for the last 8000 years. *Diversity and Distributions* 18:47-
786 59.
787 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
788 *Bioinformatics* 25:1754-1760.
789 Li H, Durbin R. 2011. Inference of human population history from individual whole-genome
790 sequences. *Nature* 475:493-496.
791 Li WH, Yi S, Makova K. 2002. Male-driven evolution. *Curr Opin Genet Dev* 12:650-656.
792 Lister AM, A.J. S. 2008. The impact of climate change on large mammal distributionand extinction:
793 Evidence from the last glacial/interglacial transition. *C.R. Geoscience* 340:615-620.
794 Liu GE, Matukumalli LK, Sonstegard TS, Shade LL, Van Tassell CP. 2006. Genomic divergences
795 among cattle, dog and human estimated from large-scale alignments of genomic sequences. *BMC*
796 *Genomics* 7:140.
797 Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y, et al.
798 2009. Bos taurus genome assembly. *BMC Genomics* 10:180.
799 Lorenzen ED, Nogues-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, Ugan A,
800 Borregaard MK, Gilbert MT, Nielsen R, et al. 2011. Species-specific responses of Late Quaternary
801 megafauna to climate and humans. *Nature* 479:359-364.

802 Lynch M. 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates
803 from high-coverage genome-sequencing projects. *Mol Biol Evol* 25:2409-2419.
804 Macleod IM, Larkin DM, Lewin HA, Hayes BJ, Goddard ME. 2013. Inferring demography from runs
805 of homozygosity in whole-genome sequence, with correction for sequence errors. *Mol Biol Evol*
806 30:2209-2223.
807 Mailund T, Dutheil JY, Hobolth A, Lunter G, Schierup MH. 2011. Estimating divergence time and
808 ancestral effective population size of Bornean and Sumatran orangutan subspecies using a coalescent
809 hidden Markov model. *PLoS Genet* 7:e1001319.
810 Mailund T, Halager AE, Westergaard M, Dutheil JY, Munch K, Andersen LN, Lunter G, Prufer K,
811 Scally A, Hobolth A, et al. 2012. A new isolation with migration model along complete genomes
812 infers very different divergence processes among closely related great ape species. *PLoS Genet*
813 8:e1003125.
814 Mazet O, Rodriguez W, Grusea S, Boitard S, Chikhi L. 2015. On the importance of being structured:
815 instantaneous coalescence rates and human evolution-lessons for ancestral population size inference?
816 *Heredity (Edinb)*.
817 Merceron G, Hofman-Kamińska E, Kowalczyk R. 2014. 3D dental microwear texture analysis of
818 feeding habits of sympatric ruminants in the Białowieża Primeval Forest, Poland. *Forest Ecology and*
819 *Management* 328:262-269.
820 Mikkelsen TS, Hillier LH, Eichler EE, Zody MC, Jaffe DB, Yang SP, Enard W, Hellmann I,
821 Lindblad-Toh K, Altheide TK, et al. 2005. Initial sequence of the chimpanzee genome and comparison
822 with the human genome. *Nature* 437:69-87.
823 Mona S, Catalano G, Lari M, Larson G, Boscato P, Casoli A, Sineo L, Di Patti C, Pecchioli E,
824 Caramelli D, et al. 2010. Population dynamic of the extinct European aurochs: genetic evidence of a
825 north-south differentiation pattern and no evidence of post-glacial expansion. *BMC Evol Biol* 10:83.
826 Nevado B, Ramos-Onsins SE, Perez-Enciso M. 2014. Resequencing studies of nonmodel organisms
827 using closely related reference genomes: optimal experimental designs and bioinformatics approaches
828 for population genomics. *Mol Ecol* 23:1764-1779.
829 Penck A, Brückner E. 1901-1909. Die Alpen im Eiszeitalter. Leipzig: Tauchnitz.
830 Pucek Z. 2004. European Bison. In: Pucek Z, Belousova Z, Krasińska M, Krasiński ZA, Olech W,
831 Group ISBS, editors. Status Survey and Conservation Action Plan. Gland, Switzerland and
832 Cambridge, UK: IUCN-The World Conservation Union. p. 54.
833 Raczyński J. 2015. European bison pedigree book 2014: Białowieża National Park.
834 Reissmann M, Ludwig A. 2013. Pleiotropic effects of coat colour-associated mutations in humans,
835 mice and other mammals. *Semin Cell Dev Biol* 24:576-586.
836 Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MT,
837 Barnes I, Binladen J, et al. 2004. Rise and fall of the Beringian steppe bison. *Science* 306:1561-1565.

838 Shimomura Y, Wajid M, Petukhova L, Kurban M, Christiano AM. 2010. Autosomal-dominant woolly
839 hair resulting from disruption of keratin 74 (KRT74), a potential determinant of human hair texture.
840 *Am J Hum Genet* 86:632-638.
841 Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt HJ, Torroni A, Richards MB. 2010.
842 The archaeogenetics of Europe. *Curr Biol* 20:R174-183.
843 Sommer RS, Nadachowski A. 2006. Glacial refugia of mammals in Europe:evidence from fossil
844 records. *Mammal rev.* 36:251-265.
845 Tokarska M, Marshall T, Kowalczyk R, Wojcik JM, Pertoldi C, Kristensen TN, Loeschcke V,
846 Gregersen VR, Bendixen C. 2009. Effectiveness of microsatellite and SNP markers for parentage and
847 identity analysis in species with low genetic diversity: the case of European bison. *Heredity (Edinb)*
848 103:326-332.
849 Tokarska M, Pertoldi C, Kowalczyk R, Perzanowski K. 2011. Genetic status of the European bison
850 Bison bonasus after extinction in the wild and subsequent recovery. *Mammal Review* 41:151-162.
851 Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain AT, Sykes
852 BC, Bradley DG. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*
853 410:1088-1091.
854 Trut L, Oskina I, Kharlamova A. 2009. Animal evolution during domestication: the domesticated fox
855 as a model. *Bioessays* 31:349-360.
856 Verkaar EL, Nijman IJ, Beeke M, Hanekamp E, Lenstra JA. 2004. Maternal and paternal lineages in
857 cross-breeding bovine species. Has wisent a hybrid origin? *Mol Biol Evol* 21:1165-1170.
858 Wilkins AS, Wrangham RW, Fitch WT. 2014. The "domestication syndrome" in mammals: a unified
859 explanation based on neural crest cell behavior and genetics. *Genetics* 197:795-808.
860 Wójcik JM, Kawałko A, Tokarska M, Jaarola M, Vallenback P, Pertoldi C. 2009. Post-bottleneck
861 mtDNA diversity in a free-living population of European bison: implication for conservation. *Journal*
862 *of Zoology* 277:81-87.
863 Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. 2006. KaKs_Calculator: calculating Ka and Ks
864 through model selection and model averaging. *Genomics Proteomics Bioinformatics* 4:259-263.
865
866

867 # FIGURES

868

869 **Figure 1**

870 **Figure 2**

871 **Figure 3**

Population Size History in real units (assuming μ_y= 2.1 x10^{-9} and g=6)