**Reviewed Preprint**
**v1** • December 4, 2024
Not revised

**Reviewed Preprint**
**v2** • June 10, 2026
Revised by authors

# Most Beefalo cattle have no detectable bison genetic ancestry

**Beth Shapiro**[1,2] ✉, **Jonas Oppenheimer**[3], **Michael P Heaton**[4], **Kristen L Kuhn**[4], **Richard E Green**[3], **Harvey D Blackburn**[5], **Timothy PL Smith**[4]

[1]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, United States • [2]Howard Hughes Medical Institute, University of California, Santa Cruz, Santa Cruz, United States • [3]Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, United States • [4]USDA, ARS, U.S. Meat Animal Research Center, Clay Center, United States • [5]USDA, ARS, National Animal Germplasm Program, Fort Collins, United States

✉ **For correspondence:**
bashapir@ucsc.edu

**Competing interests:** No competing interests declared

**Funding:** See page 20

**Reviewing editor:** David A. Paz-Garcia, Centro de Investigaciones Biológicas del Noroeste (CIBNOR), Mexico

## eLife Assessment

This **important** study used whole-genome data to investigate Beefalo ancestry for the first time, providing insight into the genetics of Beefalo cattle and challenging the long-held claim of 37.5% bison ancestry reported by the American Beefalo Association. Despite some limitations regarding sequencing depth and sampling, the expert use of a comprehensive set of population-genomic methods allowed the authors to demonstrate **convincingly** that Beefalo and bison hybrid ancestry profiles are consistent with repeated backcrossing to either parental species. The work will be of significant interest to evolutionary biologists, population geneticists, animal breeders, and those involved in the conservation genetics of bovine species.

https://doi.org/10.7554/eLife.102750.2.sa4

## Abstract

Hybridization is common among lineages in the genus *Bos*, often mediated through human management for the selection of adaptive or desirable traits. A recent example is the American Beefalo cattle breed, which was developed in the 1970s and defined as a hybrid between American bison (*Bison bison*) and cattle (*Bos taurus*). The American Beefalo Association (ABA) typically require ⅜ bison ancestry to qualify as Beefalo. Here, we sought to characterize admixed ancestry among Beefalo as a component of a larger project to understand the role of hybridization in shaping present-day diversity in bison and cattle. We generated genomic data from 50 historical and present-day Beefalo and bison hybrids, including several important founding animals, as well as from 10 bison originating from commercial herds that represent potential sources of bison ancestry in Beefalo. We found that most Beefalo did not contain detectable bison ancestry. No individual Beefalo within our data set satisfies the ancestry requirements specified by the ABA, although several Beefalo had smaller proportions of bison ancestry (2-18%). Some beefalo had detectable zebu cattle ancestry (2-38%), suggesting that hybridization of taurine and zebu cattle may contribute to morphological similarity between some Beefalo and bison. Overall, ancestry profiles of Beefalo and bison hybrid genomes are consistent with repeated backcrossing to either parental species rather than the breeding between hybrids themselves, implying significant barriers to gene flow between bison and cattle. Our results call into question the ⅜ bison ancestry targeted by the breed association and demonstrate the value of genomic information in examining claims of interspecies gene flow among *Bos* species.

# Introduction

Gene flow has been common among lineages within the genus *Bos*, including among cattle (*Bos taurus*), bison (*Bison bison*), yak (*Bos mutus*), and gaur (*Bos gaurus*) (Wu et al. 2018; Zhang et al. 2020). Much of this gene flow was facilitated by human livestock management and breeding. For example, hybridization led to gene flow between taurine (*Bos taurus*) and zebu (*Bos indicus*) cattle breeds in the Near East during the Bronze Age (Verdugo et al. 2019), and later in Africa (Kim et al. 2020) and North America (McTavish et al. 2013). Human-mediated gene flow also occurred between more deeply diverged *Bos* lineages, including Chinese cattle and banteng (Chen et al. 2018) and yak and Tibetan cattle (Wu et al. 2018). In each case, gene flow is believed to have conferred genetic benefits that led to easier management of hybrid herds or enabled adaptation to local environmental conditions (Zhang et al. 2020).

While admixture is common among *Bos* lineages, genomic incompatibilities often induce reproductive challenges to the hybrids. These incompatibilities are probably due to the considerable divergences across the *Bos* clade, with the oldest split between cattle and the other lineages dating to around three million years ago (Wang et al. 2018). Except for hybrids of taurine and zebu cattle, hybrids of all other *Bos* species follow Haldane's rule, in which male offspring of F1 crosses are sterile (Zhang et al. 2020). While taurine-zebu cattle hybrids are fertile, hybrid African cattle have uneven proportions of zebu nuclear and mitochondrial ancestry (Kwon et al. 2022; Ward et al. 2022), which may be due to mitonuclear incompatibilities. Analysis of the dynamics of speciation and recurrent gene flow among *Bos* would improve understanding of the evolutionary process of species divergence.

Beefalo are a purported example of interspecies *Bos* hybridization. The American Beefalo Association (ABA) defines Beefalo as a stable hybrid cross with ⅝ cattle and ⅜ bison ancestry ("American Beefalo Association"). The Beefalo breed was established in the early 1970s by Mr. D.C. "Bud" Basolo through complex interbreeding of the two species (Miller 1982). The establishment of this breed followed a long history of attempts to crossbreed bison and cattle for commercial production (Boyd 1908; Goodnight 1914). As early 20th century breeders had been unable to create a viable hybrid population due to reproductive difficulties (Deakin et al. 1942), the establishment of the Beefalo breed was both surprising and celebrated (Miller 1982; "Business" 1973).

Basolo did not reveal the pedigree for his original Beefalo, but subsequent analyses have allowed better understanding of the breed's origins. Paim et al. (2020) illustrated two pedigree approaches that could create a ⅜ – ⅝$^{\text{th}}$ combination, both of which require first generation hybrid males to be fertile. Analyses of Beefalo karyotypes showed that all Beefalo possess taurine Y chromosomes (Lenoir and Lichtenberger 1978), such that the initial cross would have to have been a bison cow and a taurine bull. Intriguingly, Stormont et al. used blood typing to show that foundational Beefalo almost completely lack bison-specific markers (Stormont et al. 1986). Beefalo ancestry has not, however, been evaluated with genomic information, though the breed was granted a special roll stamp for voluntary federal inspection by the United States Department of Agriculture (USDA, 1984, 9 CFR part 352) and the ongoing sale and purchase of pedigreed animals.

Here, we present results of the first genome-wide investigation of the ancestry composition of the Beefalo breed. We generate whole-genome shotgun (WGS) data from 47 Beefalo, three bison hybrids reported to carry majority bison ancestry, and ten commercial bison that represent a potential source of bison ancestry in Beefalo, and co-analyze this data along with publicly available genomes from a range of bovids.

These samples were donated to the USDA Agricultural Research Service (ARS) National Animal Germplasm Program (NAGP) collection and include ancestors from the early establishment of the breed, such as Joe's Pride, a foundational Beefalo bull that Mr. Basolo sold to a Canadian breeder's firm for $2.5 million in 1975 ("Most Expensive Cattle"). We find that most Beefalo, including key founding individuals, contain no detectable bison ancestry, while bison hybrids have approximately their reported bison ancestry. Zebu ancestry is common among Beefalo, suggesting that breeding with zebu cattle, which share a number of phenotypic similarities with bison in commercially-relevant traits (Zeder 2006), was an early strategy in the establishment of the Beefalo breed. Finally, we find that some Beefalo contain small amounts of bison ancestry, ranging from 2-18%. While this is less bison ancestry than the targeted 37.5%, these genomes prove that bison and cattle can hybridize in some circumstances. However, we always infer minor parental ancestry to be in a heterozygous state, providing evidence only of backcrossing to either parent, rather than of breeding between hybrids themselves. This is inconsistent with a scenario in which a stable bison-cattle hybrid Beefalo population had been established.

# Results

## Sequencing Beefalo genomes

We generated genomic data from 47 registered Beefalo and three bison hybrids using preserved semen samples. Beefalo semen samples were obtained from the USDA, ARS, NAGP (Supplementary Table 1). Most of the NAGP Beefalo samples were collected during the 1970s and 80s, around the time when the breed was founded, and donated to NAGP around 2007. We retrieved bison ancestry composition for each sample from metadata accompanying NAGP samples. These 50 Beefalo and bison hybrids had reported bison ancestry ranging from 75% in a bison hybrid to 12.5% in several Beefalo, with the bison ancestry of most Beefalo reported as the breed-standard 37.5%. We sequenced each sample to 2.7x median genomic coverage (range: 1.1-39x; Supplementary Table 1). Seven prominent Beefalo were selected for higher coverage sequencing (>30x), including animals from the Basolo foundation herds, from which Beefalo initially originated, and others with different proportions of reported bison ancestry.

We generated high coverage (>17.5x) genomes of ten bison from commercial herds (Supplementary Table 1). We added these to a larger data set of previously published bison genomes (Yang et al. 2020; Stroupe et al. 2022; Wu et al. 2018; Shirazi et al.2022) and genomes from several breeds of zebu and taurine cattle (Heaton et al. 2016); Supplementary Table 1), as well as the genome of Buzz, an F1 Yellowstone bison-Simmental cattle hybrid (Oppenheimer et al. 2021). Our final data set included 41 bison, 26 cattle, and 51 Beefalo and bison hybrid genomes. For some analyses we also incorporated published genomes from outgroups, including yak (Qiu et al. 2012, 2015), gaur (Heaton et al. 2016), banteng (Heaton et al. 2016), and water buffalo (Sun et al. 2020). Pseudohaploid genotypes were then called on 5.29M biallelic autosomal variants ascertained in gaur, an outgroup to bison and cattle, for downstream analyses.

## Estimating bison ancestry in Beefalo

Most Beefalo samples cluster closely with taurine cattle in principal component analysis (PCA) conducted on bison and both cattle subspecies, indicating they do not contain appreciable bison ancestry (Fig. 1A). This includes several foundational individuals, such as Joe's Pride (NAGP5887). The first principal component (PC1) separates bison from cattle, with the F1 hybrid Buzz falling halfway along this axis, while PC2 distinguishes taurine and zebu cattle. While Beefalo samples group with taurine cattle, the three bison hybrids (>50% bison ancestry) fall in an intermediate position between bison and cattle (Fig. 1A). These individuals fall closer to bison than does Buzz, confirming their majority bison ancestry. Eight Beefalo are slightly shifted toward bison in the PCA, demonstrating genetic affinity with bison, though the position of these individuals on PC1 suggests they are unlikely to contain the breed-standard 37.5% bison ancestry (Figs. 1B, C). Beefalo formed a cline between taurine and zebu cattle on PC2, suggesting zebu ancestry is a widespread and variable ancestry component found in Beefalo.

One animal labeled as Beefalo in the NAGP database, NAGP9109, fell within the zebu cluster and is likely of 100% zebu origin. The reported pedigree in the NAGP for this animal lists its composition as ½ Brahman, ¼ Charolais, ⅛ bison, 1/16th Hereford, and 1/16th Shorthorn, but the American Brahman Breeders Association records this animal (#309519) as purebred Brahman, which is a zebu breed (5 of the other 6 zebu individuals analyzed here are Brahman cattle). We infer that

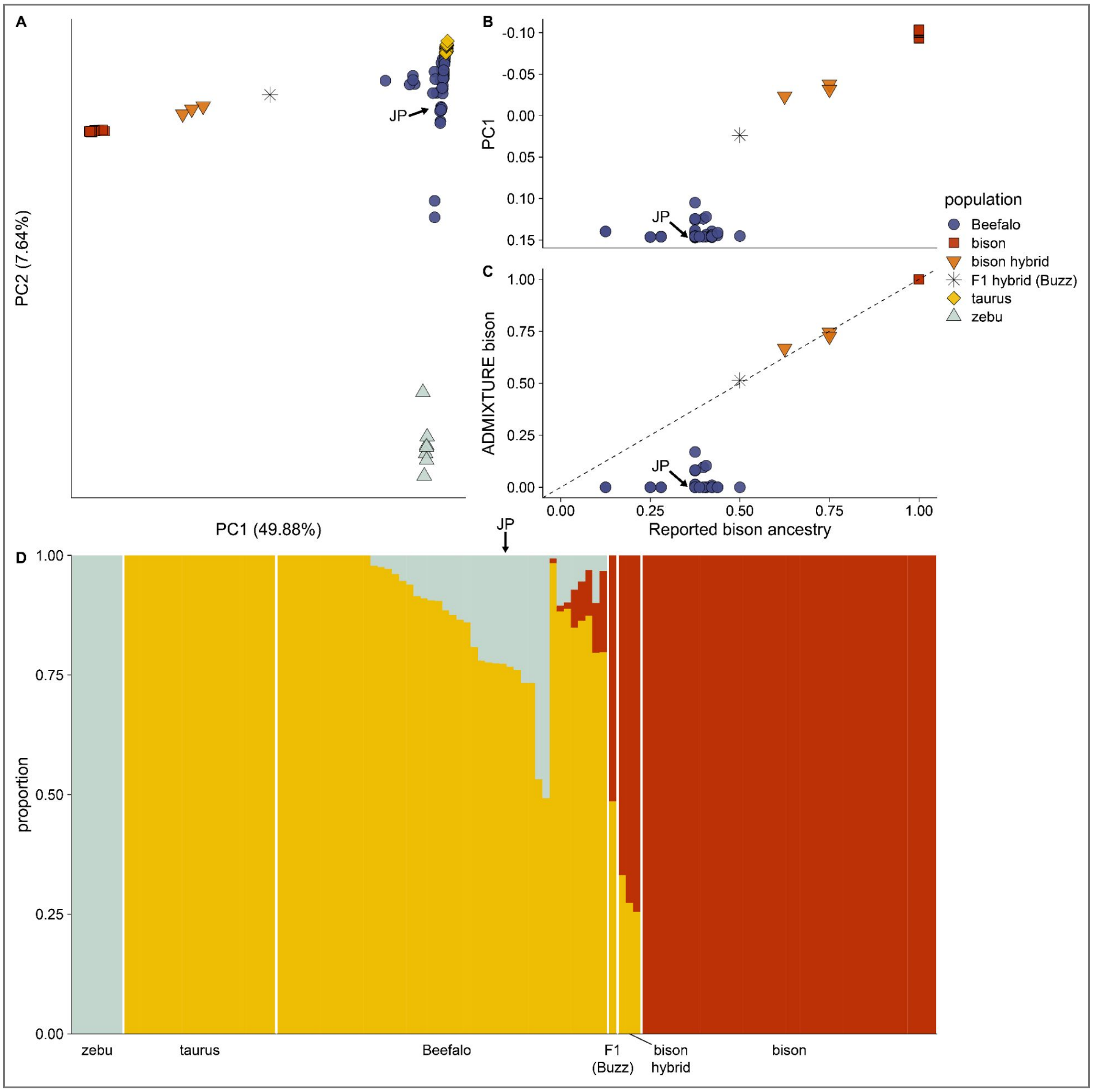


**Figure 1.**

**A)** Principal component analysis with Beefalo (blue), bison-cattle hybrids (orange), and a bison-cattle F1 hybrid (black star) projected onto axes computed from bison (red) and taurine (yellow) and zebu (gray) cattle. PC1 separates bison and cattle while PC2 separates the two cattle groups. **B)** Comparison of Beefalo reported bison ancestry and position on PC1. Placement of bison hybrids on PC1 is correlated with reported bison ancestry, though Beefalo do not follow this trend. **C)** depicts the same as in **B)** though compares inferred bison proportions from ADMIXTURE to reported bison ancestry. Reported and inferred bison proportions match well for bison hybrids, but all Beefalo were inferred to have less bison ancestry than reported. **D)** Autosomal ADMIXTURE analysis of bison, taurine and zebu cattle, bison hybrids, and Beefalo. The position of Joe's Pride (JP) is depicted with an arrow in **A-D**.

NAGP9109 was erroneously labeled as Beefalo by the contributors. This result highlights that PCA analyses of even low coverage genomes can uncover inconsistencies between genomic ancestry and animal metadata.

An unsupervised ADMIXTURE analysis supports the conclusion that the majority of Beefalo have no bison ancestry. This analysis estimates the ancestry profiles present across all bison, cattle, bison hybrid, and Beefalo individuals examined for a given value of $K$ source populations. At $K = 3$, bison, taurine cattle, and zebu cattle are divided into homogenous groups, with Beefalo assigned variable levels of these three ancestry types, mirroring the major patterns seen in the PCA (Fig. 1D). As in the PCA, the three bison-cattle hybrid backcrosses are inferred to have majority bison ancestry, consistent with their pedigrees, while 39 out of the 47 sampled Beefalo, including Basolo founding individuals, possess no bison ancestry. The remaining eight Beefalo tested are estimated to have a small amount (<18%) of bison ancestry, although less than was indicated by pedigrees and below the breed standard defined by the ABA (Fig. 1C).

These eight individuals are those shifted toward bison along PC1 (Figs. 1A, B). The unsupervised ADMIXTURE analysis correctly assigned the F1 bison-Simmental genome equal amounts of taurine and bison ancestry, demonstrating that this analysis is sensitive to the presence of large fractions of bison ancestry. Similar estimates of bison ancestry are observed at other values of $K$ (Supplementary Fig. S1), which supports the modeling of Beefalo ancestry as coming from three main components, corresponding to taurine and zebu cattle, and bison (Supplementary Fig. S2). Supervised ADMIXTURE clustering, which models Beefalo ancestry as coming specifically from bison, taurine, and zebu source panels, provides similar estimates to those obtained from the unsupervised ADMIXTURE analysis (Supplementary Fig. S3).

Examining allele sharing between Beefalo and bison relative to taurine cattle confirmed the results from PCA and ADMIXTURE, underscoring that 39 out of 47 examined Beefalo have no appreciable bison ancestry, while those that do have far less than their reported amounts. *D*-statistics of the form *D(Beefalo, taurus; bison, water buffalo)* again show that 39 Beefalo have no excess affinity with bison compared to taurine cattle ($-13.04 < Z < 3.14$), although the same eight Beefalo identified in PCA and ADMIXTURE as having bison ancestry also have an excess of bison alleles ($6.16 < Z < 34.86$), confirming their bison ancestry (Fig. 2A). f4-ratios estimate that these Beefalo derive between 2-18% of their genomes from bison (Fig. 2B), consistent with estimates from ADMIXTURE and less than their reported bison ancestry, which ranged from 37.5-50%. *D*-statistics also confirm that 18 Beefalo individuals have significantly more zebu alleles compared to taurine cattle ($3.64 < Z < 21.03$), demonstrating that zebu ancestry is widespread across Beefalo (Fig. 2C). These 18 individuals with excess zebu affinity had no evidence of bison ancestry using *D*-statistics, suggesting that allele sharing with bison, which are deeply diverged from both taurine and zebu cattle, does not complicate the inference of zebu ancestry in these individuals. Allele sharing results are similar using yak as an outgroup, instead of water buffalo (Supplementary Fig. S4).

Local ancestry inference across individual Beefalo and bison-cattle hybrid genomes provides similar estimates of overall Beefalo ancestry, inferring an absence of bison ancestry across the 37 Beefalo that lacked evidence for such ancestry in previous analyses (Fig. 3). Three bison hybrids are inferred to have ~75% bison ancestry, while eight Beefalo have detectable bison ancestry, ranging from 2-18%. In the 8 Beefalo with bison ancestry, that ancestry tends to be present in large contiguous blocks, often tens of megabases in size, indicative of recent admixture (Fig. 3A,B). Bison ancestry in Beefalo is always found in a heterozygous state, consistent with a scenario in which these individuals are the product of repeated backcrossing to cattle of an initial F1 hybrid. Thirty-eight Beefalo also have zebu ancestry (Fig. 3D) at variable levels ranging from 2-38%, with all but two Beefalo having between 2-18%. As with bison ancestry, zebu ancestry is found in large blocks, but with many smaller blocks than seen with bison ancestry, suggesting an earlier date for admixture (Fig. 3C).

Several lines of evidence attest to the efficacy of using this local ancestry inference approach for modeling Beefalo ancestry. The F1 bison-taurine cattle hybrid was inferred to be almost entirely heterozygous for bison and taurine cattle across the genome (Supplementary Fig. S5),

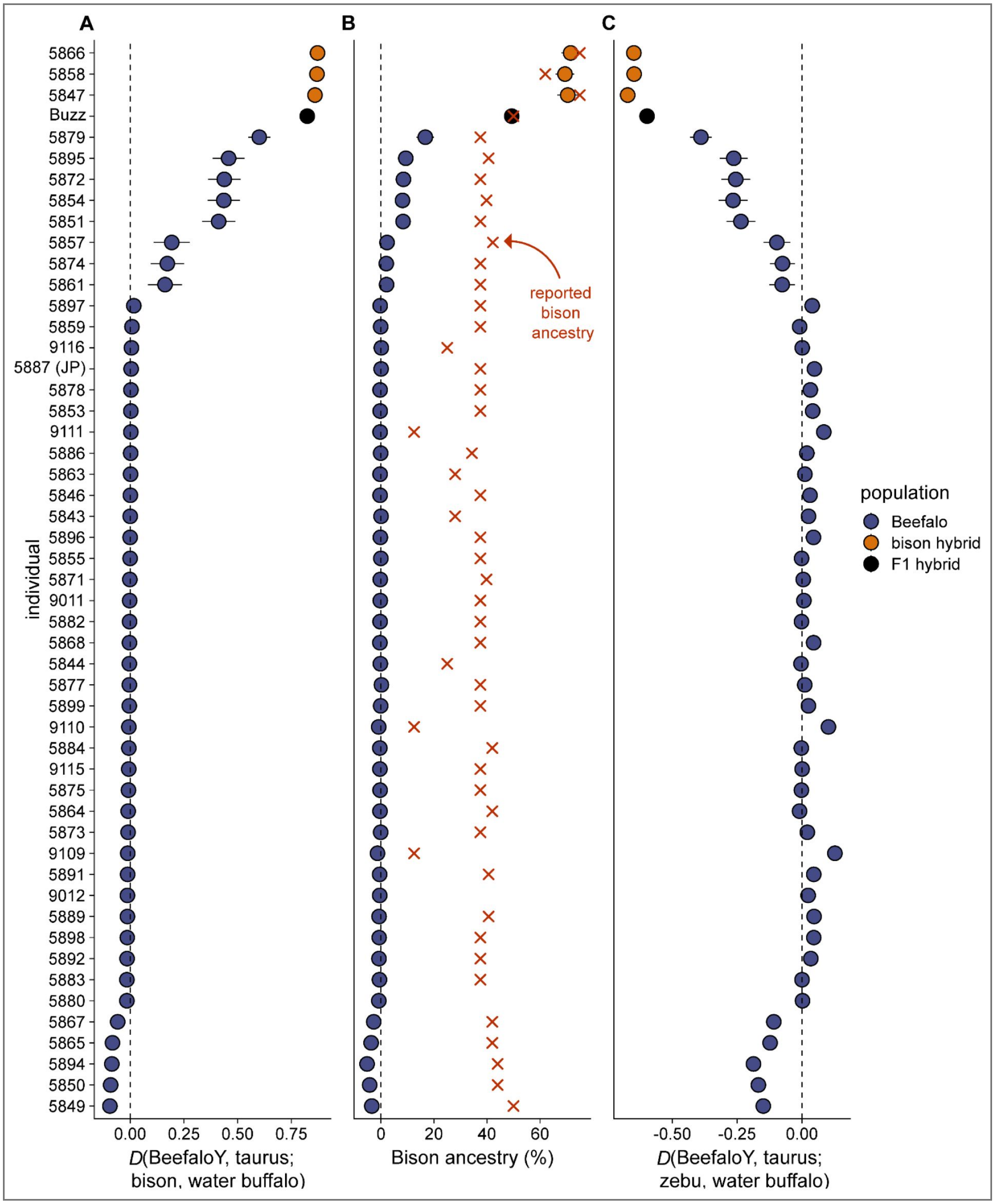


**Figure 2.** **Allele sharing statistics using individual bison hybrids and Beefalo.**

Individuals are shown on the Y axes. Beefalo are shown in blue and bison-cattle hybrids in orange, with a bison-cattle F1 shown in black. JP = Joe's Pride. **A)** Testing for the presence of bison ancestry using statistics of the form D(Beefalo, taurine cattle; bison, water buffalo). **B)** Quantification of the proportion of bison ancestry using f4-ratios. All bison hybrids have at least 50% bison ancestry, while only 8 Beefalo are inferred to have bison ancestry, ranging from 2-18%. Reported bison ancestry is shown in red. **C)** Testing for allele sharing between Beefalo and zebu cattle, relative to other taurine cattle. Many individual Beefalo are inferred to have significant allele sharing with zebu cattle. All Beefalo which were inferred to have bison ancestry in panels **A)** and **B)** display significantly negative values (due to the presence of bison ancestry, which is an outgroup to both cattle breeds), showing that any inferred zebu ancestry is not due to the presence of bison ancestry. For all panels, error bars depict 3 standard deviations.

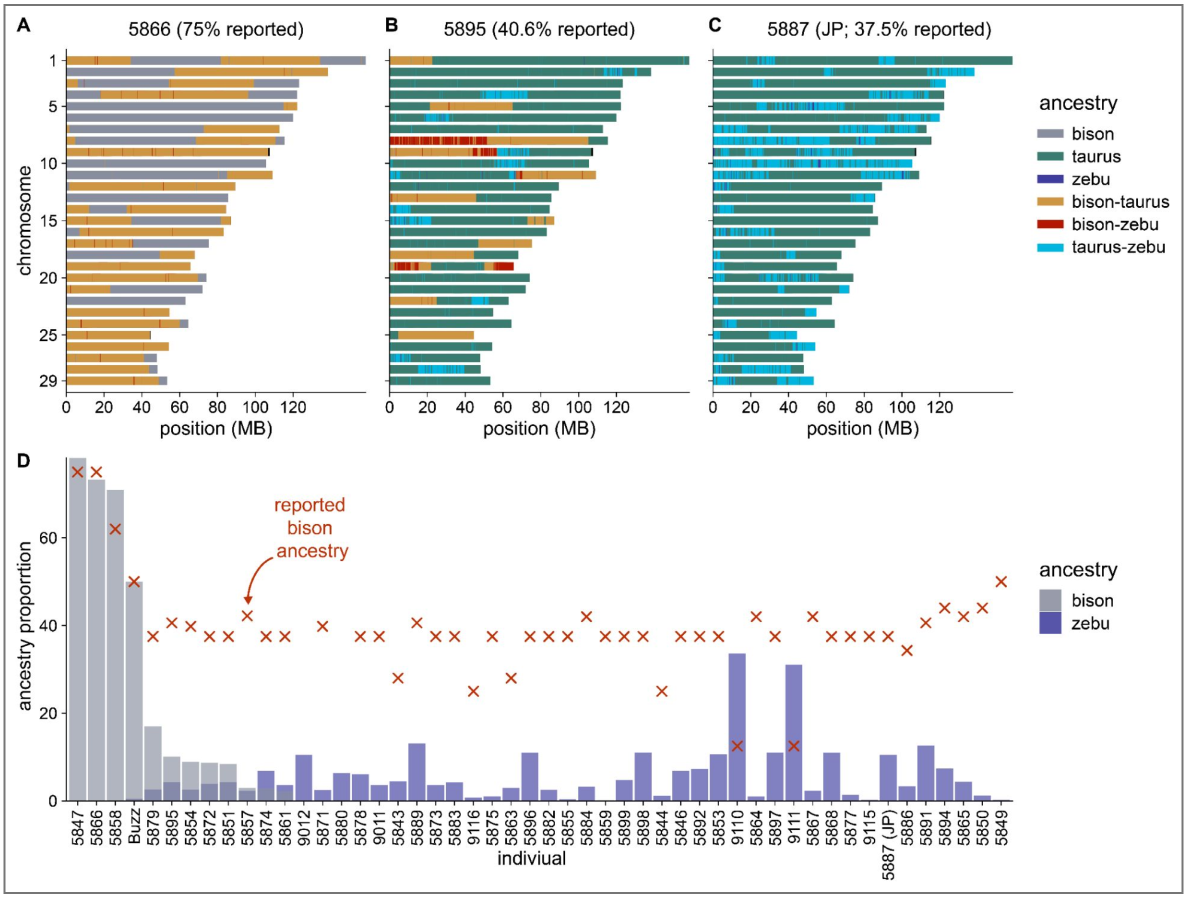


**Figure 3. Local ancestry inference of Beefalo and bison hybrid genomes, using bison and taurine and zebu cattle as potential sources.**

Inferred ancestry across the all autosomes are shown for **A**) a bison hybrid with reported 75% bison ancestry, **B**) a Beefalo with 37% reported, and 12% detected, bison ancestry, and **C**) Joe's Pride, a foundational Beefalo with 40.6% reported, and no detected, bison ancestry. Ancestry along the genome is colored by inferred source: homozygous bison, taurine, and zebu ancestry are shown in gray, green, and blue, respectively, with heterozygous bison-taurine, bison-zebu, and taurine-zebu ancestry shown in yellow, red, and light blue. Panel **D**) shows the inferred global bison and zebu ancestry proportions for each bison hybrid and Beefalo individual, with bison ancestry proportion shown in gray and zebu proportion shown in blue.

demonstrating the sensitivity of this approach to distinguishing diploid ancestry states between these groups. Further, the method assigned largely uniform and correct ancestry proportions to individual bison and cattle that were not part of the source panels, while Brangus cattle, a known taurine-zebu hybrid breed, were assigned large fractions of ancestry to both cattle groups in expected amounts (Supplementary Fig. S6). For the seven Beefalo sequenced to high coverage, we also modeled ancestry using downsampled read counts and called genotypes, which yielded nearly identical results (Supplementary Fig. S7).

### Beefalo sex chromosomal ancestry

Analyses of sex chromosomes reveal that the main mechanism for introducing bison ancestry into Beefalo was breeding with bison bulls. Beefalo with evidence of bison ancestry have less bison ancestry on X chromosomes compared to autosomes (Fig. 4A,C). Further, all Beefalo have Y chromosomes associated with taurine cattle, while the bison hybrid backcrosses have bison Y chromosomes (Fig. 4B). These patterns are the opposite of what would be expected if bison ancestry were introduced maternally, which would result in an excess of bison rather than cattle ancestry on the X chromosome (Lenoir and Lichtenberger 1978). Unfortunately, as semen was the source of genomic DNA, we could not assemble mitochondrial genomes to assess whether the Beefalo maternal line would provide any additional insight into these patterns.

## Discussion

Only eight of the 47 Beefalo that we sampled contained detectable bison ancestry, and those eight had substantially less (2-18%) than the 37.5% specified as the breed standard set by the ABA. Our sampling represents a complete survey of individuals in the USDA NAGP biobank, and includes important foundational animals that were involved in establishing Beefalo, such as those from original Basolo herds. Notably, important foundational individuals, including Joe's Pride (NAGP5887), which was sold for $2.5 million to a Canadian breeders group ("Most Expensive Cattle"), lack bison ancestry. While these results show that interbreeding between bison and cattle is possible, they also prove bison ancestry encompasses a much smaller portion of the Beefalo breed than has been claimed.

Our finding of little to no bison ancestry in founding Beefalo individuals is aligned with challenges hybridizing bison and cattle (Goodnight 1914; Murdoch 2018). Artificial insemination of taurine cows with bison semen results in 77% calf mortality and sterile male calves (Anstey 1986). In fact, Canadian government research reported that no functional males carrying more than 12.5% bison ancestry were ever observed (Brower 2008). In Beefalo individuals with bison ancestry, such ancestry was found exclusively in a heterozygous state, implying that repeated backcrossing to parental species, rather than breeding of hybrids themselves. A depletion of bison ancestry on the X chromosome and the presence of taurine Y chromosomes in all Beefalo, with bison hybrids exclusively having bison Y haplotypes, is consistent with this scenario. Therefore, it seems that extensive reproductive barriers exist to establishing a hybrid bison-cattle population.

While bison ancestry was surprisingly underrepresented in our Beefalo sample, the majority of the Beefalo genomes we sequenced contained some zebu cattle ancestry, suggesting zebu/taurine interbreeding may have been used as a mechanism to manipulate the Beefalo phenotype. Zebu cattle are known for heat and drought tolerance (Kumar et al. 2016; Vajana et al. 2018; Hansen 2004), lower nutritional demands relative to taurine cattle (Hunter and Siebert 1985; Hennessy, Williamson, and Darnell 2000) and their humped appearance (Heath 1979), and so share a number of desirable attributes with bison. The Beefalo breed standard does not require cattle ancestry to be of taurine origin, though if early breeders intentionally incorporated zebu ancestry while creating the breed, they left this detail out of the animals' reported pedigrees. For example, the Basolo founder individual Joe's Pride has a recorded pedigree that attests Hereford, Charolais, and bison (37.5%) ancestry, but we estimated that he had ~10.5% zebu ancestry.

This study is the first examination of Beefalo ancestry using whole genome data, but it is not the first to question claims surrounding bison ancestry in the breed. In an investigation of the paternal origins of Beefalo, Lenoir and Lichtenberger (1978) used karyotyping to determine that

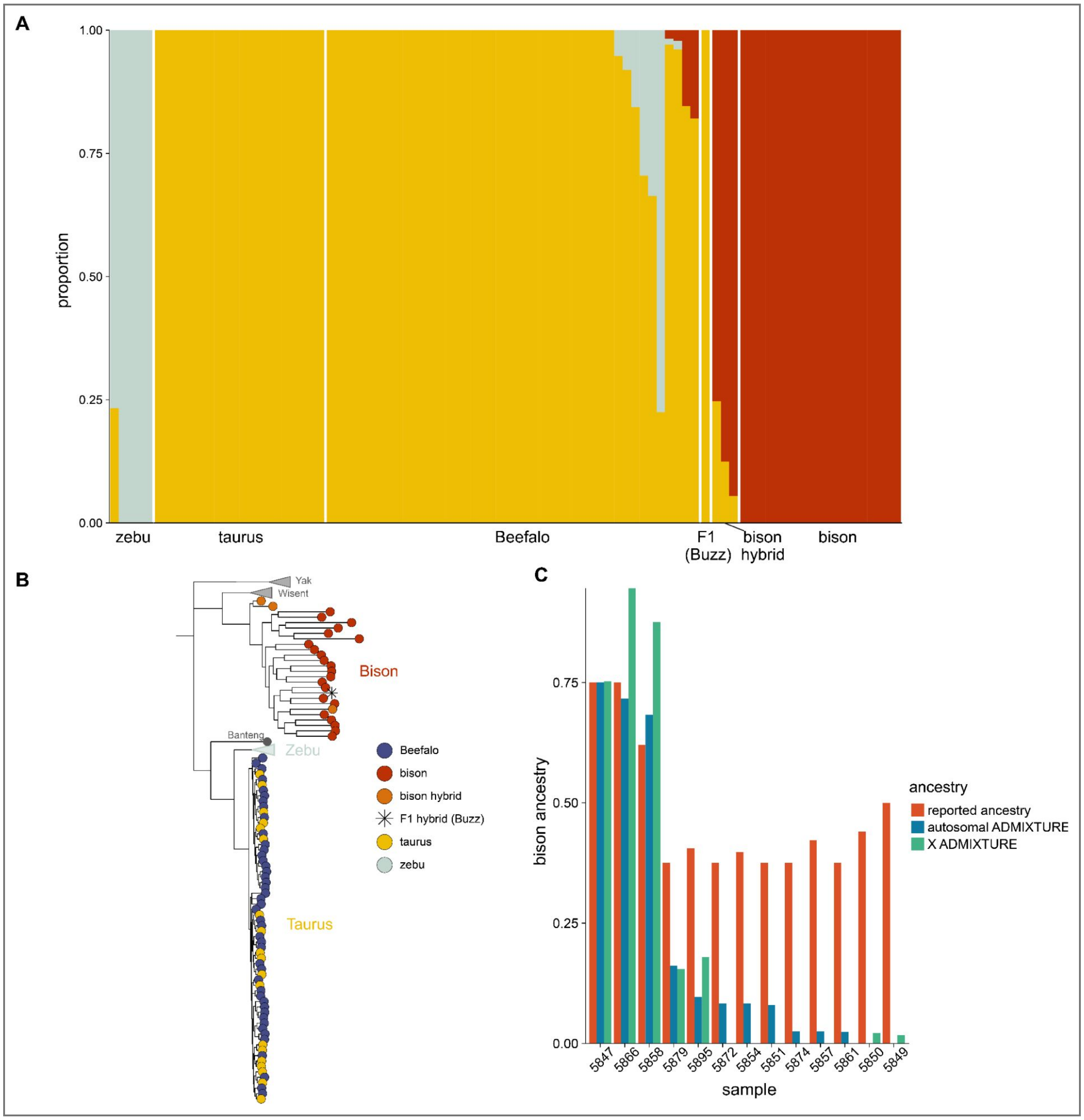


**Figure 4.**

**A)** ADMIXTURE analysis of bison, taurine and zebu cattle, bison hybrid, and Beefalo X chromosomes. Beefalo are largely inferred to have X chromosomes derived entirely from taurine cattle, though variable amounts of zebu and bison ancestry are also present in some Beefalo. The position of Joe's Pride (JP) is indicated with an arrow. Bison hybrids have majority bison ancestry. **B)** Bovid Y chromosomal phylogeny. The zebu clade is collapsed, as are those containing yak and wisent. Water buffalo (not shown) was used as an outgroup. Beefalo all have taurine Y chromosomes, while bison hybrids have bison Y chromosomes. **C)** Comparison of the reported (red) and ADMIXTURE inferred bison ancestry proportion on the autosomes (blue) and X chromosome (green) for the three bison hybrids and eight bison which had detectable autosomal bison ancestry.

all 12 Beefalo bulls they examined had cattle Y chromosomes, in concordance with our results. While there are plausible scenarios for Beefalo with 37.5% bison ancestry to have cattle Y chromosomes, these require early-generation crosses to be fertile. In a later study, Stormont et al. (1986) used species-specific blood group markers to show that only one of the 148 Basolo Beefalo they tested had any alleles found predominantly in bison. However, they also note that some later Beefalo, separate from the initial founding herds, did display some bison-associated markers, perhaps agreeing with our finding of a small fraction of Beefalo possessing appreciable bison ancestry. Blood typing provides limited insight into ancestry proportions, however, demonstrating the utility of genomic information in validating specific breeding claims.

Our genome-wide analyses were primarily conducted on ~2x coverage genomes with Beefalo samples obtained from the USDA NAGP, but we believe our Beefalo ancestry estimates are robust and representative of the breed at large. We derive highly concordant ancestry estimates across a range of approaches, including ADMIXTURE, f4-ratios, and local ancestry inference techniques, and for animals whose ancestry is known, such as an F1 bison-cattle hybrid, these methods assign the correct ancestry proportions. Additionally, we derive similar ancestry estimates in high coverage (30-42x) and downsampled data for the seven individuals we sequenced more deeply. We did not sample present-day sources of Beefalo exhaustively, instead choosing to focus largely on animals that originate at or near the founding of the breed. Beefalo herds are generally established by breeding with fullblood (37.5% bison) Beefalo individuals, rather than through backcrosses with bison ("American Beefalo Association"), so surveying breed founders is an effective way of documenting the bison ancestry across the breed. However, a larger sample of Beefalo across current producers would provide a more comprehensive understanding of the breadth and variation of bison ancestry in Beefalo.

The extent of gene flow and the existence of reproductive barriers among *Bos* species remain incompletely explored. Breeders worked throughout the 20th century to create commercially-viable bison-cattle hybrids using bulls and cows with different ancestries. All of these efforts, including Beefalo, failed to create a stable hybrid population, implying that considerable barriers to gene flow exist between these two species.

Though genomic incompatibilities among bison and cattle are hinted at by these breeding efforts themselves, which reported sterility in male hybrids even after backcrossing to cattle for several generations (Deakin et al. 1942; Brower 2008), they have not been detected directly. The general absence of bison ancestry in Beefalo also calls attention to several other examples of assumed gene flow among *Bos* species that have yet to be fully characterized. This includes introgression from cattle into wisent (Wecek et al. 2017; Soubrier et al. 2016) and American bison, the latter of which has been suggested to be widespread over the past two centuries and has led to the presence of cattle ancestry in all bison today (Stroupe et al. 2022; Halbert and Derr 2007), though has not been fully examined with genomic evidence.

# Materials and Methods

## Genomic data collection

Beefalo semen samples (n = 47) were obtained from the USDA, ARS, NAGP collection at Ft. Collins Colorado in the USA (Supplementary Table 1). The three bison hybrid semen samples were also obtained from the NAGP. The majority of the NAGP samples were collected in the 1970s and 1980s, stored by breeders, and donated to NAGP circa 2007 as a geographically diverse set of Beefalo. Purebred bison meat samples (n = 10, tongue) were purchased from commercial processors and represent animals from three different commercial bison herds (Supplementary Table 1).

DNA was extracted from semen samples with a modified standard phenol-chloroform method. Briefly, one 0.5-ml straw of semen was diluted with 1.5 ml of a solution containing 10 mM TrisCl, 100 mM NaCl, 1 mM EDTA, pH 8.0) with 1% wt/vol sodium dodecyl sulfate, 1 mg proteinase K, and 40 mM dithiothreitol. This lysis solution was incubated overnight at 37°C in a 15 ml conical tube preloaded with 2 ml of high-vacuum grease in preparation organic phase extractions. The lysed and digested solution was extracted twice with 1 vol of phenol:chloroform:isoamyl alcohol

(25:24:1), and once with 1 vol of chloroform. For each extraction the sample was centrifuged for 10 minutes in a swinging bucket rotor at 3210 x g at 23C to partition the organic phase below the band of high-vacuum grease, while to the aqueous phase was held above. The DNA was precipitated with 0.1 vol of 3 M sodium acetate (pH 5.2) and 2 vol of 100% ethanol. The precipitated DNA was washed once in 70% ethanol, briefly air dried, and dissolved in a solution of 10 mM TrisCl, 1 mM EDTA (pH 8.0). The bison tongue DNA was extracted with the Qiagen Blood and Cell Culture DNA Mini Kit according to the manufacturer's instructions (Qiagen, Venlo, The Netherlands).

We sheared extracted DNA using a Covaris ultrasonicator (Covaris, Inc. Woburn, MA) prepared into Illumina sequencing libraries using either the NEB Ultra II FS kit (NEB, Ipswich, MA) or the TruSeq PCR-free DNA kit (Illumina, San Diego, CA). We quantified libraries with a Qubit fluorometer using the 1x HS kit (Thermo Fisher, Waltham, MA) or, for PCR-free libraries, via qPCR with the Kapa Complete Universal Kit (Roche Sequencing, Santa Clara, CA). Library fragment length distribution was assessed using either a Tapestation 2200 (Agilent, Santa Clara, CA) or a Fragment Analyzer (Advanced Analytical Technologies Inc., Ames, IA). For whole genome sequence (WGS), 1 µg of genomic DNA was fragmented and used to make indexed, 350 bp, paired-end libraries.

Beefalo and bison libraries were sequenced with Illumina instruments using 2 × 150 bp paired-end kits, either on the NovaSeq 6000 for bison or on the NextSeq 2000 platform for Beefalo and bison hybrids.

## Variant calling and genotyping

Remnant adapter sequences were trimmed from reads using Trimmomatic (v0.39) (Bolger, Lohse, and Usadel 2014), requiring a minimum length of 30bp. Trimmed reads were then mapped to the cattle genome ARS-UCD1.2 with the Y chromosome from Btau5.1 appended (ARS-UCD1.2_Btau5.0.1Y) using BWA (v0.7.17-r1188) mem (Li 2013) with default parameters, except for 9 Beefalo samples that were mapped with BWA aln (Li and Durbin 2009). These samples had a lower fraction of reads that were properly paired, with higher rates of interchromosomal mapping within read pairs, possibly suggesting chimera formation. As BWA aln conducts end-to-end alignment, it mitigates the effects of any spurious mapping of incompletely clipped chimeric sequences. Duplicate reads were removed with Picard MarkDuplicates (*https://broadinstitute.github.io/picard/*).

We called genotypes in four medium-to-high coverage (>10x) gaur genomes (Heaton et al. 2016; Verdugo et al. 2019; Wu et al. 2018) using GATK HaplotypeCaller (v4.1.8.1) (DePristo et al. 2011) to ascertain a set of variants from an outgroup to both cattle and bison for use in downstream analyses. Variants were filtered variants for a minimum genotype quality of 30 and minimum and maximum depths of 1/3rd and 2x the mean coverage on a per-sample basis. We also performed mappability filtering with SNPable (*https://lh3lh3.users.sourceforge.net/snpable.shtml*), using a *k*-mer length of 35 and stringency of 0.5 (gen_mask -l 35 -r 0.5).This yielded a set of 5,291,534 high-quality autosomal variants.

To accommodate the low sequencing depth we obtained from most Beefalo, we used psuedohaploid genotypes, in which a random read covering each gaur-ascertained variant was selected to represent genotypes at those sites. This approach mitigates biases arising from differential coverage across individuals (Barlow et al. 2020). Pseudohaploid genotypes were called for all individuals using PileupCaller using samtools (v1.9) mpileup (Li et al. 2009) with BAQ disabled (-B) and pileupCaller (*https://github.com/stschiff/sequenceTools.git*), requiring a minimum map quality of 25 and minimum base quality of 30.

## Modeling Beefalo ancestry

We visualized the relationships of Beefalo and bison hybrid individuals to bison and cattle using a Principal Component Analysis (PCA). PCA was conducted by computing principal components with *smartpca* (v18140) (Patterson, Price, and Reich 2006) using medium and high coverage bison, taurine cattle, and zebu cattle, and then projecting all Beefalo and bison hybrids onto these axes, using our set of pseudohaploid genotypes. This approach allows for directly understanding the ancestry of Beefalo individuals relative to these three groups while mitigating the effects of the low sequencing depth obtained for many Beefalo.

We used ADMIXTURE (v1.3.0) (Alexander, Novembre, and Lange 2009) to further model Beefalo ancestry. ADMIXTURE was run both in supervised and unsupervised modes, using the pseudohaploid dataset filtered for missingness (--geno 0.5) and minor allele frequency (--maf 0.05), and pruned for linkage disequilibrium (--indep-pairwise 50 5 0.5) using plink (v1.9) (Chang et al. 2015). Cross-validation was used to select the optimal value of K. For the supervised ADMIXTURE analysis, Beefalo and bison hybrid ancestry was fit as potentially coming from bison, taurine, or zebu sources.

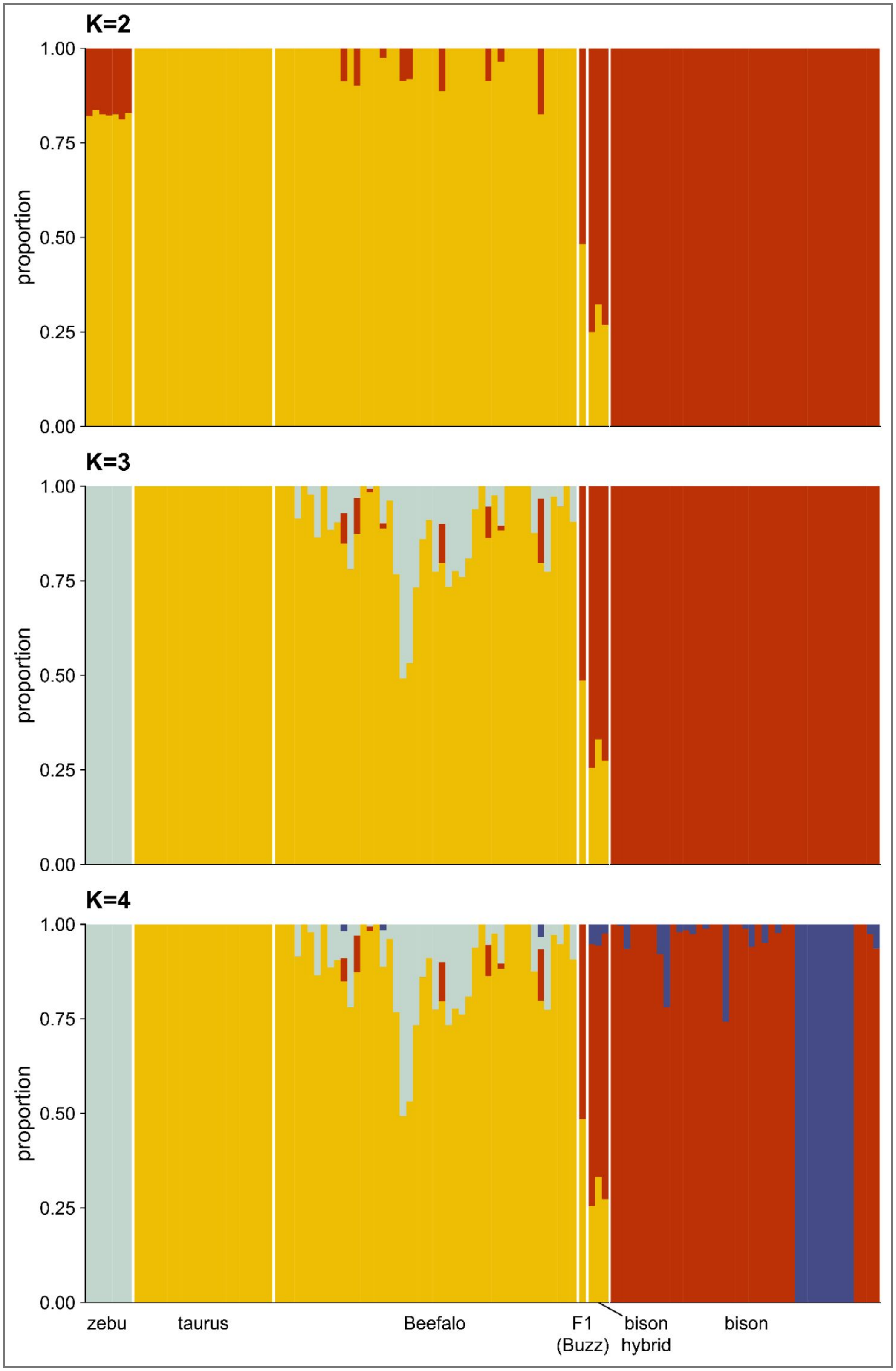
We also sought to investigate Beefalo ancestry on the sex chromosomes. For the X chromosome, there were 649,441 total variants, which were filtered to require that they were called in 66 individuals and had a minor allele frequency of at least 0.05, with 97,198 variants remaining. We used the 93 males within the overall set of samples with a level of missingness less than 25% and ran ADMIXTURE in haploid mode. We estimated the Y chromosomal phylogeny using males in the sample set with several additional species added as outgroups, including yaks, European bison, gaur, and banteng. The phylogeny was estimated from 18,056 variants with IQ-TREE (1.6.12) (Nguyen et al. 2015), using the GTR+ASC model and obtaining branch support values using ultrafast bootstrapping (Hoang et al. 2018).

After exploring overall autosomal Beefalo ancestry using model-free (PCA) and model-based (ADMIXTURE) approaches, we sought to specifically test for the presence of bison ancestry in Beefalo and bison hybrid genomes. The D-statistic *D(Beefalo, taurus; bison, water buffalo)* was used for this, which tests for excess allele sharing between Beefalo and bison, relative to taurine cattle, with water buffalo as an outgroup. We had observed that Beefalo formed a cline between taurine and zebu cattle in PCA, so we also used the statistic *D(Beefalo, taurus; indicus, water buffalo)* to test for zebu ancestry among individual Beefalo. Finally, we quantified the proportion of bison ancestry in Beefalo genomes using the f4-ratio *f4(yak, water buffalo; Beefalo, taurine)/f4(yak, water buffalo; bison, taurine)*, which estimates the proportion of ancestry, α, in Beefalo that comes from lineage related to bison, relative to a yak outgroup (the closest outgroup to bison). D- and *f*-statistics were calculated using ADMIXTOOLS2 (Maier et al. 2023) for each individual Beefalo and bison hybrid, grouping bison and taurine and zebu cattle.

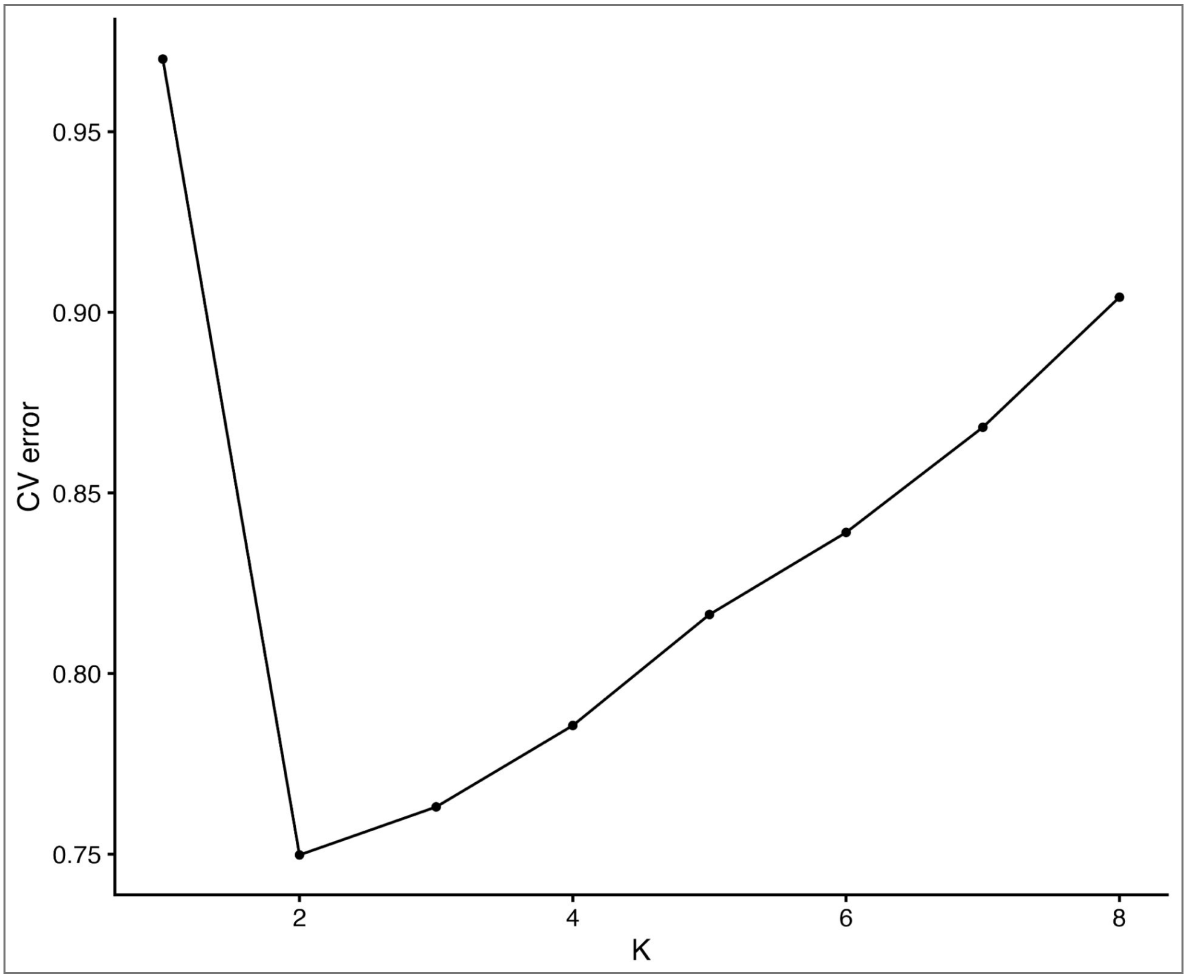
## Locating segments of differential ancestry within Beefalo

Diploid ancestry state across individual Beefalo genomes was modeled as coming from bison and cattle-related sources by using AncestryHMM (v1.0.2) (Corbett-Detig and Nielsen 2017). We used medium-to-high coverage genomes (>10x) from bison (n=20), taurine cattle (n=20), and zebu (n=6) each as source panels. We called genotypes for each species separately using the same filtering procedure as with our outgroup-ascertained variants. We then merged these panels, requiring sites to be called in at least 6 bison, 6 taurine cattle, and 3 zebu, have a frequency difference of at least 90% between any two source panels, and be at least 500bp from the nearest variant. This gave a final set of 2,322,535 ancestry informative markers. We modeled local ancestry across the genome for each Beefalo using AncestryHMM with the parameters '-e 0.02 -a 3 0.02 0.90 0.08 -p 0 -5 0.02 -p 1 10000 0.98 -p 2 -20 0.08'. This models two gene flow events, one 20 generations ago from zebu cattle at 8% and the second from bison 5 generations ago at 2%, though allows the exact timing of these admixture pulses to be inferred. We used read counts at ancestry informative markers for all Beefalo as input for ancestry inference. For seven individuals which were sequenced to high coverage, we performed additional ancestry inference using both read data downsampled to ∼2x coverage and genotypes at ancestry informative positions. Genotypes at ancestry informative markers were called using GATK HaplotypeCaller. We filtered posterior ancestry probabilities using a 0.9 threshold.

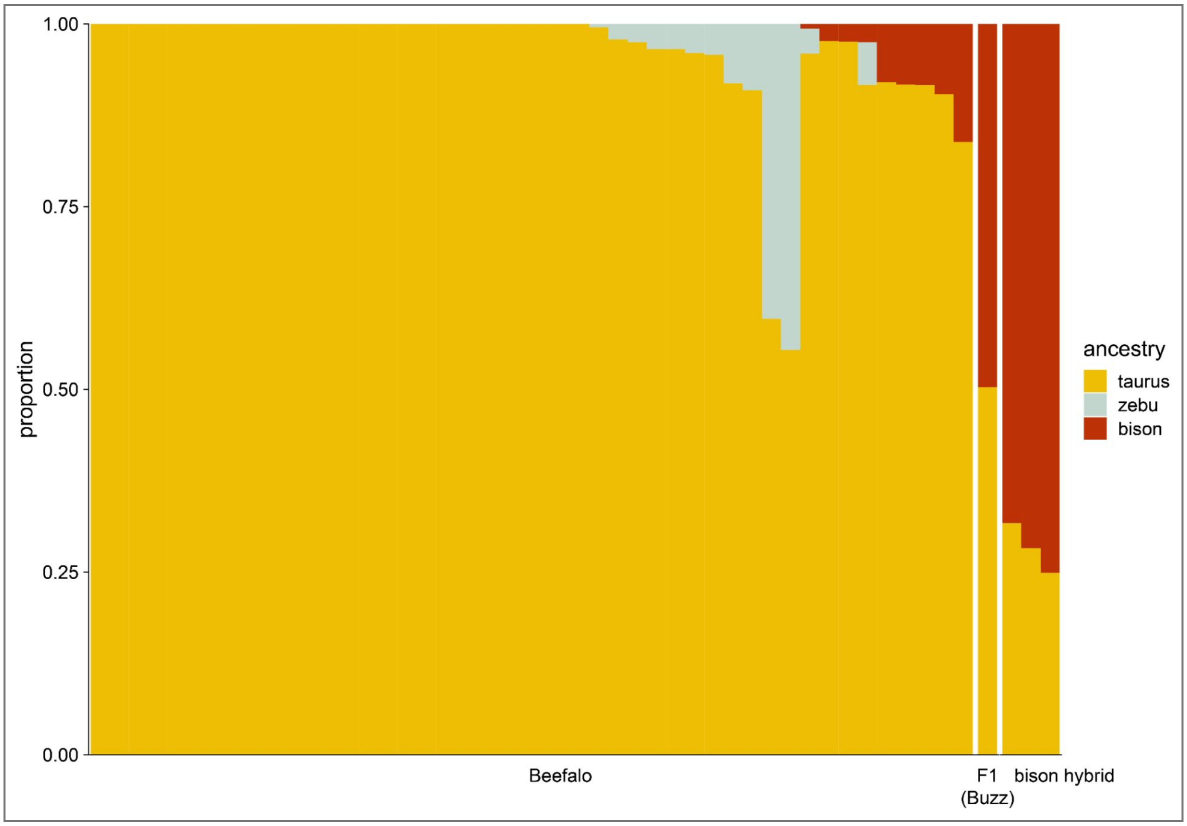
## Supplementary Figures



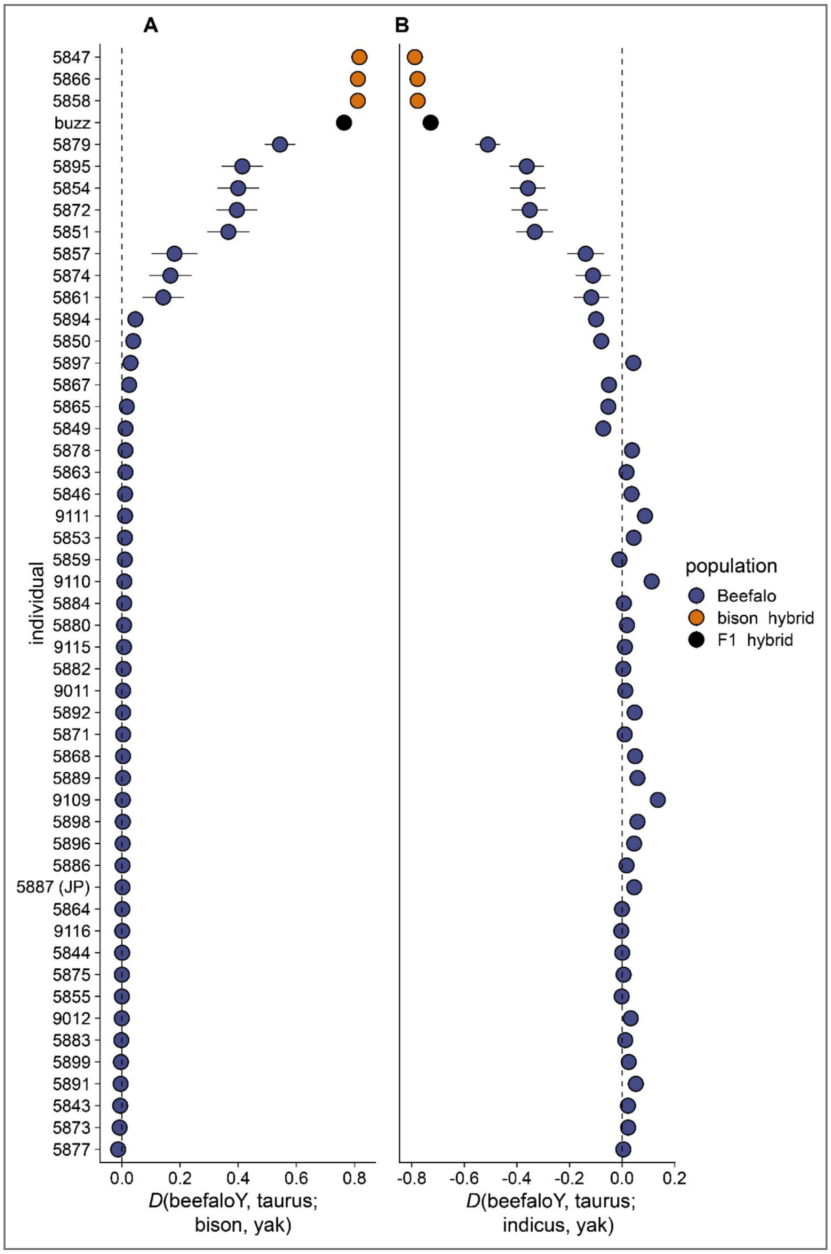
**Supplementary Figure S1. Unsupervised ADMIXTURE analysis of cattle, bison, bison-cattle hybrids, and Beefalo at different values of K, from K=2 to K=4.** At K=4, bison are split into two groups, which correspond to bison subspecies (wood and plains bison).



**Supplementary Figure S2.** **Cross-validation results comparing unsupervised ADMIXTURE using cattle, bison, bison-cattle hybrids, and Beefalo across different values of K, ranging from 1 to 8.**

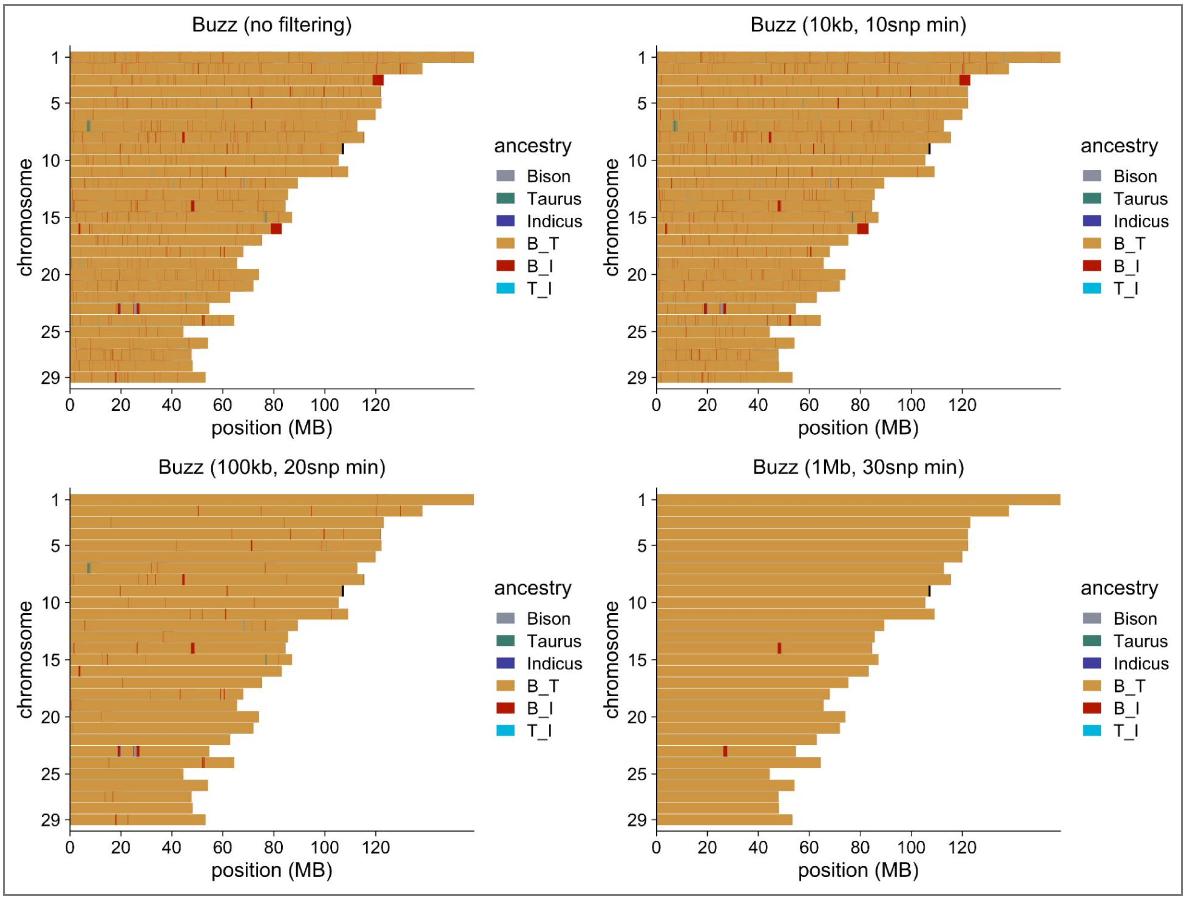


**Supplementary Figure S3.** **Supervised ADMIXTURE modeling of Beefalo and bison hybrid ancestry, using panels of bison and taurine and zebu cattle.**



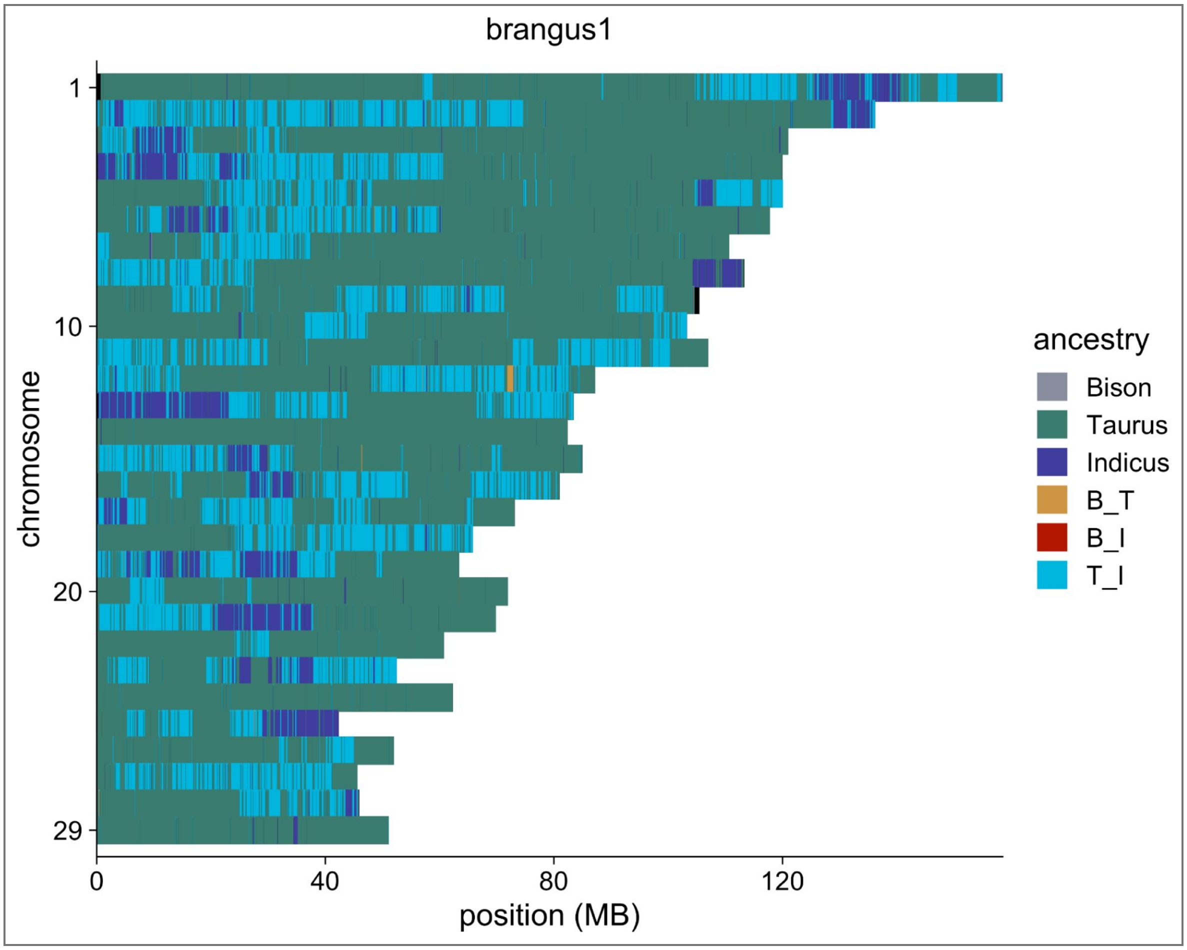
**Supplementary Figure S4.** **Allele sharing statistics using individual bison hybrids and Beefalo.**

This is the same as Fig. 2, except yak is used as an outgroup instead of water buffalo. **A)** D-statistics testing for allele sharing between individual Beefalo and bison hybrids, relative to taurine cattle. **B)** D-statistics testing for allele sharing between Beefalo and zebu cattle, relative to taurine cattle. For all panels, error bars depict 3 standard deviations.



**Supplementary Figure S5.** **Local ancestry inference results from a F1 bison-taurine cattle hybrid.**

Ancestry was correctly inferred to be almost completely heterozygous for bison and taurine ancestry across the genome, though with some small segments incorrectly assigned to be heterozygous for bison-zebu bison, likely because of the low divergence between cattle subspecies. These segments are all small and are increasingly removed with more stringent filtering (for window size and minimum SNPs to call a window, shown in panel titles).



**Supplementary Figure S6. Local ancestry inference results from a Brangus individual, showing the presence of both taurine and zebu (indicine) ancestry.**

**Supplementary Figure S7.** **Comparison of local ancestry inference using either downsampled (~2x) read data (A, D, J, M, P, S), high coverage (30-42x) read data (B, E, H, K, N, Q, T), or called genotypes (C, F, L, O, R, U) on seven individual Beefalo and bison hybrids which were sequenced to high coverage: 5851 (A-C), 5854 (D-F), 5861 (G-I), 5887 (J-L), 5895 (M-O), 5872 (P-R), 5874 (S-U).**

# Data availability

All Beefalo, bison hybrid, and bison sequencing data generated in this study are available in NCBI BioProject PRJNA1152308.

# Acknowledgements

We thank USMARC technicians H.R. Sadd and N.J. Allison for assistance with laboratory procedures. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer.

# Additional files

***Supplemental Table 1*** List of samples used.

# Additional information

## Funding

| Funder | Grant reference number | Author |
|---|---|---|
| Howard Hughes Medical Institute (HHMI) | | Beth Shapiro |

## Author ORCID iDs

**Beth Shapiro**: https://orcid.org/0000-0002-2733-7776
**Jonas Oppenheimer**: https://orcid.org/0000-0001-7973-6173

# References

1. **Alexander David H.,** Novembre John, Lange Kenneth (2009) Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Research* **19**:1655-64 https://doi.org/10.1101/gr.094052.109 | PubMed
2. **Anstey T.H** (1986) *One hundred harvests: Research Branch, Agriculture Canada, 1886-1986* Ottawa, Ontario: Research Branch Agriculture Canada.
3. **American Beefalo Association** (2024) American Beefalo Association. https://americanbeefaloassociation.com/
4. **Barlow Axel,** Hartmann Stefanie, Gonzalez Javier, Hofreiter Michael, Paijmans Johanna L. A. (2020) Consensify: A Method for Generating Pseudohaploid Genome Sequences from Palaeogenomic Datasets with Reduced Error Rates. *Genes* **11** https://doi.org/10.3390/genes11010050 | PubMed
5. **Bolger Anthony M.,** Lohse Marc, Usadel Bjoern (2014) Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **30**:2114-20 https://doi.org/10.1093/bioinformatics/btu170 | PubMed
6. **Boyd Mossom M** (1908) A Short Account of an Experiment in Crossing the American Bison with Domestic Cattle. *The Journal of Heredity* **1**:324-31 https://doi.org/10.1093/jhered/os-4.1.324
7. **Brower Jennifer** (2008) *Lost Tracks: National Buffalo Park, 1909-1939* Athabasca University Press.
8. **Chang Christopher C.,** Chow Carson C., Tellier Laurent Cam, Vattikuti Shashaank, Purcell Shaun M., Lee James J. (2015) Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *GigaScience* **4**:7 https://doi.org/10.1186/s13742-015-0047-8 | PubMed
9. **Chen Ningbo,** Cai Yudong, Chen Qiuming, Li Ran, Wang Kun, Huang Yongzhen, Hu Songmei, et al. (2018) Whole-Genome Resequencing Reveals World-Wide Ancestry and Adaptive Introgression Events of Domesticated Cattle in East Asia. *Nature Communications* **9**:2337 https://doi.org/10.1038/s41467-018-04737-0 | PubMed

10. **Corbett-Detig Russell,** Nielsen Rasmus (2017) A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLoS Genetics* **13**:e1006529 https://doi.org/10.1371/journal.pgen.1006529 | PubMed
11. **Deakin A.,** Muir G. W., Smith A. G., Lellan A. S. (1942) *Hybridisation of Domestic Cattle and Buffalo (Bison Americanus). Progress Report off the Wainwright Experiment 1935—41* Canada Department of Agriculture.
12. **DePristo Mark A.,** Banks Eric, Poplin Ryan, Garimella Kiran V., Maguire Jared R., Hartl Christopher, Philippakis Anthony A., et al. (2011) A Framework for Variation Discovery and Genotyping Using next-Generation DNA Sequencing Data. *Nature Genetics* **43**:491-98 https://doi.org/10.1038/ng.806 | PubMed
13. **Goodnight Charles** (1914) My Experience With Bison Hybrids. *The Journal of Heredity* **5**:197-99
14. **Hansen P. J** (2004) Physiological and Cellular Adaptations of Zebu Cattle to Thermal Stress. *Animal Reproduction Science* **82-83**:349-60 https://doi.org/10.1016/j.anireprosci.2004.04.011 | PubMed
15. **Heath E** (1979) Relationships of Musculus Rhomboideus, Ligamentum Nuchae and Vertebrae Thoracis and Lumbales in Bos Indicus. *Acta Anatomica* **105**:56-60 https://doi.org/10.1159/000145107 | PubMed
16. **Heaton Michael P.,** Smith Timothy P. L., Carnahan Jacky K., Basnayake Veronica, Qiu Jiansheng, Simpson Barry, Kalbfleisch Theodore S. (2016) Using Diverse U.S. Beef Cattle Genomes to Identify Missense Mutations in *EPAS1*, a Gene Associated with Pulmonary Hypertension. *F1000Research* **5** https://doi.org/10.12688/f1000research.9254.2
17. **Hennessy D. W.,** Williamson P. J., Darnell R. E. (2000) Feed Intake and Liveweight Responses to Nitrogen And/or Protein Supplements by Steers of Bos Taurus, Bos Indicus and Bos Taurus × Bos Indicus Breed Types Offered a Low Quality Grass Hay. *The Journal of Agricultural Science* **135**:35-45 https://doi.org/10.1017/s0021859699007923
18. **Hoang Diep Thi,** Chernomor Olga, von Haeseler Arndt, Minh Bui Quang, Vinh Le Sy (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**:518-22 https://doi.org/10.1093/molbev/msx281 | PubMed
19. **Hunter R. A.,** Siebert B. D. (1985) Utilization of Low-Quality Roughage by Bos Taurus and Bos Indicus Cattle. 2. The Effect of Rumen-Degradable Nitrogen and Sulphur on Voluntary Food Intake and Rumen Characteristics. *The British Journal of Nutrition* **53**:649-56 https://doi.org/10.1079/bjn19850074 | PubMed
20. **Kim Kwondo,** Kwon Taehyung, Dessie Tadelle, Yoo Dongahn, Mwai Okeyo Ally, Jang Jisung, Sung Samsun, et al. (2020) The Mosaic Genome of Indigenous African Cattle as a Unique Genetic Resource for African Pastoralism. *Nature Genetics* **52**:1099-1110 https://doi.org/10.1038/s41588-020-0694-2 | PubMed
21. **Kumar Anil,** Waiz Syma Ashraf, Sridhar Goud T., Tonk R. K., Grewal Anita, Singh S. V., Yadav B. R., Upadhyay R. C. (2016) Assessment of Adaptability of Zebu Cattle (Bos Indicus) Breeds in Two Different Climatic Conditions: Using Cytogenetic Techniques on Genome Integrity. *International Journal of Biometeorology* **60**:873-82 https://doi.org/10.1007/s00484-015-1080-0 | PubMed
22. **Kwon Taehyung,** Kim Kwondo, Caetano-Anolles Kelsey, Sung Samsun, Cho Seoae, Jeong Choongwon, Hanotte Olivier, Kim Heebal (2022) Mitonuclear Incompatibility as a Hidden Driver behind the Genome Ancestry of African Admixed Cattle. *BMC Biology* **20**:20 https://doi.org/10.1186/s12915-021-01206-x | PubMed
23. **Lenoir F.,** Lichtenberger M. J. (1978) The Y Chromosome of the Basolo Hybrid Beefalo Is a Y of Bos Taurus. *The Veterinary Record* **102**:422-23 https://doi.org/10.1136/vr.102.19.422 | PubMed
24. **Li Heng** (2013) Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. http://arxiv.org/abs/1303.3997
25. **Li Heng,** Durbin Richard (2009) Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Bioinformatics* **25**:1754-60 https://doi.org/10.1093/bioinformatics/btp324 | PubMed

26. **Li Heng,** Handsaker Bob, Wysoker Alec, Fennell Tim, Ruan Jue, Homer Nils, Marth Gabor, Abecasis Goncalo, Durbin Richard (2009) The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **25**:2078-79 https://doi.org/10.1093/bioinformatics/btp352 | PubMed

27. **Maier Robert,** Flegontov Pavel, Flegontova Olga, Isildak Ulas, Changmai Piya, Reich David (2023) On the Limits of Fitting Complex Models of Population History to F-Statistics. *eLife* **12**:April https://doi.org/10.7554/eLife.85492 | PubMed

28. **McTavish Emily Jane,** Decker Jared E., Schnabel Robert D., Taylor Jeremy F., Hillis David M. (2013) New World Cattle Show Ancestry from Multiple Independent Domestication Events. *Proceedings of the National Academy of Sciences of the United States of America* **110**:E1398-1406 https://doi.org/10.1073/pnas.1303367110 | PubMed

29. **Miller Bryan** (1982) Beefalo: There Are Few On The Range. The New York Times. https://www.nytimes.com/1982/01/06/garden/beefalo-there-are-few-on-the-range.html

30. **Most Expensive Cattle** (2024) Guinness World Records. https://www.guinnessworldrecords.com/world-records/70817-most-expensive-cattle

31. **Murdoch Brenda M** (2018) 34. Meiotic Recombination and Chromosomal Defects in Hybrid Beefalo and Ruminant Livestock Spermatocytes. *Cancer Genetics* **224**:63-64 https://doi.org/10.1016/j.cancergen.2018.04.037

32. **Nguyen Lam-Tung,** Schmidt Heiko A., von Haeseler Arndt, Minh Bui Quang (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**:268-74 https://doi.org/10.1093/molbev/msu300 | PubMed

33. **Oppenheimer Jonas,** Rosen Benjamin D., Heaton Michael P., Vander Ley Brian L., Shafer Wade R., Schuetze Fred T., Stroud Brad, et al. (2021) A Reference Genome Assembly of American Bison, Bison Bison Bison. *The Journal of Heredity* **112**:174-83 https://doi.org/10.1093/jhered/esab003 | PubMed

34. **Paim T. do P.,** Hay E. H. A., Wilson C., Thomas M. G., Kuehn L. A., Paiva S. R., McManus C., Blackburn H. D. (2020) Dynamics of Genomic Architecture during Composite Breed Development in Cattle. *Animal Genetics* **51**:224-34 https://doi.org/10.1111/age.12907 | PubMed

35. **Patterson Nick,** Price Alkes L., Reich David (2006) Population Structure and Eigenanalysis. *PLoS Genetics* **2**:e190 https://doi.org/10.1371/journal.pgen.0020190 | PubMed

36. **Qiu Qiang,** Wang Lizhong, Wang Kun, Yang Yongzhi, Ma Tao, Wang Zefu, Zhang Xiao, et al. (2015) Yak Whole-Genome Resequencing Reveals Domestication Signatures and Prehistoric Population Expansions. *Nature Communications* **6**:10283 https://doi.org/10.1038/ncomms10283 | PubMed

37. **Qiu Qiang,** Zhang Guojie, Ma Tao, Qian Wubin, Wang Junyi, Ye Zhiqiang, Cao Changchang, et al. (2012) The Yak Genome and Adaptation to Life at High Altitude. *Nature Genetics* **44**:946-49 https://doi.org/10.1038/ng.2343 | PubMed

38. **Shirazi Sabrina,** Broomandkhoshbacht Nasreen, Oppenheimer Jonas, Metcalfe Jessica Z., Found Rob, Ives John W., Shapiro Beth (2022) Ancient DNA-Based Sex Determination of Bison Hide Moccasins Indicates Promontory Cave Occupants Selected Female Hides for Footwear. *Journal of Archaeological Science* **137**:105533 https://doi.org/10.1016/j.jas.2021.105533

39. **Soubrier Julien,** Gower Graham, Chen Kefei, Richards Stephen M., Llamas Bastien, Mitchell Kieren J., Ho Simon Y. W., et al. (2016) Early Cave Art and Ancient DNA Record the Origin of European Bison. *Nature Communications* **7**:13158 https://doi.org/10.1038/ncomms13158 | PubMed

40. **Stormont C. J.,** Morris B. G., Suzuki Y., Dodd J. (1986) Blood Typing Beefalo Cattle. In: 3rd World Congress on Genetics Applied to Livestock Production.

41. **Stroupe Sam,** Forgacs David, Harris Andrew, Derr James N., Davis Brian W. (2022) Genomic Evaluation of Hybridization in Historic and Modern North American Bison (*Bison bison*). *Scientific Reports* **12**:6397 https://doi.org/10.1038/s41598-022-09828-z | PubMed

42. **Sun Ting,** Shen Jiafei, Achilli Alessandro, Chen Ningbo, Chen Qiuming, Dang Ruihua, Zheng Zhuqing, et al. (2020) Genomic Analyses Reveal Distinct Genetic Architectures and Selective Pressures in Buffaloes. *GigaScience* **9**:giz166 https://doi.org/10.1093/gigascience/giz166 | PubMed

43. **Time** (1973) Business: Have a Slice of Roast Beefalo. https://time.com/archive/6841372/business-have-a-slice-of-roast-beefalo/

44. **Vajana Elia,** Barbato Mario, Colli Licia, Milanesi Marco, Rochat Estelle, Fabrizi Enrico, Mukasa Christopher, et al. (2018) Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Frontiers in Genetics* **9**:385 https://doi.org/10.3389/fgene.2018.00385 | PubMed

45. **Verdugo Marta Pereira,** Mullin Victoria E., Scheu Amelie, Mattiangeli Valeria, Daly Kevin G., Delser Pierpaolo Maisano, Hare Andrew J., et al. (2019) Ancient Cattle Genomics, Origins, and Rapid Turnover in the Fertile Crescent. *Science* **365**:173-76 https://doi.org/10.1126/science.aav1002 | PubMed

46. **Wang Kun,** Lenstra Johannes A., Liu Liang, Hu Quanjun, Ma Tao, Qiu Qiang, Liu Jianquan (2018) Incomplete Lineage Sorting rather than Hybridization Explains the Inconsistent Phylogeny of the Wisent. *Communications Biology* **1**:169 https://doi.org/10.1038/s42003-018-0176-6 | PubMed

47. **Ward James A.,** McHugo Gillian P., Dover Michael J., Hall Thomas J., Ng'ang'a Said Ismael, Sonstegard Tad S., Bradley Daniel G., Frantz Laurent A. F., Salter-Townshend Michael, MacHugh David E. (2022) Genome-Wide Local Ancestry and Evidence for Mitonuclear Coadaptation in African Hybrid Cattle Populations. *iScience* **25**:104672 https://doi.org/10.1016/j.isci.2022.104672 | PubMed

48. **Wecek Karolina,** Hartmann Stefanie, Paijmans Johanna L. A., Taron Ulrike, Xenikoudakis Georgios, Cahill James A., Heintzman Peter D., et al. (2017) Complex Admixture Preceded and Followed the Extinction of Wisent in the Wild. *Molecular Biology and Evolution* **34**:598-612 https://doi.org/10.1093/molbev/msw254 | PubMed

49. **Wu Dong-Dong,** Ding Xiang-Dong, Wang Sheng, Wójcik Jan M., Zhang Yi, Tokarska Małgorzata, Li Yan, et al. (2018) Pervasive Introgression Facilitated Domestication and Adaptation in the Bos Species Complex. *Nature Ecology & Evolution* **2**:1139-45 https://doi.org/10.1038/s41559-018-0562-y | PubMed

50. **Yang Tianfu,** Miller Michelle, Forgacs David, Derr James, Stothard Paul (2020) Development of SNP-Based Genomic Tools for the Canadian Bison Industry: Parentage Verification and Subspecies Composition. *Frontiers in Genetics* **11**:585999 https://doi.org/10.3389/fgene.2020.585999 | PubMed

51. **Zeder Melinda A** (2006) *Documenting Domestication: New Genetic and Archaeological Paradigms* University of California Press.

52. **Zhang K.,** Lenstra J. A., Zhang S., Liu W., Liu J. (2020) Evolution and Domestication of the Bovini Species. *Animal Genetics* **51**:637-57 https://doi.org/10.1111/age.12974 | PubMed

**Shapiro B,** Oppenheimer J, Heaton MP, Kuhn KL, Green RE, Blackburn HD, Smith TPL (2024) Investigating the ancestry of Beefalo cattle. NCBI BioProject. ID PRJNA1152308 https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1152308

**Heaton MP** (2016) WGS data from cetartiodactyla. NCBI BioProject. ID PRJNA325061 https://www.ncbi.nlm.nih.gov/bioproject/325061

**USDA** (2016) WGS data from diverse types of U.S. cattle. NCBI BioProject. ID PRJNA324822 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA324822/

**Verdugo MP** (2019) Ancient cattle genomics of the Near East. ENA BioProject. ID PRJEB31621 https://www.ebi.ac.uk/ena/browser/view/PRJEB31621

**Wu D-D** (2018) Bison bison, Bison bonasus, Bos gaurus, Bos javanicus, Bos frontalis. NCBI BioProject. ID PRJNA427536 https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA427536

**Yang T** (2020) American bison raw sequence reads. NCBI BioProject. ID PRJNA658430 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA658430/

**Stroupe S** (2022) Genomic evaluation of hybridization in historic and modern North American Bison (Bison bison). NCBI BioProject. ID PRJNA658430 https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA658430

**Sun T** (2020) Genomic Analyses Reveal Distinct Genetic Architectures and Selective Pressures in Buffaloes. ENA BioProject. ID PRJNA547460 https://www.ebi.ac.uk/ena/browser/view/PRJNA547460

**Oppenheimer J** (2020) Paternal genome assembly from F1 hybrid of male American plains bison crossed with female Simmental cattle. NCBI BioProject. ID PRJNA677946 https://www.ncbi.nlm.nih.gov/nuccore/PRJNA677946

**Shirazi S** (2021) Determining sex of archaeological bison materials using shotgun sequencing. NCBI BioProject. ID PRJNA748091 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA748091

# Peer reviews

## Reviewer #1 (Public review):

Summary:

This study used whole genome data to investigate Beefalo ancestry for the first time, filling the gap in the field of Beefalo ancestry. The authors used preserved semen samples to generate genomic data on 47 registered Beefalo and 3 bison hybrids, further questioning the ABA's stated goal of ⅜ bison ancestry. In addition, the authors also show that ancestry profiles of Beefalo and bison hybrid genomes are consistent with repeated backcrossing to either parental species, demonstrate the value of genomic information in examining gene flow between species in the genus Bison. Overall, these data thus demonstrate the utility of genomic information in validating specific breeding claims for a more complete understanding of gene flow and genetic variation among bovine species. This is an interesting study, but there are still some major weaknesses that exist.

Strengths:

Numerous genetic analysis methods such as PCA, ADMIXTURE, F4 ratios, and local ancestry inference techniques revealed that no single Beefalo set meets the ancestry requirements set by the American Beefalo Association (ABA) and some beefalo had detectable indicine cattle ancestry.

Comments on revised version:

The authors have made further revisions in the revised manuscript, and these revisions have undoubtedly helped improve the article. No further comments.

https://doi.org/10.7554/eLife.102750.2.sa3

## Reviewer #2 (Public review):

Summary:

Shapiro et al. set out to verify the American Beefalo Association's claim that Beefalo cattle possess 37.5% bison ancestry. They employ a comprehensive range of well-established population genomics methods to estimate ancestry in these hybrid populations, including PCA, ADMIXTURE, D and F statistics, and local ancestry inference. Their findings conclusively demonstrate that most Beefalo lack the claimed bison ancestry, with only 8 out of 47 samples showing any detectable bison ancestry, ranging from 2-18%.

Strengths:

The primary strength of this analysis lies in the comprehensive dataset available to the authors, which includes important foundational Beefalo individuals and various reference populations. The rigorous and multi-faceted methodological approach employs several well-established techniques in population genomics for detecting and measuring admixture. Each method used has a firm basis in the field, providing consistent and robust results. The authors' approach of using PCA to initially assess the data within a global context, followed by more specific analyses using ADMIXTURE and D-statistics, provides a clear and logical progression of evidence. The presentation of these results in figures is particularly effective, clearly illustrating the key findings of the study. Additionally, the examination of both autosomal and sex chromosome ancestry offers a more complete understanding of Beefalo genetic composition and the mechanics of bison-cattle hybridisation.

Weaknesses:

One limitation of this analysis is the relatively low coverage (~2x) of many Beefalo samples. However, the authors have taken steps to mitigate biases that may arise from this, and their downsampling experiment demonstrates that this level of coverage is appropriate for summarising species-level ancestry across Bos. Another potential weakness is the limited sampling of contemporary Beefalo populations, as the study focuses primarily on historical samples. The authors have justified this choice on the grounds that contemporary Beefalo breeding involves no further bison input, so founder-era individuals are the most informative samples for addressing the study's central question.

Appraisal:

The authors have clearly achieved their primary aim using a rigorous and comprehensive methodology. Their extensive dataset and multi-faceted analytical approach provide strong support for their conclusions. The study not only addresses its main research question but also reveals unexpected insights into Beefalo genetics, particularly the presence of zebu ancestry, predominantly from Brahman cattle.

Discussion:

This study is valuable for several reasons beyond its primary findings. First, it definitively addresses and refutes the claim of 37.5% bison ancestry in Beefalo, providing crucial information for those studying these interspecies hybrids and the viability of their offspring. Second, it reveals the unexpected presence of zebu ancestry, predominantly from Brahman cattle, in many Beefalo, raising intriguing questions about the breed's development and the potential role of zebu cattle in achieving desired traits. This finding suggests that the distinctive appearance of Beefalo may be due in part to zebu admixture rather than bison ancestry. Third, the study highlights the significant barriers to admixture between bison and cattle, both in controlled breeding programs and potentially in wild populations. This has important implications for conservation genetics and our understanding of gene flow between these species. Lastly, the study demonstrates the power of genomic analysis in verifying breed claims and understanding the complex history of domestic animal breeds. These findings open new avenues for research in bovine genomics, breed development, and the dynamics of interspecies hybridisation.

Comments on revised version:

Thanks for the responses, which address my comments in full. I have no further concerns.

https://doi.org/10.7554/eLife.102750.2.sa2

## Reviewer #3 (Public review):

Summary:

The American beefalo cattle breed was developed as a mixture of 5/8 domestic cattle and 3/8 (or 37.5%) bison ancestry. The authors sequenced 50 genomes from bison and hybrids (historical and present-day). They found that most animals did not carry any detectable bison ancestry, with only a few between 2-18%, while other beefalo had taurine/zebu cattle ancestry, which may explain morphological traits. Breeding design was likely each time to a parental instead of to other admixtures.

The authors utilize whole genome sequence data to explore the ancestry of beefalo with respect to expected and possible contributions from cattle lineages. Using molecular and analytical methods central to questions exploring genomic ancestry and identity, the authors very nicely show evidence that calls into question ability of ancestry to be deduced from breed club documentation without considering reproductive challenges that are known in hybridization between cattle lineages.

Comments on revised version:

The authors have addressed all my comments to help improve presentation of specific details, results, and readability. Thank you!

https://doi.org/10.7554/eLife.102750.2.sa1

## Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *This study used whole genome data to investigate Beefalo ancestry for the first time, filling the gap in the field of Beefalo ancestry. The authors used preserved semen samples to generate genomic data on 47 registered Beefalo and 3 bison hybrids, further questioning the ABA's stated goal of ⅜ bison ancestry. In addition, the authors also show that ancestry profiles of Beefalo and bison hybrid genomes are consistent with repeated backcrossing to either parental species, demonstrating the value of genomic information in examining gene flow between species in the genus Bison. This is an interesting study that still has some major weaknesses that exist, but overall, the work demonstrates the utility of genomic information in validating specific breeding claims for a more complete understanding of gene flow and genetic variation among bovine species.*

We thank the reviewer for their thoughtful assessment of our work.

> *Strengths:*
>
> *Numerous genetic analysis methods such as PCA, ADMIXTURE, F4 ratios, and local ancestry inference techniques revealed that no single Beefalo set meets the ancestry requirements set by the American Beefalo Association (ABA) and some beefalo had detectable indicine cattle ancestry.*
>
> *Weaknesses:*
>
> *While this study contributes to our knowledge of Beefalo ancestry, there are some key issues that need to be addressed in terms of analysing the specific results as well as writing the article.*

We have followed the reviewer's suggestions for improving our study in detail (specified below), and appreciate their close reading of the manuscript.

> ***Reviewer #2 (Public review):***

> *Summary:*
>
> *Shapiro et al. set out to verify the American Beefalo Association's claim that Beefalo cattle possess 37.5% bison ancestry. They employ a comprehensive range of well-established population genomics methods to estimate ancestry in these hybrid populations, including PCA, ADMIXTURE, D and F statistics, and local ancestry inference. Their findings conclusively demonstrate that most Beefalo lack the claimed bison ancestry, with only 8 out of 47 samples showing any detectable bison ancestry, ranging from 2 - 18%.*

We thank the reviewer for their thoughtful assessment of our work.

> *Strengths:*
>
> *The primary strength of this analysis lies in the comprehensive dataset available to the authors, which includes important foundational Beefalo individuals and various reference populations. The rigorous and multi-faceted methodological approach employs several well-established techniques in population genomics for detecting and measuring admixture. Each method used has a firm basis in the field, providing consistent and robust results. The authors' approach of using PCA to initially assess the data within a global context, followed by more specific analyses using ADMIXTURE and D-statistics, provides a clear and logical progression of evidence. The presentation of these results in figures is particularly effective, clearly illustrating the key findings of the study. Additionally, the examination of both autosomal and sex chromosome ancestry offers a more complete understanding of Beefalo genetic composition and the mechanics of bison-cattle hybridisation.*
>
> *Weaknesses:*
>
> *One limitation of this analysis is the relatively low coverage (~2x) of many Beefalo samples. However, the authors have taken steps to mitigate biases that may arise from this. Another weakness is the limited sampling of contemporary Beefalo populations, as the study focuses primarily on historical samples. This may limit our understanding of how Beefalo genetics may have changed over time.*

The reviewer is correct that the low coverage obtained for many Beefalo is one potential limitation, although we believe that the downsampling experiment we performed (Fig. S4) shows that this level of coverage is appropriate for summarizing species-level ancestry across *Bos*, as the reviewer notes.

Sampling contemporary Beefalo individuals would be valuable, though as the focus of our study was to understand the origins of bison ancestry in Beefalo, we prioritized sampling individuals which played an important role in establishing the breed. We also note that contemporary Beefalo breeding involves crossing between Beefalo individuals or backcrossing to cattle, with no additional bison ancestry input since the formation of the Beefalo. As such, sampling individuals that existed close to the breed’s founding should provide the most insight into bison ancestry in Beefalo.

> *Appraisal:*
>
> *The authors have clearly achieved their primary aim using a rigorous and comprehensive methodology. Their extensive dataset and multi-faceted analytical approach provide strong support for their conclusions. The study not only addresses its main research question but also reveals unexpected insights into Beefalo genetics, particularly the presence of zebu ancestry.*
>
> *Discussion:*

> *This study is valuable for several reasons beyond its primary findings. First, it definitively addresses and refutes the claim of 37.5% bison ancestry in Beefalo, providing crucial information for those studying these interspecies hybrids and the viability of their offspring. Second, it reveals the unexpected presence of zebu ancestry in many Beefalo, raising intriguing questions about the breed's development and the potential role of zebu cattle in achieving desired traits. This finding suggests that the distinctive appearance of Beefalo may be due in part to zebu admixture rather than bison ancestry. Third, the study highlights the significant barriers to admixture between bison and cattle, both in controlled breeding programs and potentially in wild populations. This has important implications for conservation genetics and our understanding of gene flow between these species. Lastly, the study demonstrates the power of genomic analysis in verifying breed claims and understanding the complex history of domestic animal breeds. These findings open new avenues for research in bovine genomics, breed development, and the dynamics of interspecies hybridisation.*
>
> ***Reviewer #3 (Public review):***
>
> *Summary:*
>
> *I really like this topic and study. But I think much can be more focused and tightened up. All the components are here - just some more refining to really make the storyline clear, the journey of discovery, and the impact of such knowledge.*

We thank the reviewer for their thoughtful assessment of our work.

> *Strengths:*
>
> *The authors dive directly into the question of genomic ancestry as compared to the breed club's reported ancestry with heavy, quantitative data and critical analytical methods. The questioning line is direct and does not meander. The reader learns about the challenges of breeding associations, and values of understood ancestry, and presents a clear need of re-evaluating the breed standards and expectations of beefalo (if ancestry is indeed the primary goal instead of a phenotype-driven breed mission).*
>
> *Weaknesses:*
>
> *Much of the quantitative results are only referred to in the main text with qualitative language. Please incorporate more written quantitative results to highlight evidence that underlines the study narrative because it is quite an interesting study!*

The reviewer highlights an important point, and we agree that the qualitative language used to describe the results was generally lacking. We have now described the results quantitatively throughout the manuscript where possible.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations for the authors):***
>
> *(1) This study is not the first to question claims surrounding bison ancestry in the breed and is the sample size too small to be representative of the entire genetic structure of Beefalo?*

The reviewer correctly points out that this study is not the first to address uncertainty in the amount of bison ancestry present across beefalo. All earlier studies, to our knowledge, have been highlighted in the introduction and discussion (Lenoir and Lichtenberger, 1978 and Stormont et al, 1986). However, these studies examined a narrow range of Beefalo sources and used older methods (karyotyping and blood typing), such that comprehensive statements about the proportion of bison ancestry in Beefalo could not be made.

We also agree that an appropriate sampling scheme is crucial for making definitive statements about Beefalo ancestry across the breed. As Beefalo breeding typically involves breeding select “full-blood” individuals with cattle, the ancestry across contemporary Beefalo is likely complex, with the cattle component coming from a wide range of breeds. Therefore, our sampling emphasized “full-blood” representatives, especially those that were involved in the founding of the breed and from which later Beefalo descend. This involved an exhaustive survey of the Beefalo individuals contained within the USDA’s National Animal Germplasm Program. Although we did not extensively evaluate current Beefalo diversity, we believe this approach is most suited for characterizing bison ancestry within Beefalo, as bison ancestry is maintained primarily through the continued use of genetic material from these “full-blood” individuals rather than repeated hybridization between bison and cattle.

> *(2) Although genomic information is important for breeding research, this requires quality of data. The coverage of the data used in this study was mainly ~2X, and although multiple methods of analysis gave similar results, the ability to identify rare variants (e.g. insertions or deletions of long segments of the genome) may be limited at low coverage, affecting the confidence of the results.*

This is an important consideration, and we agree with the reviewer that the sequencing depth obtained for most individuals in our study precludes accurate genotype calling. Therefore, we did not attempt to perform traditional genotype calling. Rather, we used a pseudohaploid calling approach in which a random base was selected to represent the genotype at each position for each individual, using a pre-ascertained set of variants discovered in gaur, a closely related outgroup to bison and cattle. This pseudohaploid approach is common in other situations where coverage is low, for example in analyzing ancient DNA.

Furthermore, our ancestry analyses focused on biallelic SNPs which were discovered in gaur and we did not attempt to call structural variants, given the limitations in coverage. As this outgroup ascertainment approach seeks to target SNPs which were polymorphic in the ancestor of both bison and cattle, which should yield unbiased results in population genetic analyses, we were less interested in discovering rare variation within the species and populations we examined here.

Finally, we performed downsampling experiments comparing low coverage read data to genotypes called from high coverage data, and obtained consistent results between low and high coverage analyses using read-level data and called genotypes (Fig. S7).

> *(3) Missing from the conclusions is the very important presentation of the results of genomic calling, the basics of what these data look like, coverage histograms, number of SNPs, categorization, annotations, and so on. These are necessary prerequisites for subsequent population analysis.*

The reference to “5.29M” on page 14 has been replaced with the exact number of SNPs used in analyses (5,291,534). The average sequencing depth for each sample is also included in Table S1.

> *(4) The manuscript mentions "most" in a number of places, but can the authors give an accurate number based on the current data? "Most" is not a rigorous description. Based on the simulations of genomic data, how many Beefalo cattle were not detected as hybridized? This may be related to both sample size and where the authors sampled.*

We thank the reviewer for this important suggestion. We have now replaced vague summaries of results with precise numbers. However, we are unsure what “simulations” means in this context, as all results were obtained by analyzing empirical data from Beefalo, bison, cattle, and other bovines, rather than simulations.

> *(5) The information in the third and fourth paragraphs of the Introduction is not sufficiently coherent and could be further consolidated into a more logical presentation.*

We have now condensed these paragraphs and edited them for clarity.

> *(6) "For some analyses we also incorporated published genomes from outgroups". The description here is unclear as to what criteria were used to select these data, and it is possible that the choice of outgroups could lead to different conclusions from the analyses. In addition, ancient DNA data from cattle may be useful for this study and the authors are encouraged to consider it.*

Outgroup choice can certainly have a large impact on population genetic analyses. For the species examined in our study, we considered other *Bos* species, including yak, gaur, and banteng, as suitable outgroups, along with water buffalo, which is the closest outgroup outside of *Bos*. We have added comparisons of *D*-statistics using yak as an outgroup as a supplementary figure (Fig. S4), in addition to those using water buffalo as the outgroup which were presented in Figure 2.

As we were examining species-level ancestry, and given the high level of divergence between bison and cattle, relative to that between published ancient and modern cattle genomes, we believed that it was most appropriate to use high quality modern cattle data, rather than poorer quality ancient cattle genomes, for analyses. Additionally, as any hybridization which took place between bison and cattle in the formation of Beefalo would have occurred within the past ~50 years, modern cattle are likely to be the most appropriate proxy for the cattle ancestry in Beefalo, especially given the lack of published historical North American cattle genomes.

> *(7) The coordinates of the PCA plot need to be further supported by providing values.*

We have now updated axis labels for the PCA in Fig. 1A to include the proportion of variance explained for the first two components.

> *(8) In Figure 1, Beefalo has one individual, NAGP9109, which belongs exclusively to the indicine group. For this individual, wouldn't it be nicer to label it separately in the PCA and ADMIXTURE plots, like Joe's Pride (JP), to make the presentation of the results clearer?*

This individual was one which was determined to be mislabeled as Beefalo within the NAGP and is actually a Brahman cattle. Therefore, we have relabeled it as zebu, rather than Beefalo, throughout the figures.

> *(9) As the sex chromosome data do not fully support the authors' claims, some caution may be needed in describing the results.*

We interpret the sex chromosomal results as being fully consistent with patterns seen in the autosomes. However, they do shed some light on the dynamics of bison-cattle hybridization, and suggest male-mediated gene flow in which bison ancestry in Beefalo was introduced primarily through bison bulls.

> *(10) Would it be appropriate to analyse the results at K = 3 only? The admixture analysis of all bison, cattle, bison hybrids, and buffalo individuals at different K values should further refine the results.*

We now also show ADMIXTURE results at K=2 and K=4 (Fig. S2) and present the cross-validation results from ADMIXTURE (Fig. S3).

> *(11) The conclusions of this article about bison ancestry in Beefalo individuals are completely inconsistent with the American Beefalo Association, and should a description of possible reasons for this discrepancy be added to the discussion?*

Our analyses make it clear that there was much less hybridization between bison and cattle leading to the formation of the Beefalo that was previously believed. As the genetic data does not provide insight into exactly why this might be the case, we can only speculate on the precise reasons bison-cattle hybridization did not take place, which we have avoided here.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *The manuscript is well written, the figures are easily understandable, and the claims made are justified by the results obtained.*
>
> *It is need to clarify cattle breeding terminology, particularly concerning breeds like the Brahman. While often described as zebu-taurine hybrids, Brahman cattle typically show over 90% zebu ancestry when analysed using ADMIXTURE against panels including European Bos taurus, African Bos taurus, and Bos indicus animals. This context would help explain why "NAGP9109" clusters with the Zebu group.*

We thank the reviewer for this useful context, and agree that most Brahman cattle have a high proportion of zebu ancestry. In fact, the zebu group we included primarily consists of Brahman individuals, which we have now clarified in the text, which now reads:

“The reported pedigree in the NAGP for this animal lists its composition as 1/2 Brahman, 1/4 Charolais, 1/8 bison, 1/16th Hereford, and 1/16th Shorthorn, but the American Brahman Breeders Association records this animal (#309519) as purebred Brahman, which is a zebu breed (5 of the other 6 zebu individuals analyzed here are Brahman cattle).”

> *I suggest three other improvements:*
>
> *(1) Standardise terminology: The manuscript alternates between "zebu" and "indicine" when referring to these cattle. While both terms are correctly defined in the introduction as "indicine (zebu; Bos indicus)" using one term consistently throughout would improve readability. I prefer "zebu" but leave this choice to the authors.*

We agree that this mixed terminology was confusing and have replaced all instances of “indicine” with “zebu.”

> *(2) Add PCA metrics, including the percentage of variance explained by each principal component would demonstrate the genetic distinctiveness between bison and cattle, and between Taurus and zebu cattle. This would also support the selection of K=3 for the ADMIXTURE analysis.*

The axis labels for the PCA have been updated to include the proportion of variance explained for each component. We now also show ADMIXTURE results at K=2 and K=4 (Fig. S2) and present the cross-validation results from ADMIXTURE (Fig. S3).

> *(3) Improve quantitative precision: The authors could improve precision by replacing qualitative statements with exact counts. For example "39 of 47 Beefalo showed no detectable bison ancestry." The same suggestion applies when describing how many Beefalo had zebu ancestry.*

We thank the reviewer for this useful suggestion, and agree that the manuscript used imprecise language in describing the results of certain analyses. We have now added quantitative detail throughout the Results section.

> ***Reviewer #3 (Recommendations for the authors):***
>
> *(1) Introduction*
>
> *The introduction sets a tone that is heavily focused on the genetic revelation that the economics of beefalo are somewhat of a facade. Beefalo are indeed not part-buffalo (bison). It is unclear to me if the introduction also could benefit from motivating this with more of a theoretical framework based on evolution, inheritance, or trait transmission. If this is really meant to be an economics-focused article, then lean more heavily into that. As it stands, it straddles a bit of economics, a bit of legacies that appear false (beefalo are not part bison at all!), and a bit of admixture genetics theory.*

We intended the focus of this study to be on documenting the species-level ancestry of Beefalo, and concentrated the information presented in the Introduction on this topic. Given that less hybridization between bison and cattle appears to have taken place to form the Beefalo breed than was previously described, we believe that broader theoretical statements about admixture are less relevant here, beyond highlighting examples of successful and failed interspecies hybridization in *Bos*. We also avoided speculating on the history of the establishment of the breed beyond what could be understood from the genetic data.

> *Can the authors give a bit more details about beefalo breeding? Did the breeders select for any quantitative traits and is there a targeted phenotype for beefalo they used as a standard?*

Limited information exists about the precise origins of Beefalo, which were never publicly shared—possibly in part for reasons this manuscript addresses. The only criteria defining Beefalo is the proportion of bison ancestry, and so no quantitative traits or specific phenotypes are related to breed standard.

> *Can the authors provide a few examples of what is known about the incompatibilities and reproductive challenges? What is known from past research or from the Beefalo Association documenting the breeding history?*

We provided a general summary of hybridization and incompatibility across *Bos*, but unfortunately cannot provide details about incompatibilities in Beefalo specifically. Though there is a long history of challenges interbreeding bison and cattle (referenced in the third paragraph of the Introduction), to our knowledge no examination has been carried out of Beefalo specifically and little is known about Beefalo pedigrees (again, perhaps for reasons related to information presented in this study).

> *(2) Results Section Sequencing Beefalo genomes*
>
> *Please report the number of polymorphic sites to accompany the genomic read depth averages. It seems the authors could include a larger summary of the genomic data that was used for downstream analyses (like the PCA in the next section). Also, does this dataset include the sex chromosomes? How many variants that are retained for analyses are autosomal, sex-linked, or haploid? Please provide more characteristics of the data that was generated after QC and filtering.*

We have now replaced “5.29M” on page 14 with the exact number of SNPs (5,291,534) and added a description of genotype calling to the Results section. We have also included the number of SNPs used for sex chromosomal analyses.

> *(3) Results section Estimating bison ancestry in beefalo*
>
> *What is a "foundational" individual? Is this a beefalo pedigree founder, a common sire, or an individual with remarkably high bison content? I see in the introduction Joe's Pride*

> *was the "most expensive cattle" but there are surely other aspects of "foundational" that the reader should understand as the results are presented.*

We agree that this terminology was imprecise, and have now clarified that we use foundational to mean an early individual that was important in the founding of the Beefalo breed, such as those that were first bred by Bud Basolo.

> *For the sentence "The reported pedigree in the NAGP for this animal [NAGP9109] lists its composition as 1⁄2 Brahman, 1⁄4 Charolais, 1⁄8 bison, 1/16th Hereford, and 1/16th Shorthorn, but the American Brahman Breeders Association records this animal (#309519) as purebred Brahman.", this is difficult for a reader with limited cattle breed knowledge to infer significance of this. What is the origination of Brahman breed cattle? Does Brahman ancestry come from another mixed origin that could explain this discrepancy? Does the PCA have references to resolve the origin of Brahman? I realize this may sound extraneous but if membership to a breed that is recently formed from several other lineages or breeds, could you be seeing the deeper parts that compose Brahman cattle? How could one validate that the contributors erroneously labeled this individual as a beefalo?*

We have now noted that the Brahman breed has primarily zebu ancestry. The placement of this individual in the PCA supports the American Brahman Breeders Association metadata, and suggests that the NAGP labeling is incorrect:

“The reported pedigree in the NAGP for this animal lists its composition as 1/2 Brahman, 1/4 Charolais, 1/8 bison, 1/16th Hereford, and 1/16th Shorthorn, but the American Brahman Breeders Association records this animal (#309519) as purebred Brahman, which is a zebu breed (5 of the other 6 zebu individuals analyzed here are Brahman cattle). We believe NAGP9109 was erroneously labeled as Beefalo by the contributors.”

> *Figure 1A: Please add % explained by each PC.*

We have now updated axis labels for the PCA to include the proportion of variance explained for each component.

> *Figures 1B and 1C are identical except for the Y axis. Please combine them into a graph with 2 Y-axes (one for PC1 and one for ADMIXTURE). Also, please include the bison in this panel as well.*

We have now updated these panels to include bison, although have kept the labeling so that they may be referenced separately in the text.

> *I see that the authors did both unsupervised and supervised. Can the main text have the supervised graphical result instead of the unsurprised? That is more relevant for ancestry proportions via an assignment probability to ancestry groups. Or, if possible, could the authors consider STRUCTURE to also obtain the probability of assignment to a prior defied parental up to 2-generations back? This is by far the best way to leverage the ancestry information of the cattle and bison parental references in addition to the known F1/bison hybrids. Swap the Supplementary Figure 1 with Figure 1D!*

The supervised and unsupervised ADMIXTURE results are highly consistent, as could be expected given the high levels of divergence between species. We prefer to show the unsupervised results in the main text, as this makes the fewest assumptions about the ancestry of the examined individuals, and so also shows that the panels used to represent each species (taurine cattle, zebu cattle, and bison) do not contain individuals which were themselves highly admixed, which could have influenced the supervised ADMIXTURE analyses.

> *For the unsupervised ADMIXTURE analyses, what were the cross-validation values per K value tested? How did the authors decide that K=3 was the best one to show?*

We now also show ADMIXTURE results at K=2 and K=4 (Fig. S2) and present the cross-validation results from ADMIXTURE (Fig. S3).

> *Regarding "D-statistics ..... are consistent with 0 for most individual Beefalos....", I have two comments. First, by "consistent with", do you mean "are not significantly different from 0", indicating that (explain what this means in your words). Next, "most individual beefalos" means how many? Please provide numbers and values to highlight points or specific findings.*

The interpretation of the D-statistics has been clarified and Z-scores and numbers of individuals to quantitatively describe these results have been added. The text now reads:

“*D*-statistics of the form *D* (*taurus*, *Beefalo*; *bison*, *water buffalo*), which test whether Beefalo share more alleles with bison than taurine cattle, again show 39 Beefalo have no excess affinity with bison compared to taurine cattle ($-13.04 < Z < 3.14$), although the same eight Beefalo identified in PCA and ADMIXTURE as having bison ancestry also have an excess of bison alleles ($6.16 < Z < 34.86$), confirming their bison ancestry (Fig. 2A).”

> *"In Beefalo with bison ancestry, that ancestry tends to be present in large contiguous blocks, often tens of megabases in size, indicative of recent admixture (Figure 3A, B)". Please display the quantitative results (mean, max, range, standard deviation, etc.) in the main text and point the reader to the table that contains the values for each individual. The rest of this paragraph also uses the words "most' or "always" - please provide numbers. Is most 30/46 beefalo? Is it always exactly all 47 beefalo? Readers want to see numbers!*

The reviewer is correct that this section lacked specificity. We have now provided the exact number of individuals identified with bison and zebu ancestry.

> *The section starting "Several lines of evidence attest to the efficacy of using these source panels..." could realistically come first in the Results section and before beefalo results are presented. This would build confidence for the reader that this panel of samples passes a QC and will indeed be able to resolve ancestry-based questions.*

This section specifically refers to the local ancestry analyses, which we have now clarified in the text.

> *Figure 3A-C: Please include on each of these figure panels the documented (breeder association) ancestry percentage and the percentage of bison ancestry you obtained from your genomic analyses. Moving it from the legend to the figure is more immediately powerful for the reader. If the authors dated the admixture events as well, please include the meta-data of the association pedigree reporting when bison entered the target individual's genome versus the genome-estimated number of generations since admixture.*

Figure 3 has now been updated to include the reported bison ancestry. No attempt was made to date the admixture event or compare with reported pedigrees, as documented Beefalo pedigrees are typically very sparse (and may be unreliable, as our results suggest).

> *Figure 3 legend: Move the following text from the figure legend to the Results section: "Three bison hybrids are inferred to have ~75% bison ancestry, while eight Beefalo have detectable bison ancestry, ranging from 2-18%. Indicine ancestry is detected in most Beefalo at variable levels, ranging from 2-38%, with most Beefalo having between 2-18%.".*

This sentence has been removed from the legend and is now worked into the main text. The corresponding paragraph in the results now reads:

“Local ancestry inference across individual Beefalo and bison-cattle hybrid genomes provides similar estimates of overall Beefalo ancestry, inferring an absence of bison ancestry across the 37 Beefalo that lacked evidence for such ancestry in previous analyses (Fig. 3). Three bison hybrids are inferred to have ~75% bison ancestry, while eight Beefalo have detectable bison ancestry, ranging from 2-18%. Zebu ancestry is detected in 38 Beefalo at variable levels, ranging from 2-38%, with all but two of Beefalo having between 2-18%.”

> *(4) Results section Beefalo sex chromosome ancestry*
>
> *Check that the authors do not reference Figure 4B before Figure 4A.*

Thank you to the reviewer for noticing this, it has now been corrected.

> *Figure 4A: Could this panel be considered to merge with the autosomal admixture plot? It helps with comparison. Not a firm request - but it is nice to see what is consistent versus what is discordant.*

To avoid cluttering the figure with two highly similar plots, we preferred to separate the autosomal and sex chromosomal results.

> *Figure 4C: Could this panel be merged with the autosomal ancestry bar graph to help the reader with visual comparisons?*

We thank the reviewer for this suggestion, but do not understand exactly which figures they are suggesting to be merged.

> *(5) Materials and Methods: Modeling Beefalo ancestry:*
>
> *The language used in this sentence "This approach allows for directly understanding the ancestry of Beefalo individuals relative to these three groups while mitigating the effects of the low sequencing depth obtained for many Beefalo." conflicts with a sentence later in this paragraph which called PCA a model-free analysis. Please correct.*

Unfortunately, we are unsure what the reviewer refers to here and believe that this sentence does not conflict with the characterization of PCA as a model-free analytical approach.

https://doi.org/10.7554/eLife.102750.2.sa0