Gonzalez Hernandez, J., Boehm, F., Martin, J., Gonda, M., Urso, P. (2026) Advancing bison (*Bison bison*) genomics to strengthen herd and metapopulation health and resiliency, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23973: https://doi.org/10.31274/wcgalp.23973 ©2026 Gonzalez Hernandez, et al. CC BY-NC-ND

Advancing bison (*Bison bison*) genomics to strengthen herd and metapopulation health and resiliency

J.L. Gonzalez Hernandez[1*], F. Boehm[1], J.M. Martin[1], M.G. Gonda[1], and P. M. Urso[1]

[1]*South Dakota State University, Brookings, USA;* [*]*Corresponding author:* Jose.Gonzalez@sdstate.edu

Introduction

With growing importance in sustainable food production, the US bison industry is estimated to be a ~$1 billion/year business (McFarlane, 2025). Despite their deep cultural and historical roots, fewer modern genomic and herd improvement tools are available for bison than many other livestock industries. Unlike cattle and pigs, bison producers do not have access to breeding values or DNA tests that predict production, health, or longevity. Existing genetic tools for bison are based on microsatellite loci and single nucleotide polymorphism (SNP) loci identified from only a few bison (*Bison bison*) or in other species. Microsatellite loci are effective for parentage testing but have limited efficacy for many other applications and are relatively more laborious to genotype than SNP. A bison SNP panel has been developed (Stroupe and Derr, 2024), which is a major step forward for developing bison genetic tools. However, this panel only includes SNP identified first in other species (primarily cattle) or the bison reference genome. This project aimed to close this gap in cataloging genetic variation between bison and many other livestock species. Here we report two new near-complete genome assemblies (from male and female animals) for bison annotated with RNASeq data. We also report the sequencing of 93 individual bison from the South Dakota Custer State Park (CSP) herd using short reads, the discovery of high-quality single nucleotide polymorphisms (SNPs) across the genome, and the corresponding genetic diversity analysis in the CSP herd using those SNPs.

Materials & Methods

Tissue samples. All animal sampling procedures were approved by the SDSU Institutional Animal Care and Use Committee. We obtained liver samples from an adult male and two adult female bison from private herds in South Dakota and 93 ear clips from the CSP herd. From one of the female bison, we also obtained other samples for RNAseq analysis.

DNA and RNA extraction. DNA and RNA extractions were done using the DNeasy-Blood and Tissue kit and the RNeasy Plus Mini kit from Qiagen following manufacturer's protocol.

Sequencing. All libraries and sequencing were done by the South Dakota State University Genomics Sequencing Facility. Sequencing libraries for Oxford Nanopore for the gDNA samples from liver were generated using ligation-based chemistry (LSK114 v.14) and sequenced using an PromethIon sequencer (Oxford Nanopore. Oxford, UK) to obtain ~130-140 Gb of data (~40-45X coverage) for each sample.
Short-read DNA sequencing libraries for 95 (93 plus the 2 samples used for genome assembly) individuals were prepared using Illumina DNA Prep (Illumina. San Diego, CA) kit f and sequenced in 8 lanes (2 lanes per set of 24 samples) of a 25 billion reads flow cell (2x150bp)

Illumina NovaSeq X Plus. RNAseq libraries were prepared using Illumina stranded mRNA kit and sequenced in a partial lane of an Illumina NovaSeq X Plus sequencer in a 2x150bp configuration.

De-novo genome assemblies. Nanopore long reads were QC in CLC Genomics Workbench 24.0.1. Two assemblers were tested: CLC Genomics Workbench's De Novo Assemble Long Reads tool and Hifiasm 0.25.0 (Cheng *et. al,* 2021; Cheng *et al.,* 2024). Assemblies were evaluated for continuity and completeness using QUAST 5.2.0 and BUSCO 5.6.1. Assemblies from Hifiasm were selected and aligned to GCA_030254855.1_TAMU 3.0 bison genome assembly (Stroupe *et al*., 2023) originated from a beefalo (beef cattle-bison hybrid) for scaffolding, orientation and chromosome assignment using D-Genies (Cabanettes and Klopp, 2018). The final assemblies were evaluated for presence of the telomere repeat (TTAGGG) using Tidk-0.2.65 (Brown *et al.,* 2025) and the presence of the bovine centromere repeat using BLAST against NCBI accession V00125 (Pluciennicza*k et al.,* 1982) in CLC Genomics Workbench.

Genomes annotation. Annotation was done using CLC Genomics Workbench 26.0.1 Transcript discovery tool using RNAseq reads large-gap mapped to the newly assembled genomes. A total of 1,948,883,422 paired-reads from NCBI SRR1659047-SRR1659070 (2x100bp) from 6 tissues and three additional samples sequenced (2x150bp) as part of this project from 3 tissues were large-gapped mapped to both male and female assemblies using minimum overlap of 90% and a minimum similarity fraction of 90%.

SNP calling and filtering. After QC in CLC Genomics Workbench Illumina paired reads from each of the 95 individuals from the herd in question were mapped to the male genome assembly described in this manuscript and variants called using a custom workflow in the LightSpeed module in CLC Genomics. Individual VCF files were merged using VCFtools (Danecek *et al.,* 2011), variants were filtered preserving only highest quality biallelic SNPs with no missing data.

Population genomics analysis. Runs of homozygosity (ROH) were detected using a custom sliding-window algorithm implemented in R with genotype data accessed via the SNPRelate package. Quality-controlled SNP genotypes were converted from PLINK BED format to GDS, and non-autosomal chromosomes were excluded using a regular expression filter. For each individual, the algorithm scanned genotypes along each chromosome, initiating a candidate ROH segment at each homozygous site and extending it until the run was terminated by exceeding the tolerance thresholds: a maximum of 1 heterozygous genotype and 1 missing genotype within a run, or an inter-SNP gap exceeding 1 Mb. A segment was retained as a ROH only if it contained at least 50 consecutive SNPs and spanned a minimum of 500 kb. The genomic inbreeding coefficient (F_ROH) for each individual was calculated as the sum of all ROH segment lengths divided by an assumed autosomal genome length of 2.5 Gb.

Results & Discussion

De novo genome sequencing and assembly. Sequencing of ONT libraries for the male and female individuals yielded 133 Gb and 139 Gb of data, respectively, providing approximately 43x and 46x coverage. We compared two long-read assembly strategies: CLC Genomics Workbench (Raven-based) and Hifiasm. For the male sample, Hifiasm produced a more contiguous primary assembly with 1,676 contigs compared to 1,875 from CLC. Similarly, for the female sample, Hifiasm generated 1,404 contigs versus 2,072 from CLC. Based on these metrics, Hifiasm assemblies were selected for downstream processing. Primary contigs were aligned to

the *Bison bison* reference genome (GCS_030254855.1_TAMU 3.0; Stroupe *et al*., 2023) for scaffolding and orientation. Following curation, we obtained a 3.08 Gb male assembly and a 3.05 Gb female assembly. Both genomes demonstrated exceptional contiguity; the male assembly achieved an of 104.4 Mb (auN= 95.7 Mb) and the female reached 105.5 Mb (auN = 98.2 Mb). Notably, the L90 values (30 for male, 29 for female) were highly concordant with the expected haploid chromosome number for *Bison bison* (n=30), indicating that nearly all autosomes and sex chromosomes were recovered as single, chromosomal-scale scaffolds. Both assemblies exhibited high gene-space completeness, with 93.6% of mammalian BUSCO genes identified as complete. Checking for potential contaminants in the assemblies using FCS-GX returned one and two hits in the male and female genome assemblies respectively; all three hits were small unaligned contigs (1,572 bp in the male and 48,710 bp and 1,616 bp in the female) that were removed from the assembly.

The male bison assembly yielded 31 major scaffolds, effectively resolving the 29 autosomes and the X/Y sex chromosome complement (2n=60). Based on the presence and density of telomeric repeats ($TTAGGG_n$), scaffolds were categorized into three distinct structural groups, complemented by a complete mitochondrial genome (Table 1). **Group 1** comprises 13 Telomere-to-Telomere (T2T) chromosomes, including the X-specific body (chrX_1), representing fully resolved sequences capped at both termini. **Group 2** consists of 16 Telomere-to-Centromere (T2C) fragments; these represent nearly complete q-arms of acrocentric autosomes where the assembly successfully traversed the distal telomere but was arrested by highly repetitive centromeric regions or truncated p-arms. **Group 3** includes two internal fragments (2 and chrX_2). The lack of telomeric caps and the presence of low-frequency internal telomeric hits on suggest interference from Internal Transcribed Spacers (ITS) or large ribosomal DNA (rDNA) arrays. Finally, (~21.14 Mb) is identified as a likely collapsed, chimeric assembly of the X-chromosome Pseudoautosomal Region (PAR) and the Y-chromosome. The absence of telomeric caps in this region is consistent with the high recombination rates and sequence homology shared between the X and Y chromosomes.

Table 1. Summary of chromosome assembly status.

Group	Count	Chromosome / Scaffold Names	Status & Evidence
Group 1: T2T	**13**	**5, 10, 11, 12, 19, 21, 22, 23, 25, 26, 28, 29, X_1**	**Complete.** These have telomeric hits at both ends. Under a broad definition, **X_1** and **11, 21, 25** are included because they show hits at both distal and proximal ends, even if distances are slightly higher.
Group 2: T2C	**16**	**1, 3, 4, 6, 7, 8, 9, 13, 14, 15, 16, 17, 18, 20, 24, 27**	**Large Fragments.** These are "one-ended" (Distance=0 at one tip, the q-arm in bovids acrocentric chromosomes) with high hit counts (>150). These are missing their centromeric "p-arm" caps.
Group 3: internal fragment	**2**	**2, X_2**	**Incomplete/Noise. chr2** is a massive internal scaffold with no caps (~10 hits only likely an internal transcribed spacers (ITS)). **X_2** is a small fragment (~21 Mb) with low hits, likely representing a difficult part of the X chromosome or a collapse assembly of the Pseudoautosomal Region (PAR) common to the X chromosome and the Y chromosome.

Preliminary examination of linkage disequilibrium maps revealed that LD varied within and among chromosomes. Approximately 180,331 high-quality SNP (QUAL ≥ 99) with ≤ 2 individuals missing data per locus were identified from the 95 bison that were sequenced. The average inbreeding coefficient based on runs of homozygosity was 0.58 (minimum bp length = 100,000 bp). This level of homozygosity suggests loss of genetic diversity of the bison herd at Custer State Park. This loss of genetic diversity is concerning because of the potential effects of inbreeding depression and lower genetic variation, which would lead to slower rate of genetic change over time. This level is considerably impactful as Custer State Park is considered a closed herd in which no animals enter the breeding pool. This is a common practice in the industry within exceptionally large herds as risk of disease transmission through outside animals is a major risk to all bison in the U.S. This discovery in CSP shows that large herds are still susceptible to genetic bottlenecks. The loss of genetic diversity also may lead to less resilience of this bison herd in the face of environmental or disease challenges. One way to combat this loss of diversity is by introducing bison from other herds that are genetically the most distant from bison at Custer State Park. This strategy would necessitate sampling the bison genomes in conservation herds throughout North America. This work shows that understanding the interrelatedness across herds, including smaller producers, is necessary to improve the decision making of all bison producers. Using these sequenced herds, producers are more freely able to trade animals with a distinct purpose of cycling unrelated animals, thus improving the genetic diversity of their herd.

Conclusions

These results showcase the need to develop genomics tools in bison and their application in herd management decisions. Extensive ROH and the suggested loss of genetic diversity can lead, among other consequences, to increased herd vulnerability to diseases. Our future work includes the development of a phased genome, and extended pangenome by adding data from several more herds across the country and the development of low-cost assays to assess genetic diversity of herds in key chromosomal regions.

Acknowledgements

This work was funded by a grant from the South Dakota Game, Fish, and Parks Department and by the South Dakota Agricultural Experiment Station.

References

Astashyn, A., Tvedte E. S., Sweeney D., *et al*. (2024) Genome Biol. 25:60. https://doi.org/10.1186/s13059-024-03198-7

Brown M.R., Gonzalez de La Rosa P.M., and Blaxter M. (2025) Bioinformatics 41:btaf049. https://doi.org/10.1093/bioinformatics/btaf049

Cabanettes F., and Klopp C. (2018) PeerJ 6:e4958. https://doi.org/10.7717/peerj.4958

Danecek P., Auton A., Abecasis G., *et al*. (2011) Bioinformatics 27:2156-2158. https://doi:10.1093/bioinformatics/btr330.

Danecek P., Bonfield J.K., Liddle J., *et al*. (2021) Gigascience 10:giab008. https://doi.org/10.1093/gigascience/giab008.

Plucienniczak A., Skowroński J., and Jaworski J. (1982) J. Mol. Biol. 158:293-304. https://doi:10.1016/0022-2836(82)90434-x.

Stroupe S., Martone C., McCann B., (2023) G3 (Bethesda) 13:jkad156. https://doi:10.1093/g3journal/jkad156.

Stroupe S., and Derr J.N. (2024) Evol. Appl. 17:e13658. https://doi:10.1111/eva.13658.

Vaser R., and Šikic M. (2021) Nat. Comput. Sci. 1:332-336. https://doi:10.1038/s43588-021-00073-4.