Fall 2018

Genetic Structure and Diversity of the Medano-Zapata Bison Herd using Microsatellite Data

Torrey E. Davis
University of Colorado Boulder, toda9483@colorado.edu

Genetic Structure and Diversity of the Medano-Zapata Bison Herd using Microsatellite Data

By:
Torrey Davis
Ecology and Evolutionary Biology, University of Colorado at Boulder

October 23rd, 2018

Thesis Advisors:
Dr. Andrew Martin, Ecology and Evolutionary Biology
Dr. Cheryl Pinzone, Ecology and Evolutionary Biology

Defense Committee:
Dr. Andrew Martin, Ecology and Evolutionary Biology
Dr. Cheryl Pinzone, Ecology and Evolutionary Biology
Dr. Barbara Demmig-Adams, Ecology and Evolutionary Biology
Dr. Erin Huebert, Political Science
Dr. Chris Pague, The Nature Conservancy

ABSTRACT

In the late 1800s, the American plains bison (*Bison bison*) suffered from a severe reduction in population size, known as a bottleneck event. Their numbers, once in the millions, were reduced to less than a thousand individuals, but populations were able to recover owing to the diligence of several private ranchers and other private and governmental groups. While bison populations have improved today, their genetic diversity has been impacted by the bottleneck event, habitat fragmentation, and hybridization with domestic cattle (*Bos taurus*). With advances in genetic technology, it is now possible to gather genetic information of a species to aid in their conservation. Here, using genetics software, I analyzed nuclear genetic data previously used to detect cattle genes in the Medano-Zapata bison herd located in Colorado, USA. I was able to determine (1) the number of genetic clusters and (2) calculate different indices of genetic variation to estimate population viability. I hypothesized that more than one cluster would occur within the population due to its large size. My results rejected this hypothesis, with the strongest likelihood supporting a single cluster. This information is important for understanding the genetic diversity of the herd. It is also important for understanding the impacts of cattle genes on bison populations and what they mean for the species' future conservation.

INTRODUCTION

The North American plains bison (*Bison bison*) is an iconic species of the American west that historically ranged from Canada to Mexico, and from the west to east coasts of North America (List *et al*. 2007; Sanderson *et al*. 2008). In the late 1800s, bison suffered from a significant population bottleneck, or severe reduction in their population. Once numbering in the millions, bison populations were reduced to less than a thousand individuals (Hedrick 2009) through overexploitation, habitat loss, and disease (Isenberg 2000). The species was saved from near extinction in the early 1900s through the diligence of five private ranching herds, a herd at the New York Zoological Park (Hedrick 2009), and a small, isolated herd in Yellowstone National Park (Meagher 1973). During this time, some of the private ranchers experimented with crossing bison and domestic cattle (*Bos taurus*) to promote disease resistance, increase meat quality and quantity, and increase feeding efficiency and hardiness in their cattle stock (Boyd 1914; Goodnight 1914). Female cows were often bred with male bison, and the hybrid offspring were repeatedly backcrossed with male bison, leading to cattle gene introgression in the bison herds (Hedrick 2009). Hybridization between two species may lead to extinction of one of the species (Allendorf *et al.* 2001), and hybridization with a species that has been selected for domestic traits (such as cattle) may impact local adaptation of the threatened or endangered species (Randi 2008).

Animals from these hybridized herds were then used to stock many public and private herds (Coder 1975; Dary 1989), including several herds that are contemporarily managed for conservation. Many of these animals had detectable cattle gene introgression in both their mitochondrial (Polziehn *et al*. 1995; Ward *et al*. 1999) and nuclear (Halbert *et al*. 2005) genomes. Mitochondrial DNA (mtDNA) is unilaterally passed down through the mother, and this pattern of inheritance leads to an excess of cattle mtDNA in bison today. Nuclear, or autosomal, DNA is passed down by each parent contributing 50% of their genetic information to the next generation. A first-generation bison-cow hybrid would have 50% cattle nuclear DNA, and if a female from this generation was backcrossed with a male bison, the offspring would have 25% cattle autosomal DNA (see Hedrick 2009). The differences in how these types of DNA are inherited is important to consider for conservation herds if managers would like to reduce cattle gene introgression.

Of the estimated 500,000 animals that exist in North America today, only about 4% are managed for conservation and ecological purposes, such as preferential grazing, nutrient cycling, altering fuel loads, and terrain disturbance (Boyd 2003: Knapp *et al.* 1999; see Sanderson *et al*. 2008). The remainder are used for commercial meat production and are artificially selected for more desirable domestic traits, such as ease of handling (Freese *et al*. 2007; Sanderson *et al*. 2008). As such, bison are typically considered and managed as domestic livestock rather than as wildlife. The herds that are managed for conservation are considered semi-wild, meaning they roam within an enclosed space, and are spatially fragmented to sanctuaries and parks. They only cover approximately 1% of their historical range and do not fulfill their previous, large-scale roles in prairie ecosystems (Freese *et al*. 2007). In addition to not fulfilling their ecological roles, wild bison face challenges to their genetic diversity with small population sizes, isolation from each other, cattle gene introgression, and intensive management and culling practices (Freese *et al*. 2007). These factors can contribute to the effects of genetic drift, or the random loss of genetic diversity, by reducing the amount of genetic variation that is present.

Genetic drift is proportional to the effective population size (N_e). The effective population size (N_e) is a population parameter reflective of evolutionary potential, susceptibility to random events, and the population's ability to survive and reproduce (Beebee and Rowe 2008). It assumes an ideal population of equally and randomly reproducing males and females. This is not always the case in real populations, as males and females will vary in their reproductive success, hence decreasing the effective population size (Kliman *et al*. 2008). Information on understanding bison genetic variation and how it can be impacted by mechanisms such as effective population size, genetic drift, levels of cattle gene introgression, effects of the bottleneck event, and inbreeding is critically important for their long-term management and maintaining their genetic diversity (Gates *et al.* 2010).

Approaches to quantify genetic characteristics of bison have been developed by researchers, including looking at multiple microsatellite loci in the nuclear genome (Halbert *et al.* 2005). Microsatellites, also known as short tandem repeats (STRs), are short, repetitive DNA sequences that occur throughout the genomes of many eukaryotes (Bhargava and Fuentes 2010). They are commonly used in population genetics analyses because they are thought to evolve

neutrally (i.e. they are not influenced by natural selection) and are highly polymorphic, meaning they can have multiple variations, or alleles, at each locus (Putman and Carbone 2014).

The Medano-Zapata bison herd is a population of about 1,700 animals that is managed by The Nature Conservancy in Colorado, USA. It is routinely rounded up and hair and blood samples are collected. A recent project report by Halbert and Derr (2010) analyzed the genetic diversity from a subpopulation of the herd using 185 individuals and 11 polymorphic microsatellite markers. The report determined that the herd had high levels of genetic diversity (based on heterozygosity and number of alleles per locus) and was derived from at least three different source herds.

A separate study of the herd used sampled 1,700 of the individuals and analyzed different 14 microsatellite markers to detect levels of cattle genetic introgression (12.59%) in the nuclear genome (Hamilton *et al*. 2009). This study did not, however, look at levels of genetic diversity (heterozygosity and number of alleles) using the 14 markers or assess whether any genetic structure in the entire herd existed. To better assess population viability for genetic management of this larger sample of the herd, here I analyzed the larger dataset to test for the presence of genetic substructure and genetic diversity. I hypothesized that multiple clusters would be found owing to the population being large and having a higher likelihood of different genetic groups forming. Additionally, I used the expanded samples size to obtain a more accurate estimate of the levels of genetic diversity present.

METHODS

Data Collection

The microsatellite genetic data for the Medano-Zapata Ranch herd had previously been collected and archived at Texas A&M University. Previously established protocols (Halbert *et al*. 2005) were used to assess levels of cattle gene introgression at 14 diagnostic and polymorphic (containing more than one allele) microsatellite markers. The markers analyzed were: AGLA17, AGLA293, BM1314, BM4307, BM4513, BM7145, BMS2270, BMS4040, CSSM36, CSSM42, RM185, RM500, SPS113, and TGLA227. The markers spanned 10 chromosomes (of 30 total chromosomes in bison) and were used to either diagnose or confirm whether cattle gene introgression had occurred (Halbert *et al*. 2005).

Data Analysis

To analyze the structure of the population, I used STRUCTURE v. 2.3.4 (Pritchard *et al*. 2000) with the 14 microsatellite loci for all 1,700 individuals. I used the admixture model and ran seven repetitions for each K value (1-4), with K being the number of assumed genetic clusters and calculated the mean of logarithmic probability (Cao *et al*. 2017). I used a burn-in period of 500 000 steps and Monte Carlo Markov Chain (MCMC) rep length of 500 000 steps and chose the K genetic cluster values based on methods published in the literature (Pinzone and Dyer 2013).

The two indices of genetic diversity that I used were (1) expected heterozygosity (H_E) and (2) number of alleles and their associated frequencies. Expected heterozygosity is an indicator of genetic variation and provides information on the frequency of heterozygotes (markers containing different alleles, rather than the same alleles). For calculating expected heterozygosity in this study, I used the formula from Nei and Roychoudhury (1974) and used the allele frequencies calculated from my STRUCTURE analysis. I also summarized allelic data from the STRUCTURE analysis and summarized the number of alleles at each locus (including the cattle alleles) and their frequencies of occurrence within the population.

RESULTS

Genetic Clustering

From the STRUCTURE analysis, the most likely number of genetic clusters was K = 1 based on the highest average natural logarithm (Ln) of the likelihood across runs (average Ln = -14547.2). **Table 1** summarizes the results from the analysis, and **Supplementary Figure S1** plots the K values and shows the most probable likelihood value.

K	Mean Estimated Ln Probability	Standard Error of Mean Estimated Probability
1	**-14547.2**	7.43E-13
2	-15003.2	13.09
3	-16017.1	99.84
4	-17142.6	137.86

Table 1. Results from the STRUCTURE analysis of the 14 microsatellite loci and associated standard error values. The value in bold indicates the highest support (K = 1) of the K clusters based on likelihood alone.

Indices of Genetic Diversity

The expected heterozygosity (H_E) is a measure of genetic variation based on the frequencies of each allele at each locus. The average H_E values for all 14 loci was close to 0 (H_E = 0.1860), indicating the diversity of alleles was low. **Table 2** gives the H_E values for each locus.

The number of alleles at each locus ranged from 1 to 4. The total number of alleles identified was 36, and the mean number of alleles per locus was 2.57, which gives the expected number of alleles at any of the 14 given loci. Of the 14 markers used, 13 were polymorphic and RM500 was monomorphic with only one allele. The percentage of cattle alleles to the total number of alleles identified was 25%. From the analysis, I was able to determine the allele frequencies of both the bison and cattle alleles. Allelic frequencies give the prevalence of each allele within the population. The frequency of cattle allele occurrence at each locus ranged from 0.1% to 2.2% and the standard error was 0.069 (see **Supplementary Table S1**). **Table 2** provides a summary of the allelic information at each of the 14 loci, including which chromosomes they are located on and which alleles are considered indicative of domestic cattle gene introgression.

Locus	**Number of alleles**	**Alleles (bp)**	**H_E**	**Chromosome (n = 30)**
AGLA17	2	215, 218	0.004	1
AGLA293	4	218, 215, 220, 228*	0.080	5
BM1314	2	137, 157*	0.033	26
BM4307	4	185, 187, 197*, 189*	0.340	1
BM4513	2	132, 134	0.220	14
BM7145	3	108, 110, 116*	0.307	1
BMS2270	4	66, 68, 70, 98**	0.553	24
BMS4040	2	75, 95**	0.002	1

CSSM36	2	132, 158	0.000	27
CSSM42	3	167, 171, 169	0.570	2
RM185	2	92, 100*	0.004	23
RM500	1	123	0.000	5
SPS113	3	130, 132, 128	0.476	10
TGLA227	2	73, 94**	0.012	18

Table 2. A summary of information on the 14 polymorphic microsatellite loci that were used for the analysis. * Is a confirmed domestic cattle allele and ** is a suspected diagnostic cattle allele (see Halbert *et al*. 2005). Chromosome location and domestic cattle allele information are adapted from Halbert *et al*. (2005).

DISCUSSION

Genetic Clustering

The results of this study most strongly supported genetic clustering around a single cluster and rejected my hypothesis that there would be multiple clusters. This indicates that the individuals are part of one interbreeding, or admixing, population and that genes are randomly distributed among individuals. Admixture occurs when individuals from two or more genetically differentiated populations interbreed and gene flow occurs (Rius and Darling 2014). In addition, the effects of the historic bottleneck are likely to have reduced the genetic variation within the loci. The difference in mean estimated natural log probabilities between K = 1 and K = 2, however, is closer than the other clusters, hinting that perhaps two clusters are forming in the population. Future studies could potentially investigate this effect using more genetic loci.

The previous report on the Medano-Zapata by Halbert *et al*. (2010) identified that the herd had a total of eight core genetic clusters based on previous cluster identification. Their analysis indicated that the Medano-Zapata herd was mainly derived from three genetically distinct herds: the Wind Cave National Park herd, the National Bison Range herd, and the Wichita Mountains National Wildlife Refuge herd, with potential mixture from the Custer State Park bison herd. The differences in findings are likely due to Halbert *et al*. (2010) using a different analysis and different markers than the ones used in my study. The markers used in their report were developed for parentage testing and were chosen based on high heterozygosity values and for containing many alleles (Schnabel *et al*. 2000).

Indices of Genetic Diversity

The mean expected heterozygosity was relatively low for the markers used in this study and the total number of alleles at the 14 markers was also low. Heterozygosity is commonly used as a measure of genetic variation and the values are expressed as the frequency of heterozygotes (i.e. the alleles are different at each locus). It is likely that the low value in the Medano-Zapata bison population is a consequence of the historical bottleneck. Typically, after a bottleneck event, heterozygosity levels are expected to decline due to a random loss of alleles (Allendorf and Leary 1986; Nei *et al*. 1975). A population reduction can also result in inbreeding, or the mating between closely related individuals. Inbreeding increases the proportion of homozygotes (thereby reducing heterozygosity) and can also increase the probability of deleterious alleles. Additionally, the sudden loss of genetic variation can potentially affect the population's effective population size and decrease its evolutionary potential by reducing variability that can be passed on to the next generation (Nei *et al*. 1975).

The frequencies of occurrence of the cattle alleles were low but did vary depending on the locus. AGLA293 contained a cattle allele that occurred the most frequently, while BMS4040 had an allele that occurred at a very low frequency. This variation in frequencies is likely due to chance genetic drift, which allows the alleles to persist at low levels (see Hedrick 2009). Whether these nuclear alleles have any effect on the phenotype, or physical expression of the genes, is unclear. In a study by Derr *et al*. (2012), the researchers found that high levels of mitochondrial cattle introgression resulted in smaller body sizes in bison. Whether the smaller body size had any affect on the fitness of the bison was less clear, but generally there is a positive relationship between body size and fitness (Robertson 1955). Future studies could address if there are any positive (advantageous) or negative (deleterious) phenotypic effects on nuclear cattle alleles in bison populations.

Conclusions

While the 14 loci used in this study are indicative of cattle gene introgression levels, when analyzed at the genetic structure level, they provide interesting clustering results and indicators of genetic diversity that are worth investigating and understanding further. To provide a better resolution of genetic diversity in future genetic analyses, it is worth looking at multiple microsatellite loci distributed across the genome (Chambers and MacAvoy 2000).

The Medano-Zapata herd is a relatively large herd with more than 1,000 individuals that roam a large landscape. As such, it is one of the few large conservation herds that is managed as a semi-wild herd, and the protection and evaluation of its genetic features is worth continued investigation. This includes understanding and increasing the herd's effective population size by assessing differential breeding of males and females (i.e. is one male mating more often than other males and disproportionately contributing to the next generation?) and maintaining roughly equal sex ratios (Gates *et al*. 2010). In addition, it is important to continue protecting the bison's habitat and increase the connectivity between different populations. By intermixing bison populations, their genetic diversity can increase overall. Plains bison historically covered large territories, and they likely had high rates of gene flow when they were able to intermix (Berger and Cunningham 1994; Wilson and Strobeck 1999). Lastly, continued investigation of the levels of cattle gene introgression and how to effectively manage introgression levels is worth considering in future research and management.

ACKNOWLEDGEMENTS

I'd like to thank my advisors, Andrew Martin and Cheryl Pinzone, for their invaluable assistance and support that went into this project. I'd also like to thank Barbara Demmig-Adams for her continual feedback and guidance throughout the creation and writing of this thesis. Another thanks to Sean Streich for taking the time to help format the genetic dataset and help run my analysis in STRUCTURE. Lastly, I'd like to thank Chris Pague and Anya Byers from The Nature Conservancy for their help with creating this project and for allowing me to use the Medano-Zapata bison herd dataset.

REFERENCES

Allendorf FW, Leary RF (1986) Heterozygosity and fitness in natural populations of animals. In: *Conservation Biology: the Science of Scarcity and Diversity*. Sinauer Associates, Incorporated, New York.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613-622.

Berger J, Cunningham C (1994) Bison: mating and conservation in small populations. In: *Methods and Cases in Conservation Science.* Columbia University Press, New York.

Beebee T, Rowe G (2008) *An Introduction to Molecular Ecology*. Oxford University Press, Oxford.

Bhargava A, Fuentes FF (2010) Mutational dynamics of microsatellites. *Molecular Biotechnology*, **44**, 250–266.

Boyd, DP (2003) Conservation of North American bison: status and recommendations. MS Thesis, University of Calgary.

Boyd, MM (1914) Crossing bison and cattle. *Journal of Heredity*, **5**, 189–197.

Cao J, Li X, Du X, Zhao S (2017) Microsatellite based genetic diversity and population structure of nine indigenous Chinese domestic goats. *Small Ruminant Research*, **148**, 80–86.

Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **126(4)**, 455–476.

Coder GD (1975) The national movement to preserve the American buffalo in the United States and Canada between 1880 and 1920. PhD Thesis, Ohio State University.

Dary DA (1989) *The Buffalo Book: the Full Saga of the American Animal*. Swallow Press, Chicago.

Derr JM, Hedrick PW, Halbert ND *et al*. (2012) Phenotypic effects of cattle mitochondrial DNA in American bison. *Society for Conservation Biology*, **26**(6), 1130-1136.

Freese CH, Aune KE, Boyd DP *et al*. (2007) Second chance for the plains bison. *Biological Conservation*, **136**, 175–184.

Gates CC, Freese CH, Gogan PJP, Kotzman M (2010) American bison: status survey and conservation guidelines. IUCN, Gland, Switzerland.

Goodnight C (1914) My experience with bison hybrids. *Journal of Heredity,* **5**, 197–199.

Halbert ND, Derr JN (2010) Bison Genetics of the Great Sand Dunes National Park. Final Project Report, Texas A&M University.

Halbert ND, Ward TJ, Schnabel RD *et al.* (2005) Conservation genomics: disequilibrium mapping of domestic cattle chromosomal segments in North American bison populations. *Molecular Ecology,* **14**, 2343– 2362.

Hamilton RG, Halbert ND, Derr JN (2009) Bison herd genetic architecture and management – working toward a species-wide conservation approach. Final Project Report, The Nature Conservancy.

Hedrick PW (2009) Conservation genetics and North American bison (*Bison bison*). *Journal of Heredity,* **100**, 411–420.

Isenberg AC (2000) *The Destruction of the Bison: an Environmental History, 1750-1920.* Cambridge University Press, Cambridge.

Kliman R, Sheehy B, Schultz J (2008) Genetic drift and effective population size. *Nature Education,* **1(3)**, 3.

Knapp AK *et al* (1999) The Keystone Role of Bison in North American Tallgrass Prairie. *BioScience,* **49**(1), 39-50.

Meagher MM (1973) *The bison of Yellowstone National Park*. National Park Service, Government Printing Office, Washington, D.C.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution,* **29(1)**, 1-10.

Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. *Genetics,* **76**, 379-390.

Pinzone CA, Dyer KA (2013) Association of polyandry and sex-ratio drive prevalence in natural populations of *Drosophila neotestacea*. *Proceedings of the Royal Society B,* **280**, 1-8.

Polziehn RO, Strobeck CM, Sheraton J, Beech R (1995) Bovine mtDNA discovered in North American bison populations. *Conservation Biology,* **9**, 1638–1643.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155**, 945–959.

Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, **4(22)**, 4399–4428.

Randi E (2008) Detecting hybridization between wild species and their domestic relatives. *Molecular Ecology*, **17**, 285-293.

Robertson A (1955) Selection in animals: synthesis. *Cold Springs Harbor Symposium on Quantitative Biology*, **20**, 225-230.

Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonizing populations? *Trends in Ecology & Evolution,* **29(4)**, 233–242.

Sanderson EW, Redford KH, Weber B *et al.* (2008) The ecological future of the North American bison: conceiving long-term, large-scale conservation of wildlife. *Conservation Biology,* **22**, 252–266.

Schnabel RD, Ward TJ, Derr JN (2000) Validation of 15 microsatellites for parentage testing in North American bison, *Bison bison* and domestic cattle. *Animal Genetics,* **31**, 360-366.

Ward TJ, Bielawski JP, Davis SK *et al.* (1999) Identification of domestic cattle hybrids in wild cattle and bison species: a general approach using mtDNA markers and the parametric bootstrap. *Animal Conservation,* **2**, 51–57.

Wilson GA, Strobeck CM (1999) Genetic variation within and relatedness among wood and plains bison populations. *Genome,* **42**, 483-496.

SUPPLEMENTARY MATERIALS

Table S1. Summary of the allele frequencies in percentages at each microsatellite marker.

* Indicates a domestic cattle allele (from Halbert *et al*. 2005)

Marker	Allele	Frequency	Marker	Allele	Frequency
AGLA17	215	99.8%	BMS2270	66	37.1%
	218	0.2%		68	55.2%
AGLA293	218	95.9%		70	6.7%
	215	0.1%		98*	1.0%
	220	1.8%	BMS4040	75	99.9%
	228*	2.2%		95*	0.1%
BM1314	137	98.3%	CSSM36	132	0.0%
	157*	1.7%		158	100.0%
BM4307	185	79.0%	CSSM42	167	56.0%
	187	19.0%		171	32.0%
	197*	1.4%		169	12.0%
	189*	0.7%	RM185	92	99.8%
BM4513	132	87.4%		100*	0.2%
	134	12.6%	SPS113	130	62.0%
BM7145	108	81.3%		132	37.3%
	110	17.8%		128	0.7%
	116*	0.9%	TGLA227	73	99.4%
RM500	123	100.0%		94*	0.6%

Figure S1. A plot of the mean estimated Ln probability for the K (1-4) clusters. Values closest to zero (the least negative values) indicate the highest average likelihood for clustering.