Widespread Mitochondrial Disease in North American Bison

Thomas H. Pringle

Sperling Foundation, Eugene, Oregon 97405 USA tom@cyber-dyne.com

Abstract

North American bison have rebounded from near-extinction in the nineteenth century but from such small inbred founding populations that once-rare deleterious nuclear gene alleles and mitochondrial haplotypes are now be at high frequencies. The initial bottleneck was compounded by decades of unnatural selection affecting bison conservation genomics and undercutting restoration initiatives. The genomics era began in late 2010 for bison and sister species yak with the release of 102 whole mitochondrial genomes, displacing earlier control region and microsatellite data not extending to coding regions. This allows detection of both sporadic and sub-clade level mutations in mitochondrially encoded proteins and tRNAs by comparative genomics methods: deleterious mutations in both cytochrome b (V98A) and ATP6 (I60N) occur within a single common bison haplotype. Since similar mutations in human and dog cause clinical impairment of mitochondrial oxidative phosphorylation, these bison are predicted significantly impaired in aerobic capacity, disrupting highly evolved cold tolerance, winter feeding behaviors, escape from predators and competition for breeding. Because Yellowstone National Park bison are subjected to genetically uninformed culls and surplus animals used to seed new conservation herds, mutational status has significant implications. Continuing take of the remaining bison with wildtype mitochondria may recapitulate errors of nineteenth century bison stewardship bringing bison conservation to the point of no return.

Introduction

Recovery of a species from a severe bottleneck requires consideration of both nuclear and mitochondrial genomics (1, 2) because inbred reduced populations may have lost much of their former genetic diversity and harbor unnaturally high frequencies of deleterious alleles (3). Inbreeding depression in Florida panthers (4), collapse of the pygmy rabbit captive breeding program (5), facial tumors in tasmanian devil (6) and required rescue of the Texas State Bison Herd (7) have put such concerns on center stage.

In the case of bison, natural selection has not been fully operative on deleterious alleles for decades, having been largely displaced by predator control, genetically uninformed culls, trophy bull hunts, winter hay feeding, and selection for docility. Recovery of large herds of animals outwardly resembling bison serves no authentic conservation purpose if these bison are hobbled by inherited disease and no longer function as they had evolved up to the era of human interference.

Genetic isolation of small bison populations has been mitigated in the past by animal exchanges but these have sometimes spread tuberculosis and brucellosis (8) and brought in descendants of hybridization experiments with domestic cattle (9). The focus today is restoration of pre-settlement bison genetic diversity without inadvertent spread of inherited disease (1).

For the nuclear genome, the 1000 Human Genome Project concluded (10) that each individual human nuclear genome carries 275 loss-of-function variants and 75 variants previously implicated in inherited disease (both classes typically heterozygous), additionally differing from the reference human proteome at 10,488 non-synonymous sites. The deleterious alleles include 200 in-frame indels, 90 premature stop codons, 45 splice-site-disrupting variants and 235 deletions shifting reading frame. In bison, in view of the recent extreme bottleneck and subsequent small herd size history, the overall genetic load should be worse, with deleterious nuclear genes more commonly homozygous.

Based on extensive clinical testing, a recent study estimates that 1 in 5000 human live births has inherited mitochondrial disease, with an additional 1 in 200 asymptomatic carriers (11), subject to the complexities of heteroplasmic inheritance of mitochondrial genomes discussed below. Bison mitochondrial genomes may be affected at even higher frequencies; indeed, significant anomalies in mitochondrial gene products have already been reported (2).

The limited assessments of bison genetic status provided by sequencing functionally uninformative regions (homopolymers, control regions, synonymous and non-coding SNPs) have been superseded by direct sequencing of whole nuclear and mitochondrial proteomes. The cattle nuclear genome was released in October 2007 and soon assembled, aligned and refined (12-14); yak and water buffalo genome are underway (15) and bison proposed.

Mitochondrial genomes are available for over 214 species of mammals (16), sometimes in high multiplicities for individual species. This multiplicity allows rare private polymorphisms and features of small clades to be distinguished from an appropriate reference (or inferred ancestral) genome. Certain mitochondrial genes have been separately sequenced, with over 5,000 GenBank entries for 1,250 mammalian species in the case of cytochrome b. For ancient DNA, some 44 control region sequences are available from fossil bison and 298 for extinct steppe bison (17). Whole genome sequencing from a frozen carcass is feasible but not yet begun (18).

Data availability is thus quite favorable for comparative genomics though mitochondrial DNA evolves quite rapidly diminishing the utility of distant outgroups. However the phylogenetic tree is well-established for close-in pecoran ruminants, with ((((bison, steppe bison), yak), cow), water buffalo) the relevant phylogenetic tree topology here (43).

Below, analysis of complete bison mitochondrial genomes in this comparative genomics context establishes that a widespread bison haplotype carries substitutions in both cytochrome b (V98A: valine at position 98 changed to alanine) and ATP6 (I60N: isoleucine at position 60 changed to asparagine). Both variants are unambiguously deleterious according to numerous bioinformatic criteria and clinical observations in other species; together they strongly imply that these bison are affected by significant mitochondrial disease.

Methods

Relevant GenBank entries were acquired by Blastn of reliably annotated seed sequence queries. Some entries needed re-curation when the mitochondrial DNA arose from hybridization with another species (such as GenBank accession AB177774). Other entries were not adequately annotated by haplotype, herd origin, or domestication status in the accompanying publication. Entries lacking bison source data (or journal publication) has diminished value but was still used after quality control. Older GenBank entries containing multiple alignment anomalies were discarded as sequencing or submission error. It is difficult to distinguish between sporadic mutation implied by a single sequence from outright sequencing error when raw read data is not deposited at GenBank and indeed may not have been saved. Exceptional outcomes are rarely validated by re-sequencing (46).

Wild cattle such as gaur and banteng and fossil aurochsen data were used in place of modern domestic cattle breeds to improve reconstruction of ancestral sequence nodes and bison mitochondrial protein assessment because inbred domestic cattle have numerous derived characters (13, 14).

Sequence data never submitted to GenBank, such as the control region sequence fragments for Yellowstone and Grand Teton National Parks bison located in an unpublished but approved 2007 dissertation (19), was accepted after passing quality review; haplotypes were mapped into current whole genome based nomenclature (reference 2 and Table 8).

Haplotype re-mapping was also conducted for ancient bison DNA but not for extinct steppe bison because of sequence divergence and DNA damage possibly resulting in sequence error (17, 20, 21). However alignment of steppe bison DNA at cytosines strictly conserved in extant yak and bison did not exhibit the expected excess of C to T transitions resulting from putative postmortem cytosine deamination, though alignment indel abundances are compatible with a substantial error rate in homopolymer run length determination (data not shown). Full length mitochondrial genomes from steppe bison are very relevant to contemporary bison conservation genomics but none had been posted to GenBank as of January, 2011.

Fasta headers of sequences directly downloaded from GenBank begin with uninformative accession numbers, yet alignment programs are generally restricted as to the number of characters displayed. As this leads to unintelligible output in alignments of thousands of sequences from multiple species, headers were replaced with a simple flat database of concatenated attributes such as genus, species, haplotype, domestic, wild, and disease allele status.

NCBI provides the main gateway to sequence data via Blast servers (61). While the algorithm itself is stable, formatting options are not. Understanding these is critical when querying mitochondrial gene products because the enormous amount of available data overwhelms current display capacity. This causes shorter sequences from bison and yak to score lower than longer sequences from diverged species and even nuclear numts, resulting in their loss in output. This can be partly remedied by a careful choice of query, setting the hidden output parameter to its current maximum of 20,000 matches, and making full use of taxonomic inclusion and exclusion settings. This still does not address GenBank entries filed under the wrong taxon, a serious issue when hybridizations was not recognized at the time of sequence submission, resulting in spurious genetic diversity when interpreted by a bioinformatics pipeline.

Output format from a Blast query were manually adjusted to raise the default to the maximum number of alignments allowed, withthe 'alignment view' option re-set to 'flat query anchored with dots for identities' because this accommodates indels in either query or matches, reducing clutter by exhibiting only differences relative to the query. However to speed Blast searches, GenBank now collapses multiplicities into single precomputed matches in the original output (that is, identical sequences -- even from different species -- are represented by a single proxy). Those multiplicities are lost upon reformatting but can be captured by using the taxonomy link from the initial output.

Because output line width is restricted to 80 characters even for a 16,000 base pair whole genome query, this may result in hundreds of screens of output, almost all dots. To extract information, output was reformatted into a single row and columnized at each position using a desktop spreadsheet or after realignment of collected accessions in Multalin (22) set at one character break. Uninformative columns can be moved aside or deleted and the manageable remainder assessed for phylogenetic characters, haplotype classification, or amino acid variational statistics (reduced alphabet). This approach is far easier to adjust after a formatting change is made at NCBI than a formal mining algorithm code.

The new 'find related data' feature at GenBank proves critical to managing thousands of accession numbers -- a central issue in bioinformatics because ten thousand vertebrate genomes are being deposited, each with twenty thousand coding genes (15). A long list of match accessions can be recast as species multiplicities for a non-redundant list of taxa, ordered in a provided phylogenetic tree. Similarly, all PubMed publication identification numbers associated with the accessions can be requested, allowing quick recovery of all applicable journal abstracts and those with open access. The thirteen proteins from thousands of mammalian

mitochondrial genomes can be extracted into fasta format and recompiled as separate sets with a database sort on the header. tRNA compilation still requires screen scrapes or complicated grep procedures -- however these have been conveniently compiled elsewhere (25).

Multalin proved to be the most useful online alignment program (22) because it accepts very large input databases, allows arbitrary display widths, optionally maintains input order, allows character separation and an option to display all residues or only differences relative to a canonical first sequence.

Bison and yak tRNA variations were evaluated using MitoMap (33), which maintains a complete list of human tRNA mutations and associated disease, and the curated phylogenetic tRNA dataset Mamit (23). Comparative genomics takes a few twists because mammalian mitochondrial tRNAs evolve quite rapidly in comparison to nuclear tRNAs. Generalized nucleotide alignment programs do not reliably place gaps, in part because they do not recognize -- or anchor the alignment to -- reverse complementary stem base pairs and other non-local constraints. Even if they did, homological (evolutionary descent) alignment may differ from cloverleaf-based structural alignment. Parsing tRNA sequences into standard folding cloverleaf subdomains and reverse-complementing one member of each stem pair allows a single difference alignment to display all exceptions to stem base pairing. This proved not indicative of mitochondrial disease because presumptive wildtype mitochondrial tRNAs in many species already show imperfect base pairing.

Seven methods are used here to classify observed amino acid variation in bison and yak mitochondrial proteins as either dysfunctional, near-neutral, or potentially adaptive. While non-synonymous SNP interpretation never attains 100% sensitivity and selectivity, the massive data set of mitochondrial proteins (12,603 sequences in 1,637 mammals utilized for cytochrome b) and relatively slow evolution (83% identity between bison and platypus) favor reliable interpretation. Prediction accuracy -- as tested on clinically and biochemically validated human disease alleles -- is very high on the moderate to radical change at conserved sites considered here (31, 32).

1. Validity of the amino acid substitution: the reported change may be an artifact when an older sequencing technology is used or read coverage is low, when ambient DNA base composition is anomalous or a homopolymer read length error could occur, when too-complex a mutation is required to produce the substitution, when ambiguity in base determination occurs elsewhere in the submitted sequence, when multiple anomalous substitutions are reported within the same animal, when nuclear pseudogenes (numts) and hybridization with another species are not considered, or when no re-sequencing or raw read reevaluation is reported in the case of a surprising change in a large population survey of inbred animals.

2. Unique aspects of mitochondrial disease: the inheritance of the mitochondrial genome is quite different from that of nuclear genes because of heteroplasmy, undercutting decades of assumptions in population genetics (60), challenging the accuracy of that data (47-49), and here affecting methods for associating amino acid variation with mitochondrial disease:

- Mitochondrial DNA is maternally inherited in bison. The bull may carry mitochondrial mutations or even whole cattle mitochondria but these do not pass on to descendants (negligible paternal leakage). On the other hand, cattle mitochondrial DNA from a distant hybridization persists indefinitely without dilution or fragmentation despite subsequent backcrosses to bison.

- Each maturing bison oocyte has some 260,000 copies of mitochondrial DNA with up to ten copies in a single mitochondrion (24). These may differ in initial sequence (heteroplasmy) and acquire additional differences over time which are not shuffled onto composite haplotypes because of the effective absence of recombination.

- Replication occurs unevenly in maturing oocytes as does subsequent stochastic segregation of mitochondria to generations of daughter cells and later mitochondrial replication in descendent stem cells. Post-natal selection may affect haplotype ratios. A disease haplotype, initially rare and functionally compensated by healthy haplotypes, can surge either ontogenetically (causing late disease onset in a somatic cell lineage) or phylogenetically (leading to effective inheritance in descendants and earlier onset) within a single generation, a haplotype bottleneck effect.

- Bison DNA sequences reported to GenBank do not originate from oocytes but rather from blood, muscle, skin, hair, or intestinal sloughing. Although heteroplasmy implies these tissues should often contain multiple haplotypes, only one haplotype per sample has ever been reported for bison, with the rest apparently discarded as sequencing or cloning error. In contrast, recent careful re-sequencing of a 6,738 year old aurochs established a single heteroplasmic control region site (59). Consequently bison data does not directly report on oocyte haplotype abundances nor predict disease status of either parent or descendants. This applies in particular to sporadic mutations reported for a single animal (which could equally be sequence error, somatic mutation, or heteroplasmic surge). However when a substantial sub-clade with a given set of amino acid variations is inherited over many generations, that provides evidence for complete oocyte penetration of that haplotype and conventional inheritance of the amino acid change in all descendants. On the other hand, frameshift and internal stop codon mutations -- lethal because oxidative phosphorylation is essential -- imply internal compensation by unreported wildtype haplotypes must be occurring at a significant level given the essentiality of oxidative phosphorylation.

- The mutation rate in mitochondria is some tenfold higher than in chromosomal DNA, often attributed to reactive oxygen species generated by oxidative phosphorylation, sub-optimal functioning of imported nuclear replication and repair genes or lack of protective histones (26). Mutational hotspots may also exist because of local base compositional anomalies, homopolymer run susceptibility or regional propensity to fold into tertiary structures. Despite this, observed recurrent mutation is rare in protein coding regions: the same amino substitution seldom arises de novo in oocyte DNA, much less rises from low heteroplasmy status to full heritability, as seen from human mitochondrial disease statistics (27) and the lack of phylogenetic reoccurrence in unrelated clades despite voluminous data (next section). This implies most observed amino acid substitution in bison arose from unique events and their presence today in descendants is phylogenetically informative.

- If multiple haplotypes in a population happen to wax and wane across a species divergence, the resulting mitochondrial lineage sorting could give a quite different phylogenetic tree of that from nuclear genes, an issue compounded by small sampling sizes chosen for sequencing. Mitochondrial heteroplasmic persistence and resurgence makes the determination -- indeed definition -- of amino acid ancestral state somewhat problematic. It further complicates establishment of the reduced alphabet (set of fully functional amino acids at a given position) which may drift over time according to clade because of varying co-evolving residues.

- Functional compensation can occur if single mitochondrion carries multiple haplotypes, one of them wildtype which can arise via mitochondrial fusion if not initially present. Compensation by a nuclear genes imported into mitochondria may occur in yeast (28). Non-compensable mitochondria may turn over more rapidly or replicate more slowly (be selected against). These considerations make it very difficult to predict the impact on oxidative phosphorylation a deleterious variation in a mitochondrially encoded protein unless only this haplotype class occurs in the cell and the amino acid substitution is at a site not physically interacting with or influenced by any nuclear gene product. Cytochrome b resides in the bc1 complex which involves 11 gene products, all but one nuclear encoded. Thus a bison cytochrome b variation cannot be internally compensated nor by a secondary variant of ATP6 (or any other mitochondrially encoded protein). Compensation by an altered nuclear gene could give another form of mitochondrial disease when cross-matched with a conventional bison mitochondrial haplotype.

- Despite these opportunities for compensation, it is rarely observed experimentally except as yeast petite suppressors (billions of cells plated out under strong selection, bison populations in the thousands). Mitochondrial disease is surprisingly common in species far less inbred than bison:

    "One in 4000 individuals is at risk of developing a mitochondrial disease sometime in their lifetime. Half of those affected are children who show symptoms before age five, and approximately 80% of them will die before age 20. The mortality rate is roughly that of cancer... The mutation rate of the mitochondrial genome is 10–20 times greater than of nuclear DNA, and mtDNA is more prone to oxidative damage than is nuclear DNA. Mutations in human mtDNA cause premature aging, severe neuromuscular pathologies and maternally inherited metabolic diseases, and influence apoptosis (29)." [WC Copeland: Chief, Laboratory of Molecular Genetics, NIH]

3. Physical-chemical nature of amino acid change: statistics drawn from millions of amino acid changes studied over the last sixty years (30) in conjunction with in vitro assessment of the resulting protein and correlation to clinical disease status show mild changes in properties such as hydrophobicity, polarity, charge, aromaticity, branched side chains, hydrogen-bonding and disulfide capacity, and beta sheet, 3-10 turn or alpha helix forming or ending propensity are vastly less probable to give rise to dysfunctional protein than radical substitutions. This conservatism is reflected in mitochondrial genetic code in that the most common form of mutation often give no change at all (synonymy) or cause a mild change (same cell, row or column). The risk that substitutions cause protein dysfunction can be quantitated by Grantham distances, Blosum matrices, or the more recent SIFT and PolyPhen2 algorithms used in part here (31). The latter relies on a curated protein sequence and structural data set covering only a small fraction of available GenBank data and that out of phylogenetic tree topological context (e.g. sea urchin proteins are not equally relevant as cow to bison).

4. Comparative genomics: amino acids vary greatly in their tolerance of substitutions according to site position within the protein and its importance to function, correlating with selective pressure to maintain it. Should a peptide containing the residue be post-translationally removed and degraded, there may be no constraints whatsoever provided by the physical-chemical nature of the substitution. At the other extreme, a residue critical to an enzyme active site tolerates no change at all over trillions of years of observable branch length. Most site positions are intermediate, allowing mild change as defined by a restricted alphabet specific to that site via comparative genomic alignment, along the lines of phylogenetic tree-aware TreeSAAP (32). Most nsSNPs involve a single nucleotide change with transitions predominating over transversions. If an amino acid change -- mild or not -- is a mere transition away from a conserved residue yet never observed in thousands of species despite a high mutation rate, it likely has arisen numerous times but never attained a foothold because selective pressure eliminated it (when drift alone did not), even as other reduced alphabet changes did get fixed in the same protein. Homoplasic recurrence of the change elsewhere in mammals are detected here via clade pattern analysis: if the change does occur uncommonly elsewhere, are these concentrated in sub-clades (indicating persistence and so fitness for reduced alphabet) or scattered randomly in the taxonomic sense (indicting unfixed recurrent mutation). This sharpening provided by clade pattern analysis -- quantitated by supportive branch length summing -- becomes critical as the amount of data becomes astronomical. Although moderate frequency of a polymorphism is sometimes used in medical genomics to argue against deleterious nature, that is not applicable here because of the founder effect and subsequent inbreeding in

bison, in effect making them the counterpart of a inherited disease pedigree; allele frequency is in fact extremely low when taken in the comparative genomics sense.

5. Protein structure: amino acid changes can be placed within a high resolution x-ray structure of a homologous protein (or domain) when available at PDB. No such data is directly available for bison or yak proteins, but quite commonly for cattle. In such cases, it may be possible to interpret the structural significance of a substitution in terms of correct folding, protein stability, interactions with other residues in the same or hetero-oligomeric partners and evaluate in silico the impact of the substitution on stability energetics by molecular dynamics simulations. High sequence conservation (bison to cow) and extreme fold conservation allows reliable annotation transfer. Surprisingly few of the thirteen mitochondrially encoded proteins have any match whatsoever (Blastp against PDB), cytochrome b and cytochrome oxidase components being the exceptions. The latter are irrelevant because no bison variation occurs in them (next section). The best match of bison ATP6 -- to the E. coli crystallographic structure 1C17 -- is uninformative because of excessive divergence.

6. Orthologous disease transfer: a comparable allele to a bison variation may have been previously studied in human mitochondrial disease with direct biochemical and clinical assessment of disease phenotype. The outcome of thousands of such studies are tabulated at the MitoMap web site (33). Additional data from veterinarian studies of mitochondrial disease is compiled at OMIA (34). The bison alleles considered here do not have exact counterparts. Even if they did, divergence elsewhere in the primary sequence would preclude direct interpretative transfer. However the bison variant V98A of cytochrome b corresponds in orthologous position to canine V98M. Affected Shetland sheepdogs have significant mitochondrial disease (35). Since alanine is a more extreme substitution than methionine for branched chain aliphatic valine, V98A bison are predicted to be similarly impaired. Unlike nuclear genes, the many hundreds of known human mitochondrial disease alleles -- dispersed over 13 different proteins and 22 tRNAs -- all present with related phenotypes despite different nomenclatures such as lactic acidosis, cardiomyopathy, exercise intolerance. The outcome of all mutations is the same: impaired oxidative phosphorylation. This is the case for all 21 human disease alleles of ATP6 (33), none of which correspond to position 60 of bison I60N. However specifics of human disease manifestation vary because of heteroplasmy -- different cell lineages affected at different stages of development, altering location and age of disease onset (29). Putative disease alleles in bison need direct biochemical validation; that could be provided rom bison muscle biopsies bythe same laboratories that routinely evaluate human mitochondria (35).

7. Experimental model systems: yeast and hamster cell lines provide surprisingly relevant model systems (28, 37-39). Mutant alleles of human genes are knocked into the nucleus, suitably altered to use the nuclear gene translation table and carry mitochondrial targeting signals, with mitochondrial transcription of the competing homolog knocked down by ethidium bromide. This results in a heterologous oxidative phosphorylation system (allotopic expression) of one human protein interacting with 12 mitochondrial and many nuclear proteins from the other species. Because little has changed in oxidative phosphorylation over the last billion years, this characterizes structural and functional attributes of known mutations despite heterologous protein-protein interactions. The yeast or hamster system could be adapted to specifically study the effect of mutations in bison mitochondrial proteins to supplement strictly bioinformatic methods reported here.

8. The accuracy of SNP interpretation is high but still imperfect. The rate of false positives is routinely estimated by blinded prediction of allele status of intensively studied human genes such as hemoglobin, cystic fibrosis, and rhodopsin where large numbers of variants from patients, controls and large scale SNP surveys have been clinically evaluated (31 and references therein). Accuracy can exceed 90%, with errors overwhelmingly concentrated in borderline changes in amino acid physical-chemical properties (32). Methods used here on bison V98A and I60N were validated using unblinded test suite of known pathogenetic mutations in human cytochrome b and ATP6 (51). In the special case here of two significant but not catastrophic changes in two mitochondrial proteins in different non-complementing oxidative phosphorylation complexes with demonstrated heritability (low or no wildtype heteroplasmy), prediction accuracy in the mitochondrial disease setting can be conservatively estimated as the complement of the probability of both being innocuous: $1- (1 - 0.85)(1 - 0.85) = 98\%$.

9. Fossil bison DNA represents an important control because animals operating under the full force of natural selection are not expected to exhibit high levels of deleterious alleles. Here 44 fossil bison control region sequences are available at GenBank (17). Note these have widely variable Holocene dates and do not represent a population sample in the sense of contemporary bison data.

Results

Bison sequencing effort initially focused on control region polymorphisms, homopolymer run lengths (microsatellites), and determination of base at primarily intergenic SNPs (43), data perhaps relevant to overall genetic diversity and past hybridization but not to proper functioning of the 20,000-odd nuclear and mitochondrial genes where the bison genetic burden can be expected to be far worse than in less inbred species which are already serious enough (10). Steppe bison could be helpful here but only the last 33 bp of tRNA Pro are available (eg AY748559). Since the entire tRNA pro is identical in all bison and yak, the observed 100% identity of steppe bison comes as no surprise. There is additional sporadic variation but this probably arises from dna damage that affects sequencing accuracy.

However new population-level sequencing of yak (40-42) and bison whole mitochondrial genomes (2) now makes it feasible to search for mitochondrial disease in these species. Indeed, Douglas and coworkers note an alarming number of anomalies and the

possibility of mitochondrial dysfunction in bison (2), no surprise given the historic bison bottleneck, subsequent inbreeding in small founder herds, restriction of gene flow between herds, and decades of unnatural selection (3,7) given the high incidence of mitochondrial disease in other mammals (29, 50).

Here, variation in all 13 mitochondrially encoded proteins and 22 tRNAs in all published bison and yak mitochondrial genomes (Table 1) was collected and evaluated by comparative genomics and comparison to human mutational databases (see methods section). Extensive data both from bison and yak and their immediate outgroups (wild cattle, water buffalo) proved necessary to sort out sporadic private variation affecting individual animals from variation affecting whole subclades and to assign synapomorphies (derived characters relative to ancestral state) to the appropriate lineage using the known ruminant phylogenetic tree (43).

Thus a variation in bison with respect to cattle is not a bison mutation issue if yak and water buffalo outgroups are in concordance with bison but instead the ancestral value. Should only yak support bison, again the variation is not plausibly a mutation because of fixation and persistence over millions of years in the face of natural selection but better viewed as a shared synapomorphy of these sister taxa.

As a byproduct of this investigation, informative synapomorphies of Bos taurus were found shared in every case by all other species of wild cattle (Tables 1, 2), establishing that neither banteng (Bos javanicus) or gaur (Bos gaurus) should be allied with bison + yak in contrast to a weakly supported outcome from a nuclear bead chip SNP array (43). It is not at all unusual for nuclear and mitochondrial phylogenies to differ. Using the latter, the taxonomic incongruity of yak nomenclature (Bos grunniens sistered to Bison bison) is easily remedied by a change to Bison grunniens (or Poephagus grunniens). Note that the divergence of bison and cattle exceeds that of bison and yak, taken as 2.5 myr after reviewing fossil evidence and prior estimates (40), so distinction by genus is warranted both by time of divergence and extent of accrued divergence.

Mitochondrial disease can results from tRNA mutations (33) through inefficient or incorrect production of mitochondrial proteins with clinical symptoms variable but ultimately similar to protein mitochondrial disease as both affect oxidative phosphorylation. Bison and yak tRNAs differed at 11 sites in 9 tRNAs though Pro, Trp, Ile, Leu (CUN), and Arg tRNAs were conserved in the 102 mitochondrial genomes considered (51). None of the tRNA substitutions had an exact counterpart to any linked to human mitochondrial disease; the degree of overall sequence divergence is too high in many cases for reliable annotation transfer.

Bison have no tRNA changes that track with sub-clades, only private alleles (51). None of these affect the anti-codon loop or other strongly conserved domains, disfavoring the errant nuclear gene hyper-mutagenizing scenario (46). Analysis of a manually curated alignment of 127 species including yak and 41 other cetartiodactyls shows high natural variability at all of the affected sites, suggesting the substitutions may be innocuous or nearly so. Thus tRNA alleles have no implications for bison conservation genomics management (51).

Sporadic amino acid mutations are not heritable if they represent somatic mutation, tissue-specific surge in heteroplasmic abundance of a haplotype not significantly represented in their own oocytes or occur in a bull. These have little significance for herd conservation genomics management in contrast to deleterious variations observed in multiple animals (which rules out sequencing error and a sub-dominant heteroplasmy ratio). Consequently amino acid variants observed in single bison are considered further in supplemental material (51).

Eight of the thirteen bison mitochondrially encoded proteins lack non-sporadic coding variation in the 33 complete bison genomes available, namely COX1-3, ND1-3, ND4L and ND6. Yak are similar with subclade level variation in ND1, ND3, ND6, COX1, COX3 and ATP8. These results accord with the relative conservation of these mitochondrial proteins within mammals (44, 45) and are again inconsistent with an errant imported hyper-mutating nuclear gene such as POLG replicase being the source of mitochondrial proteome variation (46).

However 8 sites in the remaining 5 bison proteins (9 in 7 yak proteins) do represent heritable nsSNP variation that cannot be explained as sequence error, somatic mutation or heteroplasmy artifact (Tables 3, 4). The nine approaches described in the methods section were applied in turn to each variant to distinguish between inconsequential neutral drift within the phylogenetically acceptable range of the site (reduced alphabet wobble), a derived adaptive shift, and structurally deleterious changes at the protein level.

These latter are not necessarily maladaptive because they could reflect balanced polymorphisms (eg malarial disease resistance of heterozygous sickle cell hemoglobin E6V nuclear gene (30), corresponding to a persistent heteroplasmy ratio in mitochondrial genes) or be fully compensated by co-evolving amino acid residues within the same protein or one that interacts with it (as seen for oxidative phosphorylation in yeast suppressors of the petite phenotype). These possibilities must be considered in evaluating a structurally deleterious substitution for mitochondrial disease.

The two strongest candidates for mitochondrial disease in bison are V98A in the CYTB gene (cytochrome b) and I60N in ATP6 (ATP synthase Fo component). These are systematically evaluated below using nine quasi-independent methods.

Neither variant can be explained by sequencing error. Both are observed in 17 whole genome determinations and targeted partial sequences from different laboratories. Neither occurs in a compositionally or structurally anomalous region that might be prone to repeated independent sequence error. Both are one-step base changes, the most common type of mitochondrial mutation.

Because both are observed in multiple animals in physically separated herds, neither somatic mutation nor heteroplasmy ratio surging are likely explanations. Despite high read coverage with modern sequencing technology, no heteroplasmy was reported even subsequent to enhanced awareness following its discovery in aurochsen (59). Although clearly heritable as the dominant heteroplasmic haplotype, no descendent genealogies have been directly studied: a heteroplasmy contribution from residual co-existing wildtype haplotypes may continue, with current limits of detection sensitivity being about 5% (46). Mitochondrial disease still results in humans in the heteroplasmic state (27, 29).

Both V98A and I60N represent potentially significant changes in the physical-chemical nature of the substituted amino acid not commonly found accepted in statistical compilations of observed amino acid changes. Valine and isoleucine are branched chain aliphatic residues with bulky hydrophobic side chains, typically deeply shielded in the interior of proteins (or as transmembrane regions) from the aqueous milieu. The side chain of alanine consists of a single small methyl substituent, that of asparagine is larger but strongly polar and hydrogen bonding. As substitutions, these are energetically unfavorable in typical proteins with the former leaving an unfilled hole and the latter disengaging a polar component from hydrogen-bonding solvent.

The site-specific context can strongly refine evaluation by the physical-chemical nature of the change alone -- not all residues operate under the same evolutionary constraints. Table 5 shows the outcome of very large scale comparative genomics alignment, made possible by thousands of GenBank entries for phylogenetically close-in mitochondrial proteins. Neither alanine or asparagine has any place in the normal reduced alphabet at these sites in any species -- despite the fact that both are just a simple base change away from the original amino acid and would not only have arisen frequently but also been fixed in some species as were changes to bona fide reduced alphabet elements in the same or flanking amino acids. Billions of years of branch length are sampled by the data.

Exceptions inevitably arise in massive data sets as allele frequencies sampled drop below 0.1%, matching sequence error rates especially for older data. Those exceptions can be objectively evaluated by examining their taxonomic coherence. In the case of both V98A and I60N, the exceptions are phylogenetically scattered (51) and do not form subclades indicative of acceptance into that subclade's local reduced alphabet. These exceptions could instead represent somatic mutation, heteroplasmy sampling effects, or even an independent reoccurrence of the mutation seen in bison. All of these are expected at some level given the high incidence of mitochondrial disease reported from human (11).

As a further control to other variant loci of bison and yak (51), variation relative to cattle mitochondrial genomes was assessed in all 13 mitochondrially encoded proteins of a Holocene aurochs sequenced at heteroplasmy depth (59). Here A23T in cytochrome b, T90I in ND5 and V55L in ND4 initially appear anomalous. However upon large-scale alignment within ruminants, all three changes are both common and taxonomic sub-clade coherent (51). This leaves no candidate loci for mitochondrial disease in the thirteen proteins of this aurochs.

In terms of structural impacts of the two bison amino acid substitutions, ATP6 and cytochrome b act in different structural complexes, namely the F_1F_O ATP synthase and complex III. These do not physically interact, implying that I60N and V98A cannot induce cross-compensatory change. Their effects on decreased ATP production could be additive or alternatively one mutation be rate limiting rendering the other moot.

ATP6, like many multi-pass membrane proteins, does not have a satisfactory crystallographic structural determination. To locate I60N within the tertiary structure, it is currently necessary to use E. coli PDB model 1C17 for bison, even though the alignment is quite weak overall at 27% identity and does not extend informatively to the I60N site (Table 6A). The corresponding residue, position 108, lies within the second transmembrane helix (56). However this is not consistent with secondary structure of mammalian orthologs.

The secondary structure analysis of yak ATP6 (SwissProt accession Q7YCA5) shows I60N is located in a loop region connecting the first two transmembrane helices as it does in cattle (accession P00847). The yak protein is 97% identical to bison and has the expected wildtype value isoleucine at position 60. Other secondary structure prediction tools give the same result (data not shown).

Conservation within the primary sequence may have structural implications, for example periodic conservation of 3 or 4 residues suggests an alpha helix with one side constrained. However phylogenetic conservation of a 5 residue window on each side of I60N does not show a clear pattern (Table 6B). Several the residues have broad reduced alphabets while others are near-invariant. Proline, a helix-breaker, occurs at position 63 in the bison protein. Thus the structural implications of asparagine at position 60, which arises from from wildtype valine via a T --> A transversion, cannot be determined, though it is clearly a radical substitution (branched aliphatic hydrophobic to polar) in terms of the constraints of this site, not being seen in 1635 other mammalian ATP6 sequences despite numerous substitutions in position 60 and nearby residues gaining traction.

The structural effect of V98A on bison cytochrome b can be assessed using the available high resolution data for the orthologous bovine complex (57). V98 occurs at the end of transmembrane helix B, sandwiched between six strongly conserved residues. The preceding residue, H97, is one of two critical axial histidine ligands to the high-potential heme. Two residues later, a strictly invariant

arginine R100 provides an essential salt bridge to a buried proprionate side chain of this same heme. Structural models of this region, adapted from previous publications (57, 58), are shown in Figure 1.

An alignment of 5000 cytochrome b sequences establishes that 99.6% of mammalian species have valine or isoleucine (and less commonly methionine). This cannot be because opportunities for fixation of substitutions are rare -- 18 of the 20 possible amino acids occur at position 62 of ATP6. V98A arises from a simple T to C transition. Thus strong selective pressure enforces the restriction to isoleucine and valine here.

A structural perturbation of the histidine axial ligand might change heme redox potential (optimized for the respiratory chain long ago); disruption of the proprionate salt bridge could affect folding, overall thermodynamic stability or production of reactive oxygen species. Thus it is no surprise that position 98 is under strong selection -- but why isoleucine and valine and no others?

The explanation lies in the post-beta carbon side chain, not the alpha carbon or peptide bond -- any amino acid provides these. A beta carbon alone cannot suffice -- in that case glycine and proline might be excluded from the reduced alphabet but nothing else. The side chain of alanine, the bison mutation under consideration, consists of just a methyl group. Here the complete absence of observed leucine at position 98 (Table 5) is informative because like isoleucine and valine, it is a branched chain aliphatic amino acid more similar than alanine. However the branching comes later, at the gamma carbon. This rules out insertion in the hydrophobic membrane milieu because leucine could do that equally well. Similarly phenylalanine, tryptophan and tyrosine do not occur at position 98, establishing the hydrophobicity, while necessary, is not sufficient. The reduced alphabet then consists of only isoleucine and valine because the constraint is a packing issue in a hydrophobic environment. Alanine is functionally unacceptable because it leaves a packing void that is energetically unfavorable to protein folding, stability or functionality at the axial or proprionate ligands (57).

Mutations at orthologous sites known to cause (or not) mitochondrial disease in other species can be informative, even though sequence divergence elsewhere in the protein means results cannot be literally carried over to bison. However establishment of pathogenicity and its clinical manifestation are still useful because bison are inconvenient to study in a veterinary context.

In the case of V98A in bison cytochrome b, a very similar mutation V98M causing severe mitochondrial disease has been studied in dog (35). A human polymorphism V98L was found in a large scale SNP survey but the individual's disease status (if any) was not determined. Mutations at fourteen other positions have been documented (27) for human cytochrome b, producing exercise intolerance (S35P, S151P, G290D), cardiac or skeletal muscle myopathy (G34S, G251S, N255H, G339E, multi-system disorder (Y278C), neonatal polyvisceral failure (S279P), and optic neuropathy from swollen ganglia (A29T, D171N, V356M). These are still relevant to bison V98A because at some level all mitochondrial disease affects the same process (cellular energy production via oxidative phosphorylation).

Another 125 human amino acid changing polymorphism of cytochrome b have been detected but rarely is a disease association ruled out. This is unfortunate because changes proven to be neutral could supplement the test suite provided by known pathogenetic mutations and allow double blinded assessment of prediction accuracy statistics. Lacking that, the first site occurrence in each phenotypic class (S35P, G34S, Y278C, S279P, and A29T), studied along with ATP6 described below by the same methods used here to assess bison V98A, all classified as pathogenic (51).

In the case of bison I60N in the ATP6 protein, no other mammal has a known pathogenic mutation or neutral polymorphism at position 60. However pathogenic human mutations are known at twenty sites along the protein. Three sites were selected for addition to the methods test suite, H90Y, L156P, and L220P to represent exercise fatigue alleles (50). The reported symptoms have an uncanny similarity to those of cytochrome b (which resides in a separate protein complex not interacting with ATP6): exercise endurance, muscle weakness, ataxia, lactic acidemia, optic neuropathy, and so forth (27). Thus the phenotype for bison I60N would likely be similar to these and indeed to V98A.

Experimental model systems allow biochemical testing of specific mutations. These are applicable to bison but have not yet been applied to the specific sites considered here. However five cytochrome b sites pathogenic in human have been tested in allotopic yeast, including two causing exercise intolerance (S151P, G290D) and one causing skeletal muscle myopathy (G34S). These gave results concordant with clinical expectations despite the great divergence of human and yeast but more importantly give insight into the specific chain of events leading to protein malfunction, for example failure to assembly into a respiratory sub-complex. The human ATP6 mutation L156P brings about a 45% reduction in ATP production relative to controls (53); L217P also had reduced ATP production though normal coupling efficiencies (54). Hamster cell lines offer less sequence mismatch; an allotopic system there has evaluated L156R in ATP6x (55). These studies indicate that bison mutations V98A and I60N could reduce mitochondrial ATP production by a wide variety of mechanisms. Together in the same haplotype, these could be additive, synergistic, or one site be rate limiting and the other irrelevant in the combination. Even with a low level of residual wildtype heteroplasmy, ATP production would still be lowered (53).

The prevalence of bison haplotypes hosting V98A / I60N is not directly known, other than from the 32 complete whole genomes sequenced and fragmentary cytochrome b (and many fewer ATP6) sequences. Researchers have not generally specified whether bison were chosen at random for sequencing nor indicated familial relationships or even gender. Sample sizes are quite small relative to herd size. With older samples, the bison would not still be alive. It is thus problematic to extrapolate to overall incidence in North American conservation herds at this time.

However a great deal of bison control region data exists. While not directly extendable to the coding regions of interest, the haplotype class defined by the control region can be mapped into the new global haplotype classes defined by whole genome sequencing (2). This mapping is necessarily ambiguous (one to many) given changes outside the control region are not observed and the full set of global haplotypes might not yet be known.

However the five global haplotypes containing the V98A / I60N double feature, namely Bhap2, Bhap8, Bhap10, Bhap11, and Bhap17 in earlier terminology (2) are very similar to each other overall according to all-vs-all Blastn (Table 7), identical in the control region fragment commonly sequenced and reliably distinguished there from the other known global haplotypes by conserved diagnostic residues within positions 15894-15965 of the Yellowstone bison GU947004 (Table 8).

Thus under the assumption that the control region alone serve as a satisfactory proxy to positions 98 of CYTB and 60 of ATP6, the status of an additional 133 bison (44 of them fossil, only two with potential V98A I60N) can be inferred from GenBank data. Gardipee determined haplotypes for 179 additional bison in a 2006 masters thesis (21). The data for contemporary are combined into an overall estimate of 145 bison with the V98A / I60N haplotype and 34 without in Table 8 and Table 9. All 29 bison sequenced from Grand Tetons National Park carry the deleterious allele, as do 117 of 151 bison tested from Yellowstone (72%). Rates are much lower in overall bison data (42%, Table 7) and were a few percent at most in fossil bison.

The two fossil bison at issue may carry either, both or neither of V98A I60N assuming the double haplotype arose in sequential steps over time. These bison, dated to 170 (Anchorage AY748509) and 3250 years ago (Natural Trap Cave, Wyoming AY748518) are priority targets for re-sequencing as they could reflect intermediate stages in the development of the disease haplotype. Steppe bison control region sequences are available (17) but not easily mapped to bison haplotypes. Here too it would be valuable to sequence entire mitochondrial genomes.

Discussion

Applying the nine methods for evaluating amino acid substitutions, observed bison protein variants V98A and I60N are both predicted to have significantly sub-optimal energy production from oxidative phosphorylation. Two mutations acting synergistically in the same heritable haplotype strengthens the conclusion that affected bison have mitochondrial disease to an estimated 98% confidence (see methods). Biochemical verification of reduced ATP production by allotropic knock-in and standard laboratory protocols is important and feasible (36).

Mitochondrial disease is common in humans so it comes as no great surprise to find another species affected by it. The alarming frequency of occurrence in bison can be attributed (3, 7) to the severe bison bottleneck of the nineteenth century followed decades of inbreeding and suppression of natural selection. Mitochondrial disease in dog breeds has a similar history (35).

Based on the available evidence, the disease haplotype was uncommon in pre-contact bison but widespread today, affecting bison in numerous discrete herds including Yellowstone and Grand Tetons national parks. Ironically, Yellowstone bison are used to found new herds and improve genetics of existing herds.

While symptoms of mitochondrial disease vary somewhat according to the specific mutation, the common denominator is inadequate ATP production from loss of oxidative phosphorylation capacity. Exercise intolerance, lactic acid buildup in blood, and ragged red muscle fiber can be expected in affected bison (2, 27). While not lethal at birth, these bison may be significantly impaired in escape from predators, winter cold tolerance, brushing snow aside for feeding, combat for breeding opportunities and similar aerobic activities.

If the goal in bison management is restoration of pre-contact genomics (1), then mutations in the mitochondrial genome need more extensive survey sequencing, direct biochemical and behavioral evaluation, genetic testing prior to culls and introductions to other herds and similar consideration of disease alleles in the companion nuclear genome. The nuclear genome may carry valuable genetic diversity (even in an animal with mitochondrial disease) but may also have very substantial genetic disease issues of its own (10).

To date, genetic management of bison conservation herds has largely eliminated bovine mitochondrial DNA introduced a century ago in hybridization experiments. Bison mitochondrial disease could also be managed away with retention of nuclear genetic diversity since only the latter is passed on by bulls. However this is not occurring with the present system of quasi-random culls of animals of unknown genetic status.

The present emphasis on genetically pure herds may be a misplaced priority if genetic disease is widespread in both nuclear and mitochondrial genomes of non-introgressed herds. Indeed, based on time of divergence and sequenced genomes from related laurasiatheres, the pre-contact bison genome will be 99% identical to nineteenth century cattle. These changes will be overwhelmingly concentrated in intergenic and intronic gene deserts and in synonymous sites in coding exons, with amino acid change primarily at intrinsically variable inconsequential sites (52). Thus a bison nuclear proteome, briefly introgressed by an introduced cow in 1900 and diluted for the subsequent century (7) will not exhibit a single significant change in a typical protein of 450 residues relative to a pre-contact bison, the real issue being exacerbation of the formerly heterozygous genetic burden by

inbreeding, small herd sizes and genetically uninformed culls. Mildly introgressed animals may preserve important and authentic pre-contact genetic diversity no longer otherwise represented in extant pure bison.

References

1. Marris E. Conservation: The genome of the American West. Nature 2009 Feb19;457(7232):950-2. Abstract

2. Douglas KC, Halbert ND et al. Complete mitochondrial DNA sequence analysis of Bison bison and bison-cattle hybrids: function and phylogeny. Mitochondrion 2011 Jan;11(1):166-75. Abstract

3. Hedrick PW. Conservation genetics and North American bison (Bison bison). JHered 2009 Jul-Aug;100(4):411-20. Abstract

4. Johnson WE, Onorato DP et al. Genetic restoration of the Florida panther. Science 2010 Sep24;329(5999):1641-5. Abstract

5. Saylor R, Shipley L Considering the possible extinction of the Columbia Basin pygmy rabbit. WSU Investigators Report 2004 Jan: 1-6. Pdf

6. McCallum H, Jones M et al. Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. Ecology 2009 Dec;90(12):3379-92. Abstract

7. Hedrick PW. Cattle ancestry in bison: explanations for higher mtDNA than autosomal ancestry. Mol Ecol 2010 Aug;19(16): 3328-35. Abstract

8. Halbert ND, Derr JN. A comprehensive evaluation of cattle introgression into US federal bison herds. J Hered 2007 Jan-Feb;98(1): 1-12. Abstract

9. Geist V. Phantom Subspecies: The Wood Bison (Bison bison athabascae Rhoads 1897) is not a valid taxon, but an ecotype. Arctic 1991 Dec;44(4):283-300. Pdf

10. Durbin RM, Abecasis GR et al. A map of human genome variation from population-scale sequencing. Nature 2010 Oct28;467 (7319):1061-73. Abstract

11. Cree LM, Samuels DC et al. The inheritance of pathogenic mitochondrial DNA mutations. Biochim Biophys Acta 2009 Dec;1792 (12):1097-102. Abstract

12. Miller W, Rosenbloom K et al. 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. GenomeRes 2007 Dec;17(12):1797-808. Abstract

13. Zimin AV, Delcher AL et al. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol 2009 10(4):R42. Abstract

14. Liu Y, Qin X et al. Bos taurus genome assembly. BMC Genomics 2009 Apr 24;10:180. Abstract

15. Genome 10K Community of Scientists Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate specie. J Hered 2009 Nov-Dec;100(6):659-74. Abstract

16. Shen YY, Shi P et al. Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. Genome Res 2009 Oct;19(10):1760-5. Abstract

17. Shapiro B, Drummond AJ et al. Rise and fall of the Beringian steppe bison. Science 2004 Nov 26;306(5701):1561-5. Abstract

18. Axelsson E, Willerslev E et al. The effect of ancient DNA damage on inferences of demographic histories. Mol Biol Evol 2008 Oct;25(10):2181-7. Abstract

19. Rambaut A, Ho SY et al. Accommodating the effect of ancient DNA damage on inferences of demographic histories. Mol Biol Evol 2009 Feb;26(2):245-8. Abstract

20. Zuzula GD, MacKay M et al. A late Pleistocene steppe bison (Bison priscus) partial from Tsiigehtchic, Northwest Territories, Canada. Quaternary Science Reviews 2009 28;2734–2742. Pdf

21. Gardipee MF. Development of fecal dna sampling methods to assess genetic population structure of Greater Yellowstone bison. University of Montana 2007 MS thesis. Pdf

22. Corpet F Multiple sequence alignment with hierarchical clustering. NucleicAcids Res 1988 Nov 25;16(22):10881-90. Abstract

23. Pütz J, Dupuis B et al. Mamit-tRNA, a database of mammalian mitochondrial tRNA primary and secondary structures. RNA 2007 Aug;13(8):1184-90. Abstract

24. Michaels GS, Hauswirth WW et al. Mitochondrial DNA copy number in bovine oocytes and somatic cells. Dev Biol 1982 Nov;94(1):246-51. Abstract

25. Helm M, Brulé H et al. Search for characteristic structural features of mammalian mitochondrial tRNAs. RNA 2000 Oct;6(10):1356-79. Abstract

26. Zhao H, Jie S et al. Length heteroplasmies in human mitochondrial DNA control regions and breast cancer risk. Int J Mol Epidemiol Genet 2010;1(3):184-192. Pdf

27. Brandon MC, Lott MT et al. MITOMAP: a human mitochondrial genome database--2004 update. Nucleic Acids Res. 2005 Jan 1;33:D611-3. Abstract

28. Steinmetz LM, Curt Scharfe. Systematic screen for human disease genes in yeast. Nat Genet. 2002 Aug;31(4):400-4. Abstract

29. Copeland WC. Mitochondrial DNA Replication Group: Mitochondrial Genetic Diseases. NIEHS NIH 2011 Jan. Html

30. Ingram VM. The chemical difference between normal and sickle cell haemoglobins. Biochim Biophys Acta 1959 Dec;36:402-11. Abstract

31. Adzhubei IA, Schmidt S et al. A method and server for predicting damaging missense mutations. Nat Methods 2010 Apr;7(4):248-9. Abstract

32. McClellan DA, Ellison DD. Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. Int J Bioinform Res Appl. 2010;6(2):120-33. Abstract

33. MitoMap Html

34. Sayers EW, Barrett T et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2011 Jan;39:D38-51. Abstract

35. Li FY, Cuddon PA et al. Canine spongiform leukoencephalomyelopathy is associated with a missense mutation in cytochrome b. Neurobiol Dis. 2006 Jan;21(1):35-42. Abstract

36. Giulivi C, Zhang YF et al. Mitochondrial dysfunction in autism. JAMA 2010 Dec 1;304(21):2389-96. Abstract

37. Zullo SJ, Parks W Tet al. Stable transformation of CHO Cells and human NARP cybrids following transfer of ATPase6 gene to the nuclear genome. Rejuvenation Res. 2005 Spring;8(1):18-28. Abstract

38. Hao HX, Rutter J. Revealing human disease genes through analysis of the yeast mitochondrial proteome. Cell Cycle. 2009 Dec 15;8(24):4007-8. Abstract

39. Perocchi F, Mancera E et al. Systematic screens for human disease genes, from yeast to human and back. Mol Biosyst. 2008 Jan;4(1):18-29. Abstract

40. Wang Z, Shen X et al. Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal. Journal of Biogeography 2010 Dec;v37(12)2332–2344. Abstract

41. Gu Z, Zhao X et al. Complete sequence of the yak (Bos grunniens) mitochondrial genome and its evolutionary relationship with other ruminants. Mol Phylogenet Evol. 2007 Jan;42(1):248-55. Abstract

42. XB Q, Jianlin H et al. Assessment of cattle genetic introgression into domestic yak populations using mitochondrial and microsatellite DNA markers. Anim Genet. 2010 Jun 1;41(3):242-52. [Abstract](#)

43. Decker JE, Pires JC et al. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. PNAS 2009 Nov 3;106(44):18644-9. [Abstract](#)

44. da Fonseca RR, Johnson WE et al. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics. 2008 Mar 4;9:119. [Abstract](#)

45. Shen YY, Shi P, Sun YB, Zhang YP. Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. Genome Res. 2009 Oct;19(10):1760-5. [Abstract](#)

46. Wang W, Shen P et al. Identification of rare DNA variants in mitochondrial disorders with improved array-based sequencing. Nucleic Acids Res. 2011 Jan 1;39(1):44-58. [20843780](#)

47. Schon EA, Gilkerson RW. Functional complementation of mitochondrial DNAs: mobilizing mitochondrial genetics against dysfunction. Biochim Biophys Acta. 2010 Mar;1800(3):245-9. [Abstract](#)

48. Poe BG, Duffy CF et al. Detection of heteroplasmy in individual mitochondrial particles. Anal Bioanal Chem. 2010 Aug;397(8):3397-407. [Abstract](#)

49. Osborne A, Reis AH et al. Single-Molecule LATE-PCR Analysis of Human Mitochondrial Genomic Sequence Variations. PLoS ONE. 2009; 4(5): e5636. [Abstract](#)

50. Elliott HR, Samuels DC et al. Pathogenic mitochondrial DNA mutations are common in the general population. Am J Hum Genet. 2008 Aug;83(2):254-60. [Abstract](#)

51. Supporting supplemental information is provided at the UCSC comparative genomics web site. [Html](#)

52. Schuster SC, Miller W et al. Complete Khoisan and Bantu genomes from southern Africa. Nature. 2010 Feb 18;463(7283):943-7. [Abstract](#)

53. Kucharczyk R, Rak M et al. Biochemical consequences in yeast of the human mitochondrial DNA 8993T>C mutation in the ATPase6. Biochim Biophys Acta. 2009 May;1793(5):817-24. [Abstract](#)

54. Kucharczyk R, Ezkurdia N et al. Consequences of the pathogenic T9176C mutation of human mitochondrial DNA on yeast mitochondrial ATP synthase. Biochim Biophys Acta. 2010 Jun-Jul;1797(6-7):1105-12.[Abstract](#)

55. Mattiazzi M, Vijayvergiya C et al. The mtDNA T8993G (NARP) mutation results in an impairment of oxidative phosphorylation that can be improved by antioxidants. Hum Mol Genet. 2004 Apr 15;13(8):869-79. [Abstract](#)

56. Tsaousis GN, Tsirigos KD et al. ExTopoDB: a database of experimentally derived topological models of transmembrane proteins. Bioinformatics. 2010 Oct 1;26(19):2490-2. [Abstract](#)

57. Iwata S, Lee JW et al. Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. Science 281:64-71 (1998). [Abstract](#)

58. Esser L, Gong X et al. Surface-modulated motion switch: capture and release of iron-sulfur protein in the cytochrome bc1 complex. PNAS 103:13045-13050(2006). [Abstract](#)

59. Edwards CJ, Magee DA et al. A Complete Mitochondrial Genome Sequence from a Mesolithic Wild Aurochs (Bos primigenius). PLoS One. 2010 Feb 17;5(2):e9255. [Abstract](#)

60. White DJ, Wolff JN et al. Revealing the hidden complexities of mtDNA inheritance. Mol Ecol. 2008 Dec;17(23):4925-42. [Abstract](#)

61. NCBI Blast server [Html](#)

Table 1. Data set of complete cytochrome b and whole mitochondrial genome sequences used. The first two columns provides genus, species and common name according to NCBI taxonomy. The third column gives the number of of complete genomes; the fourth auxiliary cytochrome b partial sequences. Bison and yak accession numbers are provided elsewhere (2, 40).

Bos	grunniens	yak	72	53
Bison	bison	plains bison	33	7
Bison	athabascae	woods bison	2	3
Bison	bonasus	wisent	4	9
Bison	priscus	steppe bison	0	0
Bos	sauveli	kouprey	0	5
Bos	frontalis	mithun gayal	0	16
Bos	gaurus	gaur	0	17
Bos	javanicus	banteng	2	39
Bos	taurus	cattle	168	500
Bos	primigenius	aurochs	1	17
Bos	indicus	zebu	3	387
Bubalus	bubalis	water buffalo	4	342

Table 2. Informative derived characters of Bos taurus shared by Bos javanicus and Bos gaurus: amino acid variation at otherwise conserved sites that are fixed in all wild cattle sequences but differ from ancestral ruminant value (as exhibited in water buffalo, bison and yak). None of the sites support a sister relationship of any species of wild cattle to bison + yak to the exclusion of Bos taurus; no evidence lineage sorting was found. Bos frontalis and Bos sauveli are not included because of insufficient data.

	Bos indicus		Bos primigenius		Bos javanicus		Bos gaurus	
	Bos	Bison	Bos	Bison	Bos	Bison	Bos	Bison
CYTB	4	0	4	0	4	0	4	0
ATP8	1	1	2	0	2	0	–	–
ND5	14	0	14	0	14	0	–	–
ND6	4	0	–	–	4	0	–	–
ND4L	2	0	–	–	2	0	–	–
ND3	0	0	0	0	0	0	–	–
COX1	–	–	1	0	1	0	–	–
COX3	5	0	5	0	5	0	–	–
COX2	2	0	2	0	2	0	2	0
ND2	5	0	5	0	5	0	–	–
ND1	2	0	2	0	2	0	–	–
ATP6	3	1	4	0	4	0	–	–
ND4	8	0	8	0	8	0	–	–
totals	50	1	47	0	47	0	6	0

Table 3. Bison non-sporadic mitochondrial protein variation. The first column provides GenBank accession, haplotype nomenclature (2) and bison provenance (YP: Yellowstone NP, BR: National Bison Refuge, MT: private Montana herd, EI: Elk Island NP, TX: Texas State Bison Herd, IT: Italian zoo). The second column provides multiplicity (number of bison with a given haplotype). The following columns provide the gene name and amino acid at a specified site (numbering relative to GU970004 proteins). V98A in CYTB (cytochrome b) and I60N in ATP6 are deleterious (see text).

Accession_Haplo_Loc	mult	CYTB	ATP6	ATP6	ND5	ND4	ATP6	ND4	ATP8
GU947004_bHap17_YP	1	A98	N60	M182	H159	T314	A177	L442	E38
GU947001_bHap2__BR	11	A98	N60	M182	H159	T314	A177	L442	E38
GU947000_bHap10_FN	3	A98	N60	M182	H159	T314	A177	L442	E38
GU946994_bHap11_MT	1	A98	N60	M182	H159	T314	A177	L442	E38
GU946988_bHap8__MT	1	A98	N60	M182	H159	T314	A177	L442	E38
GU946979_bHap3__MT	2	V98	I60	T182	Y159	A314	A177	L442	K38
GU946998_bHap12_MT	1	V98	I60	T182	Y159	A314	A177	L442	K38
GU946980_bHap4__MT	1	V98	I60	T182	Y159	A314	A177	L442	E38
GU946985_bHap6__MT	1	V98	I60	T182	Y159	A314	A177	L442	E38
GU946989_bHap9__MT	2	V98	I60	T182	Y159	A314	A177	M442	E38
GU946982_bHap5__MT	1	V98	I60	T182	Y159	A314	A177	M442	E38
GU947006_wHap14_EI	1	V98	I60	T182	Y159	A314	A177	L442	E38
EU177871_bHapX__IT	1	V98	I60	T182	Y159	T314	A177	L442	E38
GU946987_bHap7__MT	1	V98	I60	T182	Y159	A314	A177	L442	E38
GU947002_bHap13_TX	2	V98	I60	T182	Y159	A314	T177	L442	E38
GU947003_bHap16_TX	1	V98	I60	T182	Y159	A314	T177	L442	E38
GU947005_wHap15_EI	1	V98	I60	T182	Y159	A314	T177	L442	E38

Table 4. Yak non-sporadic mitochondrial protein variation. The first column provides GenBank accession, haplotype nomenclature (grouped from those of reference 40 which also provides Tibetan location). The following columns provide the gene name and amino acid variation at a specified site (numbering relative to GQ464262 proteins). M383V is possibly sub-functional (see text).

Accession_Haplo	ND2	ND2	ND4	ATP8	ND5	ND5	ND4L	COX3	CYTB
GQ464262_hapA	I29	S318	V398	S63	I62	M383	V17	V192	I192
GQ464261_hapA	I29	S318	V398	S63	I62	M383	V17	V192	I192
GQ464257_hapA	I29	S318	V398	S63	I62	M383	V17	V192	I192
GQ464251_hapA	I29	S318	V398	S63	I62	M383	V17	V192	I192
GQ464250_hapA	I29	S318	V398	S63	I62	M383	V17	V192	I192
GQ464249_hapA	I29	S318	V398	S63	I62	M383	V17	V192	I192
GQ464260_hapB	V29	P318	M398	L63	I62	M383	V17	V192	I192
GQ464266_hapB	V29	P318	M398	L63	V62	V383	V17	V192	I192
GQ464263_hapB	V29	P318	M398	L63	V62	V383	V17	V192	I192
GQ464259_hapB	V29	P318	M398	L63	V62	V383	V17	V192	I192
GQ464264_hapB	V29	P318	M398	L63	V62	V383	V17	V192	T192
GQ464246_hapB	V29	P318	M398	L63	V62	V383	V17	V192	T192
GQ464253_hapB	V29	P318	M398	L63	V62	V383	V17	V192	I192
GQ464252_hapB	V29	P318	M398	L63	V62	V383	V17	V192	I192
GQ464247_hapB	V29	P318	M398	L63	V62	V383	V17	V192	I192
GQ464255_hapB	V29	P318	M398	L63	V62	V383	A17	V192	I192
GQ464254_hapB	V29	P318	M398	L63	V62	V383	A17	V192	I192
GQ464265_hapB	V29	P318	M398	L63	V62	V383	A17	V192	I192
GQ464248_hapB	V29	P318	M398	L63	V62	V383	V17	I192	I192
GQ464258_hapB	V29	P318	M398	L63	V62	V383	V17	I192	I192
GQ464256_hapB	V29	P318	M398	L63	V62	V383	V17	I192	I192

Table 5. Reduced alphabet for each nsSNP determined by large-scale comparative genomics alignment. The first two rows provide mitochondrial gene name and variation relative to ancestral. The columns under each variant provide the abundances of other amino acids at that position in other ruminant data. The set of major alternative residues at a given site form its reduced alphabet; uncommon residues generally reflect sequence error or mutation (possibly somatic or heteroplasmic) in the individual chosen for sequencing but are sometimes adaptive or a clade-specific addition to the reduced alphabet.

CYTB		ATP6		ATP6		ND5		ND4		ATP6		ND4		ATP8	
V98A		I60N		T182M		Y159H		A314T		A177T		L442M		E38K	
4522	V	531	M	553	S	225	Y	281	A	998	A	221	L	215	E
430	I	392	I	286	M	73	H	8	T	88	T	46	I	34	K
34	M	106	T	98	T			5	I	4	S	19	M	17	S
11	A	37	V	92	L			3	V	2	V	8	T	15	M
2	G	6	A	57	I					1	P	2	V	5	T
1	A	5	N	10	A					1	A	1	F	5	G
		4	L	4	V									3	V
		2	P	2	F									2	A
		1	S	1	M										
				1	C										

Table 7. Haplotype clustering by all-vs-all Blastn of all complete bison mitochondrial genomes. The first column provides global haplotypes using the terminology of reference 2; the second a reduced set of haplotypes defined in Table 7 by the control region diagnostic domain; the third a representative accession number. The fourth column gives multiplicities (example: first data row shows 11 complete mitochondrial genomes are identical to GU94700). The remaining columns show the number of base pair differences out of the 16,323 compared. Shading indicates blocks of closely related global haplotypes. The matrix is not quite symmetric because of gap alignment effects.

Haplo	mHap	Accession	#	U01	U04	U00	994	988	997	U06	982	980	871	985	987	992	998	U05	U02
BNBR1	hapA	GU947001	11	0	2	3	4	4	20	21	21	22	22	25	25	26	26	23	23
BYNP1586	hapA	GU947004	1	2	0	5	6	6	22	23	23	24	24	25	25	28	28	25	25
BFN5	hapA	GU947000	3	3	5	0	1	1	19	20	20	21	21	24	24	25	25	22	22
B1031	hapA	GU946994	1	4	6	1	0	2	20	21	21	22	22	25	25	26	26	23	23
B973	hapA	GU946988	1	4	6	1	2	0	20	21	21	22	22	25	25	26	26	23	23
B1091	hapB	GU946997	2	20	22	19	20	20	0	11	1	8	16	11	11	12	12	13	13
wElkI14	hapB	GU947006	1	21	23	20	21	21	11	0	12	13	17	16	16	17	17	14	16
B897	hapB	GU946982	1	21	23	20	21	21	1	12	0	9	17	12	12	13	13	14	14
B877	hapB	GU946980	1	22	24	21	22	22	8	13	9	0	18	7	10	10	10	11	9
Euro	hapB	EU177871	1	22	24	21	22	22	16	17	17	18	0	21	21	22	22	19	21
B935	hapB	GU946985	1	25	25	24	25	25	11	16	12	7	21	0	13	13	13	14	12
B961	hapB	GU946987	1	25	25	24	25	25	11	16	12	10	21	13	0	14	14	15	15
B1018	hapB	GU946992	2	26	28	25	26	26	12	17	12	10	22	13	14	0	1	15	15
B1191	hapB	GU946998	1	27	26	24	25	25	13	18	14	11	23	14	15	1	0	16	16
wElk1	hapC	GU947005	1	23	25	22	23	23	13	14	14	11	19	14	15	15	15	0	10
BTSBH1001	hapD	GU947002	3	23	25	22	23	23	13	16	14	9	21	12	15	15	15	10	0

Table 6. A. Phylogenetic conservation of ATP6 in a ten amino acid window about I60N. Light shading shows amino acids found in other mammals at greater than 1% frequency. Dark shading indicates bison sequence (not always the highest frequency)>The neighborhood of I60N exhibits variable levels of conservation and no overt periodicity. The structure of the loop region containing I60N is thus not constrained by available data. B. Alignment against the only homologous protein with determined 3D structure (PDB 1C17|M from E. coli) is not informative in the I60N region so the effect of the asparagine substitution cannot be modeled.

Phylogenetic Conservation of ATP6 in Neighborhood about I60N

55		56		57		58		59		60		61		62		63		64		65	
K	1356	Q	1582	M	1011	M	1189	A	325	M	535	H	1083	N	1336	T	603	K	1151	G	1600
Q	161	H	13	L	515	F	126	S	238	P	525	L	414	S	197	K	212	G	212	A	11
N	37	N	12	I	75	I	7	L	196	I	391	I	72	T	38	P	193	N	54	S	5
H	19	E	11	V	17	T	4	M	92	T	108	V	29	G	21	Q	173	P	52	V	3
S	17	R	2	F	5	V	2	I	90	Q	35	M	24	D	17	L	107	Q	31	M	2
Y	12	Y	1	T	2	S	2	Q	58	V	9	T	1	H	6	H	71	T	28	T	1
G	10	T	1			Y	1	N	49	L	8	Q	1	P	4	V	54	A	28	P	1
R	6	Q	1					G	25	A	6	H	1	K	4	I	49	E	25	K	1
T	3	I	1					V	12	S	1	A	1	R	1	S	41	S	18	G	1
M	1	F	1					F	11	N	1			N	1	A	37	L	8	E	1
L	1							H	4	H	1			L	1	N	25	R	6		
K	1							S	1							Y	24	M	4		
I	1							R	1							M	14	I	3		
E	1															R	13	D	2		
A	1															F	4	K	1		
																E	3	H	1		
																P	2				
																Y	1				

B. Alignment of bison ATP6 with homologous protein PDB 1C17|M from E. coli

```
                         *
Bison: KQMMSNHNPKGQTWTLMLMLMSLILFIGSTNLLGLLPHSFTPTTQLSMNLGMA
                        L+ + L+ +I    ++LGL       P+   +++ L MA
Ecoli: PLALTIFVWVFLMNLMDLLPIDLLPYIAE–HVLGLPALRVVPSADVNVTLSMA
```

Table 8. Predicting health status from control region haplotypes. The first column provides GenBank accession, the second the sequence of the diagnostic portion of the control region (positions 15894-15965 of Yellowstone bison GU947004), the third haplotype notation here, with columns four through six providing the observed or predicted amino acid of position 98 in cytochrome b and position 60 in ATP6. Column seven correlates diagnostic region haplotype with haplotype terminology of reference 2. Light red: observed disease alleles; dark red: predicted disease. Light green: observed healthy alleles; dark green: predicted health. Other light/dark colors pairs show observed/predicted healthy alleles in less common haplotypes.

Accession	TTGCAAACACCACTAGCTAACGTCACTCACCCCCAAAAATGCATTACCCAAACGGGGGGGAAATATACATAAC	Haplo	Status	CYTB	ATP6	Haplo1
GU946996	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU946995	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU946981	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU946976	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU946983	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU946984	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU946993	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU946986	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU946977	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU946978	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU947001	C.......................................−................................	hapA	observed	98A	60N	bHap2
GU947004	C.......................................−................................	hapA	observed	98A	I60	bHap17
GU946991	C.......................................−................................	hapA	observed	98A	60N	bHap10
GU946990	C.......................................−................................	hapA	observed	98A	60N	bHap10
GU947000	C.......................................−................................	hapA	observed	98A	60N	bHap10
GU946994	C.......................................−................................	hapA	observed	98A	60N	bHap11
GU946988	C.......................................−................................	hapA	observed	98A	60N	bHap8
AY748509	C.......................................−................................	hapA	predicted	98A	60N	.....
AY748518	C.......................................−................................	hapA	predicted	98A	60N	.....
AF083362	C.......................................−................................	hapA	predicted	98A	60N	.....
AF083357	C.......................................−................................	hapA	predicted	98A	60N	.....
EF693810	C.......................................−................................	hapA	predicted	98A	60N	.....
U12948..	C.......................................−................................	hapA	predicted	98A	60N	.....
U12946..	C.......................................−................................	hapA	predicted	98A	60N	.....
U12945..	C.......................................−................................	hapA	predicted	98A	60N	.....
U12934..	C.......................................−................................	hapA	predicted	98A	60N	.....
U12933..	C.......................................−................................	hapA	predicted	98A	60N	.....
EU272058	C.......................................−................................	hapA	predicted	98A	60N	.....
EU272056	C.......................................−................................	hapA	predicted	98A	60N	.....
DQ452027	C.......................................−................................	hapA	predicted	98A	60N	.....
AY428859	C.......................................−................................	hapA	predicted	98A	60N	.....
GU946997	....−.............−−.................	hapB	observed	V98	I60	bHap9
GU946989	....−.............−−.................	hapB	observed	V98	I60	bHap9
GU947006	....−.............−−.................	hapB	observed	V98	I60	wHap14
GU946982	....−.............−−.................	hapB	observed	V98	I60	bHap5
GU946980	....−.............−−.................	hapB	observed	V98	I60	bHap4
EU177871	....−.............−−.................	hapB	observed	V98	I60	.....
GU946985	....−.............−−.................	hapB	observed	V98	I60	bHap6
GU946987	....−.............−−.................	hapB	observed	V98	I60	bHap7
GU946992	....−.............−−.................	hapB	observed	V98	I60	bHap3
GU946979	....−.............−−.................	hapB	observed	V98	I60	bHap3
GU946998	....−.............−−.................	hapB	observed	V98	I60	bHap12
AY748745	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748477	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748530	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748519	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748630	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748495	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748693	....................A...................−.............−−.................	hapB	predicted	V98	I60	.....
AY748690	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748689	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748687	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748702	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748739	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748747	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748700	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748685	...............................−..........A..−−.................	hapB	predicted	V98	I60	.....
AY748678	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748676	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748677	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748701	....−.............−−.................	hapB	predicted	V98	I60	.....
AY748680	....−.............−−.................	hapB	predicted	V98	I60	.....

ID	Sequence	Hap	Status			
AY748671	....−.................−−..............	hapB	predicted	V98	I60	.....
AY748672	....−.................−−..............	hapB	predicted	V98	I60	.....
AY748679	....−.................−−..............	hapB	predicted	V98	I60	.....
AY748617	........T........................−.......T.........−−..............	hapB	predicted	V98	I60	.....
AY748699	....−.................−−..............	hapB	predicted	V98	I60	.....
AY748620	..........................T.......................−.................−−..............	hapB	predicted	V98	I60	.....
AY748748	....−.................−−..............	hapB	predicted	V98	I60	.....
AY748742	....−.................−−..............	hapB	predicted	V98	I60	.....
AY748758	....−.................−−..............	hapB	predicted	V98	I60	.....
AY748757	....−.................−−..............	hapB	predicted	V98	I60	.....
AY748618	....−.................−−..............	hapB	predicted	V98	I60	.....
AF083364	....−.................−−..............	hapB	predicted	V98	I60	.....
AF083363	....−.................−−..............	hapB	predicted	V98	I60	.....
AF083360	....−.................−−..............	hapB	predicted	V98	I60	.....
AF083361	....−.................−−..............	hapB	predicted	V98	I60	.....
U12941..	....−.................−−..............	hapB	predicted	V98	I60	.....
U12943..	....−.................−−..............	hapB	predicted	V98	I60	.....
U12960..	....−.................−−..............	hapB	predicted	V98	I60	.....
GU947005	....−.................−−.............T	hapC	observed	V98	I60	wHap15
AY748478	....−.................−−.............T	hapC	predicted	V98	I60	.....
AY748476	..A..−.................−−.............T	hapC	predicted	V98	I60	.....
AY748695	............................T.......−.................−−.............T	hapC	predicted	V98	I60	.....
AY748675	....−.................−−.............T	hapC	predicted	V98	I60	.....
AY748673	....−.................−−.............T	hapC	predicted	V98	I60	.....
AY748674	....−.................−−.............T	hapC	predicted	V98	I60	.....
AY748529	....−.................−−.............T	hapC	predicted	V98	I60	.....
AF083359	....−.................−−.............T	hapC	predicted	V98	I60	.....
U12959..	....−.................−−.............T	hapC	predicted	V98	I60	.....
U12942..	....−.................−−.............T	hapC	predicted	V98	I60	.....
U12937..	....−.................−−.............T	hapC	predicted	V98	I60	.....
U12951..	....−.................−−.............T	hapC	predicted	V98	I60	.....
GU946999	....−.......−....G.......T	hapD	observed	V98	I60	bHap13
GU947002	....−.......−....G.......T	hapD	observed	V98	I60	bHap13
GU947003	....−.......−....G.......T	hapD	observed	V98	I60	bHap16
AY748682	....−−..............	hapE	weak pred	V98	I60	.....
AY748669	....−−..............	hapE	weak pred	V98	I60	.....
AY748521	....−−..............	hapE	weak pred	V98	I60	.....
AF083358	....−−..............	hapE	weak pred	V98	I60	.....
AY748681	....−.................−..............	hapF	weak pred	98A	60N	.....
U12864..	....−.................−..............	hapF	weak pred	98A	60N	.....

Table 9. Sequence data of Gardipee (22) for Yellowstone and Grand Tetons National Parks. First and second columns provide herd locations. The third column shows number of bison carrying the disease haplotype; the fourth the numbers of bison with wildtype mitochondrial DNA. The final column shows the percentage of healthy bison varies with geographic location of the herds (which have little observed mixing).

Park	Herd	V98A I60N	V98V I60I	% Healthy
YNP	Hayden Valley	88	6	6%
YNP	Lamar Valley	19	22	54%
YNP	Mirror Plateau	10	6	38%
GTNP	Antelope Flats	20	0	0%
GTNP	Wolf Creek	8	0	0%

Figure 1. Structural models of the V98A region of bison cytochrome b adapted from previous crystallographic models of the bovine ortholog (57, 58). Here H97 is an axial ligand to heme iron and R100 forms a salt bridge to heme proprionate.