1

Development of a novel mule deer genomic assembly and species-diagnostic SNP panel for assessing introgression in mule deer, white-tailed deer, and their interspecific hybrids

[†]Ty Russell, [†]Catherine Cullingham , [‡]Arun Kommadath, [‡]Paul Stothard, [‡]Allen Herbst, [†]David Coltman

[†] University of Alberta, Department of Biological Sciences, Edmonton, Alberta, Canada, T6G 2E9

[‡] University of Alberta, Department of Agricultural Food and Nutritional Science, Edmonton, Alberta, Canada, T6G 2E9

Abstract

Mule deer (*Odocoileus hemionus*) are endemic to a wide variety of habitats in western North America, many of which are shared in sympatry with their closely related sister-species white-tailed deer (*Odocoileus virginianus*), whom they hybridize with in wild populations. Although mule deer meet many ideal conditions for a molecular ecological research species, such as high abundance, ecological importance, and broad dispersal and gene flow, conservation genetic studies have been limited by a relative lack of existing genomic resources and inherent difficulties caused by introgression with white-tailed deer. Many molecular tools currently available for the study of cervids were designed using reference assemblies of divergent model species, specifically cattle (*Bos taurus*). Bovidae and Cervidae diverged approximately 28 million years ago, therefore, we sought to ameliorate the available resources by contributing the first mule deer whole genome sequence draft assembly with an average genome-wide read depth of 25X, using the white-tailed genome assembly (Ovir.te_1.0) as a reference. Comparing the two assemblies, we identified ~33 million single nucleotide polymorphisms (SNPs) and insertion/deletion variants. We then verified fixed SNP differences between the two species and developed a 40-loci SNP assay capable of identifying pure mule deer, white-tailed deer, and interspecific hybrids. Assignment capacity of the panel, which was tested on simulated datasets, is reliable up to and including the third backcross hybrid generation. Identification of post-F1 hybrids will be necessary for hybrid zone population studies going forward, and the new mule deer assembly will be a valuable resource for genetic and comparative genomics studies.

Introduction

Hybridization is not uncommon among plant and animal taxa worldwide (Schwenk *et al.* 2008). Mallet (2005) estimated that 10-30% of multicellular plant and animal species hybridize regularly and that, of those species, between 1 in 100 and 1 in 10 000 sympatric individuals are hybrids. Introgression has been identified as a mechanism for generating progressive evolutionary events such as novelty, divergent selection, and speciation (Dowling and Secor 1997; Lamer *et al.* 2015; Seehausen 2004) though it can also complicate management and conservation by compromising coadapted gene complexes (Edmands *et al.* 2009), morphological discernment (Leary *et al.* 1996), local adaptation (Martinsen *et al.* 2001), and the genetic integrity of unique phylogenetic lineages (Rhymer and Simberloff 1996). For these reasons, genetic tools to monitor current distributions and degrees of hybridization will be valuable for future researchers, policy-makers, hunters, and conservationists.

Introgressive hybridization tends to occur most frequently among sympatric, closely related species in rapidly diversifying adaptive radiations (Abbott *et al.* 2013; Gourbière and Mallet 2010; Price and Bouvier 2002) such as between mule deer (MD) (*Odocoileus hemionus*) and its sister species, white-tailed deer (WT) (*O. virginianus*), in the prairies of Western North America (Bradley *et al.* 2003 ; Cronin 1991; Hornbeck and Mahoney 2000; Stelfox and Adamczewski1993). Cervids constitute cornerstone taxa in ecological, economic, and cultural sectors of Western Canada. The output value of elk and deer farms, before indirect spillover effects into other industries, is estimated to be more than $43 million in Canada with Alberta being responsible for over a quarter of that total (Petigara *et al.* 2011). As well, total big game hunting expenditures by Canadian residents in 2012 exceeded $1 billion with $169 million coming from Albertans (Federal, Provincial, and Territorial Governments of Canada, 2014). As of 2018, Canadian federal and provincial laws do not recognize hybrid deer as a separate entity from the parental species, which can create confusion as to hunting and harvesting regulations. Access to tools capable of identifying hybrid individuals and characterizing the rate of hybridization in wild populations could facilitate the implementation of management standards. Hybridization may also play a role in the spread of chronic wasting disease (CWD), a transmissible spongiform encephalopathy (TSE) that affects both focal *Odocoileus spp.* (Miller *et al.* 2012; Williams 2005). CWD has been reported to be more prevalent in MD relative to WT in areas of sympatry (Habib *et al.* 2011; Miller *et al.* 2000). Behavioral differences between species, in part, account for this asymmetry (Cullingham *et al.* 2010; Cullingham *et al.* 2011) but genetic polymorphisms have also been implicated (Wilson *et al.* 2009). Considering that hybrids potentially provide an opportunity to bridge disease transmission across species, further research of CWD susceptibility and pervasiveness among hybrids may provide insight into the transmission dynamics.

Identification of hybrid deer using morphological traits alone is rarely done with confidence. Coloration and antler shape are not always intermediate between parental phenotypes. The most consistent and accessible morphological marker appears to be the metatarsal gland, which is intermediate in both position and size (Carr *et al.* 1986; Wishart 1980). The biomechanics of the escape gait may serve as another indicator. While the stot of the MD is highly differentiated from the gallop of the WT, the bound of the hybrid is highly variable, even between strides of the same individual, and seems to be wholly inefficient (Lingle 1993). Molecular markers used to identify hybrids include serum albumin electrophoresis (Hornbeck and Mahoney 2000), a ribosomal 28S DNA marker (Bradley *et al.* 2003), and mitochondrial endonuclease recognition site mapping (Carr *et al.* 1986). None of these methods, however, are reliably informative of backcrosses after the F1 generation. Single nucleotide polymorphisms (SNPs) have proven effective in investigations of admixture between divergent taxa in several hybrid systems (Stephens *et al.* 2009; Twyford and Ennos 2011; Wiley *et al.* 2009). This is due, in part, to their high abundance in the genome which improves discriminating power and facilitates the recognition of varying levels of introgression. Furthermore, SNPs are more consistently reproducible, easier to automate, and more stable in mammals than microsatellites (Fernandez *et al.* 2013; Tokarska *et al.* 2009). An added benefit of SNPs is that their biallelic nature is convenient for the discernment of two different species.

A species-discriminating SNP assay will effectively describe two traits of hybridization: the classification of hybrids based on the presence of species-specific alleles and a measure of hybridity based on the proportions of those alleles. Quantification of the varying introgression depths of hybridized populations will provide a snapshot with more resolution than would be available from observation of the F1 individuals alone. Detection power of this magnitude is necessary to explore hybrid zone structure because current hypotheses predict reduced fitness in F1 individuals (Hornbeck and Mahoney 2000); therefore, the frequency and abundance of F1's in the wild is not a suitable proxy for the occurrence of advanced-generation backcrosses. The ability to resolve backcross generation status will be informative of hybrid productivity and directionality. By differentiating advanced-generation backcrosses from pure breeding individuals, the rate of false negative hybrid diagnoses will drastically decrease, effectively increasing hybrid sample size and lineage diversity as well as our understanding of hybridization frequency and geographical distribution (Lamer *et al.* 2015). The identification of the degree of introgression is crucial for determining the structure and stability of the population and the entire hybrid zone in which it resides. For example, whether selection is working for or against hybrids can help to elucidate the type of suture zone present and, from that information, we can draw inferences and predictions about the hybrid zone itself (Hu 2005). By better understanding the structure of the populations that make up the hybrid zone,

we improve our ability to forecast its future.

Currently, molecular resources suitable for application in *Odocoileus spp.* exist as sets of various individual loci but none capture mass sequence data capable of providing a reference to which polymorphisms can be mapped. Along with assorted microsatellite loci (Bishop *et al.* 1994; DeWoody *et al.* 1995; Jones *et al.* 2000; Wilson *et al.* 1997), current SNP datasets have been engineered by subjecting MD, black-tailed deer (*O. h. columbianus*), and WT to genotyping on the BovineSNP50 Bead Chip (Haynes and Latch 2012); bovine exon-targeted resequencing (Powell *et al.* 2016); and mapping high-throughput next-generation sequencing data to an existing bovine reference genome assembly (Brauning *et al.* 2015). All of these datasets, however, relied heavily upon previously available bovine genomic resources for mapping and/or template purposes, thereby losing data in regions not conserved after the divergence of Cervidae and Bovidae some 28 million years ago (Hassanin *et al.* 2012). A whole genome sequence (WGS) assembly of MD will function as a more suitable reference for future polymorphism mining in Cervids and increase accessibility to regions of the genome not conserved from Bovidae.

In this study, we present two novel genomic resources targeted for the study of MD and MD X WT hybrids: a draft WGS assembly of the MD and a species-diagnostic SNP panel. The MD WGS data is a versatile and informative dataset for cervid researchers investigating the genetic basis of traits, disease susceptibilities, and variants by providing a reference genome to which polymorphisms can be mapped (Ng and Kirkness 2010). The species-diagnostic SNP panel provides a method to reliably differentiate between MD and WT as well as to identify hybrids and quantify introgressive geneflow, even in the limited amounts present in advanced-generation backcrosses.

Materials and methods

Reference-guided MD Genome Assembly and Annotation

Total DNA was isolated from the ear tissue of a hunter harvested MD. The ear was shaved to remove hair and subsequently powdered by grinding under liquid nitrogen with a mortar and pestle. DNA was extracted from the powdered ear tissue (0.1g) for one hour at 68 degrees centigrade with 500μl of an extraction buffer containing 2% cetrimonium bromide, 100mM TrisCl pH7.5, 1.4M NaCl, 20mM EDTA, 2% polyvinylpyrrolidine, 0.2% mercaptoethanol, 100μg/mL proteinase K, and 20μg/mL RNase A. An equal volume of chloroform/isoamyl alcohol (24:1) was added, the sample centrifuged for 5 minutes at 14,000 x g and the aqueous phase collected. The aqueous phase was extracted with tris saturated phenol/chloroform, centrifuged and again the aqueous phase collected. Finally, the DNA was precipitated with 2.5 volumes of 95% ethanol and 0.1 volume of sodium acetate. The pellet was washed with 70% ethanol, briefly dried to remove residual ethanol, and dissolved in water. DNA quality was assessed by UV-Vis spectroscopy and agarose gel electrophoresis.

Table 1. Assignment criteria and category definition of white-tailed (WT) and mule deer (MD) hybrids. Heterozygous and homozygous genotype frequencies are required as input parmeters in NewHybrids to specify the number and configuration of hybrid classes.

NewHybrid category		Genotype frequency			NewHybrid category expanded
		WT (AA)	H (AB)	MD (BB)	
WT	Parental WT	1	0	0	WT
MD	Parental MD	0	0	1	MD
F1	First-generation hybrid	0	1	0	(WT x MD)
F2	Second-generation hybrid	0.25	0.5	0.25	((WT x MD) x (WT x MD))
BxWT	First-generation backcross	0.5	0.5	0	(WT x (WT x MD))
BxMD	First-generation backcross	0	0.5	0.5	(MD x MD x WT))
Bx2WT	Second-generation backcross	0.75	0.25	0	(WT x (WT x (WT x MD)))
Bx2MD	Second-generation backcross	0	0.25	0.75	(MD x (MD x (MD x WT)))
Bx3WT	Third-generation backcross	0.875	0.125	0	(WT x (WT x (WT x (WT x MD))))
Bx3MD	Third-generation backcross	0	0.125	0.875	(MD x (MD x (MD x (MD x WT))))

Genomic DNA was sequenced on an Illumina HiSeq X Ten sequencer (shotgun PCR free library; paired-end sequencing; 150 bp reads). Read quality was assessed using FastQC software (available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) prior to and after performing quality control (QC) steps. The QC included quality based read trimming and adapter removal using Trimmomatic (Bolger *et al.* 2014) followed by adapter and phiX read removal using Cutadapt (Martin 2011) with default settings for both software. Reads that passed QC were mapped to the WT deer reference genome assembly (Ovir.te_1.0; GenBank assembly accession: GCA_002102435.1) using the BWA-MEM algorithm of the Burrows-Wheeler Alignment tool (Li 2013) with default settings. Mapped reads were sorted and indexed using SAMtools (Li *et al.* 2009). Duplicate reads were tagged using MarkDuplicates software from Picard tools (available at: http://broadinstitute.github.io/picard). Variant calling and MD consensus sequence generation were performed using a combination of mpileup, bcftools and vcfutils.pl scripts from SAMtools. Identified SNPs were annotated using SnpEff (Cingolani *et al.* 2012). Regions with no coverage on the consensus sequence were filled with N's.

Selection and Validation of Species-discriminating SNPs

DNA was extracted from a sampling group of multiple species of deer as part of another study whose purpose was to aggregate and identify markers that would be useful to the New Zealand deer industry for parentage assignment and sub-species differentiation (Rowe *et al.* 2015). Samples were genotyped on Illumina 50K CervusSNP50, 24-sample Bead Chips (Illumina, San Diego), a high-throughput SNP assay developed for deer (*Cervus* genus) by comparing genomic sequence data from red deer, wapiti (subspecies of *Cervus elaphus*), and sika (*Cervus nippon*) by Brauning *et al.* (2015). A set of 44,448 SNPs were genotyped in 396 individuals, including 17 WT and 8 MD from Alberta and Saskatchewan, Canada. Loci and samples exhibiting non-

autosomal and non-Mendelian inheritance patterns, low call rates, and duplication were excluded (Rowe *et al.* 2015). The GenABEL package in R software (Aulchenko *et al.* 2007) was used to carry out genotype quality control measures. We identified 129 loci that had call rates >0.7 in both WT and MD and also appeared to show fixed differences between the two species, ie. biallelic loci where one allele is homozygous in WT and the other in MD.

A preliminary validation step was implemented by aligning the 201 bp sequences that include the SNP and its 100 bp flanking sequences to the new MD genome assembly generated here and to the WT genome assembly (Ovir.te_1.0). This allowed us to eliminate 10 SNPs that were non-discriminating or otherwise unable to be mapped prior to assay design and further validation, leaving 119 species-discriminating loci, from which a subset of 40 were randomly selected for validation. Because the morphology of hybrids varies and their frequency in wild populations as of now can only be approximated, the genetic purity of deer from Alberta should not be assumed. Even minute levels of introgression in individuals believed to be pure could compromise the efficacy of species delimitation. As a precaution, we minimized the risk of ancestral hybridization by importing samples from regions of allopatry, where hybridization is less common or non-existent. MD are native to western North America; populations are rarely found further east than western Minnesota or Iowa (Patterson *et al.* 2003). For this reason, we used WT samples collected in Ontario (n=10). WT distribution spans continental North America but excludes most of Utah and Nevada (Hewitt 2011). MD samples used for validation were collected from Utah (n=2), Nevada (n=5), and Montana (n=3). Note that Montana is a known region of sympatry and thus should be considered as quantitative support of the other two MD sampling locations but would otherwise not be used independently. As well, samples of *a priori* hybrid heredity (n=10) were obtained from Alberta Environment and Parks. Pedigree records were available for these individuals, as they were intentionally bred as hybrids as part of a long-term study in the mid-1980's (Wishart *et al.* 1988). DNA was extracted using a Qiagen 96 DNeasy® Blood and Tissue kit following the manufacturer's instructions (Qiagen, Mississauga, Ontario, Canada) and genotyped on the final panel. Genotypes of these individuals are listed in supplementary table S1.

SNP Assay Assignment Efficacy

HybridLab software (Nielsen *et al.* 2006) uses classical genetics principles to simulate offspring genotypes from a set of observed parental genotypes. Two populations of pure and hybrid genotypes were independently simulated in *HybridLab* using the parental genotypes of 10 pure MD from Utah, Nevada, and Montana and 10 pure WT from Ontario. Simulated populations consisted of 100 individuals in each of 10 hybrid generation classes: parental WT, parental MD, F1, F2, and three backcross generations of each species. Posterior probabilities of assignment for simulated populations were computed by *NewHybrids* version 1.1 (Anderson and Thompson 2002) using 500,000 burnin iterations preceding 500,000 sweeps. Genotype frequency

classes were set in *NewHybrids* using Jeffrey's-like priors (table 1). Repetitions were also done using two and four backcross generations (data not shown): fourth backcrosses could not be distinguished from third backcrosses or pure-species and, upon retraction of the third backcross as an assignment option, the post-secondary backcross individuals were erroneously assigned to the second backcross generation. For these reasons, the third backcross generation was deemed as the most advanced introgression level detectable. *NewHybrids* assignments were tested for efficacy using the R package *hybriddetective* (Wringe *et al.* 2017). *hybriddetective* is designed to analyze hybrid assignments as deep as one backcross generation, therefore, some code was amended to account for two additional levels of introgression (available upon request). Assignment probabilities for simulated populations, as calculated by *NewHybrids,* were evaluated by *hybriddetective* using three metrics: accuracy, the proportion of assignments to a particular category that were correct ([correct assignments] / [total assignments of a particular category]); efficiency, the proportion of individuals in a particular category that were assigned correctly ([correct assignments] / [total individuals of a particular category]); and power, the product of accuracy and efficiency (Vähä and Primmer 2006).

Results

MD Genome Assembly, Variant Identification and Annotation

A total of 378,654,417 paired-end reads were obtained in fastq format files. Following quality control steps, 334,409,387 paired-end reads remained that were then aligned to the WT reference genome assembly. A total of 536,357,624 reads were mapped (mapping rate of 75%) of which 81.69% were properly paired. The average genome-wide coverage and GC content of mapped reads were 25X and 40% respectively. Variant calling on mapped reads compared to the WT reference resulted in approximately 30 million SNPs and 2.9 million INDELs. After filtering at SNP mapping quality Q20, the variants ranged from 30.1M at read depth 10 to 32.9M at read depth 2. The genome-wide ratio of transitions to transversions at read depth 5 and Q20 was 2.4, which is similar to the value reported in human studies (DePristo *et al.* 2011; Jun *et al.* 2015).

Species-discriminating SNP Assay

Based on aligning the selected loci and their 100 bp flanking sequences to the genome assembly of the Red Deer (CerEla1.0, GenBank: GCA_002197005.1), which is finished at chromosomal level, we determined that the 40 loci were distributed across 24 of the 35 Red deer chromosomes (Fig.1), indicating good genome-wide distribution. SNP genotypes of 30 individuals, including pure MD, pure WT, and *a priori* hybrids, were successfully called at all 40 loci. Because all loci were highly discriminating, pure-breeding, F1, and F2 individuals were assigned with predictable
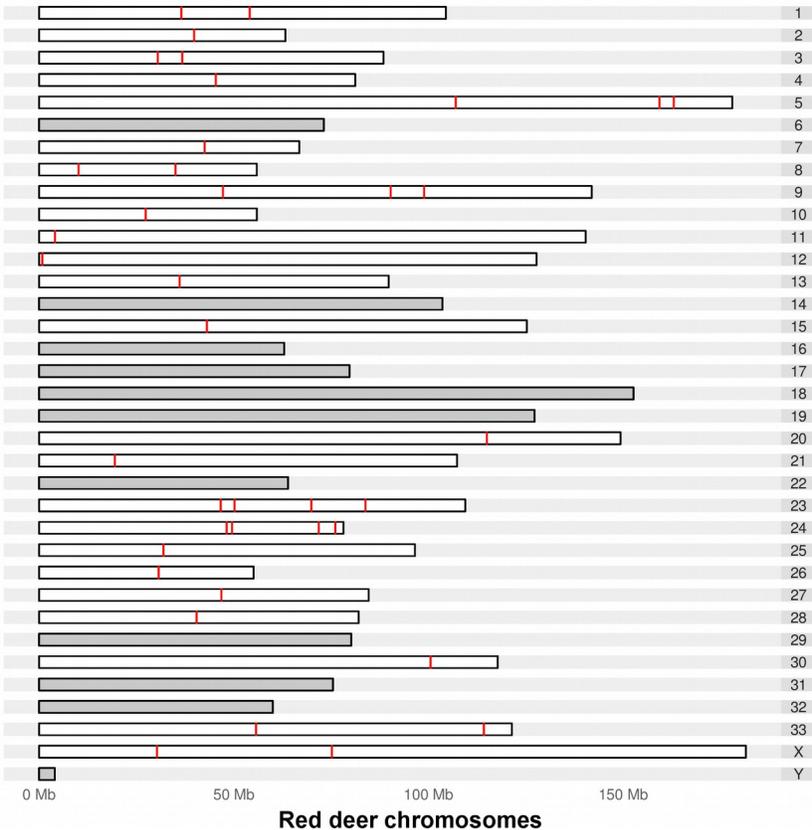
Figure 1. MD/WT species discriminating SNP loci mapped to Red Deer genome assembly. Red bars indicate the positions on the Red deer chromosomes that the 40 loci SNP assaymapped to. The mapping positions were determined by alignments of the SNP loci along with the 100 bp flanking sequences onto the Red deer genome assembly (CerEla1.0) using BLAT. The loci covered 24 of the 35 chromosomes, with some chromosomes harboring multiple loci (chromosomes not covered are shaded grey). One of the loci mapped to an unplaced contig (not depicted here).

advanced-generation WT backcross assignments were accurate and efficient at a rate of ~85%, while the same MD assignments achieved rates of ~75%. Accuracy increased with the probability threshold, but efficiency and sample size decreased (Fig.2). Both of these trends agree with expected outcomes. The critical probability threshold will be subjective based on the user's research question; by increasing the threshold, some samples assigned with a lower probability will be excluded but those that remain are more likely to be assigned correctly. For all samples to be assigned to a hybrid category, a probability threshold of 0.5 should be used. The power of assignment for a specific hybrid category slowly declined as the probability threshold was increased from 0.5 to 0.8, then decreased more drastically. Alternatively, the power of assignment for hybrids in general (ie. all hybrid categories combined) remained relatively stable over a wider range of probability thresholds: 0.80 at a critical probability threshold of 0.5 and 0.69 at a threshold of 0.9.

confidence: all reached 100% efficiency and greater than 93% accuracy at a critical posterior probability threshold of 0.50. Likewise, first-generation backcrosses were detected at ~95% efficiency and accuracy at the same threshold. Second- and third-generation WT backcrosses were assigned with slightly more proficiency than their MD counterparts;

Discussion

The development of the MD genome assembly helped us to identify fixed differences between WT and MD, and to develop a 40-loci SNP assay capable of reliably classifying pure WT, pure MD, and their interspecific hybrids into introgression categories. This diagnostic panel overcomes several limitations encountered by previous methods used in deer hybridization analyses, including: lack of assignment power due to low marker abundance, ambiguity of distinguishing morphological traits, inconsistent reproducibility, and inability to resolve advanced-generation backcrosses. Along with a solution to these deficiencies, the whole genome sequence, to which these and other polymorphisms can be mapped, improves accessibility to regions of the MD genome previously withheld from analyses by divergence from model research species. Together, these resources bear valuable conservation potential for researchers, hunters, farmers, and environmentalists.

Hybrid populations simulated in duplicate using genotypes of pure-species from regions of allopatry demonstrated that the SNP assay diagnostic strength is reliable up to and including the third backcross generation. Simulations were composed of 10 hybrid classes (n=100 for each class), including both pure species, F1 and F2 hybrids, and 3 backcross generations of each species (Table 1). We chose not to include simulations for individuals with pedigrees containing multiple hybrids (eg. backcross X backcross or backcross X F1/F2 scenarios) because the differences in individual genotype frequencies are very low between respective classes and any attempt to distinguish those classes with medium-throughput genotype data would be impractical. Furthermore, Wishart *et al.* (1988) found male F1 hybrids to be almost exclusively sterile, significantly decreasing the likelihood of encountering, in a wild population, an F2 or other product of hybrid X hybrid pairings. This apparent adherence to Haldane's rule (Haldane 1922), which predicts infertility in the heterogametic sex of interspecific hybrids, is just one opportunity for future research in a complex hybrid system that now has the molecular resources for further investigation. Another question of interest is the directionality of hybridization. In the same study, more viable offspring were produced from crosses under controlled laboratory conditions between MD males and WT females than in the reciprocal cross (Wishart *et al.* 1988); however, studies of hybrids in wild populations have both supported this finding (Carr *et al.* 1986) and refuted it (Carr and Hughes 1993; Cathey *et al.* 1998; Stelfox and Adamczewski 1993). Behavioral traits such as boldness and promiscuity likely play major roles in this dynamic (Lingle *et al.* 2007a; Lingle, *et al.* 2007b) but genetic incompatibilities may also contribute, as indicated by increased spontaneous abortions of fetuses generated by MD female X WT male matings (Wishart *et al.* 1988). This assay will be instrumental in validating observations such as these with empirical data from the field by alleviating reliance on pedigree records. By confirming or disproving the persistence of these trends in wild populations, we can explore the structural dynamics of the hybrid zone with more resolution than has been available in the past. Furthermore,
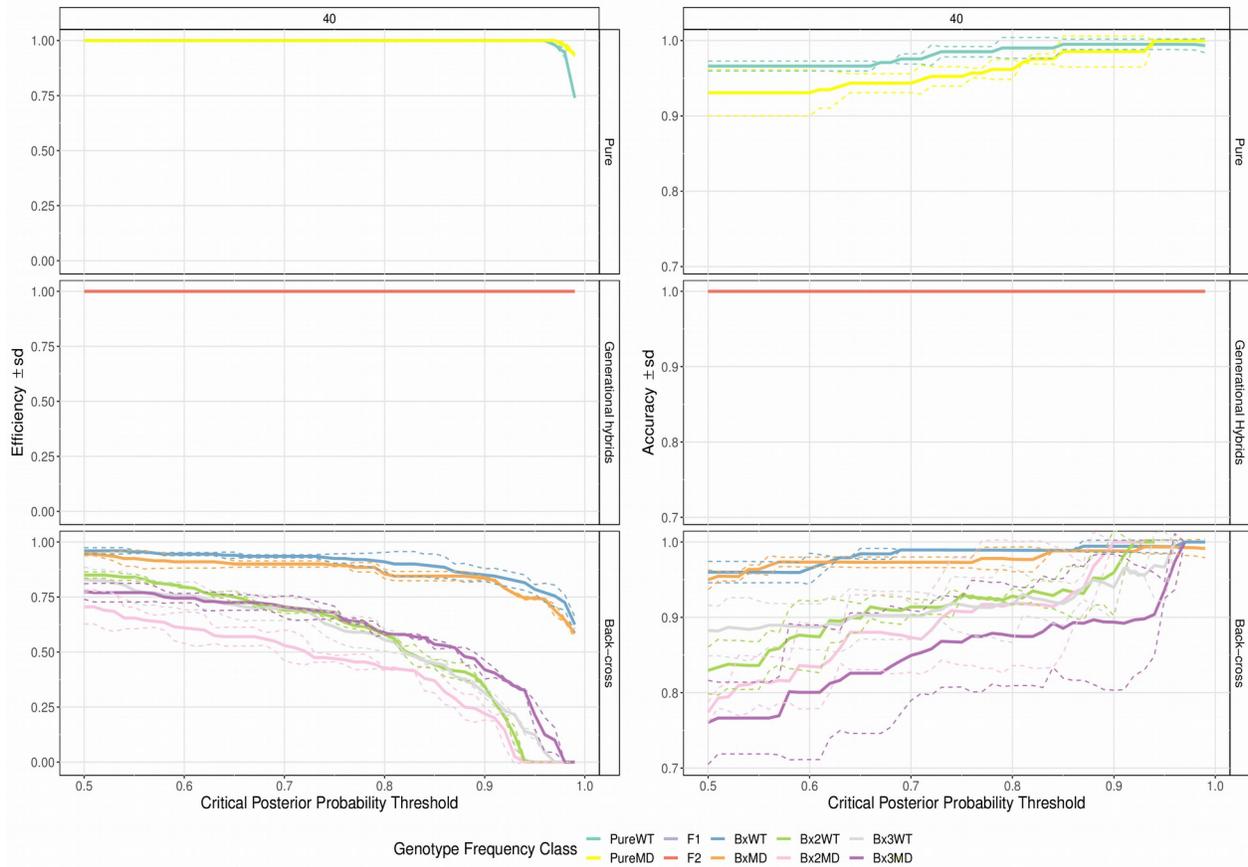
Figure 2. Hybrid assignment efficacy as determined by the hybridPowerComp function in the R package hybriddetective. Facets labelled as Pure indicate parental WT and MD, Generational Hybrids refers to F1 and F2 individuals (lines overlaid), and Back-cross includes first, second, and third generation backcrosses of both species. Accuracy was calculated as [correct assignments] / [total assignments of a particular category]. Efficiency was calculated as [correct assignments] / [total individuals of a particular category].

the focal hybrid system is not confined to western Canada; studies have reported MD x WT hybridization in West Texas and Montana (Carr *et al.* 1986; Cronin 1991). Whether these zones are consistent or clinal in hybridization frequency is yet to be investigated. Rusek *et al.* (2015) noted that, although hybrid complexes are not uncommon in nature, their evolutionary inception and subsequent persistence remains relatively ambiguous and that

studies targeted towards fine-scale genetic analyses of such systems are largely limited by lack of resources capable of identifying and quantifying introgression.

Previous sets of molecular markers and a few diagnostic morphological traits were successful at identifying F1 hybrids with some degree of accuracy but none had the diagnostic power to differentiate subsequent backcross generations; many backcross hybrids were mistaken either as

F1's or as pure-species. Without the ability to interrogate more advanced generations, the underlying factors balancing the co-existence of hybrids with parental species has remained inaccessible to researchers and conservationists (Hornbeck and Mahoney 2000). Such factors are both ecological and genetic in nature and typically include reproductive barrier strength, vigour and fertility of hybrids, pathogen pressure, and selection by predation (Daum *et al.* 2012; Duenez-Guzman *et al.* 2009; Hall *et al.* 2006; Wolf *et al.* 2001). The first step to elucidating the specific conditions of these factors as they pertain to the focal system is to delve deeper into the structure and distribution of hybridized populations via a large-scale study. With access to advanced-generation backcrosses, we now have the detection power necessary to capture a robust snapshot of the hybrid landscape. Population data will be imperative to resolve broad questions about the hybrid zone in which the *Odocoileus* system resides. For example, which model does it most resemble? Endogenous selection against hybrids (heterozygote breakdown) is indicative of a tension zone, whereas exogenous selective pressure potentially driven by niche divergence in the parental species is more consistent with the geographical selection-gradient model (Johnson *et al.* 2016). A third possibility would be that hybrids fill some ecological niche and are being selected for. Until a population-scale study is undertaken, we can only theorize. The results of that study will not only advance the collective body of knowledge of hybrid zones, but more generally, that of cervid evolution and ecology in western North America at a time when they are faced with the looming threat of chronic wasting disease. CWD is a fatal neurodegenerative disorder that has proven highly transmissible in wild populations (Williams *et al.* 2002). At the time of this writing, CWD is endemic in two Canadian provinces, Alberta and Saskatchewan, and at least 23 states in the continental United States, including some near the Canadian border such as Minnesota, Wisconsin, Michigan, Pennsylvania, New York, and both Dakotas (Haley and Hoover 2015; Richards 2018; see also https://www.cdc.gov/prions/cwd/occurrenc e.html). Containing the spread of CWD continues to be a high priority for governmental bodies, researchers, and conservationists because of the cultural, economic, and ecological importance of cervids in Canada and the United States.

The MD whole genome sequence presented here may also serve to progress the study of CWD. Genetic polymorphisms of the *Prnp* gene, which encodes the protein that is eventually converted into the infectious agent, have been associated with varied disease incubation periods (Wilson *et al.* 2009). Our assembly provides an ideal reference to which these and other variable sites can be mapped, while also serving as a localizing template for structures of interest, such as primer design. The utility of this functional, all-purpose *Odocoileus* reference lies in alleviating the reliance on model-species genomes that may be less recently diverged. In doing so, ease of access to novel regions of the genome, including those not conserved in Bovidae and undocumented subtleties such as promoters, splice sites, and introns, will be vastly improved. Facilitating access to

these untapped domains will effectively foster the development of more specific molecular advances applicable to a range of research questions throughout the *Odocoileus* genera.

Data accessibility

See Supplementary Table 1 for species-discriminating loci names and genotypes of individuals used for validation. All SNP loci are available from Brauning et al. (2015). This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession RCHL00000000. The version described in this paper is version RCHL01000000.

Acknowledgments

Supplemental material

TableS1. Deposited in Figshare under the title "Supplemental material for Russell *et al.*, 2018" at https://figshare.com/s/450e040d89fd19f1659d

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology, 26*(2), 229-246. doi: 10.1111/j.1420-9101.2012.02599.x

Anderson, E. C., and Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics, 160*(3), 1217. http://www.genetics.org/content/160/3/1217.abstract

Aulchenko, Y. S., Ripke, S., Isaacs, A., and van Duijn, C. M. (2007). GenABEL: An R library for genome-wide association analysis. *Bioinformatics, 23*(10), 1294-1296. http://dx.doi.org/10.1093/bioinformatics/btm108

Bishop, M., Kappes, S., Keele, J., Stone, R., Sunden, S., Hawkins, G., … Beattie, C. W. (1994). A genetic-linkage map for cattle. *Genetics, 136*(2), 619-639. PubMed ID: 7908653

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi: 10.1093/bioinformatics/btu170

Bradley, R. D., Bryant, F. C., Bradley, L. C., Haynie, M. L., Baker, R. J., and Jones, C. A. (2003). Implications of hybridization between white-tailed deer and mule deer. *The Southwestern Naturalist, 48*(4), 654-660. doi: 10.1894/0038-4909(2003)048<0654:IOHBWD>2.0.CO;2

Brauning, R., Fisher, P., Mcculloch, A., J Smithies, ,Russell, Ward, J., Bixley, M., … McEwan, J. (2015). Utilization of high throughput genome sequencing technology for large scale single nucleotide polymorphism discovery in red deer and canadian elk. *bioRxiv 027318*. https://doi.org/10.1101/027318.

Carr, S. M., Ballinger, S. W., Derr, J. N., Blankenship, L. H., and Bickham, J. W. (1986). Mitochondrial DNA analysis of hybridization between sympatric white-tailed deer and mule deer in west texas. *Proc Natl Acad Sci USA, 83*(24), 9576. doi: 10.1073/pnas.83.24.9576

Carr, S., and Hughes, G. (1993). Direction of introgressive hybridization between species of north-American deer (*Odocoileus*) as inferred from mitochondrial-cytochrome-B sequences. *Journal of Mammalogy, 74*(2), 331-342. doi: 10.2307/1382388

Cathey, J., Bickham, J., and Patton, J. (1998). Introgressive hybridization and nonconcordant evolutionary history of maternal and paternal lineages in north American deer. *Evolution, 52*(4), 1224-1229. doi: 10.2307/2411253

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly, 6*(2), 80-92. doi:10.4161/fly.19695

Cronin, M. A. (1991). Mitochondrial and nuclear genetic relationships of deer (*Odocoileus spp.*) in western north America. *Canadian Journal of Zoology, 69*(5), 1270-1279. doi: 10.1139/z91-179

Cullingham, C. I., Cooke, J. E., K., Dang, S., and Coltman, D. W. (2013). A species-diagnostic SNP panel for discriminating lodgepole pine, jack pine, and their interspecific hybrids. *Tree Genetics and Genomes, 9*(4), 1119-1127. doi: 10.1007/s11295-013-0608-x

Cullingham, C. I., Merrill, E. H., Pybus, M. J., Bollinger, T. K., Wilson, G. A., and Coltman, D. W. (2010). Broad and fine-scale genetic analysis of white-tailed deer populations: Estimating the relative risk of chronic wasting disease spread. *Evolutionary Applications, 4*(1), 116-131. doi: 10.1111/j.1752-4571.2010.00142.x

Cullingham, C. I., Nakada, S. M., Merrill, E. H., Bollinger, T. K., Pybus, M. J., and Coltman, D. W. (2011). Multiscale population genetic analysis of mule deer (*Odocoileus hemionus hemionus*) in western Canada sheds new light on the spread of chronic wasting disease. *Canadian Journal of Zoology, 89*(2), 134-147. doi: 10.1139/Z10-104

Daum, J. M., Davis, L. R., Bigler, L., and Woodhams, D. C. (2012). Hybrid advantage in skin peptide immune defenses of water frogs (*Pelophylax esculentus*) at risk from emerging pathogens. *Infection Genetics and Evolution, 12*(8), 1854-1864. doi: 10.1016/j.meegid.2012.07.024

DePristo, M. A., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., Hartl, C., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics, 43*(5), 491-498. doi:10.1038/ng.806

DeWoody, J., Honeycutt, R., and Skow, L. (1995). Microsatellite markers in white-tailed deer. *Journal of Heredity, 86*(4), 317-319. doi: 10.1093/oxfordjournals.jhered.a111593

Dowling, T., and Secor, C. (1997). The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics, 28*, 593-619. doi: 10.1146/annurev.ecolsys.28.1.593

Duenez-Guzman, E. A., Mavarez, J., Vose, M. D., and Gavrilets, S. (2009). Case studies and mathematical models of ecological speciation. 4. Hybrid speciation in butterflies in a jungle. *Evolution, 63*(10), 2611-2626. doi: 10.1111/j.1558-5646.2009.00756.x

Edmands, S., Northrup, S. L., and Hwang, A. S. (2009). Maladapted gene complexes within populations of the intertidal copepod *Tigriopus californicus*? *Evolution, 63*(8), 2184-2192. doi: 10.1111/j.1558-5646.2009.00689.x

Federal, Provincial, and Territorial Governments of Canada. 2014. *2012 Canadian Nature Survey: Awareness, participation, and expenditures in nature-based recreation, conservation, and subsistence activities.* Ottawa, ON: Canadian Councils of Resource Ministers.

Fernandez, M. E., Goszczynski, D. E., Liron, J. P., Villegas-Castagnasso, E. E., Carino, M. H., Ripoli, M. V., … Giovambattista, G. (2013). Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred angus herd. *Genetics and Molecular Biology, 36*(2), 185-U194. doi: 10.1590/S1415-47572013000200008

Gourbière, S., and Mallet, J. (2010). Are species real? The shape of the species boundary with exponential failure, reinforcement, and the missing snowball. *Evolution, 64*(1), 1-24. doi: 10.1111/j.1558-5646.2009.00844.x

Habib, T. J., Merrill, E. H., Pybus, M. J., and Coltman, D. W. (2011). Modelling landscape effects on density-contact rate relationships of deer in eastern alberta: Implications for chronic wasting disease. *Ecological Modelling, 222*(15), 2722-2732. doi: 10.1016/j.ecolmodel.2011.05.007

Haldane, J. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics, 12*(2), 101-109. doi: 10.1007/BF02983075

Haley, N. J., and Hoover, E. A. (2015). Chronic wasting disease of cervids: Current knowledge and future perspectives. *Annual Review of Animal Biosciences, 3*(3), 305-325. doi: 10.1146/annurev-animal-022114-111001

Hall, R. J., Hastings, A., and Ayres, D. R. (2006). Explaining the explosion: Modelling hybrid invasions. *Proceedings of the Royal Society B-Biological Sciences, 273*(1592), 1385-1389. doi: 10.1098/rspb.2006.3473

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., van Vuuren, B. J., Matthee, C., Ruiz-Garcia, M., … Couloux, A. (2012). Pattern and timing of diversification of *Cetartiodactyla (Mammalia, Laurasiatheria*), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus Biologies, 335*(1), 32-50. doi: 10.1016/j.crvi.2011.11.002

Haynes, G. D., and Latch, E. K. (2012). Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus spp.*) using the BovineSNP50 BeadChip. *Plos One, 7(*5) doi: 10.1371/journal.pone.0036536

Hewitt, D. (Ed.). (2011). Biology and management of white-tailed deer (1st ed.). Boca Raton, FL: CRC Press.

Hewitt, G. M. (1988). Hybrid zones- Natural laboratories for evolutionary studies. *Trends in Ecology and Evolution, 3*(7), 158-167. doi: https://doi.org/10.1016/0169-5347(88)90033-X

Hornbeck, G. E., and Mahoney, J. M. (2000). Introgressive hybridization of mule deer and white-tailed deer in southwestern alberta. *Wildlife Society Bulletin, 28*(4), 1012-1015. doi: 10.2307/3783860

Hu, X. (2005). Tension versus ecological zones in a two-locus system. *Theoretical Population Biology, 68*(2), 119-131. doi: https://doi.org/10.1016/j.tpb.2005.02.003

Jeffery, N. W., DiBacco, C., Wringe, B. F., Stanley, R. R. E., Hamilton, L. C., Ravindran, P. N., Bradbury, I. R. (2017). Genomic evidence of hybridization between two independent invasions of european green crab (*Carcinus maenas*) in the northwest atlantic. *Heredity, 119*(3), 154-165. doi: 10.1038/hdy.2017.22

Johnson, S. E., Delmore, K. E., Brown, K. A., Wyman, T. M., and Louis, E. E., Jr. (2016). Niche divergence in a brown lemur (*Eulemur spp.*) hybrid zone: Using ecological niche models to test models of stability. *International Journal of Primatology, 37*(1), 69-88. doi: 10.1007/s10764-015-9872-y

Jones, K., Levine, K., and Banks, J. (2000). DNA-based genetic markers in black-tailed and mule deer for forensic applications. *California Fish and Game, 86*(2), 115-126. https://www.cabdirect.org/cabdirect/abstract/20013155304

Jun, G., Wing, M. K., Abecasis, G. R., and Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Research, 25*(6), 918-925. doi:10.1101/gr.176552.114

Lamer, J. T., Ruebush, B. C., Arbieva, Z. H., McClelland, M. A., Epifanio, J. M., and Sass, G. G. (2015). Diagnostic SNPs reveal widespread introgressive hybridization between introduced bighead and silver carp in the mississippi river basin. *Molecular Ecology, 24*(15), 3931-3943. doi: 10.1111/mec.13285

Lamer, J. T., Sass, G. G., Boone, J. Q., Arbieva, Z. H., Green, S. J., and Epifanio, J. M. (2014). Restriction site-associated DNA sequencing generates high-quality single nucleotide polymorphisms for assessing hybridization between bighead and silver carp in the united states and china. *Molecular Ecology Resources, 14*(1), 79-86. doi: 10.1111/1755-0998.12152

Latch, E. K., Heffelfinger, J. R., Fike, J. A., and Olin, R. J. (2009). Species-wide phylogeography of north American mule deer (*Odocoileus hemionus*): Cryptic glacial refugia and postglacial recolonization. *Molecular Ecology, 18*(8), 1730-1745. doi: 10.1111/j.1365-294X.2009.04153.x

Leary, R. E., Gould, W. R., and Sage, G. K. (2011). Success of basibranchial teeth in indicating pure populations of rainbow trout and failure to indicate pure populations of westslope cutthroat trout. *North American Journal of Fisheries Management, 16*(1), 210-213. doi: 10.1577/1548-8675(1996)016<0210:SOBTII>2.3.CO;2

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Available from: http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Martin, G. … Durbin, R. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi: 10.1093/bioinformatics/btp352

Lingle, S. (1993). Escape gaits of white-tailed deer, mule deer, and their hybrids: Body configuration, biomechanics, and function. *Canadian Journal of Zoology, 71*(4), 708-724. doi: 10.1139/z93-095

Lingle, S., Rendall, D., and Pellis, S. M. (2007). Altruism and recognition in the antipredator defence of deer: 1. Species and individual variation in fawn distress calls. *Animal Behavior. 73*(5), 897-905. doi: 10.1016/j.anbehav.2006.11.003

Lingle, S., Rendall, D., Wilson, W. F., DeYoung, R. W., and Pellis, S. M. (2007). Altruism and recognition in the antipredator defence of deer: 2. Why mule deer help nonoffspring fawns. *Animal Behavior. 73*(5), 907-916. doi: 10.1016/j.anbehav.2006.11.004

Lohr, J. N., and Haag, C. R. (2015). Genetic load, inbreeding depression, and hybrid vigor covary with population size: An empirical evaluation of theoretical predictions. *Evolution, 69*(12), 3109-3122. doi: 10.1111/evo.12802

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution, 20*(5), 229-237. doi: 10.1016/j.tree.2005.02.010

Martin, M. (2011). CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal, 17*(10-2) doi: 10.14806/ej.17.1.200

Martinsen, G., Whitham, T., Turek, R., and Keim, P. (2001). Hybrid populations selectively filter gene introgression between species. *Evolution, 55*(7), 1325-1335. doi: 10.1111/j.0014-3820.2001.tb00655.x

Miller, M. W., Wolfe, L. L., Sirochman, T. M., Sirochman, M. A., Jewell, J. E., and Williams, E. S. (2012). Survival patterns in white-tailed and mule deer after oral inoculation with a standardized, conspecific prion dose. *Journal of Wildlife Diseases, 48*(2), 526-529. doi: 10.7589/0090-3558-48.2.526

Miller, M., Williams, E., McCarty, C., Spraker, T., Kreeger, T., Larsen, C., Thorne, E. (2000). Epizootiology of chronic wasting disease in free-ranging cervids in

Colorado and Wyoming. *Journal of Wildlife Diseases, 36*(4), 676-690. doi: 10.7589/0090-3558-36.4.676

Ng, P. C., and Kirkness, E. F. (2010). Whole genome sequencing. *Genetic variation: Methods and protocols, 628*, 215-226. doi: 10.1007/978-1-60327-367-1_12

Nielsen, E. E., Bach, L. A., and Kotlicki, P. (2006). HYBRIDLAB (version 1.0): A program for generating simulated hybrids from population samples. *Molecular Ecology Notes, 6*(4), 971-973. doi: 10.1111/j.1471-8286.2006.01433.x

Patterson, B. D., Ceballos, G., Sechrest, W., Tognelli, M. F., Brooks, T., Luna, L., Ortega, P., Salazar, I., Young, B. E. (2003) Digital distribution maps of the mammals of the western hemisphere, version 1.0. NatureServe, Arlington, Virginia, USA.

Petigara, M., Dridi, C., and Unterschultz, J. (2011). The economic impacts of chronic wasting disease and bovine spongiform encephalopathy in Alberta and the rest of Canada. *Journal of Toxicology and Environmental Health-Part A-Current Issues, 74*(22-24), 1609-1620. doi: 10.1080/15287394.2011.618987

Powell, J. H., Amish, S. J., Haynes, G. D., Luikart, G., and Latch, E. K. (2016). Candidate adaptive genes associated with lineage divergence: Identifying SNPs via next-generation targeted resequencing in mule deer (*Odocoileus hemionus*). *Molecular Ecology Resources, 16*(5), 1165-1172. doi: 10.1111/1755-0998.12572

Price, T., and Bouvier, M. (2002). The evolution of F-1 postzygotic incompatibilities in birds. *Evolution, 56*(10), 2083-2089.

Rhymer, J. M., and Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics, 27*(1), 83-109. doi: 10.1146/annurev.ecolsys.27.1.83

Richards, B. (2018). Distribution of Chronic Wasting Disease in North America. U.S. Geological Survey. Accessed from *https://www.usgs.gov/media/images/distribution-chronic-wasting-disease-north-america-0*

Rowe, S., Clarke, S., Van Stijn, T., Hyndman, D., Ward, J., McEwan, K., Dodds, k., … Asher, G. (2015). Developing genomic tools in the new zealand deer industry. *New Zealand Society of Animal Production, 75*, 91-93.

Rusek, J., Ayan, G. B., Turko, P., Tellenbach, C., Giessler, S., Spaak, P., and Wolinska, J. (2015). New possibilities arise for studies of hybridization: SNP-based markers for the multi-species *Daphnia longispina* complex derived from transcriptome data. *Journal of Plankton Research, 37*(3), 626-635. doi: 10.1093/plankt/fbv028

Schwenk, K., Brede, N., and Streit, B. (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical*

Transactions of the Royal Society B-Biological Sciences, 363(1505), 2805-2811. doi: 10.1098/rstb.2008.0055

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution, 19*(4), 198-207. doi: 10.1016/j.tree.2004.01.003

Stelfox, B. J., and Adamczewski, J. (1993). *Hoofed mammals of Alberta* (2nd ed.) Edmonton, Alberta, Canada: Lone Pine Publishing.

Stephens, M., Clipperton, N., and Bernie, M. (2009). Subspecies-informative SNP assays for evaluating introgression between native golden trout and introduced rainbow trout. *Molecular Ecology Resources, 9*(1), 339-343. doi: 10.1111/j.1755-0998.2008.02407.x

Tokarska, M., Marshall, T., Kowalczyk, R., Wójcik, J. M., Pertoldi, C., Kristensen, T. N., Loeschcke, V., … Bendizen, C. (2009). Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: The case of European bison. *Heredity, 103*, 326. doi: 10.1038/hdy.2009.73

Twyford, A. D., and Ennos, R. A. (2011). Next-generation hybridization and introgression. *Heredity, 108*, 179. doi: 10.1038/hdy.2011.68

Vähä, J., and Primmer, C. (2006). Efficiency of model-based bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology, 15*(1), 63-72. doi: 10.1111/j.1365-294X.2005.02773.x

Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution : GSE, 34*(3), 275-305. doi: 10.1186/1297-9686-34-3-275

Wiley, C., Qvarnström, A., Andersson, G., Borge, T., and Saetre, G. (2009). Postzygotic isolation over multiple generations of hybrid descendants in a natural hybrid zone: How well do single-generation estimates reflect reproductive isolation? *Evolution, 63*(7), 1731-1739. doi: 10.1111/j.1558-5646.2009.00674.x

Williams, E. S., Miller, M. W., Kreeger, T. J., Kahn, R. H., and Thorne, E. T. (2002). Chronic wasting disease of deer and elk: A review with recommendations for management. *The Journal of Wildlife Management, 66*(3), 551-563. doi: 10.2307/3803123

Williams, E. (2005). Chronic wasting disease. *Veterinary Pathology, 42*(5), 530-549. doi: 10.1354/vp.42-5-530

Wilson, G., Strobeck, C., Wu, L., and Coffin, J. (1997). Characterization of microsatellite loci in caribou *Rangifer tarandus*, and their use in other artiodactyls. *Molecular Ecology, 6*(7), 697-699. doi: 10.1046/j.1365-294X.1997.00237.x

Wilson, G. A., Nakada, S. M., Bollinger, T. K., Pybus, M. J., Merrill, E. H., and Coltman, D. W. (2009). Polymorphisms at the *PRNP* gene influence susceptibility to chronic wasting disease in two species of deer (*Odocoileus spp.*) in Western Canada. *Journal of Toxicology and Environmental Health-Part A-Current Issues, 72*(17-18), 1025-1029. doi: 10.1080/15287390903084264

Wishart, W. D. (1980). Hybrids of white-tailed and mule deer in Alberta. *Journal of Mammalogy, 61*(4), 716-719. doi: 10.2307/1380318

Wishart, W. D., Hrudka, F., Schmutz, S. M., and Flood, P. F. (1988). Observations on spermatogenesis, sperm phenotype, and fertility in white-tailed X mule deer hybrids and a yak X cow hybrid. *Canadian Journal of Zoology, 66*(7), 1664-1671. doi: 10.1139/z88-240

Wolf, D., Takebayashi, N., and Rieseberg, L. (2001). Predicting the risk of extinction through hybridization. *Conservation Biology, 15*(4), 1039-1053. doi: 10.1046/j.1523-1739.2001.0150041039.x

Wringe, B. F., Stanley, R. R. E., Jeffery, N. W., Anderson, E. C., and Bradbury, I. R. (2017). hybriddetective: A workflow and package to facilitate the detection of hybridization using genomic data in r. *Molecular Ecology Resources, 17*(6), 275-284. doi: 10.1111/1755-0998.12704