MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) **16**, 1165–1172 doi: 10.1111/1755-0998.12572

SPECIAL ISSUE: SEQUENCE CAPTURE

Candidate adaptive genes associated with lineage divergence: identifying SNPs via next-generation targeted resequencing in mule deer (*Odocoileus hemionus*)

JOHN H. POWELL,*,† STEPHEN J. AMISH,‡ GWILYM D. HAYNES,§ GORDON LUIKART¶ and EMILY K. LATCH*

Behavioral and Molecular Ecology Research Group, Department of Biological Sciences, University of Wisconsin-Milwaukee, 3209 N Maryland Ave., Milwaukee, WI 53211, USA, †Inland Fish and Wildlife Department, Sault Ste. Marie Tribe of Chippewa Indians, 523 Ashmun Street, Sault Ste. Marie, MI 49783, USA, ‡Fish and Wildlife Genomics Group, University of Montana, 32 Campus Drive, Missoula, MT 59812, USA, §Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, USA, ¶Flathead Lake Biological Station, Fish and Wildlife Genomics Group, Division of Biological Sciences, University of Montana, Polson, MT 59860, USA

Abstract

Mule deer (*Odocoileus hemionus*) are an excellent nonmodel species for empirically testing hypotheses in landscape and population genomics due to their large population sizes (low genetic drift), relatively continuous distribution, diversity of occupied habitats and phenotypic variation. Because few genomic resources are currently available for this species, we used exon data from a cattle (*Bos taurus*) reference genome to direct targeted resequencing of 5935 genes in mule deer. We sequenced approximately 3.75 Mbp at minimum 20X coverage in each of the seven mule deer, identifying 23 204 single nucleotide polymorphisms (SNPs) within, or adjacent to, 6886 exons in 3559 genes. We found 91 SNP loci (from 69 genes) with putatively fixed allele frequency differences between the two major lineages of mule deer (mule deer and black-tailed deer), and our estimate of mean genetic divergence (genome-wide F_{ST} = 0.123) between these lineages was consistent with previous findings using microsatellite loci. We detected an over-representation of gamete generation and amino acid transport genes among the genes with SNPs exhibiting potentially fixed allele frequency differences between lineages. This targeted resequencing approach using exon capture techniques has identified a suite of loci that can be used in future research to investigate the genomic basis of adaptation and differentiation between black-tailed deer and mule deer. This study also highlights techniques (and an exon capture array) that will facilitate population genomic research in other cervids and nonmodel organisms.

Keywords: black-tailed deer, exon capture, hybridization, nonmodel taxa, population genomics

Received 31 October 2014; revision received 15 April 2016; accepted 23 June 2016

Introduction

A good candidate species for a landscape genomic study of local adaptation is one of large (effective) population size with limited population structure at neutral markers across different environments or selection gradients (Luikart *et al.* 2003; Schwartz *et al.* 2009; Hohenlohe *et al.* 2010). Random genetic drift is weak in such species, allowing natural selection to determine allele frequencies even at loci with small selection coefficients (Hartl & Clark 2007). Selecting species with limited population structure (and large N_e) should increase the power to detect local adaptation, because strong population structure can obscure signatures of natural selection (Lowry 2010).

Odocoileus hemionus occur across a wide diversity of habitats in North America, west of the 100th meridian between 23° and 60°N (Anderson & Wallmo 1984). Populations historically, or currently, occupy all but two or three of the more than 60 identified types of natural vegetation found in this region (Wallmo 1981). Little genetic divergence has been observed among the 11 named subspecies of *O. hemionus* (Smith *et al.* 1990; Scribner *et al.* 1991; Cullingham *et al.* 2011; Powell *et al.* 2013; Latch *et al.* 2014). The low divergence is likely due to a combination of high population sizes, current estimates of which are approximately three million individuals (Carpenter 1998) and the capacity for long distance dispersal of individual deer (Anderson & Wallmo 1984).

Correspondence: Emily K. Latch, Fax: 414-229-3926; E-mail: latch@uwm.edu

© 2016 John Wiley & Sons Ltd

As a species, *O. hemionus* have been classified into two broad evolutionary lineages, mule deer and black-tailed deer (Latch *et al.* 2011) that diverged in allopatry during the last glacial maxima (~18 000 years before present, Latch *et al.* 2009). These lineages are characterized by differences in body size (Bandy *et al.* 1970; Anderson 1981) and morphology (Wallmo 1981), but a previous attempt to develop a dichotomous key for subspecies of *O. hemionus* (Cowan 1936) failed. In addition, these lineages are not reproductively isolated, as a stable hybrid swarm exists along their contact zone (Latch *et al.* 2011).

Although many features of the biology of *O. hemionus* point to its promise as a focal species for landscape genomic research, limited genomic resources are available. Therefore, the objective of this study was to use exon capture techniques to identify single nucleotide polymorphism (SNP) loci throughout the *O. hemionus* genome. We predicted that patterns of genetic divergence at SNP loci would correlate with the already identified evolutionary lineages in this species. By identifying loci across the *O. hemionus* genome, we predicted that we would also discover loci with likely fixed allele frequency differences between the lineages of deer. Finally, we hypothesized that these newly identified loci could be used to test for signatures of natural selection across groups of genes known to be involved in specific biological processes. This study highlights a valuable technique for developing genomic resources in organisms for which a reference genome is absent, and the resources developed herein will facilitate future work investigating differentiation between evolutionary lineages of *O. hemionus*.

Methods

Tissue samples were collected opportunistically at hunter check stations in Oregon and Washington during fall hunting seasons in 1999, 2000, 2001 and 2003. Field collections were preserved as described in Latch *et al.* (2011). The seven deer included in this analysis were selected based on their posterior probability of having sole ancestry in either the black-tailed deer or mule deer lineages based on a previous hybridization screen using 10 microsatellite loci (Latch *et al.* 2011) and having sufficient DNA quantities (>2 μg) for the exon capture technique. A sex bias was present in the original samples screened for hybridization (approximately 77% of all samples were male; Latch *et al.* 2011), and the subsample used in this study consisted of six males (4 mule deer and 2 black-tailed deer) and a single female black-tailed deer. All three black-tailed deer were collected in western Oregon, whereas three of the mule deer were collected in north central Washington. The remaining mule deer was harvested in east central Oregon.

Reduced-representation sequence libraries for these seven individuals were used in an analysis of the effects of evolutionary divergence on exon capture success (Cosart 2013). A full description of the exon capture techniques can be found in that paper and the references therein. In summary, exon baits were developed from *Bos taurus* genome (BTau 4.0, The Bovine Genome Sequencing Consortium 2009) using the ExonSampler software (Cosart *et al.* 2014) for 24 524 exons targeted from 5935 genes. The majority of exons sampled were evenly distributed across the cattle genome. A subset of 2542 exons was directly targeted due to their presence in 349 candidate genes that are associated with immunity, reproduction and speciation (Cosart 2013). A modified Agilent in-solution capture protocol was used to enrich the genomic DNA for template sequences orthologous to the baits, and this enriched template was sequenced using Illumina 100-base pair (bp) paired-end reads on a HiSeq sequencer. Initial trimming and filtration were carried out as described by Cosart (2013).

The trimmed and filtered sequence reads from each sample were aligned individually to the *B. taurus* genome and a reference white-tailed deer mitochondrial genome (GenBank HQ332445) with the program Geneious, version 6.1.7 created by Biomatters (available from http://www.geneious.com/). Each alignment was performed using the default settings of the map to reference tool, with medium/low sensitivity and fine-tuning set to iterate as many as five times. The default settings allowed for a maximum gap size of 15 bp with a maximum of 10% gaps per read. Word length was set to 18 bp, with an index word length of 13 bp. Words were ignored when repeated more than 12 times. A maximum of 20% mismatches were allowed per read, and the maximum ambiguity parameter was four. We used paired read distances in assembly, and if multiple locations were the best match, the read was mapped randomly to one of those locations.

After constructing the seven individual alignments, we used the following procedure for SNP discovery. We identified all positions in which an individual genotype differed from the reference cattle sequence within the coding region of targeted exons and in a 250-bp window on either side of the open-reading frame. These positions were identified using the 'Find Variations/SNPs' tool in Geneious version 6.1.7. For this analysis, we required a minimum of 20X coverage within an individual, a minor allele frequency of no less than 0.15 and a maximum variant *P*-value of 10^{-7}(Appendix S1, Supporting information). There was no penalty for observing multiple alleles at a locus in an individual. Separate annotations were generated for each allele of each putative variant nucleotide position. Because we were identifying all loci within an individual that differed from the reference

© 2016 John Wiley & Sons Ltd

cattle genome, not identifying a locus in an individual could have been caused by either (i) an individual did not have the required 20X coverage at the locus, or (ii) an individual was homozygous for the reference allele at that locus. Therefore, to separate these possibilities, we also identified all loci with greater than 20X coverage in each individual.

After identifying putatively variable nucleotide sites within each individual alignment, we used the statistical computing program R version 3.0.2 (R Core Team 2013) to further screen identified loci. First, we removed all loci that did not have the required 20X coverage in all individuals. Next, we removed high-density SNPs, defined as >5 SNPs within 10 bp, and loci located within 3 bp of an insertion or deletion, to account for false SNP discovery due to local misalignments (Cosart 2013). Finally, we removed all loci where more than two alleles were identified within an individual deer. Following removal of loci that did not meet our quality control standards, we filtered all loci in which the identified SNP was homozygous for a nonreference allele in all seven genotyped deer. Because the 'Find Variations/SNPs' tool only identified differences from the reference genome, we assumed that an individual was homozygous for the reference allele at loci that passed our quality control standards but for which no genotype was called in that individual. We used Pearson correlation coefficients to investigate patterns of capture success across the cattle genome.

To investigate differences between the mule deer and black-tailed deer lineages, we used the R package Geneland (Guillot *et al.* 2005) to calculate Weir & Cockerham's (1984) estimates of F_{ST} and F_{IS}. To reduce the observed variance of single locus estimates when examining patterns of local F_{ST} in human populations, Weir *et al.* (2005) pooled loci within 5-Mbp windows. Due to a change in the number of chromosomes between a hybrid red deer (*Cervus elaphus*) × Père David's deer (*Elaphurus davidianus*) and cattle (Slate *et al.* 2002), we are not confident that a 5-Mbp window along a cattle chromosome remains intact along a *O. hemionus* chromosome. Therefore, to avoid calculating summary statistics, such as F_{ST}, using groups of noncontiguous SNPs, we limited the pooling of sites to only occur within a single gene. Therefore, in addition to estimating F_{ST} and F_{IS} from all polymorphic loci, we also estimated these parameters for all sites pooled within a gene.

Observed patterns of genetic differentiation between mule deer and black-tailed deer were used to investigate general patterns of conservation across the captured regions. We tested for a genomewide pattern of purifying selection by comparing F_{ST} values for loci located within exons to loci in the flanking regions. We used random permutations to test the hypothesis that loci located within exons would show lower genetic divergence between mule deer and black-tailed deer than loci within the flanking introns. For this analysis, we randomly permuted the classification of a site (exon or intron) 1000 times and calculated F_{ST} in the respective regions for each permutation. The *P*-value of this test was calculated as the proportion of permutations with as small, or smaller an F_{ST} as observed in exons, and as large, or larger an F_{ST} as observed in introns.

We scanned for signatures of selection between mule deer and black-tailed deer by testing for a statistical over- or underrepresentation of biological processes using Panther version 9.0 (Thomas *et al.* 2006; Mi *et al.* 2013a,b). We used all genes with loci that had fixed allelic differences for the over- or underrepresentation tests in Panther 9.0. This analysis was run based on the logic that if outlier loci exist, those sites with fixed allele differences ($F_{ST} = 1$) must be included in the outliers. Testing genes with putatively fixed allele frequency differences between mule deer and black-tailed deer, we expected biological processes under balancing selection to be underrepresented and biological processes under positive selection to be over-represented in this analysis. Within Panther 9.0, statistical over- or underrepresentation analyses were run by first classifying the biological processes for all 3559 genes with SNP loci that passed our quality control filters. We then classified biological processes for the 69 genes with diagnostic loci. Finally, we tested for an over- or underrepresentation of biological processes relative to the total set of identified genes.

When testing the over- or underrepresentation of ontology terms, the nested nature of the classifications makes the determination of the number of simultaneous tests difficult (Thomas *et al.* 2006). Therefore, adjusting α for a family-wise error rate based on the number of compared ontology terms will be too conservative, and an alternate Bonferroni adjustment of α for biological processes was applied (Thomas *et al.* 2006). In addition, we also applied the false discovery rate correction procedure of Benjamini & Yekutieli (2001) based on the total number of analysed ontology terms, to try and avoid the conservative nature of the Bonferroni adjustment procedure (Benjamini & Hochberg 1995). This latter method may have higher power than a traditional Bonferroni adjustment, despite likely being too conservative due to an overestimation of the true number of tests.

Results

Of a total 95 337 234 sequences across the seven individuals, 86 434 625 were aligned to the cattle reference genome. The number of aligned reads per individual ranged from a maximum of 18 456 326 to a minimum of

© 2016 John Wiley & Sons Ltd

5 910 620. On average, approximately 90.7% (individual range between 89.9% and 91.0%) of all sequences were aligned to the reference genome. Each alignment used all five possible fine-tuning iterations, with the average first iteration aligning approximately 78.3% of reads (individual range between 75.1% and 80.3%). There was a total of 15.2 Mbp of sequence located in regions spanning up to 250 bp on either side of the 24 524 targeted exons (sampled from 5935 genes). Of this targeted sequence, 3 748 225 bp (or approximately 24.7%) met our requirements of at least 20X coverage in all individuals.

We identified 103 823 SNPs which contained alleles that differed from the *B. taurus* reference genome in at least one individual deer. At 80 970 of these loci (approximately 2.16% of loci meeting our filtration criteria), all seven *O. hemionus* samples amplified allele(s) that differed from the cattle reference genome. At 351 of the 80 970 loci, no deer amplified the cow reference allele, but a polymorphism (two alleles) was observed within the seven individual deer. Because these 351 polymorphic loci also differed from the ancestral cow reference genome, approximately 0.43% of all loci with fixed differences from the cattle genome are estimated to have experienced at least two mutations since diverging from a common ancestor.

We identified a total of 23 204 loci that were polymorphic (rare variant observed >1 time) among the seven deer in this study, comprising 16 445 transitions and 6995 transversions (transition:transversion ratio of approximately 2.35). These loci were located within, or adjacent to, 6886 exons (approximately 28.1% of all targeted exons) sampled from 3559 genes (approximately 60% of all targeted genes). Of the 23 204 total polymorphic loci within *O. hemionus*, 14 777 (64.7%) were located within the open-reading frame of their associated exon. Three alleles (nucleotide variants) were detected in only 121 of the 23 204 identified SNP loci. The observed proportion of SNPs amplifying three alleles among polymorphic loci (approximately 0.52%) was elevated relative to expectations based on the per cent of identified loci with fixed differences from the cattle genome (0.43%).

Exon capture success was similar throughout the cattle genome, as indicated by the association between the number of mapped SNPs and chromosome length (r^2 = 0.83) and the number of mapped SNPs and the number of targeted exons on a chromosome (r^2 = 0.84, Fig. 1). As expected, these associations were lower than the correlation between the number of SNPs identified per cattle chromosome and the number of exons on that chromosome with at least one identified SNP (r^2 = 0.95). The distribution of the number of SNP loci identified in an exon, given that it had at least one SNP, was strongly right skewed, with a mean of approximately 3.3 SNPs per exon, a median of two SNPs per exon and a mode of a single SNP per exon (Fig. 2). For the intronic SNPs, more than 80% of all identified positions were located within 75 bp of the open-reading frame (Fig. 3).

The average minor allele frequency was similar for both mule deer (0.238) and black-tailed deer (0.226). Across all 23 204 loci, we detected moderate levels of genetic divergence between mule deer and black-tailed deer (F_{ST} = 0.123, Fig. S1, Supporting information). Genetic divergence was reduced in the exonic SNPs (F_{ST} = 0.097) relative to the intronic SNPs (F_{ST} = 0.168), with none of the 1000 permutations of SNP location producing estimates of genetic divergence more extreme than the observed value.

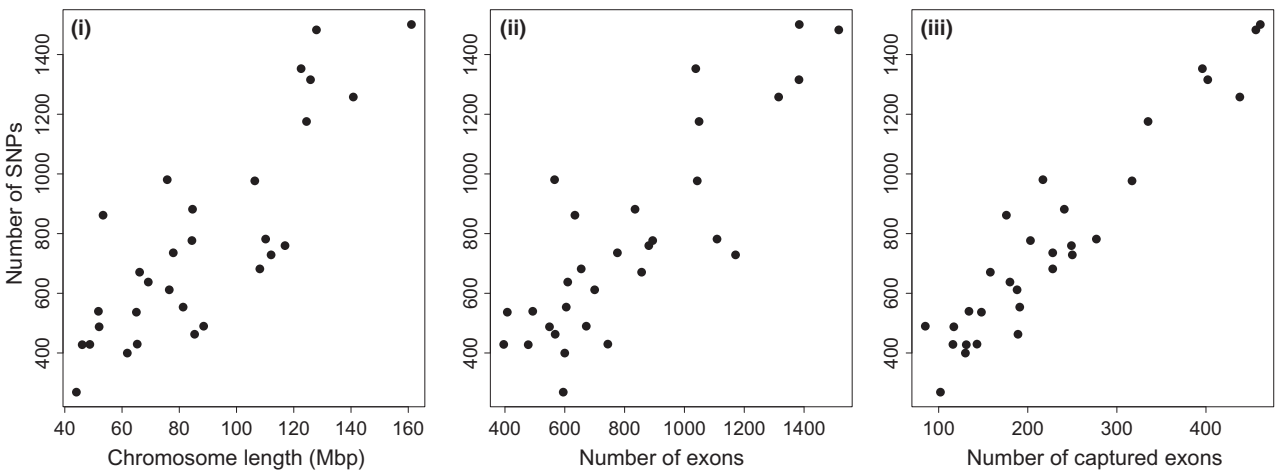

Fig. 1 The number of SNP loci identified given i) cattle chromosome length, ii) number of targeted exons per cattle chromosome and iii) number of exons with at least one SNP identified within the open-reading frame (ORF), or the 250-bp region on either end of the ORF, per cattle chromosome.

© 2016 John Wiley & Sons Ltd

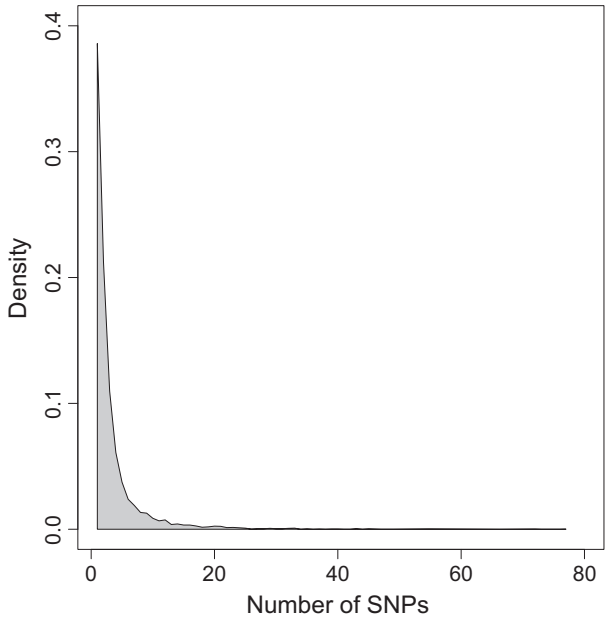

Fig. 2 Empirical probability distribution of the number of identified SNP loci per exon with at least one SNP identified within the open-reading frame (e.g. exon), or the 250-bp region on either end.

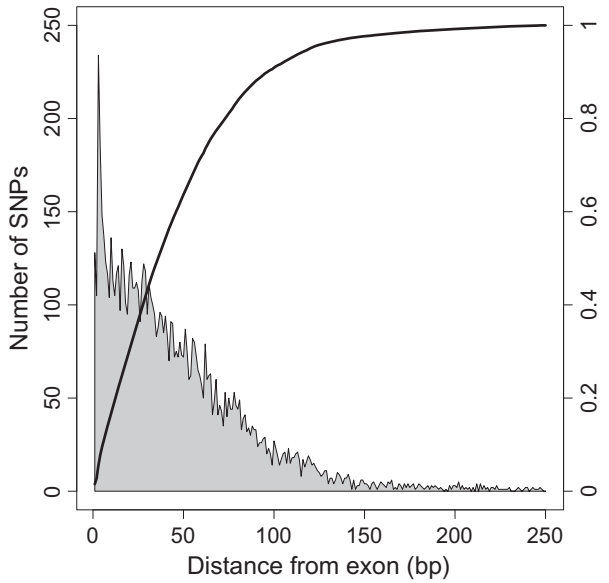

Fig. 3 The proportion of SNP loci identified at a given distance (bp) from either the 5′ or 3′ end of the exon is plotted in the shaded region. The line depicts the cumulative density function across the region.

We identified a total of 91 loci, located in 69 genes, with putatively fixed allele frequency differences between mule deer and black-tailed deer (Table S1, Supporting information). Of these diagnostic loci, 53 were identified within a coding region, and approximately 66% (35) of these loci were nonsynonymous substitutions. Thirteen total genes had more than one diagnostic SNP, with the most (8) occurring in the gene PAG19 (Table S1, Supporting information), a member of a family of duplicated genes undergoing diversification in bovids (Hughes *et al.* 2000).

Panther 9.0 was able to map the ontology of 3160 of the 3559 genes with identified SNPs in this study. We were able to map the ontology of 60 of the 69 genes with diagnostic loci. Based on a Panther 9.0 analysis with a Bonferroni adjustment of α, no biological process coded for by the genes with fixed allele frequency differences at one or more SNP loci was statistically over- or underrepresented. When a Benjamini & Yekutieli (2001) false discovery rate adjustment ($m = 209$ tests) was applied, however, both the gamete generation (P-value = 0.0061) and amino acid transport (P-value = 0.0076) biological processes were statistically over-represented in the subset of genes with fixed allele differences (Table S2, Supporting information).

Discussion

We saw consistent patterns of capture success across bovine chromosomes, and approximately one-quarter of the targeted exons had sufficient sequence depth to meet our relatively conservative (demanding) filtration criteria of at least 20X coverage in every individual. This success rate yielded 3559 genes with SNPs. The rate (~60%) was likely limited by DNA sequence changes that accrued over the approximately 30 million years since the split between *O. hemionus* and cattle (27.7 MYA Hassanin & Douzery 2003, 32.5 MYA Guha *et al.* 2007) impairing both exon capture efficiency and the accurate alignment of *O. hemionus* DNA sequences against the cattle genome.

The observed frequency of triallelic SNP loci in this study (0.52%) indicates that falsely identified alleles may remain in the data set after our quality control screens. The observed frequency of polymorphic loci amplifying three alleles was greater than what was detected in these same seven *O. hemionus* by Cosart (2013, 0.11%) and was nearly twice as high as previous studies on human populations (0.22% Hüebner *et al.* 2007, 0.20% Hodgkinson & Eyre-Walker 2010). The overall proportion of triallelic SNPs identified in this study, however, was less than what was identified using a different reduced-representation library technique (RAD sequencing) in white-tailed deer (3.07% Seabury *et al.* 2011). These indications of false-positive alleles, and the limited geographic range from which samples were collected, highlight the need to further validate loci (*e.g.* amplify loci of interest in more individuals sampled throughout the range of *O. hemionus*).

© 2016 John Wiley & Sons Ltd

Divergence patterns from cattle

In the sequences that met our filtration criteria, our estimated degree of divergence between *O. hemionus* and cattle (2.16%) is lower than what was expected given an estimated divergence time between cattle and deer of between 27.7–27.8 million years (Hassanin & Douzery 2003) and 32.5 million years (Guha *et al.* 2007). Cosart *et al.* (2011) observed 0.5% sequence divergence between *Bos taurus* and *Bison bison*, species that shared a common ancestor one to two million years ago (Hedrick 2009). Despite cattle and *O. hemionus* diverging potentially more than 30 million years prior to cattle and bison, the estimated sequence divergence is only approximately four times larger.

The low divergence between deer and cattle could be explained by our use of exon capture to create reduced-representation sequence libraries, a technique that favours the detection of conserved regions. When these seven deer were included in a larger study investigating exon capture success across evolutionary time, Cosart (2013) found that whereas a cow aligned 78% of the 3.6 Mb of targeted exonic sequence at greater than 20X coverage, coverage decreased across increasing divergence time (African buffalo (*Syncerus caffer*) 12%, bighorn sheep (*Ovis canadensis*) 10%, mule deer 20% and pigs (*Sus scrofa*) 29% fewer bp). Furthermore, previous work by Vallender (2011) and Bi *et al.* (2012) documented a negative relationship between sequence coverage and sequence divergence in exon capture studies, which means that the greater the sequence divergence between two species, the poorer the coverage.

Divergence patterns within mule deer

Across all 23 204 loci, we detected a similar level of genetic divergence between black-tailed deer and mule deer (F_{ST} = 0.123) to what was detected between these two lineages using 10 microsatellite loci (F_{ST} = 0.124, Latch *et al.* 2011), but less divergence, than was detected using 878 putatively neutral SNP loci developed in cattle (F_{ST} = 0.166; Haynes & Latch 2012). As expected, we detected reduced genetic divergence in the exonic SNPs (F_{ST} = 0.097) relative to the intronic SNPs (F_{ST} = 0.168). A preliminary analysis indicated that quality control filters removing loci in highly polymorphic regions (>5 SNP loci in ≤10 bp) primarily operated within introns. Any bias introduced by this filter would result in observed differences between intronic and exonic divergence estimates being reduced relative to the true value. The fact that we still observed higher intronic genetic divergence despite these potential biases suggests the exon capture data are of relatively high quality.

Gene function and over-representations

Gamete generation and amino acid transport biological processes were over-represented among the subset of genes with fixed allele differences between black-tailed deer and mule deer. Supporting research suggesting logical mechanisms to explain the over-representation of genes involved in gamete generation (Nielsen *et al.* 2005) and amino acid transport (Clark *et al.* 2003) is encouraging. The identification of biological processes that are frequently associated with fixed genetic variation provides a starting point for the continued investigation of genes responsible for the differentiation of mule deer and black-tailed deer using a more broadly distributed set of samples.

Conclusions

We achieved our main objective, to identify numerous putative SNP markers throughout functional regions of the *O. hemionus* genome, despite the lack of a reference genome in this species. Although only moderate genetic divergence between lineages was detected in this study, we identified many loci with putatively fixed allele frequency differences between mule deer and black-tailed deer. Because exon capture can be used on low-quality DNA, such as from museum specimens and faeces (Perry *et al.* 2010; Bi *et al.* 2013), this technique is suitable for studies (such as this one that collected wildlife from hunter check stations) that would prohibit the use of other techniques for sequencing expressed regions of the genome (such as building an expressed sequence tag (EST) library from mRNA). For nonmodel organisms with no genomic resources, exon capture and SBE (single base extension) assays designed for a closely related domestic species offer potential genotyping options. Both are likely to introduce some ascertainment bias, either by selecting for conserved regions when using exon capture, or by only selecting for loci with the identical alleles as observed in the reference species. Based on the sequence data from these seven individuals, baits can be redesigned to improve capture efficiency and to capture additional sequence around regions of interest. More work is necessary to determine whether these markers are diagnostic across the range of mule deer and black-tailed deer. Nonetheless, this technique (and the exon capture array) should be useful in many species for discovery of SNPs in genes genomewide for a wide range of applications in population genetics and landscape genomics.

Acknowledgements

This research was funded by the University of Wisconsin – Milwaukee Research Growth Initiative. Jim Heffelfinger and Don

© 2016 John Wiley & Sons Ltd

Whittaker provided assistance with sample collection, and Ted Cosart assisted with exon capture bait design and initial processing of results. We thank the anonymous reviewers for their comments that helped to improve this manuscript.

References

Anderson AE (1981) Morphological and physiological characteristics. In: *Mule and Black-Tailed Deer of North America*(ed Wallmo O. C.), pp. 27–97. University of Nebraska Press, Lincoln.

Anderson AE, Wallmo OC (1984) *Odocoileus hemionus*. *Mammalian Species*, **219**, 1–9.

Bandy PJ, Cowan IM, Wood AJ (1970) Comparative growth in four races of black-tailed deer (*Odocoileus hemionus*). Part I. Growth and body weight. *Canadian Journal of Zoology*, **48**, 1401–1410.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, **57**, 289–300.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependence. *The Annuals of Statistics*, **29**, 1165–1188.

Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*, **13**, 403.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C (2013) Unlocking the vault: next-generation museum population genomics. *Molecular Ecology*, **22**, 6018–6032.

Carpenter LH (1998) Deer in the west. In: *Proceedings of the 1997 Deer/Elk Workshop, Rio Rico, Arizona*(ed deVos J. C.), pp. 1–10. Arizona Game and Fish Department, Phoenix.

Clark AG, Glanowski S, Nielsen R *et al.* (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science*, **302**, 1960–1963.

Core Team R (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Cosart T (2013) *Evaluation of New Methods for Large-Scale and Gene-Targeted Next Generation DNA Sequencing in Nonmodel Species*. PhD Dissertation, University of Montana, Missoula, MT. 104 pp.

Cosart T, Beja-Periera A, Chen S, Ng SB, Shendure J, Luikart G (2011) Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genomics*, **12**, e347.

Cosart T, Beja-Pereira A, Luikart G (2014) ExonSampler: a computer program for genome-wide and candidate gene exon sampling for targeted next-generation sequencing. *Molecular Ecology Resources*, **14**, 1296–1301.

Cowan IM (1936) Distribution and variation in deer (Genus *Odocoileus*) of the Pacific Coast region of North America. *California Fish and Game*, **22**, 155–246.

Cullingham CI, Nakada SM, Merrill EH, Bollinger TK, Pybus MJ, Coltman DW (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**, 134–147.

Guha S, Goyal SP, Kashyap VK (2007) Molecular phylogeny of musk deer: a genomic view with mitochondrial 16S rRNA and cytochrome *b* gene. *Molecular Phylogenetics and Evolution*, **42**, 585–597.

Guillot G, Mortier F, Estoup A (2005) Geneland: a program for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Hartl DL, Clark AG (2007) *Principles of Population Genetics*, 4th edn. Sinauer Associates Inc., Sunderland.

Hassanin A, Douzery EJP (2003) Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae. *Systematic Biology*, **52**, 206–228.

Haynes GD, Latch EK (2012) Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus* spp.) using the BovineSNP50 BeadChip. *PLoS ONE*, **7**, e36536.

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

Hodgkinson A, Eyre-Walker A (2010) Human triallelic sites: evidence for a new mutational mechanism? *Genetics*, **184**, 233–241.

Hohenlohe PA, Phillips PC, Cresko WA (2010) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences*, **171**, 1059–1071.

Hüebner C, Petermann I, Browning BL, Shelling AN, Ferguson LP (2007) Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: *MDR1* (*ABCB1*) *G2677/T/A* as an example. *Cancer Epidemiology Biomarkers & Prevention*, **16**, 1185–1192.

Hughes AL, Green JA, Garbayo JM, Roberts RM (2000) Adaptive divergence within a large family of recently duplicated, placentally expressed genes. *Proceedings of the National Academy of Sciences USA*, **97**, 3319–3323.

Latch EK, Heffelfinger JR, Fike JA, Rhodes OE (2009) Species-wide phylogeography of North American mule deer (*Odocoileus hemionus*): cryptic glacial refugia and postglacial recolonization. *Molecular Ecology*, **18**, 1730–1745.

Latch EK, Kierepka EM, Heffelfinger JR, Rhodes OE (2011) Hybrid swarm between divergent lineages of mule deer (*Odocoileus hemionus*). *Molecular Ecology*, **20**, 5265–5279.

Latch EK, Reding DM, Heffelfinger JR, Alcalá-Galván CH, Rhodes OE (2014) Range-wide analysis of genetic structure in a widespread, highly mobile species (*Odocoileus hemionus*) reveals the importance of historic biogeography. *Molecular Ecology*, **23**, 3171–3190.

Lowry DB (2010) Landscape evolutionary genomics. *Biology Letters*, **6**, 502–504.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Mi H, Muruganujan A, Thomas PD (2013a) PANTHER 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, **41**, D377–D386.

Mi H, Muruganujan A, Casagrande JT, Thomas PD (2013b) Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, **8**, 1551–1566.

Nielsen R, Bustamante C, Clark AG *et al.* (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology*, **3**, e170.

Perry GH, Marioni JC, Melsted P, Gilad Y (2010) Genome-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology*, **19**, 5332–5344.

Powell JH, Kalinowski ST, Higgs MD, Ebinger MR, Vu NV, Cross PC (2013) Microsatellites indicate minimal barriers to mule deer *Odocoileus hemionus* dispersal across Montana, USA. *Wildlife Biology*, **19**, 102–110.

Schwartz MK, Luikart G, McKelvey KS, Cushman S (2009) Landscape genomics: a brief perspective. In: *Spatial Complexity, Informatics and Wildlife Conservation*(eds Cushman S. A., Huettman F.), pp. 165–174. Springer, Tokyo.

Scribner KT, Smith MH, Garrott RA, Carpenter LH (1991) Temporal, spatial, and age-specific changes in genotypic composition of mule deer. *Journal of Mammalogy*, **72**, 126–137.

Seabury CM, Bhattarai EK, Taylor JF *et al.* (2011) Genome-wide polymorphism and comparative analyses in the white-tailed deer (*Odocoileus virginianus*): a model for conservation genomics. *PLoS ONE*, **6**, e15811.

Slate J, Van Stijn TC, Anderson RM *et al.* (2002) A deer (subfamily Cervinae) genetic linkage map and the evolution of ruminant genomes. *Genetics*, **160**, 1587–1597.

Smith MH, Scribner KT, Carpenter LH, Garrott RA (1990) Genetic characteristics of Colorado mule deer (*Odocoileus hemionus*) and comparisons with other cervids. *Southwestern Naturalist*, **35**, 1–8.

The Bovine Genome Sequencing and Analysis Consortium, Elsik CG, Tellam RL, Worley KC (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, **324**, 522–528.

© 2016 John Wiley & Sons Ltd

Thomas PD, Kejariwal A, Guo N *et al*. (2006) Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Research*, **34**, W645–W650.
Vallender EJ (2011) Expanding whole exome resequencing into non-human primates. *Genome Biology*, **12**, R87.
Wallmo OC (1981) Mule and black-tailed deer distribution and habitats. In: *Mule and Black-Tailed Deer of North America*(ed Wallmo O. C.), pp. 1–25. University of Nebraska Press, Lincoln.
Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG (2005) Measures of human population structure show heterogeneity among genomic regions. *Genome Research*, **15**, 1468–1476.

E.K.L. designed the study, and S.J.A. and G.L. performed laboratory work. J.H.P. and G.D.H. analyzed and interpreted the data, and J.H.P. and E.K.L. wrote the paper with contributions from all authors.

Data accessibility

DNA sequences, assembly and SNP data: DRYAD entry doi:10.5061/dryad.pv18p

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Genes with putatively diagnsotic SNP loci.

Table S2 Results of the Panther 9.0 analysis examining over- and under-representation of biological processes in genes with diagnostic loci.

Appendix S1 Supplemental methods.

© 2016 John Wiley & Sons Ltd