bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genomic architecture of artificially and sexually-selected traits in a wild cervid

Running title: Genomic basis of sexually selected traits

S. J. Anderson[1]*, S. D. Côté [2], J. H. Richard[2], A.B.A. Shafer[1,3*]

[1] Environmental and Life Sciences Graduate Program, Trent University, K9J 7B8 Peterborough, Canada.

[2] Département de biologie, Centre d'études nordiques & *Chaire de recherche industrielle CRSNG en aménagement intégré des ressources de l'île d'Anticosti*, Université Laval, G1V 0A6, Québec, QC, Canada

[3] Forensics Department, Trent University, K9J 7B8 Peterborough, Canada.

* Corresponding authors: sanderson@trentu.ca, aaronshafer@trentu.ca

Keywords

Quantitative traits, population genetics, gene enrichment, polygenic traits

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Characterization of the complex genomic architecture underlying quantitative traits can provide valuable insights into the study, conservation, and management of natural populations. This is particularly true for fitness-related traits such as body size and male ornamentation in mammals because as indicators of quality and health, these traits are often subject to sexual and artificial selective pressures. Here we performed high-depth whole genome re-sequencing on pools of individuals representing the phenotypic extremes in our study system for antler and body size in white-tailed deer (*Odocoileus virginianus*). Samples were selected from a database containing phenotypic data for 4,466 male white-tailed deer from Anticosti Island, Quebec, with four pools representing the extreme phenotypes for antler and body size in the population. Our results revealed a largely panmictic population (F_{ST} ~ 0.01), but also detected diverged regions (F_{ST} > 0.11) between pools for both traits. These regions revealed genomic islands with signatures of positive selection and demographic expansion (negative Tajima's *D*), and contained putative genes of small-to-moderate effect. Through qPCR analysis we genotyped an intron variant on gene SRP54 that is a potential QTL for antler size, and propose two missense variants on the MYADM and SPATA31E1 genes known to influence body size in mammals and reproductive success as potential QTL. This study revealed the polygenic nature of both antler morphology and body size in white-tailed deer and identified target loci for additional analyses.

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Quantifying the genomic architecture underlying phenotypes in natural populations provides insights into the evolution of quantitative traits (Stinchcombe & Hoekstra, 2008). Some quantitative traits are correlated with metrics of fitness, and as such are particularly important because they may directly influence population viability (Kardos & Shafer, 2018). This relationship between genomic architecture and quantitative traits, however, is not easy to empirically identify (Kardos & Shafer, 2018), and often has unclear and unpredictable responses to selection (Bunger et al., 2005). Outside of a few traits in well studied systems such as horn morphology in Soay sheep (Johnston et al., 2011, 2013) and migration timing in Pacific salmon (Prince et al., 2017), there are few empirical studies that have successfully identified quantitative trait loci (QTL) in wild vertebrate populations with effects on fitness-related traits.

Genome-wide association scans (GWAS) have the potential to identify links between genomic regions and fitness relevant traits that can directly inform the management of species, and have become more prominent in non-model organisms (Ellegren 2014; Santure & Grant 2018). The variation of expressed phenotypes that are observed in wild populations is influenced by demographic, environmental and epigenetic factors (Ellegren & Sheldon 2008), which leads to difficulties when trying to calculate the direct influence of genetic polymorphisms and more generally disentangle the effects of drift from selection (Kardos & Shafer 2018). Most traits also appear to be composed of many genes of small effect, with heritability for complex traits often explained by the effects on genes outside of core pathways (Robinson et al., 2013; Santure et al., 2013; Boyle et al., 2017). Population genetic modeling allows

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

disentangling phenotypic plasticity and responses to environmental change, while expanding our understanding of the evolutionary processes in wild populations (Santure & Grant 2018). Where traits directly linked to reproductive success are also those humans select for through harvesting (Coltman et al., 2003; Kuparinen & Festa-Bianchet, 2017), there is a need to understand and predict the evolutionary outcome, which is in part dependent on the genomic architecture (Kardos & Luikart, 2019).

For traits relevant to fitness, including those under sexual selection, we might expect directional selection towards the dominant or desired phenotype. In long term studies of wild populations, however, there is often a relatively high degree of observed genetic variation underlying sexually selected traits (Johnston et al., 2013; Berenos et al., 2015; Malenfant et al., 2018), with the proposed mechanism for this genetic variation being balancing selection (Mank, 2018). For example, Johnston et al. (2013) showed that variation is maintained at a single large effect locus through a trade-off between observed higher reproductive success for large horns and increased survival for smaller horns in Soay sheep. These trade-offs, creating signatures consistent with balancing selection, are also observed in cases where social dominance structures exist as a result of male-male competition for female mate acquisition, where alternative mating strategies have been suggested as a means for younger or subordinate males to achieve greater mating success (Foley et al., 2015).

Emerging approaches to investigate the genomic architecture of complex phenotypes

The methodological approach of analyzing extreme phenotypes seeks to sample individuals representing the extreme ends of the spectrum for any observable

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

phenotype, instead of randomly sampling individuals from the entire distribution (Perez-Gracia et al., 2002; Li et al., 2011; Emond et al., 2012; Barnett et al., 2013; Gurwitz & McLeod, 2013; Kardos et al., 2016). Kardos et al (2016) utilized whole genome re-sequencing of groups representing extreme phenotypes in forehead patch size, a sexually selected trait in collared flycatchers, to study the genetic basis of this phenotypic variation. Kardos et al. (2016) highlighted the limitations that exist when seeking to reliably identify QTL in large natural populations, with a key point being large sample size and complete sampling of the genome as requisites for detecting sites under selection. However, in small populations with widespread strong linkage disequilibrium, fewer individuals are required to detect large effect loci (Kardos et al, 2016; Lowry et al., 2017).

Pooled sequencing (pool-seq) uses the resequencing of pooled DNA from individuals within a population and is a cost-effective alternative to the whole genome sequencing of every individual (Anand et al., 2016). Individual identity is lost through the pooling of DNA, but the resulting allelic frequencies are representative of the population and can be used to conduct standard population genomic analyses, including outlier detection (Schlötterer et al., 2014). Pool-seq methods have been applied with the purposes of identifying genes of moderate-to-large effect loci in a variety of taxa including colour morphs in butterflies and birds, abdominal pigmentation in *Drosophila*, and horn size of wild Rocky Mountain bighorn sheep (Kardos et al, 2015; Endler et al., 2017; Neethiraj et al., 2017).

Here, we explore the genic basis for phenotypic variation by sampling the extreme phenotypes in a non-model big game species, the white-tailed deer

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(*Odocoileus virginianus*, WTD). Two traits are of particular interest in WTD; body size and antler size, which have a degree of observable variation (Hewitt 2011) and are connected to individual reproductive success (DeYoung et al., 2009; Newbolt et al., 2016; Jones et al., 2018). Heritability for antler and body measurements are moderate to high with mean h^2 estimated to be 0.32 (0.07-0.54) for the number of points and 0.42 (0.11-0.69) for basal circumference (Michel et al., 2016) and 0.58-0.64 for body mass (Williams et al., 1994). Large antlers and body size are also sought after by hunters, both as trophies and food sources throughout North America, and thus could be subject to artificial selection (Mysterud, 2011; Ramanzin & Sturaro, 2014). Our objective was to identify QTL throughout the genome associated with variation in antler and body size phenotypes in WTD, under the hypothesis that many genes of small to moderate effect underly these phenotypes. We performed whole genome re-sequencing on pools representing the extreme distribution of antler and body size phenotypes in a wild, but intensively monitored WTD population.

Materials and Methods

Study area

Anticosti Island (49°N, 62°W; 7,943 km^2) is located in the Gulf of St. Lawrence, Québec (Canada) at the northeastern limit of the white-tailed deer range (Figure 1). The island is within the balsam fir-white birch bioclimatic region with a maritime sub-boreal climate characterized by cool and rainy summers (630mm/year), and long and snowy winters (406 cm/year; Environment Canada, 2006; Simard et al., 2010). The deer population was introduced in 1896 with ca. 220 animals and rapidly increased. Today, densities are

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

>20 deer/km^2 and can exceed 50 deer/km^2 locally (Potvin and Breton 2005) with an estimated population >160,000 individuals.

Sample Collection and Modeling

We collected tissue samples and phenotypic data on 4,466 male deer harvested by hunters from September to early December, 2002–2014. We used cementum layers in incisor teeth to age individuals (Hamlin et al., 2000). Two metrics of antler size were selected: the number of antler tines or points (>2.5 cm) and beam diameter (measured at the base; ±0.02 cm) (Simard et al. 2014). We used one metric for body size: body length which is correlated to other metrics such as body weight and hind foot length (Bundy et al. 1991). Because antler and body size of male cervids are correlated to age (Solberg et al. 2004; Nilsen & Solberg, 2006), we controlled for age before ranking males according to antler and body size. We used linear models to assess the relationship between age and each metric separately. We computed an antler and body size index based on the average rank of each individual's residual variation for the phenotypic metrics. The top and bottom 150 individuals from each group were selected from the available database for DNA extraction, and only deer with equal representation for large and small phenotypes from a given year were included for further analysis. This created four groups: large antler (LA), small antler (SA), large body size (LB), and small body size (SB) (Table S1).

DNA Extraction and Genome Sequencing

We isolated DNA from tissue using the Qiagen DNeasy Blood & Tissue Kit. The concentration of each DNA extract was determined using a Qubit dsDNA HS Assay Kit

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Life Technologies, Carlsbad, CA, USA). Equal quantities of DNA (100 ng/sample) were combined into representative pools for LA (n=48), and SA (n=48), LB (n=54), and SB (n=61) for a desired final concentration of 20 ng/ul combined DNA for each pool. We sequenced each pool to approximately 50X coverage at The Centre for Applied Genomics (Toronto, ON, Canada) on an Illumina HiSeqX with 150 bp pair-end reads (Table S2).

Genome Annotation

A newly generated draft WTD genome constructed from long (PacBio) and short-read (Illumina) data was used as a reference (Fuller et al., 2019; NCBI PRJNA420098). We performed a full genome annotation by masking repetitive elements throughout the genome using a custom WTD database developed through repeat modeler (Smit & Hubley, 2015) in conjunction with repeatmasker (Smit, Hubley & Green, 2015) using the NCBI database for artiodactyla, without masking low complexity regions. The masked genome was then annotated using the MAKER2 pipeline (Holt & Yandell, 2011). We used a three-stage process (Laine et al., 2016) utilizing available cow EST and Protein sequences available through NCBI for initial training with SNAP2. The resulting GFF output was then used as evidence for the MAKER2 prediction software, again using SNAP2. The final annotation used evidence from the prior SNAP2 trials and the available human training data set with AUGUSTUS to generate a final annotation (Dryad Accession ID XXXXXX). Gene IDs were generated using blastp (Johnson et al., 2008) on the WTD annotation protein transcripts, restricting the blast search to a list of GI numbers obtained from the NCBI proteins database.

Mapping and characterization of SNPs

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We performed initial quality filtering for all reads using fastqc. Raw reads were aligned independently to the masked WTD reference genome with BWA-mem (Li, 2013). We used samtools (Li et al., 2009) to merge and sort all aligned reads into four files for each representative pool. Using bcftools (Li et al., 2009) we generated mpileup files representative of the antler and body phenotypes (parameters = -B -q20) following standard poolseq protocol (Kolfer et al., 2011).

Genome wide differences were calculated between large and small phenotypes for antler and body size. We tested for pairwise allele frequency differences using Fisher's exact test (FET) and pairwise fixation index (F_{ST}) from the Popoolation2 software suite (Kofler et al., 2011), with parameters set to a minimum coverage of 50, maximum coverage of 200 and a minimum overall count of the minor allele of 8 for each pool. F_{ST} values were highly correlated to FET P-values (Pearson r = 0.96, Figure S1), and thus we refer only to F_{ST} for outlier detection. Manhattan plots were generated using a custom R script to plot the distribution of F_{ST} by position and scaffold throughout the entire WTD genome. We identified the 99th percentile of F_{ST} scores across the genome and calculated the genome-wide mean F_{ST}.

For all pools, we examined the proximity to genic regions for all identified SNPs using SnpEff (Cingolani et al., 2012). All SNP locations were characterized as being in an intergenic/intragenic region, 25kb up/downstream of a gene, intron, or exon. We also obtained an assessment of the putative impact of each variant (HIGH, MODERATE, LOW) and proximity to the nearest gene for every identified SNP. Numerous candidate genes for body size and head gear have been identified in ruminants (Bouwman et al., 2018; Ker et al., 2018; Wang et al., 2019); we looked specifically for SNPs and

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

divergence in a non-exhaustive list of genes known *a priori* to influence these phenotypes (Table S3).

We used a sliding window approach (10kb) to create estimates of Tajima's *D* across the entire WTD genome using the software popoolation (Kolfer et al., 2011). Here, all reads, irrespective of pool, were merged into a single pileup file then subsampled to 100X coverage per site to account for overrepresented sequences. Tajima's *D* was calculated in 10kb sliding windows, and gene only windows that were 60% covered by pre-defined parameters (min coverage 30x, max coverage 200x, min count 2, q- 20) (Kolfer et al., 2011). Resulting *D* values in the 99th percentile were used to help identify outlier regions throughout the genome, and were plotted on scaffolds of interest previously identified through the F_{ST} outlier analysis.

Transposable Elements

We used consensus transposable element (TE) sequences from the repbase database for cow to repeat mask the WTD reference genome, and subsequently merged this masked genome with the repbase TE reference sequences (Bao et al., 2015). We used the repbase database consensus sequences to re-mask the genome due to more robust information for TE identities, family and order required in this analysis. Using methods identical to those previously stated, all reads for both antler and body size phenotypes were aligned to the TE-WTD merged reference genome. We used the recommended workflow for the software suite popoolationTE2 (Kofler et al., 2016) to create a list of predicted TE insertions. From this we calculated TE frequency differences between large and small phenotypes as well as proximity to genic regions. Here, we only examined TEs that were within 25-kb up or downstream from a gene and

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

within the 95th percentile for absolute frequency differences to allow for more features to be assessed subsequently.

GO Pathways

To identify shared gene pathways among outlier regions we used an analysis of gene ontology (GO) terms. GO terms were obtained for all genes that were within 25-kb of SNPs in the 99th F_{ST} and 95% TE frequency percentile for antler and body size phenotypes independently. We created outlier gene ID lists for antler and body size that were provided independently to DAVID v6.8 (Huang, Sherman & Lempicki, 2009), while removing all duplicate gene IDs. DAVID uses a modified FET to determine the significance of enrichment for any given GO pathway based on the size of the gene list provided, the number of genes used as a background for a species, and level of enrichment for each term from the list. Functional annotation enrichments for GO terms using GOTERM_BP_ALL, GOTERM_CC_ALL, and GOTERM_MF_ALL were obtained with the top 10 enriched GO terms being plotted with gene counts and p-values. We used the program REVIGO (Supek et al., 2011) to remove redundant GO terms and to visualize semantic similarity-based scatterplots. Results from DAVID with a benjamini corrected p-value <0.05 were used with REVIGO to generate plots of significant GO terms for biological processes, molecular function, and cellular components of antler and body size phenotypes.

Validation of outliers

We selected three outlier SNPs in the 99th percentile F_{ST} values in genic regions associated with antler size for qPCR validation. Custom genotyping assays using

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

rhAMP chemistry (Integrated DNA Technologies) were designed and genotyped on the QuantStudio 3 (Thermo Fisher Scientific). Oligo sequences and reaction parameters are provided in Table S4. Using the phenotypic category as the binary response variable, we ran a logistic regression treating the genotypic data as additive (e.g. 0-2 LA/LB alleles per locus, per individual).

Results

Phenotypes

The distribution of phenotypes from all individuals in the existing database is shown in Figure S1. We only selected the top 150 individuals at the tail ends of the distribution for measurements used in our antler and body size rankings which are representative of the “extreme phenotypes”. Artist renderings and the distribution of measurements for the number of antler points, beam diameter, and body length between the groups of individuals representing each extreme phenotype pool (LA, SA, LB, SB) are shown in Figure 2a. There were no differences in mean age between the pools (Figure 2b).

Detecting genetic variants and their genomic regions

The total reads generated from resequencing are observed in Table S3 (BioProject ID PRJNA576136), and the estimated mean genome-wide coverage based on the 150 bp read length and 2.5 Gb WTD genome. After filtering for coverage, INDELS, and minimum allele counts, 834,855 and 1,016,105 SNPs were identified for the antler and body size phenotype pools respectively; the distribution for which F_{ST} values across the entire WTD genome are shown appears in Figure 3a. The mean F_{ST} across the antler

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

phenotype SNPs was calculated to be 0.019; F_{ST} values > 0.12 fell within the 99th percentile which were considered outliers for downstream analysis. For the comparison of body size phenotypes, the mean F_{ST} was calculated as 0.017 with a 99th percentile F_{ST} value cut off of 0.11. We identified 336,379 SNPs that were shared between both antler and body size pools.

The distribution of SNPs throughout the WTD genome, the classification of their annotation (exon, intron, 25-kb flank, or intron) and the SNP density based on F_{ST} category are shown in Figure 3b. The comparison of the overall count and percentage of SNPs between outlier and total SNPs are observed in Table 1 and a selection of outlier regions is shown in Figure 4 and Figure S3. From the list of 32 *a priori* candidate genes in the literature, we identified the highest F_{ST} SNP and total number of SNPs from each gene, including a 25 kb window up/downstream, from the analysis of each trait (Table S3); five genes had F_{ST} values in the 99th percentile.

Estimates of Tajima's *D*

Tajima's *D* was estimated in 10kb sliding windows across the entire WTD reference genome yielding 251,275 windows, 103,832 of these windows met the criteria previously defined for estimations of *D*. The genome-wide mean *D* was calculated to be -0.50. We identified windows that fell within the 99th percentile of positive (>0.35) and negative (<-1.75) (Figure S4) values for comparison with the F_{ST} outliers and gene regions (Table 1).

TE Insertions

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We identified 19,734 TE insertions through the joint analysis of antler phenotype sequences (both large and small). Of these TEs, we identified those that fell within the 95th percentile for the absolute difference between large and small phenotypes for further analysis (>0.198%; Figure S5). Of the 95th percentile TE insertions in the antler analysis, 169 were found to overlap with genes, with 206 within a 25-kb window up or downstream of genic regions. For body size, 364 insertions in the 95th percentile (>0.205%) overlapped with genes, and 143 within 25-kb of a genic region.

Gene Ontology Annotations

From the top SNPs identified through outlier analysis, 1,806 genes were identified for antler GO analysis after the removal of duplicate genes, and 1,983 from the body size analysis. The top 10 enriched GO terms and the gene counts from these analyses are in Figure 5. We also showed term reductions for significantly enriched GO terms (benjamini-corrected p-value <0.05) through the program REVIGO, with clustered terms relating to semantic similarity of the terms. This was only done for GO categories in which significant values existed (Figure S6-9). We chose to only display GO figures relating to biological processes as it is most relevant to our study. As there was no significant enrichment through the body size analysis, only the main findings for antler are represented in Figure S6, where we observed GO terms grouped by semantic similarity and highlighted based on calculated significance of enrichment. As the list of GO terms was reduced by REVIGO, the labels present are representative of less dispensable terms.

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In analyzing the GO term enrichments for biological processes of genes that met our inclusion criteria from the TE outlier analysis, we found no significantly enriched pathways.

Validation of Outliers

We genotyped three loci (RIMS1, PTEN, and SRP54) which resulted in 78 individuals having complete genotype and phenotypic data (n = 45 HA and 35 LA; Dryad Accession no. XXXXXX). The model showed an effect of the number of HA alleles at the SRP54 locus on antler category (ß = -1.05, p = 0.02), but not the other two loci (p > 0.05).

Discussion

The study of sexually selected quantitative traits in the Anticosti Island WTD population has provided novel insights into the underlying genomic architecture of phenotypic variation. The extensive database with phenotypic measurements for these deer (n=4,466) allowed us to select individuals from the wide range of the distribution (Figure 2; S2) that are representative of extreme phenotypes for the two antler and one body size measurements. This sampling methodology is important to maximize the additive genetic variance for each trait, increasing power and the ability to detect QTL (Kardos et al., 2016). In this sense, this is the first study to apply this methodology of extreme phenotypic sampling in a wild vertebrate population for sexually selected and hunter-targeted traits.

Anticosti Island WTD form a putatively panmictic population, where we would expect gene flow across the island and thus a lack of fixation at any specific loci, or existence of any population structure (Fuller et al., 2019). Consistent with this, genome-

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

wide F_{ST} was ~0.01; however, we identified QTL for both traits within the top percentile (antler $F_{ST} > 0.12$, maximum=0.49; body $F_{ST} > 0.11$, maximum=0.35) that are atypically large for what would be expected in a panmictic population. The top percentile SNPs and divergent TEs for antler and body traits are widely dispersed throughout the WTD genome (Figure 3a), which imply a polygenic model for these quantitative traits. This is consistent with the literature on body size and ornaments in mammals (Visscher et al., 2007; Bouwman et al., 2018; Wang et al., 2019); however, a few QTL of small-to-moderate effect appear present, at least for antlers (i.e. SRP54) which is consistent with studies on sheep (Berenos et al., 2015).

Genomic architecture of antlers and body size

For both traits, there is evidence for divergent loci represented by high F_{ST} values throughout the genome that overlap with genic regions (Figure 3b). Traditionally, variation in SNPs is focussed on non-synonymous variants, but it is becoming evident that primarily non-coding regions impact phenotypic variation (Watanabe et al., 2019), often with clear relationships to promoters and enhancer regions (Pagani & Baralle 2004; Zhang & Lupski 2015; Foote et al., 2016).

The identification of 2,690 SNPs for antler traits and 3,849 SNPs (Table S3) for body size traits with high F_{ST} that overlap with genic windows is supportive of many QTL of small effect and a polygenic model for these quantitative traits. While some are surely false positives (Whitlock & Lottheros 2015), cattle GWAS studies typically identify 100s to 1000s QTL (Cole et al., 2011; Jiang et al., 2019). Our data suggest that it is the cumulative effects from these variants, including TE insertions, and possibly their epistatic interactions that are ultimately driving the phenotypic variation in antler and

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

body sizes. Examining many of these high F_{ST} variants at the scaffold level revealed signatures of linkage and selection at specific sites (Figure 4). Many of these regions (Figure S3) show "genomic islands" surrounding highly differentiated SNPs. These islands presumably represent divergent haplotypes between extreme phenotypes and thus a putative connection to phenotypic variation. Moreover, negative Tajima's *D* values (Figure 4) are consistent with the demographic history of the island and our sampling strategy that mimics divergent selection.

Analysis of the SNPs resulting from the antler identified pathways relating to various aspects of antler growth and regeneration (Figure 5; S6). The three loci selected for further qPCR genotyping assays had a relatively high degree of differentiation and close proximity to genic regions (Figure S3). All three SNPs are intron variants for the genes RIMS1, PTEN and SRP54 respectively. RIMS1 (regulating synaptic membrane exocytosis 1) codes for proteins that are central in integrating active zone proteins and synaptic vesicles into scaffolds that control neurotransmitter release and are highly expressed in brain and testis tissues of human and mice (Lonart, 2002; Schoch et al., 2002). Phosphatase and tensin homolog (PTEN) is a tumor suppressor gene and is part of chemical pathways involved in apoptosis (Chalhoub & Baker 2009). SRP54 is a signal recognition particle that mediates the targeting of proteins to the endoplasmic reticulum (Pool et al., 2002); this protein has been found in human clinical, and zebrafish models to be associated with bone marrow failure syndromes and skeletal abnormalities (Carapito et al., 2017).

Antlers are the only completely regenerable organ found in mammals (Li et al., 2014), a unique process that involves simultaneous exploitation of oncogenic pathways

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and tumor suppressor genes, and the rapid recruitment of synaptic and blood vesicles (Wang et al., 2019). The relative high degree of differentiation found at intron regions for these genes between large and small antlered pools, in combination with their physiological connections to the underlying processes involved in antler growth and regeneration (notably PTEN and SRP54), supports their potential function in the observed variation between phenotypes. We explored this relationship through qPCR genotyping of these loci for individuals included in each respective pool. The model showed a relationship between antler size and genotype for the SRP54 SNP previously described, indicating it as a QTL of moderate effect considering our results. Further bolstering this association was the GO terms showing an enriched Reactome Pathway (Benjamini corrected $p = 0.02$) for an SRP-dependent cotranslational protein targeting to membrane (GO:0006614), that have known functions in bone abnormalities and congenital neutropenia (Carapito et al 2017),The lack of a strong RIMS1 and PTEN signal in the antler model is consistent with the polygenic nature, and the likely epistatic interactions (discussed more below), driving trait variation meaning any individual locus on its own lacks a strong signal. Increasing sample size and individual genome sampling is required to fully characterize these potential epistatic interactions (e.g. Knief et al. 2019).

Outlier analysis for SNPs identified through the comparison of extreme body size phenotype sequences also revealed an array of genes of interest. However, it was more difficult than antlers to not engage in storytelling (Pavlidis et al., 2012) as traits of this nature, for example human height, involve 100s of SNPs and a multitude of biological pathways (Marouli et al., 2017); this is reflected in our study with no statistically

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

significant enriched GO terms (Figure 5). However, a few SNPs are worth highlighting. The upstream variant (4426 bp, F_{ST}=0.31) for liver-enriched gene 1 (LEG1) showed multiple points that appear to be indicative of haplotypes for a QTL, both on and off genic regions throughout the ~106 kb scaffold. LEG1 is linked to liver function (Chang et al. 2011), and our WTD annotation results predicted two genes on this scaffold, both with similarity to LEG1. In addition to this, we aligned sequences from a previously conducted WTD transcriptome analysis (Genomic Resources Development Consortium et al., 2013), which revealed expressed sequences between this QTL and the upstream LEG1 gene. As our annotation is based off existing evidence for human and cow expressed sequences, this would suggest a novel structure for the gene in WTD and a role in body size.

Two missense variants that fall on genes for MYADM (Myeloid-associated differentiation marker) and SPATA31E1 (Spermatogenesis-associated protein 31E1) with F_{ST} values of 0.31 and 0.27. Gonzalez et al. (2013) identified a QTL in sheep that was significantly associated with decreased mean corpuscular volume, a trait correlated with body size. Importantly, the authors observed a divergent artiodactyla MYADM-like variant in strong linkage disequilibrium with the associated SNP (Gonzalez et al., 2013), as well as being in a region of strong historical selection (Kijas et al., 2012). Schrimpf et al. (2016) identified 9 high-impact SNPs absent in fertile stallions with putatively deleterious effects on fertility, that included SPATA31E1. Considering this gene's connection in spermatogenesis (Schimpf et al., 2016) and homology in a related species for fertility, there is a link to the sexually selected nature of the trait being examined and is consistent with body size being linked to reproductive success in deer

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Vanpé et al., 2010); interestingly this could reflect correlated evolution between body-size and sperm viability. Both of these markers require validation but open up the possibility for marker-assisted breeding programs and studying sexual selection in wild cervids.

Transposable elements and epistatic considerations

We generated a list of novel TE insertions for both phenotypes. Traditionally, masking these elements reduces misalignments due to the repetitive nature of their sequences and current limitations with mapping software (Treangen & Salzberg 2012), but also misses a wealth of information that encompasses a large portion of the genome. Our focus was the frequency differences of TE insertions to properly mapped reference sequences, and how these varied between phenotypes. This stems from increasing evidence showing that TE insertions impact gene expression, and thus the variation for a given phenotype (Bourque et al., 2018). We found 169 divergent TE insertions in genic regions from our antler analysis and 68 from the body size analysis. While there was no significant enrichment of pathways, the insertion of these highly variable sequences has the potential to impact gene function as studies are now starting to emerge that both validate insertions (Lerat et al., 2018) and show evidence for positive selection (Kofler et al., 2011).

Epistatic relationships of many genes of small effect, including TEs, constitutes a complex genomic architecture and is the most likely explanation for the natural variation observed for these sexually selected traits. Anticosti Island should provide a more-or-less homogenous environment, but microclimate could drive some degree of this variation. We were liberal in determining genic regions (exon, intron, 25-kb flanks and

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

99th percentile) to avoid missing potentially relevant QTL that traditional approaches might exclude (Albert & Kruglyak, 2015). This is partially observed through the enrichment of GO terms shown in Figure 5 and S6, where we see strong semantic clustering for terms relating to macromolecule localization (GO:0033036) and regulation of cellular component organization (GO:0051128), and terms relating to the ornament pathways identified by Wang et al. (2019) such as; anatomical structure morphogenesis (GO:0009653), chemical synaptic transmission (GO:0007268), localization (GO:0051179), and cellular component organization or biogenesis (GO:0071840). Although the number of terms is reduced by removing redundant terms and grouping based on semantic similarity, we still showed the diverse array of functions represented of by the divergent SNPs (Figure S6).

Lastly, in analyzing a list of 32 *a priori* genes identified for their significance in various pathways relating to antler and body development from the literature (Table S3), we generally observed low levels of variation, i.e. 27 of 32 genes had SNPs with F_{ST} <99th percentile, between phenotypes for these genes for both traits, despite their known effect. These findings are not unsurprising in the case of genes identified through Wang et al. (2019), as many of these genes are identified as being under selective pressures and are responsible for the presence or absence of a trait between species, not variation of the trait within a species per se. Conversely, Bouwman et al. (2018) specifically identified common genes that regulate body size in mammals where we would expect to see more variation between phenotypes. Many of these genes were found to be in divergent regions, with SNPs for IGF2 (F_{ST}=0.15) and PLAG1 (F_{ST}=0.17) present in the outlier analysis. PLAG1 (pleomorphic adenoma gene 1) initiates

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transcription of IGF2 (insulin like growth factor 2), with both playing a role in the variation of stature in humans and cattle (Bouwman et al., 2018; Wood et al., 2014; Pryce et al., 2011; Fortes et al., 2014). This is again suggesting a small effect of these variants for the genetic contribution to variation in body size from our analysis.

Implications for understanding of artificial and sexual selection

Antler and body size are traits that are sexually selected for and linked to reproductive success in WTD (Newbolt et al., 2016; Morina et al., 2018), and are traits desired by hunters, managers, and farmers. We have identified a few low-to-moderate effect QTL, and many genome-wide variants with clear differentiation between the phenotypic extremes for these traits. We suggest the combined effects of these variants as being the genetic drivers behind observed phenotypic variation. The measured phenotypes are expected to be under balancing selection (Mank 2018); however this does not have to be reflected in all gene pathways that ultimately contribute to the expression of a complex trait. Considering the case for an omnigenic model where all relevant peripheral genes related to a given phenotype can affect the function of core phenotype-related genes (Boyle et al., 2017), it would be inaccurate to attribute the evolution of a complex trait to any specific genomic region without examining the additive affects of the entire genomic landscape; thus more sequencing is required.

We provide evidence for positive selection at various sites in Figure 4, as defined by clear dips in *D* values (approaching values of -2), and through genic regions with significant genomic island of differentiation between phenotypes (Figure S3). Positive *D* values represent balancing selection, but our sampling design (i.e. no intermediate phenotypes) and rapid population expansion (Potvin and Breton 2005) impacts metrics

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

like Tajima's *D* (Stajich & Hahn, 2004). Thus, we predict that subsequent genotyping of targeted loci across the spectrum of phenotypes should produce values more consistent with balancing selection.

Although we have identified many potential QTL relating to the variation in phenotypes for WTD, their direct use in management or monitoring remains unclear. Considering the gene-targeted conservation road map presented by Kardos and Shafer (2018), there appears to be no current application in management or monitoring for WTD species given the existing body of knowledge and conservation status of WTD. It is conceivable that a gene panel could be developed that could be used in breeding and management programs (e.g. Quality Deer Management) or assess the effects of artificial selection by trophy hunting; however until a reasonable amount of phenotypic variation can be attributed to specific QTL, as we have begun to do with SRP54 genotyping, a gene targeted approach for management and breeding does not seem warranted at this stage.

Acknowledgments

This work was supported by Natural Sciences and Engineering Research Council of Canada Discovery Grant (ABAS and SDC); ComputeCanada Resources for Research Groups (ABAS); Canadian Foundation for Innovation: John R. Evans Leaders Fund (ABAS); The Symons Trust Fund for Canadian Studies (ABAS); Trent University start-up funds [ABAS]; and Industrial Chairs and Collaborative Research and Development Grants from the Natural Sciences and Engineering Research Council of Canada (SDC). We thank the outfitters of Anticosti Island and the Ministère des Forêts, de la Faune et des Parcs du Québec for logistical help associated with fieldwork, and all of the field

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

assistants and technicians who collected samples and made phenotypic measurements throughout the duration of this project.

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. Nature Reviews Genetics, 16(4), 197–212. https://doi.org/10.1038/nrg3891

Anand, S., Mangano, E., Barizzone, N., Bordoni, R., Sorosina, M., Clarelli, F., … De Bellis, G. (2016). Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering. Scientific Reports, 6(1), 33735. https://doi.org/10.1038/srep33735

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA, 6(1). https://doi.org/10.1186/s13100-015-0041-9

Barnett, I. J., Lee, S., & Lin, X. (2013). Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. Genetic Epidemiology, 37(2), 142–151. https://doi.org/10.1002/gepi.21699

Bérénos, C., Ellis, P. A., Pilkington, J. G., Lee, S. H., Gratten, J., & Pemberton, J. M. (2015). Heterogeneity of genetic architecture of body size traits in a free-living population. Molecular Ecology, 24(8), 1810–1830. https://doi.org/10.1111/mec.13146

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., … Feschotte, C. (2018). Ten things you should know about transposable elements. Genome Biology, 19(1), 199. https://doi.org/10.1186/s13059-018-1577-z

Bouwman, A. C., Garrick, D. J., Reecy, J., & Van Tassell, C. P. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nature Genetics, 50, 362–367 https://doi.org/10.1038/s41588-018-0056-5

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell, 169(7), 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bundy, R. M., Robel, R. J., & Kemp, K. E. (1991). Whole Body Weights Estimated from Morphological Measurements of White-Tailed Deer. Transactions of the Kansas Academy of Science (1903-), 94(3/4), 95-100. https://doi.org/10.2307/3627856

Bünger, L., Lewis, R. M., Rothschild, M. F., Blasco, A., Renne, U., & Simm, G. (2005). Relationships between quantitative and reproductive fitness traits in animals. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1459), 1489–1502. https://doi.org/10.1098/rstb.2005.1679

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Carapito, R., Konantz, M., Paillard, C., Miao, Z., Pichot, A., Leduc, M. S., … Bahram, S. (2017). Mutations in signal recognition particle SRP54 cause syndromic neutropenia with Shwachman-Diamond-like features. The Journal of Clinical Investigation, 127(11), 4090–4103. https://doi.org/10.1172/JCI92876

Chalhoub, N., & Baker, S. J. (2009). PTEN and the PI3-kinase pathway in cancer. Annual Review of Pathology, 4, 127–150. https://doi.org/10.1146/annurev.pathol.4.110807.092311

Chang, C., Hu, M., Zhu, Z., Lo, L. J., Chen, J., & Peng, J. (2011). liver-enriched gene 1a and 1b encode novel secretory proteins essential for normal liver development in zebrafish. PloS One, 6(8), e22910. https://doi.org/10.1371/journal.pone.0022910

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 w1118; iso-2; iso-3. Fly, 6(2), 80–92. https://doi.org/10.4161/fly.19695

Cole, J. B., Wiggans, G. R., Ma, L., Sonstegard, T. S., Lawlor, T. J., Crooker, B. A., … Da, Y. (2011). Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics, 12(1), 408. https://doi.org/10.1186/1471-2164-12-408

Coltman, D. W., O'Donoghue, P., Jorgenson, J. T., Hogg, J. T., Strobeck, C., & Festa-Bianchet, M. (2003). Undesirable evolutionary consequences of trophy hunting. Nature, 426(6967), 655–658. https://doi.org/10.1038/nature02177

DeYoung, R. W., Demarais, S., Gee, K. L., Honeycutt, R. L., Hellickson, M. W., & Gonzales, R. A. (2009). Molecular Evaluation of the White-tailed Deer (Odocoileus Virginianus) Mating System. Journal of Mammalogy, 90(4), 946–953. https://doi.org/10.1644/08-MAMM-A-227.1

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*(1), 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Ellegren, H., & Sheldon, B. C. (2008). Genetic basis of fitness differences in natural populations. Nature, 452(7184), 169–175. https://doi.org/10.1038/nature06737

Emond, M. J., Louie, T., Emerson, J., Zhao, W., Mathias, R. A., Knowles, M. R., … Bamshad, M. J. (2012). Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nature Genetics, 44(8), 886–889. https://doi.org/10.1038/ng.2344

Endler, L., Betancourt, A. J., Nolte, V., & Schlötterer, C. (2016). Reconciling differences in pool-GWAS between populations: A case study of female abdominal

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pigmentation in Drosophila melanogaster. Genetics, 202(2), 843–855. https://doi.org/10.1534/genetics.115.183376

Environment Canada. (2006). Climate normals and averages, daily data reports of Port-Menier's station from 1995 to 2005. Canada's National Climate Archive. Retrieved from http://climate.weatheroffice.gc.ca

Foley, A. M., DeYoung, R. W., Hewitt, D. G., Hellickson, M. W., Gee, K. L., Wester, D. B., … Miller, K. V. (2015). Purposeful wanderings: mate search strategies of male white-tailed deer. Journal of Mammalogy, 96(2), 279–286. https://doi.org/10.1093/jmammal/gyv004

Foote, A. D., Vijay, N., Ávila-Arcos, M. C., Baird, R. W., Durban, J. W., Fumagalli, M., … Wolf, J. B. W. (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. Nature Communications, 7(1), 11693. https://doi.org/10.1038/ncomms11693

Fortes, M. R. S., Kemper, K., Sasazaki, S., Reverter, A., Pryce, J. E., Barendse, W., … Lehnert, S. A. (2013). Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. Animal Genetics, 44(6), 636–647. https://doi.org/10.1111/age.12075

Fuller, J. (2019) Différenciation génétique et sélection chez le cerf de Virginie (Odocoileus virginianus) introduit à l'île d'Anticosti. [Unpublished Masters Thesis] Universite Laval. https://corpus.ulaval.ca/jspui/handle/20.500.11794/34902

Genomic Resources Development Consortium, Bensch S, Coltman DW, Davis CS, Hellgren O, Johansson T, Malenfant RM, Moore SS, Palinauskas V, Valkiūnas G (2014) Genomic Resources Notes accepted 1 June 2013 – 31 July 2013. Molecular Ecology Resources 14(1): 218. https://doi.org/10.1111/1755-0998.12166

Gonzalez, M. V., Mousel, M. R., Herndon, D. R., Jiang, Y., Dalrymple, B. P., Reynolds, J. O., … White, S. N. (2013). A Divergent Artiodactyl MYADM-like Repeat Is Associated with Erythrocyte Traits and Weight of Lamb Weaned in Domestic Sheep. PLoS ONE, 8(8), e74700. https://doi.org/10.1371/journal.pone.0074700

Gurwitz, D., & McLeod, H. L. (2013). Genome-wide studies in pharmacogenomics: harnessing the power of extreme phenotypes. Pharmacogenomics, 14(4), 337–339. https://doi.org/10.2217/pgs.13.35

Hamlin, K. L., D. F. Pac, C. A. Sime, R. M. DeSimone, and G. L. Dusek. 2000. Evaluating the accuracy of ages obtained by two methods for Montana ungulates. Journal of Wildlife Management 64, 441–449

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Hewitt, D. G. (2011). Biology and management of white-tailed deer. CRC Press. Retrieved from https://www.crcpress.com/Biology-and-Management-of-White-tailed-Deer/Hewitt/p/book/9781439806517

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics, 12(1), 491. https://doi.org/10.1186/1471-2105-12-491

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Jiang, J., Ma, L., Prakapenka, D., VanRaden, P. M., Cole, J. B., & Da, Y. (2019). A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. Frontiers in Genetics, 10, 412. https://doi.org/10.3389/fgene.2019.00412

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Research, 36(Web Server), W5–W9. https://doi.org/10.1093/nar/gkn201

Johnston SE, McEwan J, Pickering NK et al. (2011) Genome-wide associa-tion mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. Molecular Ecology, 20, 2555–2566.

Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature, 502(7469), 93–95. https://doi.org/10.1038/nature12489

Jones, P. D., Strickland, B. K., Demarais, S., Wang, G., & Dacus, C. M. (2018). Nutrition and ontogeny influence weapon development in a long-lived mammal. Canadian Journal of Zoology, 96(9), 955–962. https://doi.org/10.1139/cjz-2017-0345

Kardos, M., Luikart, G., Bunch, R., Dewey, S., Edwards, W., McWilliam, S., … Kijas, J. (2015). Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. Molecular Ecology, 24(22), 5616–5632. https://doi.org/10.1111/mec.13415

Kardos, M., Husby, A., Mcfarlane, S. E., Qvarnstrom, A., & Ellegren, H. (2016). Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. Molecular Ecology Resources, 16(3), 727–741. https://doi.org/10.1111/1755-0998.12498

Kardos, M., & Shafer, A. B. A. (2018). The Peril of Gene-Targeted Conservation. Trends in Ecology & Evolution, 33(11), 827–839. https://doi.org/10.1016/j.tree.2018.08.011

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ker, D. F. E., Wang, D., Sharma, R., Zhang, B., Passarelli, B., Neff, N., … Yang, Y. P. (2018). Identifying deer antler uhrf1 proliferation and s100a10 mineralization genes using comparative RNA-seq. Stem Cell Research & Therapy, 9(1), 292. https://doi.org/10.1186/s13287-018-1027-6

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., … Consortium, other members of the I. S. G. (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. PLoS Biology, 10(2), e1001258. https://doi.org/10.1371/journal.pbio.1001258

Knief, U., Bossu, C. M., Saino, N., Hansson, B., Poelstra, J., Vijay, N., … Wolf, J. B. W. (2019). Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. Nature Ecology & Evolution, 3(4), 570–576. https://doi.org/10.1038/s41559-019-0847-9

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., … Schlötterer, C. (2011). PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. PLoS ONE, 6(1), e15925. https://doi.org/10.1371/journal.pone.0015925

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics, 27(24), 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Kofler, R., Gómez-Sánchez, D., & Schlötterer, C. (2016). PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. Molecular Biology and Evolution, 33(10), 2759–2764. https://doi.org/10.1093/molbev/msw137

Kuparinen, A., & Festa-Bianchet, M. (2017). Harvest-induced evolution: insights from aquatic and terrestrial systems. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1712), 20160036. https://doi.org/10.1098/rstb.2016.0036

Laine, V. N., Gossmann, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J. F., … Groenen, M. A. M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. Nature Communications, 7(1), 10474. https://doi.org/10.1038/ncomms10474

Lerat, E., Goubert, C., Guirao-Rico, S., Merenciano, M., Dufour, A., Vieira, C., & González, J. (2019). Population-specific dynamics and selection patterns of transposable element insertions in European natural populations. Molecular Ecology, 28(6), 1506–1522. https://doi.org/10.1111/mec.14963

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Li, C., Zhao, H., Liu, Z., & McMahon, C. (2014). Deer antler – A novel model for studying organ regeneration in mammals. The International Journal of Biochemistry & Cell Biology, 56, 111–122. https://doi.org/10.1016/j.biocel.2014.07.007

Li, D., Lewinger, J. P., Gauderman, W. J., Murcray, C. E., & Conti, D. (2011). Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. Genetic Epidemiology, 35(8), 790–799. https://doi.org/10.1002/gepi.20628

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

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

Lonart, G. (2002). RIM1: an edge for presynaptic plasticity. Trends in Neurosciences, 25(7), 329–332. https://doi.org/10.1016/S0166-2236(02)02193-8

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular Ecology Resources, 17(2), 142–152. https://doi.org/10.1111/1755-0998.12635

Malenfant, R. M., Davis, C. S., Richardson, E. S., Lunn, N. J., & Coltman, D. W. (2018). Heritability of body size in the polar bears of Western Hudson Bay. Molecular Ecology Resources, 18(4), 854–866. https://doi.org/10.1111/1755-0998.12889

Mank, J. E. (2017). Population genetics of sexual conflict in the genomic era. Nature Reviews Genetics, 18(12), 721–730. https://doi.org/10.1038/nrg.2017.83

Marouli, E., Graff, M., Medina-Gomez, C., Lo, K. S., Wood, A. R., Kjaer, T. R., … Lettre, G. (2017). Rare and low-frequency coding variants alter human adult height. *Nature*, *542*(7640), 186–190. https://doi.org/10.1038/nature21039

Michel, E. S., Demarais, S., Strickland, B. K., Smith, T., & Dacus, C. M. (2016). Antler characteristics are highly heritable but influenced by maternal factors. The Journal of Wildlife Management, 80(8), 1420–1426. https://doi.org/10.1002/jwmg.21138

Morina, D. L., Demarais, S., Strickland, B. K., & Larson, J. E. (2018). While males fight, females choose: male phenotypic quality informs female mate choice in mammals. Animal Behaviour, 138, 69–74. https://doi.org/10.1016/J.ANBEHAV.2018.02.004

Mysterud, A. (2011). Selective harvesting of large mammals: how often does it result in directional selection? Journal of Applied Ecology, 48(4), 827–834. https://doi.org/10.1111/j.1365-2664.2011.02006.x

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Neethiraj, R., Hornett, E. A., Hill, J. A., & Wheat, C. W. (2017). Investigating the genomic basis of discrete phenotypes using a Pool-Seq-only approach: New insights into the genetics underlying colour variation in diverse taxa. Molecular Ecology, 26(19), 4990–5002. https://doi.org/10.1111/mec.14205

Newbolt, C. H., Acker, P. K., Neuman, T. J., Hoffman, S. I., Ditchkoff, S. S., & Steury, T. D. (2016). Factors Influencing Reproductive Success in Male White-Tailed Deer. Journal of Wildlife Management 81(2), 206-217 https://doi.org/10.1002/jwmg.21191

Nilsen, E. B., and E. J. Solberg. (2006). Patterns of hunting mortality in Norwegian moose (Alces alces) populations. European Journal of Wildlife Research 52:153–163

Pagani, F., & Baralle, F. E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. Nature Reviews Genetics, 5(5), 389–396. https://doi.org/10.1038/nrg1327

Pavlidis, P., Jensen, J. D., Stephan, W., & Stamatakis, A. (2012). A Critical Assessment of Storytelling: Gene Ontology Categories and the Importance of Validating Genomic Scans. Molecular Biology and Evolution, 29(10), 3237–3248. https://doi.org/10.1093/molbev/mss136

Perez-Gracia, J.P., Ruiz-Ilundain, M.G, Garcia-Ribas, I., & Carrasco, E.M. (2002). The role of extreme phenotype selection studies in the identification of clinically relevant genotypes in cancer research. Cancer, 95(7), 1605–1610. https://doi.org/10.1002/cncr.10877

Pool, M. R., Stumm, J., Fulga, T. A., Sinning, I., & Dobberstein, B. (2002). Distinct modes of signal recognition particle interaction with the ribosome. Science, 297(5585), 1345–1348. https://doi.org/10.1126/science.1072366

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., … Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Science Advances, 3(8), e1603198. https://doi.org/10.1126/sciadv.1603198

Pryce, J. E., Hayes, B. J., Bolormaa, S., & Goddard, M. E. (2011). Polymorphic regions affecting human height also control stature in cattle. Genetics, 187(3), 981–984. https://doi.org/10.1534/genetics.110.123943

Ramanzin, M., & Sturaro, E. (2014). Habitat quality influences relative antler size and hunters' selectivity in roe deer. European Journal of Wildlife Research, 60(1), 1–10. https://doi.org/10.1007/s10344-013-0744-5

Robinson, M.R. et al. (2013) Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. Mol. Ecol. 22, 3963–3980

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Santure, A. W., De Cauwer, I., Robinson, M. R., Poissant, J., Sheldon, B. C., & Slate, J. (2013). Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus major) population. Molecular Ecology, 22(15), 3949–3962. https://doi.org/10.1111/mec.12376

Santure, A. W., & Garant, D. (2018). Wild GWAS-association mapping in natural populations. Molecular Ecology Resources, 18(4), 729–738. https://doi.org/10.1111/1755-0998.12901

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nature Reviews Genetics, 15(11), 749–763. https://doi.org/10.1038/nrg3803

Schoch, S., Castillo, P. E., Jo, T., Mukherjee, K., Geppert, M., Wang, Y., … Südhof, T. C. (2002). RIM1α forms a protein scaffold for regulating neurotransmitter release at the active zone. Nature, 415(6869), 321–326. https://doi.org/10.1038/415321a

Schrimpf, R., Gottschalk, M., Metzger, J., Martinsson, G., Sieme, H., & Distl, O. (2016). Screening of whole genome sequences identified high-impact variants for stallion fertility. BMC Genomics, 17(1), 288. https://doi.org/10.1186/s12864-016-2608-3

Simard, M. A., T. Coulson, A. Gingras, and S. D. Côté. (2010). Influence of density and climate on the population dynamics of a large herbivore under harsh environmental conditions. Journal of Wildlife Management, 74, 1671–1685

Simard, M.-A., J. Huot, S. de Bellefeuille and S. D. Côté. (2014). Influences of habitat composition, plant phenology, and population density on autumn indices of body condition in a northern white-tailed deer population. Wildlife Monographs 187, 1-28

Smit, AFA, Hubley, R & Green, P. (2019, November, 8) RepeatMasker Open-4.0. 2013-2015. Retrieved from http://www.repeatmasker.org

Smit, AFA, Hubley, R. (2019, November, 8) RepeatModeler Open-1.0. 2008-2015 retrieved from http://www.repeatmasker.org

Solberg, E. J., A. Loison, J.-M. Gaillard, and M. Heim. (2004). Lasting effects of conditions at birth on moose body mass. Ecography 27:677–687

Stajich, J. E., & Hahn, M. W. (2004). Disentangling the Effects of Demography and Selection in Human History. Molecular Biology and Evolution, 22(1), 63–73. https://doi.org/10.1093/molbev/msh252

Stinchcombe, J. R. & Hoekstra, H. E. (2008). Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity*, 100(2), 58–170 https://doi.org/10.1038/sj.hdy.6800937

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS ONE, 6(7), e21800. https://doi.org/10.1371/journal.pone.0021800

Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews Genetics, 13(1), 36–46. https://doi.org/10.1038/nrg3117

Vanpé, C., Gaillard, J.-M., Kjellander, P., Liberg, O., Delorme, D., Mark, A. J., … Hewison, A. J. M. (2010). Assessing the intensity of sexual selection on male body mass and antler length in roe deer Capreolus capreolus: is bigger better in a weakly dimorphic species? Oikos, 119, 1484–1492. https://doi.org/10.1111/j.1600-0706.2010.18312.x

Visscher PM, Macgregor S, Benyamin B et al. (2007) Genome partitioning of genetic variation for height from 11,214 sib- ling pairs. American Journal of Human Genetics, 81, 1104–1110

Wang, Y., Zhang, C., Wang, N., Li, Z., Heller, R., Liu, R., … Qiu, Q. (2019). Genetic basis of ruminant headgear and rapid antler regeneration. Science, 364(6446), eaav6335. https://doi.org/10.1126/science.aav6335

Watanabe, K., Stringer, S., Frei, O., Mirkov, M., de Leeuw, C., Polderman, T. J. C., … Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nature Genetics, 51(9), 1339–1348. https://doi.org/10.1038/s41588-019-0481-0

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F ST. The American Naturalist, 186(S1), S24–S36. https://doi.org/10.1086/682949

Williams, J. D., Krueger, W. F., & Harmel, D. H. (1994). Heritabilities for antler characteristics and body weight in yearling white-tailed deer. Heredity, 73(1), 78–83. https://doi.org/10.1038/hdy.1994.101

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., … Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics, 46(11), 1173–1186. https://doi.org/10.1038/ng.3097

Zhang, F., & Lupski, J. R. (2015). Non-coding genetic variants in human disease: Figure 1. Human Molecular Genetics, 24(R1), R102–R110. https://doi.org/10.1093/hmg/ddv259

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Data Accessibility

Raw sequence data FastQ files are available on the Sequence Read Archive (Accession: PRJNA576136). All bioinformatic and analytical code available on GitLab (https://gitlab.com/WiDGeT_TrentU/PoolSeq)

Author Contributions

S.J.A., S.D.C., and A.B.A.S. designed the study; J.H.R., and S.J.A. coordinated data collection and sample curation; S.J.A., performed research and analyzed data; S.J.A. wrote the manuscript with input from J.H.R., S.D.C., and A.B.A.S.

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

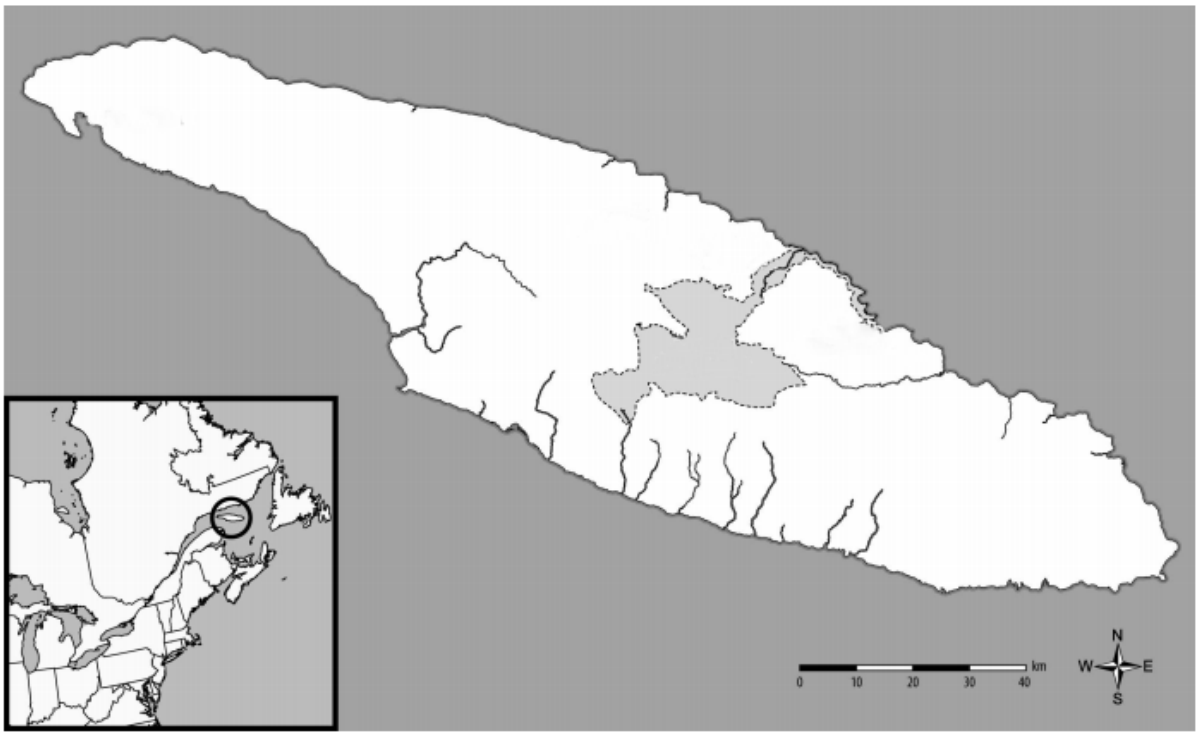

Figure 1. A map of Anticosti Island, Quebec where all white-tailed deer samples were acquired between 2002-2014. The grey area reflects Anticosti National Park where no hunting is permitted. The inset represents the larger North American landscape.

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

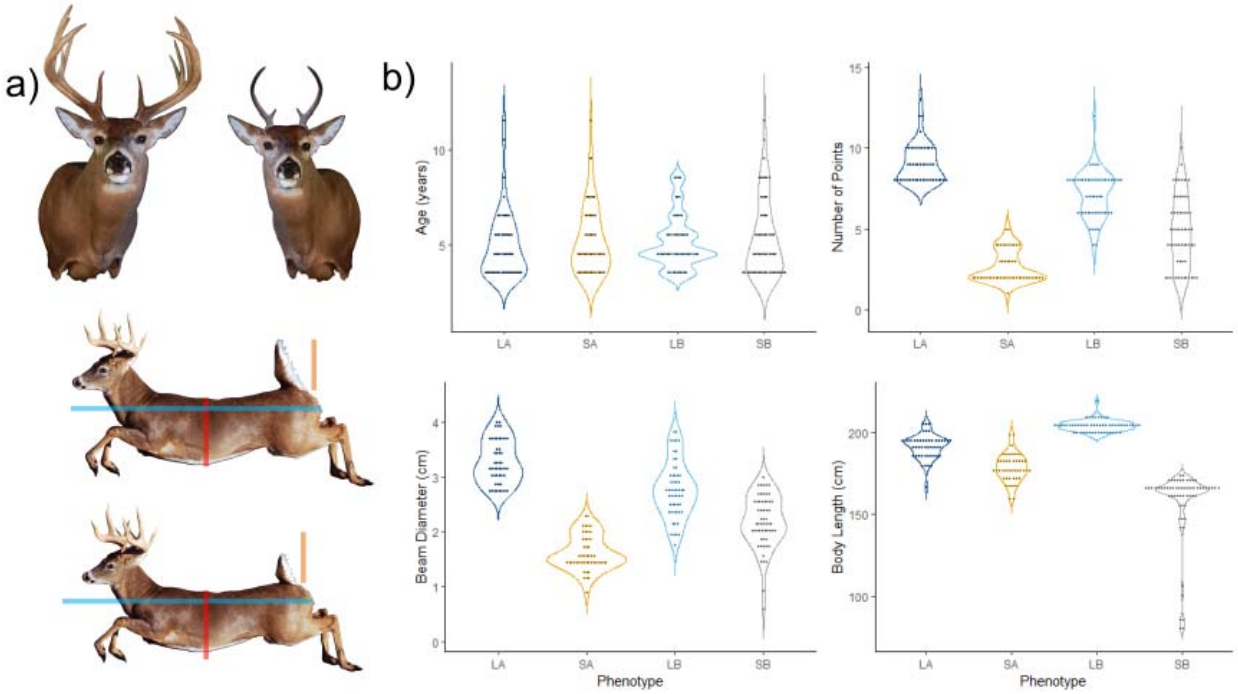

Figure 2. The phenotypic extremes used in our sampling methodology. a) Artist renderings of the average phenotypic measures for all individuals included in each pool. Measurements for antler pools (top) included mainbeam diameter and number of points, while body size (bottom) included body length. Scale bars are present for body size to better show differences in average body length, tail length, and chest circumference between large and small phenotypes. b) Violin plots displaying the original measurements of each individual, grouped by large antler (LA), small antler (SA), large body (LB), and small body (SB) pools. The phenotypes measured include age (years), number of points, beam diameter (cm), and body length (cm).

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

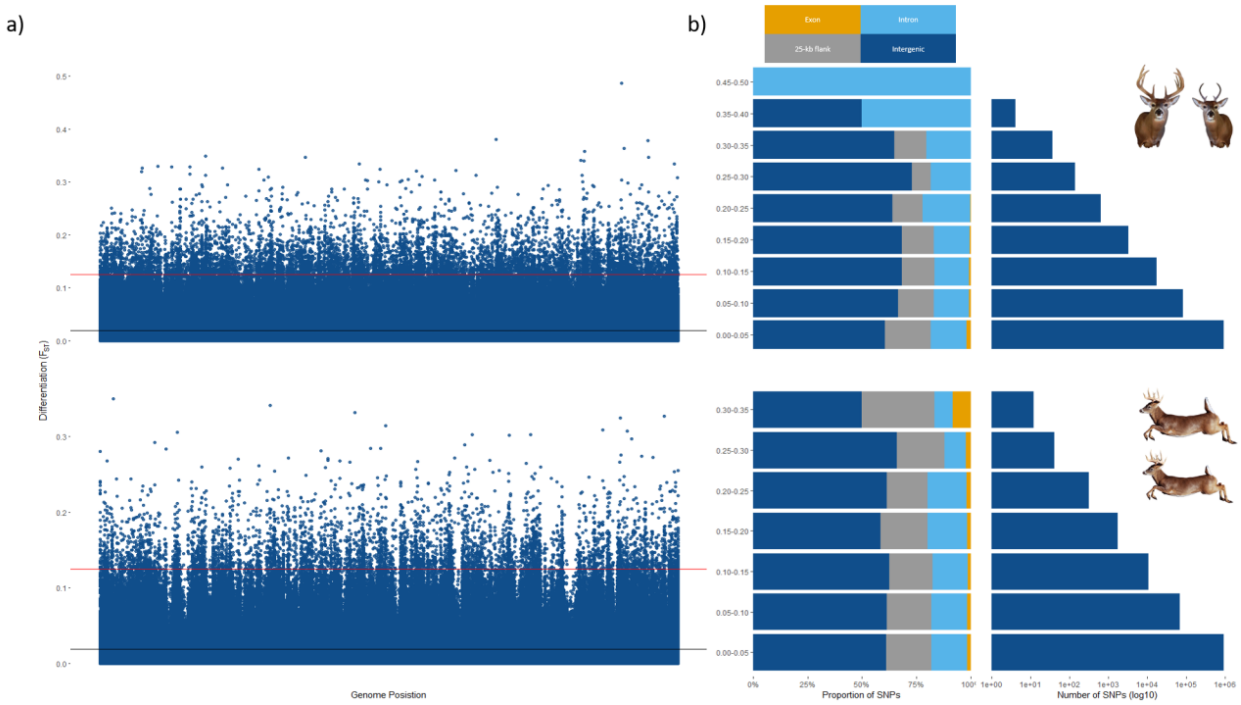

Figure 3. a) Manhattan plot representing pairwise genetic differentiation (F_{ST}) across the white-tailed deer genome between extreme phenotypes for antler (top) and body (bottom) traits. The horizontal red lines represent the 99th percentile F_{ST} and the horizontal black lines represent genome-wide mean F_{ST}. **b)** Stacked barplots representing the proportion of SNPs in different genomic regions (exon, intron, 25-kb flank, or intergenic). Each bin is grouped by F_{ST} and corresponds with the y-axis values from the Manhattan plots. Coordinating barplots represent the number of SNPs in each bin on a log scale.

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

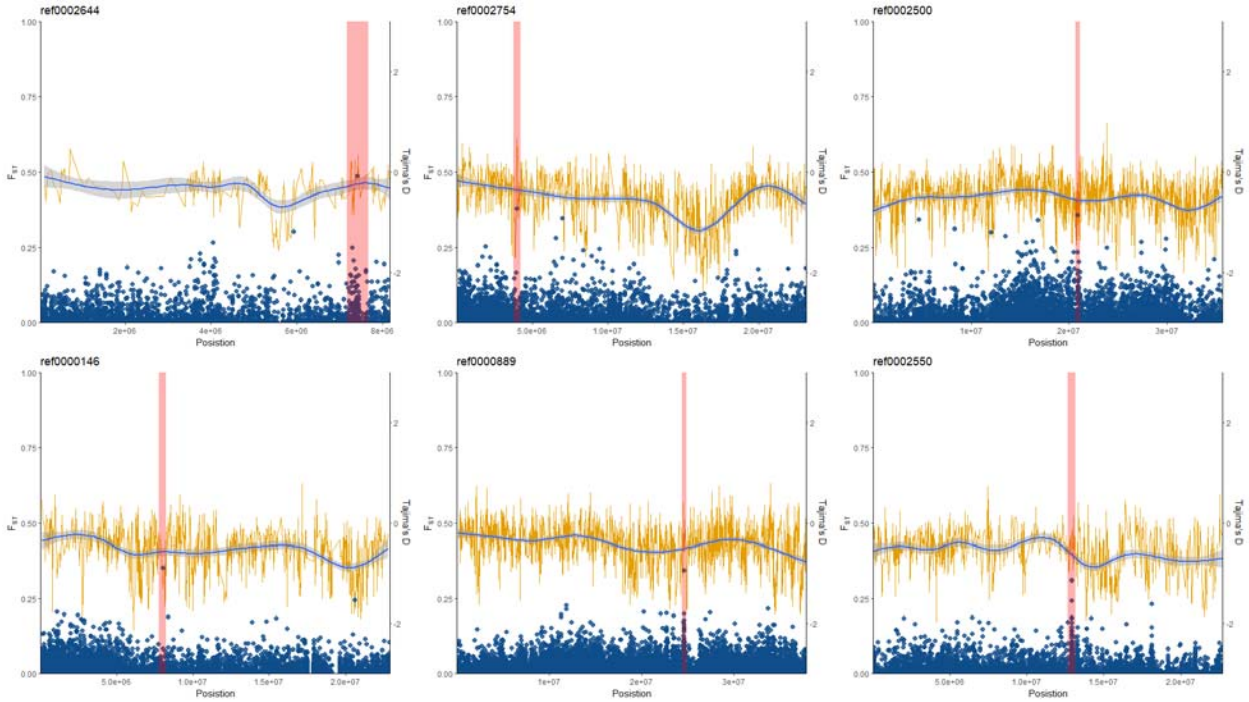

Figure 1. Manhattan plot representing pairwise genetic differentiation (F_{ST}) for scaffolds of interest (left y-axis), and Tajima's *D* values (right y-axis) in 10-kb sliding windows with a smoothed blue line function of these values (generalized additive model). Highlighted red regions represent a 50-kb window surrounding the target SNP of interest and their associated gene regions; the top panel representing RIMS1(ref0002644), PTEN (ref0002574), and SRP54 (ref0002500) from the comparison of antler phenotype pools and the bottom panel representing LRIF1 (ref0000146), ABCB1 (ref000889), and MYADM (ref0002550) from the comparison of body phenotype pools.

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

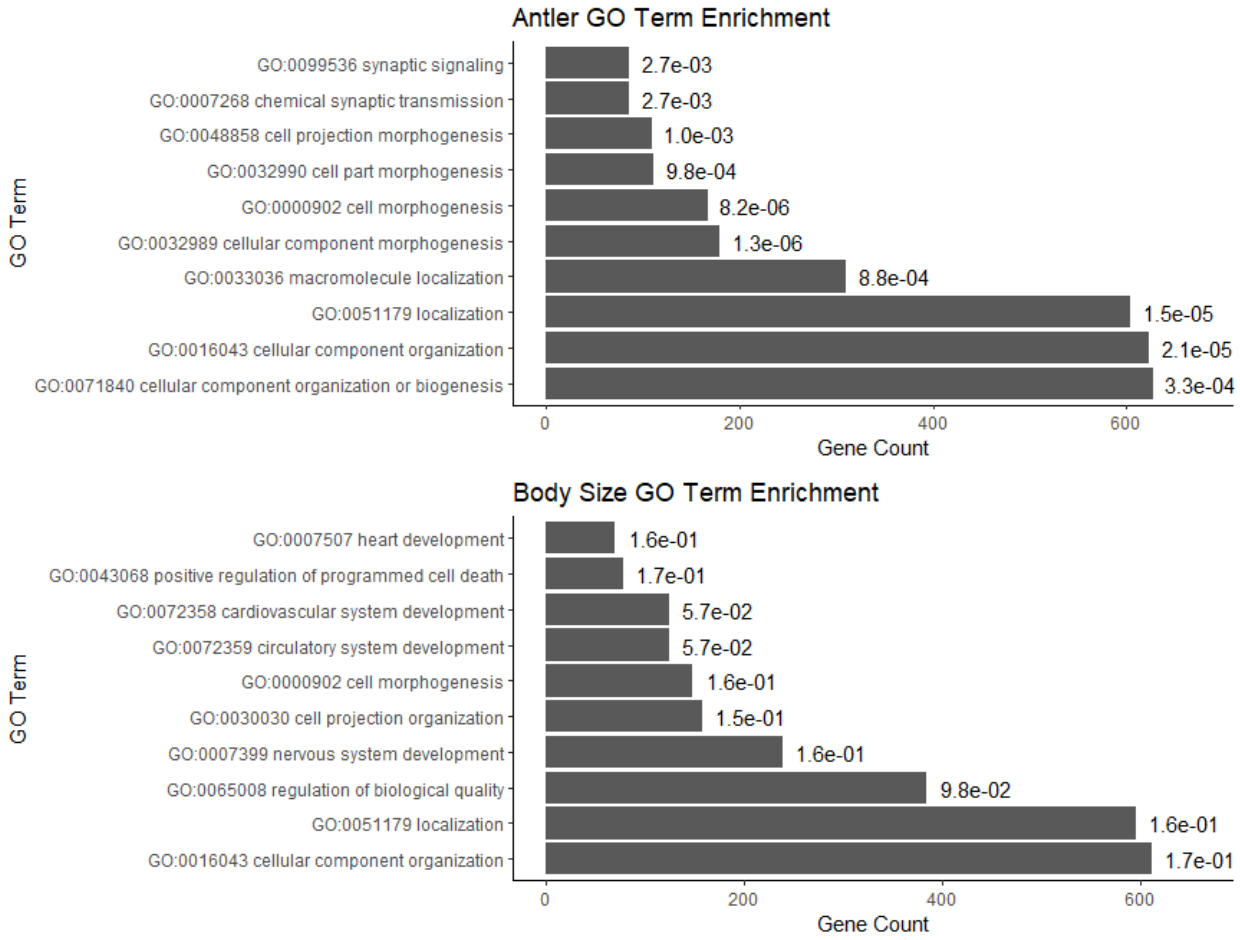

Figure 5. Histograms of the top ten enriched GO terms identified through the comparison of outlier genes in the 99th percentile for antler (top) and body size (bottom). Bars represent the number of genes included in each GO category and are labeled with the benjamini corrected p-values for enrichment.

bioRxiv preprint doi: https://doi.org/10.1101/841528; this version posted November 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. The breakdown of total SNPs identified through genome wide association scans for antler and body phenotypes compared to SNPs in the 99th percentile, categorized by gene region. Percent total provided in parentheses.

	Antler		Body	
	Total	**Outliers**	**Total**	**Outliers**
Intergenic	510,632 (*61.16%*)	5,660 (*67.78%*)	621,747 (*61.19%*)	6,313 (*62.12%*)
25-kb-flank	171,471 (*20.54%*)	1,225 (*14.67%*)	208,292 (*20.50%*)	2,018 (*19.86%*)
Intron	137,201 (*16.43%*)	1,399 (*16.75%*)	167,081 (*16.44%*)	1,666 (*16.39%*)
Exon	15,551 (*1.86%*)	66 (*0.79%*)	18,985 (*1.87%*)	165 (*1.62%*)