

RESEARCH

Open Access



Genomic architecture of phenotypic extremes in a wild cervid

S. J. Anderson^{1*}, S. D. Côté², J. H. Richard² and A. B. A. Shafer^{1,3*}

Abstract

Identifying the genes underlying fitness-related traits such as body size and male ornamentation can provide tools for conservation and management and are often subject to various selective pressures. Here we performed high-depth whole genome re-sequencing of pools of individuals representing the phenotypic extremes for antler and body size in white-tailed deer (*Odocoileus virginianus*). Samples were selected from a tissue repository 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 after controlling for age. Our results revealed a largely homogenous population but detected highly divergent windows between pools for both traits, with the mean allele frequency difference of 14% for and 13% for antler and body SNPs in outlier windows, respectively. Genes in outlier antler windows were enriched for pathways associated with cell death and protein metabolism and some of the most differentiated windows included genes associated with oncogenic pathways and reproduction, processes consistent with antler evolution and growth. Genes associated with body size were more nuanced, suggestive of a highly complex trait. Overall, this study revealed the complex genomic make-up of both antler morphology and body size in free-ranging white-tailed deer and identified target loci for additional analyses.

Keywords: Quantitative traits, Population genetics, Gene enrichment, Pool-seq, Extreme phenotypes

Background

Characterizing the genomic architecture underlying phenotypes in natural populations provides insights into the evolution of quantitative traits [1]. Some quantitative traits are correlated with metrics of fitness and are particularly important because they might directly influence population viability [2]. This relationship between genomic architecture and quantitative traits, however, is not easy to empirically identify [3], and often has unclear and unpredictable responses to selection [4]. Sample size in particular, of both the number of sequenced individuals and assayed SNPs, greatly limits the power of genome scans [5]. As such, alternative sequencing and sampling

strategies have emerged to identify the genetic basis of traits in natural populations.

One approach that has gained traction seeks to sample individuals representing the extreme ends of the spectrum for any observable phenotype, instead of randomly sampling individuals from the entire distribution (e.g. [5–10]). This sampling methodology of so-called extreme phenotypes aims to maximize the additive genetic variance for the sampled trait, increasing the power to detect quantitative trait loci or QTL [5]. Simulations have showed that sampling the whole genome increases power [11], but this varies by the effect size of each SNP [5]. Population history (i.e. degree of linkage disequilibrium) also has a profound effect on identifying QTL, essentially making the ability to identify linked or causative loci easier in small (isolated) populations, and challenging in large populations with high effective population size (N_e), and by proxy high recombination rates (see also [12]).

*Correspondence: sanderson@trentu.ca; aaronshafer@trentu.ca

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

³ Forensics Department, Trent University, Peterborough K9J 7B8, Canada
Full list of author information is available at the end of the article



In contrast to individual genome re-sequencing, pooled sequencing (pool-seq) uses DNA from many individuals within a population and is a cost-effective alternative to the whole genome sequencing of each individual separately [13]. 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 QTL mapping and association studies [14]. Pool-seq methods have successfully identified genes underlying colour morphs in butterflies and birds, abdominal pigmentation in *Drosophila*, horn size of wild Rocky Mountain bighorn sheep, and growth and reproduction in salmon [15–18]. Depending on the quality of annotation, similar pooled approaches can characterize differences in transposable element insertions [19]. The latter aspect, broadly speaking, is particularly relevant for genotype-phenotype correlations as most causative SNPs occur outside coding regions [3, 20], and there is evidence of TE insertions causing phenotypic differences [21] that are adaptive [22].

Combining pool-seq and sampling of phenotypic extremes seems promising, if not only for financial considerations [23], in studies interested in the genetic basis of traits in wild populations. Coalescent simulations by Hivert et al. [24] showed high accuracy of F_{ST} estimates from pool-seq, at times more precise than individual genomes. Similarly, Inbar et al. [25] showed that pooled approaches produced near identical allele frequency distributions to that of whole and reduced genome sequencing, with simulations reliably detecting moderate to large effect QTL. While Bastide et al. [26] showed that low frequency alleles and small effect sizes, particularly for complex traits, were difficult to detect, this is also the case for non-pooled studies (e.g. Caballero et al. [11]). Recent pool-seq work by Mohamed et al. [23] and Michelletti and Narum [17] identified the same associated regions for traits previously identified with extensive SNP arrays [27] and pedigree-based QTL mapping [28], respectively. Even when effect sizes are small, gene set and pathway analyses can be used to detect the collective effect of SNPs underlying the traits [29, 30], and provide complementary tools to detect genetic associations to a given phenotype [31].

Here, we explored the genomic basis for phenotypic variation by sampling the extreme distribution ends of a phenotype in a non-model big-game species, the white-tailed deer (*Odocoileus virginianus*, WTD). Two traits are of particular interest in WTD; body size and antler size, which have a degree of observable variation [32] and are connected to individual reproductive success [33–35]. Heritability for antler and body measurements are moderate to high for antler features and body

size [36–38]. Large antlers and body size are also sought after by hunters, both as trophies and food throughout North America. Our objective was to identify genomic windows and TE insertions associated with variation in antler and body size phenotypes in WTD; we assume a polygenic trait (e.g. [39]), but due to extreme phenotypic sampling there should be detectable outlier windows and gene pathways related to these phenotypes. This to our knowledge the first application of pooled-GWAS study of extreme phenotypes in a free-ranging population.

Methods

Study area

Anticosti Island (49°N, 62°W; 7,943 km²) is located in the Gulf of St. Lawrence, Québec (Canada) at the north-eastern limit of the white-tailed deer range (Fig. S1). 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 [40, 41]). Approximately 220 white-tailed deer were introduced between 1896 and 1897 and the population has an estimated contemporary N_e of ~1,500 [42]: the island's demographic history might have resulted in increased additive variance [43]. There is no population subdivision [42] and the mean relatedness of our database is zero [36]. Any related individuals in the data set should manifest in genome-wide patterns as opposed divergent windows between pools. Quantitative genetic analyses showed no effect of sampling year on trait heritability, which was interpreted as a uniform environmental effect across the island [36].

Sample Collection and Phenotypic Distribution

No live animals were directly involved in this study, rather we collected tissue samples stored in 95% ethanol and phenotypic data on 4,466 male deer harvested by hunters from September to early December, 2002–2014 on Anticosti Island. The sample and measurements were completed as part of a large effort led by the NSERC Industrial Research Chair in integrated resource management of Anticosti Island focussed on collecting data on the body condition of deer on the island. We used cementum layers in incisor teeth to age individuals [44]. 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) [45]. We used one metric for body size: body length which is correlated to other metrics such as body weight and hind foot length [46]. Because antler and body size of male cervids are correlated to age [47, 48], we used linear models to first assess the relationship between age and each metric separately. We then 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 ranked individuals from each group were selected from the available database for DNA extraction, with only non-overlapping samples for large and small phenotypes from the same year, were included in the final pool to limit temporal variation. This created four groups for pooled-sequencing: 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 (Life Technologies, Carlsbad, CA, USA). We aimed to sequence each pool to 50X as per recommendations of Schlötterer et al. [14]. 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. Sequencing was conducted at The Centre for Applied Genomics (Toronto, ON, Canada) on an Illumina HiSeqX with 150 bp pair-end reads (Table S2).

Genome Annotation

The draft WTD genome constructed from long (PacBio) and short-read (Illumina) data was used as a reference (NCBI PRJNA420098; Accession No. JAAVWD000000000). We performed a full genome annotation by masking repetitive elements throughout the genome using a custom WTD database developed through repeat modeler v1.0.11 [49] in conjunction with repeatmasker v4.0.7 [50] using the non-masked NCBI repeat database for *artiodactyla*, without masking low complexity regions. The masked genome was annotated using the MAKER2 v2.31.9 pipeline [51]. We used a three-stage process [52] for the generation of an initial training data set; 1) using publicly available white-tailed deer EST and Protein sequences available through NCBI an initial GFF file was generated with SNAP v2013-11-29 [53]; 2) an hmm file was generated from the initial SNAP training GFF output and was used as evidence for the MAKER2 prediction software, again using SNAP; and lastly 3) evidence from the prior SNAP trials in GFF format were used as for the generation of a training data set for AUGUSTUS v3.3.2 gene prediction software. Gene IDs were generated using blastp v2.9.0 [54] on the WTD annotation protein transcripts, restricting the blast search to human protein annotations in the uniprot database (parameters -max_hsps 1 -max_target_seqs 1 -outfmt 6 -taxids 9606).

Mapping and SNP calling

We trimmed reads for quality and adaptors using the default Trimmomatic v.0.36 settings [55]. Reads were then aligned to the unmasked WTD reference genome with BWA-mem v0.7.17 [56]. We used samtools v1.10 [57, 58] to merge and sort all aligned reads into four files for each representative pool. Prior to calling SNPs we filtered for duplicates using Picard v2.20.6 [59], kept uniquely mapped reads with samtools, and conducted local realignment using GATKv3.8 [60]. We called SNPs with samtools mpileup (minimum mapping quality (-q) = 20). INDELs were identified using a perl script in the Popoolation2 software suite [61, 62], and all INDELs as well as SNPs within 5 kb up/downstream of these regions were removed. We filtered out all masked regions and all scaffolds ≤ 50 kb.

Genome wide differences were calculated between large and small phenotypes for antler and body size. We used a sliding window approach with window sizes of 1000 bp with a step size of 500 bp and minimum covered fraction of 0.8 to test for allele frequency differences using Fisher's exact test (FET) and estimate the fixation index (F_{ST}) from the Popoolation2 software suite [61, 62]. A minimum and maximum coverage were determined based on the mode depth \pm half the mode as per Kurland et al. [63] and a minimum overall count of the minor allele of 3 for each pool was specified (Fig. S2). We calculated the mean allele frequency differences of each window in R v3.6.1. We first corrected for multiple testing by using an FDR correction (see [64]) but also set a conservative α ($1.0e-7$) as Popoolation multiples all p -values within each window by default. We ran all subsequent analyses on the FET windows, noting here a correlation to F_{ST} (Fig. S3). Manhattan plots were generated using a custom R script to plot the distribution of outliers by position and scaffold throughout the entire WTD genome. Bedtools v2.27.1 was used to characterize whether a window was within 25 kb up/downstream of a gene (i.e. regulatory) which were used for gene set analysis (below). To identify the most differentiated SNPs within each outlier region we used the modified chi-square test from Spitzer et al. [65] that accounts for overdispersion. This multi-pronged approach to identify outlier windows and SNPs was used to limit the potential for false positives.

Lastly, we generated two null populations to quantify the rate of false positives. Here, we combined all the reads from SA, SB, LA, and LB into one large pool using samtools -merge; we then generated two artificial pools by randomly sub setting reads using the samtools view -s flag that matched the coverage of our original pools. The same window-based analyses were conducted assuming any outlier would represent background noise. Note

that pool-seq uses read counts to estimate allele frequencies, hence the need for high coverage; sampling variance is expected to generate some differences in minimally diverged populations as for example 50X coverage with a minor allele count of 3 and 4 results in a 2% allelic frequency difference.

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 [66]. 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. Trimmed reads for both antler and body size phenotypes were aligned to the TE-WTD merged reference genome using *bwa-mem* as stated in previous methods for SNP analysis. We used the recommended workflow for the software suite *popoolationTE2* v1.10.04 [19] 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 within the 95th percentile for absolute frequency differences to allow for more features to be assessed subsequently. The same analyses were performed on our null pools to generate the null TE frequency distribution.

GO Pathways

To evaluate the enrichment of gene pathways in our outlier windows the program *Gowinda* v1.12 [67] was used to test for GO term enrichment, while also accounting for biases in gene length but requires input of individual SNPs. We also used TE start position as a proxy for SNP location for program compatibility. A *gtf* version of our annotation was created by removing duplicate genes and retaining only the longest transcript – resulting in 15,395 unique genes. All SNPs in outlier windows that were within 25 kb of a gene were included in this analysis and compared to all SNPs in every outlier window for antler and body size phenotypes independently. Similarly, all TEs that were within 25 kb up or downstream from a gene and within the 95th percentile for absolute frequency differences were included in a separate analysis for antler and body phenotypes independently. We used the program *REVIGO* [68] to remove redundant GO terms and to visualize semantic similarity-based scatterplots: the latter turns GO biological descriptions into numerical values that allow for aggregating similar terms. Results from

Gowinda with p -values < 0.01 were used with *REVIGO* to generate plots of significant GO terms for biological processes of antler and body size phenotypes. Only GO terms with a dispensability score < 0.20 were included in figures, representing the least redundant terms in the analysis.

We also created outlier gene ID lists (i.e. genes within 25 kb of an outlier TE and window) for antler and body size that were provided independently to *DAVID* v6.8 [69], 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. Results from *DAVID* with a p -value < 0.01 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.

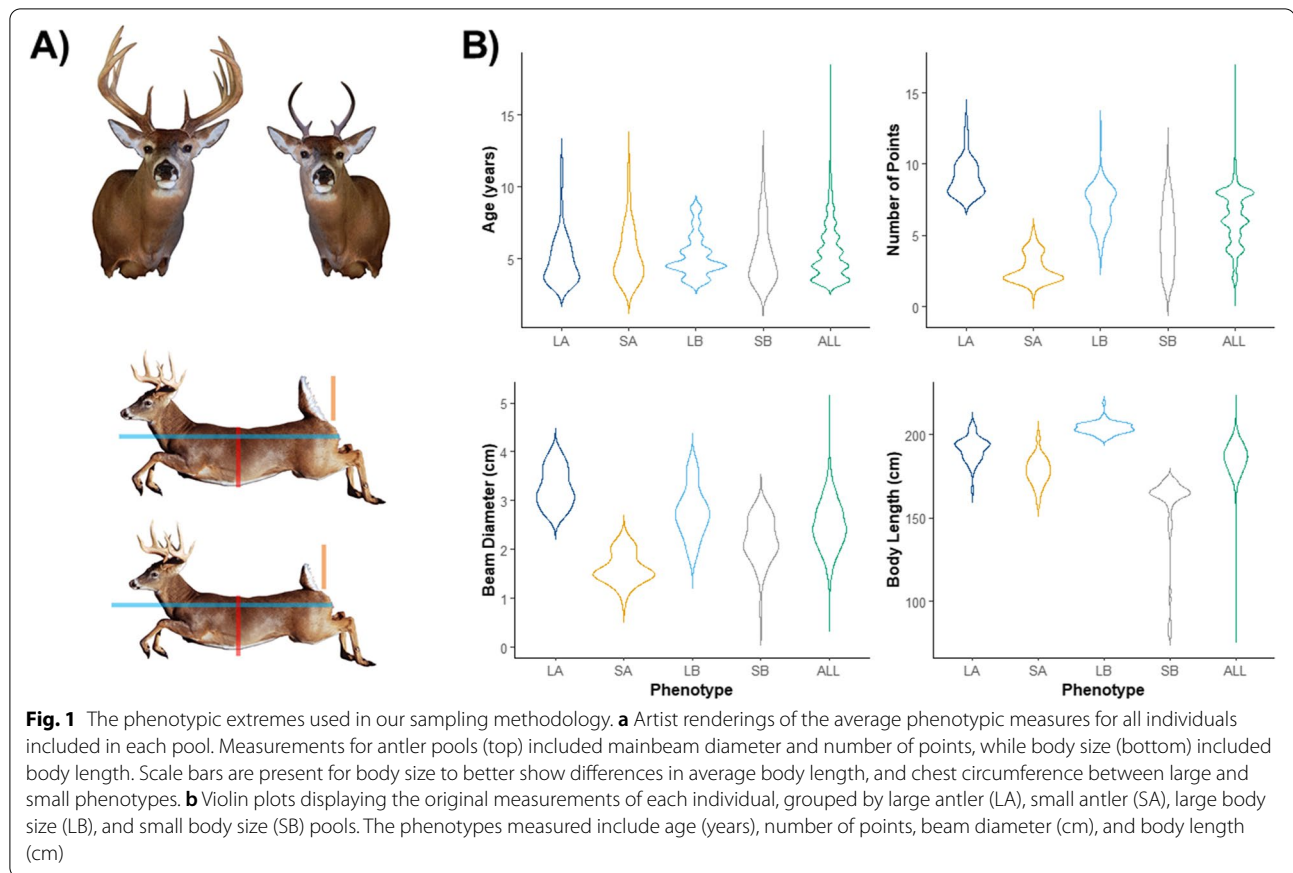
Validation of outliers

Pool-seq SNP callers can vary in their allele frequency estimates [70]; we therefore selected three candidate SNPs for qPCR validation of allele frequencies. The SNPs were selected based off significant FET and chi-square values, were near or inside a gene, and passed quality control with respect to primer design (Fig. S4 & S5). Custom genotyping assays using *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 S3. 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) to calculate the effect size.

Results

Phenotypes

We selected the top 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 Fig. 1a. There were no significant differences in mean age between the pools (Fig. 1b) and individual rankings between antler and body size were weakly correlated (Pearson $r = 0.33$, $p < 0.01$).



Detecting genetic variants and their genomic regions

The total reads generated from resequencing are observed in Table S 2 (BioProject ID PRJNA576136). The WTD genome is 2.52 GB and has a N50 of 17 MB, 727 scaffolds >50 kb (98.85% of genome), and BUSCO completeness of 91% across 303 orthologs. The genome annotation resulted in 20,750 genes and we identified 29,484,233 and 29,703,634 SNPs from the antler and body analyses respectively across the genome, with 25,116,953 shared SNPs (Fig. 2). The mean allele frequency difference across the antler outlier windows were calculated to be 14% with 5038 windows meeting the FDR and α threshold ($p \leq 1 \times 10^{-7}$); many of the top windows were adjacent to genes with known functions (Fig. 3; Tables 1 & 2; full gene list in Tables S4 & S5). For the comparison of body size, the mean allele frequency difference was 13% in 1301 windows meeting the FDR threshold. 247 windows overlapped between phenotypes. For all individual SNPs within outlier windows the modified chi-square test with FDR correction identified 5,184 and 1,166 divergent SNPs ($p < 0.01$) from the antler and body analysis, respectively. Overall genome-wide allele frequency differences in both phenotype pools were 8%; in contrast, the null pool

comparisons had 3% difference on average and only 5 windows met the FDR cut-off with a mean allele frequency difference of 5%.

TE Insertions

We identified 19,160 TE insertions through the joint analysis of antler phenotype sequences (both large and small). Of these TEs, we identified 950 insertions that fell in the 95th percentile for the absolute difference between large and small phenotypes for further analysis ($>|0.20|\%$; Fig. S6). Of the 95th percentile TE insertions in the antler analysis, 92 were found to overlap with genes, with 275 within a 25 kb window up or downstream of genic regions. For body size, 6,610 TE insertions were identified with 320 insertions in the 95th percentile ($>0.21\%$). Of these, 28 overlapped with genes, and 104 within 25 kb of a genic region. The null pool TE comparisons had considerably shifted frequency distribution, with the 95th percentile being TEs that differed by $>0.11\%$.

Gene Ontology Annotations

The results from Gowinda comparing all SNPs within outlier windows to background SNPs from all windows showed the top 10 enriched GO terms and gene counts

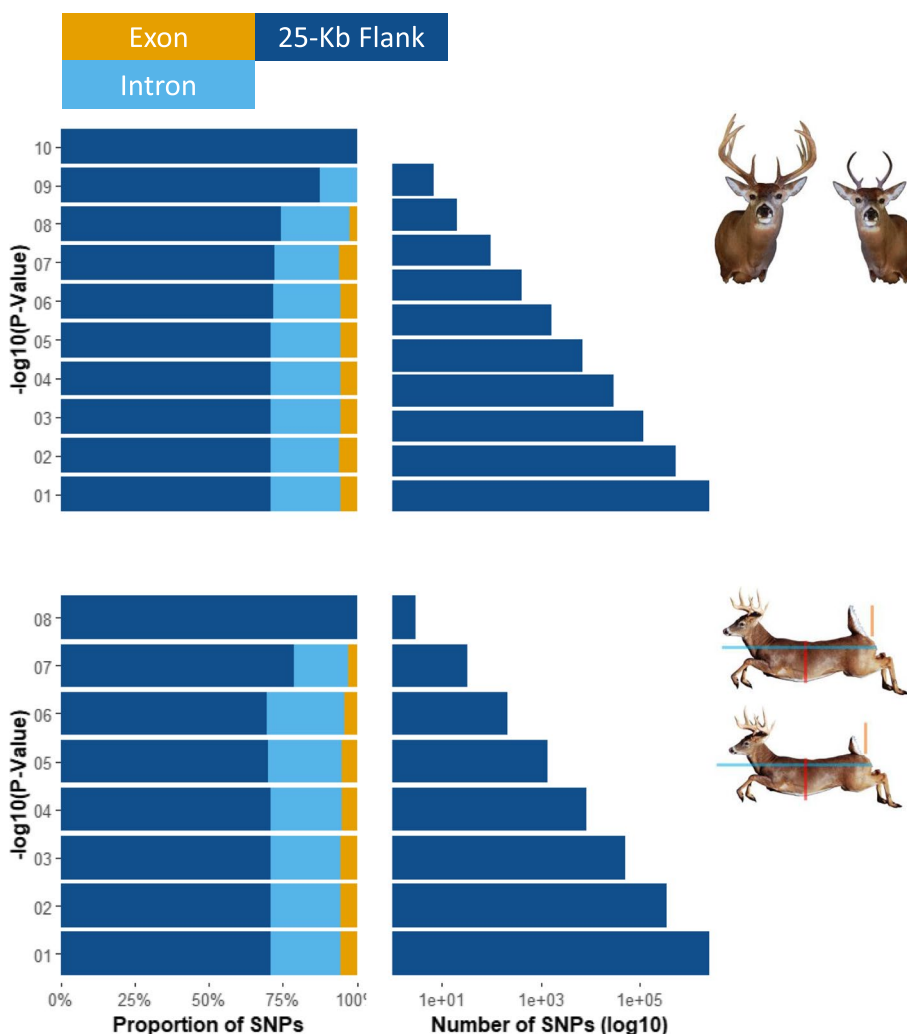


Fig. 2 Stacked barplots representing the proportion of SNPs from all windows in different genomic regions (exon, intron, or 25 kb flank). Each bin is grouped by $-\log_{10}(p\text{-values})$ calculated by Fisher’s Exact Test, and corresponds with the y-axis values from the Manhattan plots. Coordinating barplots represent the number of SNPs in each bin is presented on a log scale

(Fig. S7). Six statistically enriched GO terms (FDR < 0.01) from the antler analysis were: GO:0004888 (transmembrane signaling receptor activity), GO:0004872 (receptor activity), GO:0038023 (signaling receptor activity), GO:0004930 (G-protein coupled receptor activity), GO:0022835 (transmitter-gated channel activity), GO:0022824 (transmitter-gated ion channel activity). There were no significant GO terms identified through the body analysis following FDR corrections. We show term reductions for significantly enriched GO terms ($p\text{-value} < 0.01$) through the program REVIGO, with clustered items relating to the semantic similarity in the antler pool (Fig. 4 and Fig. S7) and body pools (Fig. S8). There were no significant GO terms identified through

the TE analysis using Gowinda and REVIGO following FDR corrections (FDR < 0.01) for both the antler and body analysis.

DAVID pathway analysis identified 6 statistically significant GO terms (FDR < 0.05); GO:0006614 (SRP-dependent cotranslational protein targeting to membrane), GO:0000184 (nuclear-transcribed mRNA catabolic process, nonsense-mediated decay), GO:0019083 (viral transcription), GO:0006413 (translational initiation), GO:0055085 (transmembrane transport), and GO:0007586 (digestion). No significant terms were identified through the body analysis. There were insufficient outliers ($n = 5$) with adjacent genes ($n = 2$) to conduct these analyses on the null pools.

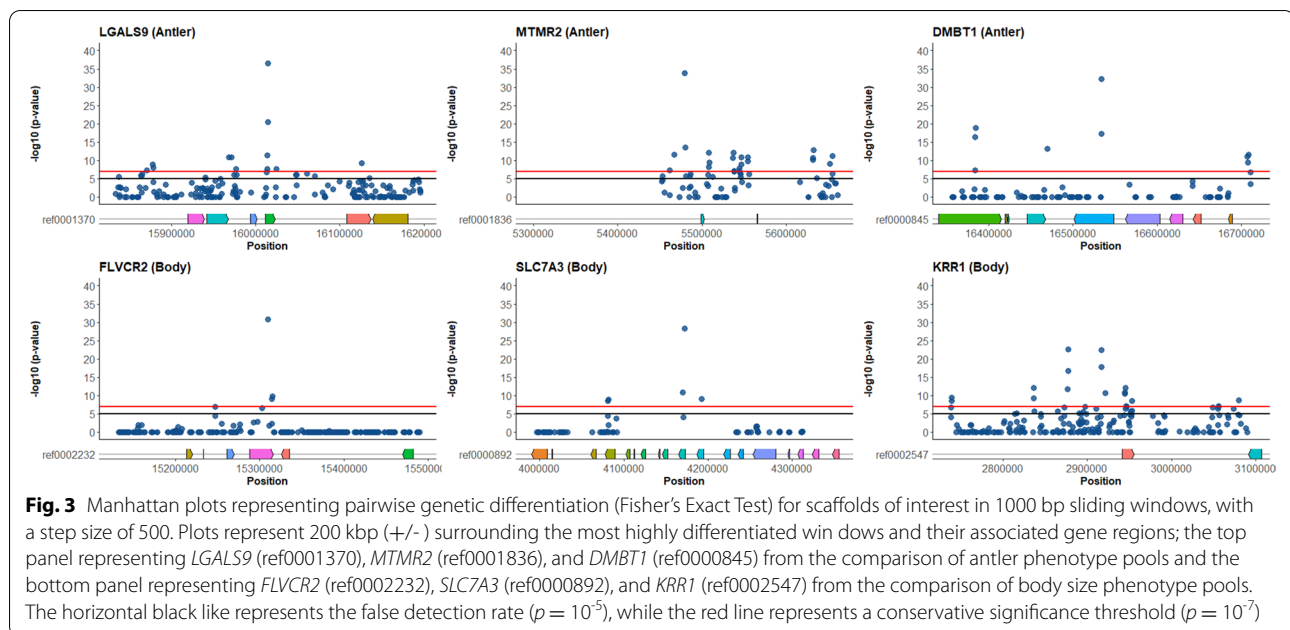


Table 1 The top 10 outlier window positions from the antler trait analysis based on corrected p -values that were within 25 kb of genes or regulatory regions up/downstream. The closest gene is listed for each window. Functions of each gene are abbreviated from RefSeq/Uniprot

Scaffold	Window Position	p	Gene	Function
ref0001370	16015000	2.66E-37	LGALS9	The galectins are a family of beta-galactoside-binding proteins implicated in modulating cell-cell and cell-matrix interactions.
ref0001836	5479500	1.24E-34	MTMR2	This gene is a member of the myotubularin family of phosphoinositide lipid phosphatases.
ref0000845	16534000	5.97E-33	DMBT1	Considered as a tumor suppressor gene. May play a role in epithelial differentiation.
ref0002391	7908000	1.75E-29	ITLN2	ITLN2 (Intelectin 2) is likely involved include carbohydrate binding.
ref0001924	2123500	1.52E-28	TRIM64	TRIM proteins are involved in pathogen-recognition and by regulation of transcriptional pathways in host defence (Ozato et al., 2008).
ref0000845	16751500	1.14E-27	CD163	Involved in the clearance and endocytosis of hemoglobin/heptaglobin complexes by macrophages. May protect tissues from free hemoglobin-mediated oxidative damage.
ref0002391	7907500	1.79E-27	ITLN2	See above.
ref0000505	9079500	1.37E-26	ATP5F1A	This gene encodes a subunit of mitochondrial ATP synthase.
ref0000881	8059500	1.91E-26	MYH9	This gene encodes a conventional non-muscle myosin.
ref0002644	7312000	2.23E-26	RIMS1	The protein encoded by this gene is a RAS gene superfamily member that regulates synaptic vesicle exocytosis.

Validation of Outliers

We genotyped putative antler SNPs (*RIMS1*, and *SRP54*) and one body locus (*LIRF1*) split between pools which resulted in 80 and 70 individuals having complete genotype and phenotypic data, respectively, that were a subset of the pooled samples. The allele frequency differences between poolseq and qPCR were as follows: *RIMS1* (15% vs 6%), *SRP54* (16% vs 10%), and *LIRF1* (12% vs 16%). There was an effect of the number of antler alleles at the *SRP54* locus on antler category ($\beta = -0.99$, $p = 0.02$) and the *LIRF1* alleles had an effect on body size.

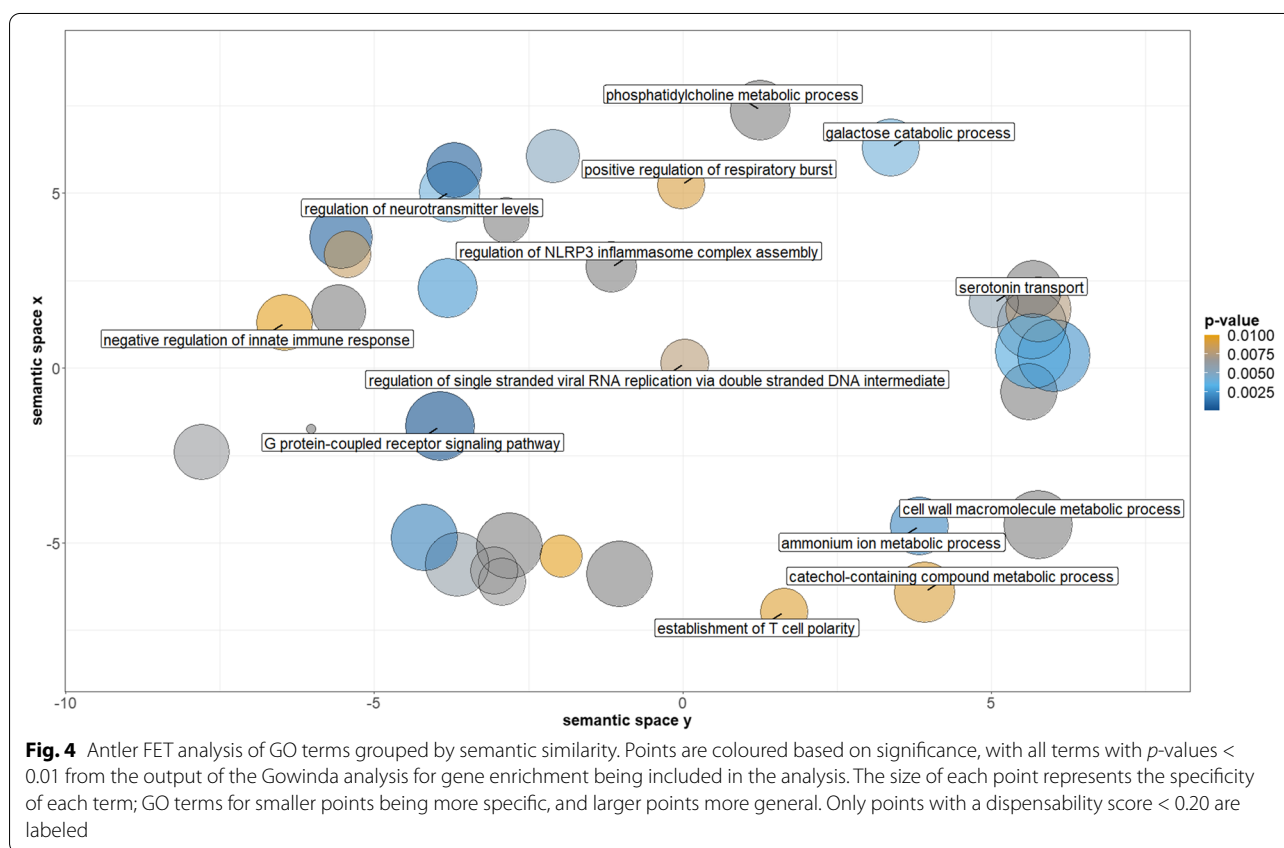
($\beta = -0.94$, $p = 0.03$; The *RIMS1* locus frequency differences were not shown to have an effect in the model ($p > 0.05$).

Discussion

The extensive database with phenotypic measurements for Anticosti Island white-tailed deer allowed us to select individuals from the full range of the distribution resulting in a sampling representative of extreme phenotypes (Fig. 1). The deer on the island form a panmictic population and have a smaller N_e [42] relative to contiguous

Table 2 The top 10 outlier window positions from the body trait analysis based on corrected *p*-values that were within 25 kb of genes or regulatory regions up/downstream. The closest gene is listed for each window. Functions of each gene are abbreviated from RefSeq/Uniprot

Scaffold	Window Position	<i>p</i>	Gene	Function
ref0002232	15310000	1.71E-31	FLVCR2	This gene encodes a transmembrane protein that is a calcium transporter.
ref0000892	4172500	4.82E-29	SLC7A3	SLC7A3 appears to be involved in amino acid transmembrane transporter activity and basic amino acid transmembrane transporter.
ref0002547	2917000	2.98E-23	KRR1	KRR1 (KRR1 Small Subunit Processome Component Homolog) appears to be involved in cytosol and Gene Expression.
ref0000845	16369000	1.26E-20	CD163	Involved in the clearance and endocytosis of hemoglobin/heptaglobin complexes by macrophages. May protect tissues from free hemoglobin-mediated oxidative damage.
ref0000438	13320500	1.64E-20	KIF3C	Microtubule-based anterograde translocator for membranous organelles.
ref0002813	16028000	8.08E-20	DOCK1	This gene encodes a member of the dedicator of cytokinesis protein family.
ref0002788	19947500	2.45E-19	SYK	This gene encodes a member of the family of non-receptor type Tyr protein kinase that is involved in coupling activated immunoreceptors to downstream signaling events that mediate diverse cellular responses, including proliferation, differentiation, and phagocytosis.
ref0000529	431000	4.31E-19	TRDV1	T cell receptors recognize foreign antigens which have been processed as small peptides and bound to major histocompatibility complex (MHC) molecules at the surface of antigen presenting cells (APC).
ref0002645	24851500	8.41E-19	RNASEH1	Endonuclease that specifically degrades the RNA of RNA-DNA hybrids
ref0002547	2916500	1.27E-18	KRR1	See above.



mainland populations [71]. Moreover, quantitative genetic assessment revealed a consistent environmental effect across traits [36]. Collectively, these factors suggest

a pooled-sequencing approach of phenotypic groups could identify key genes and pathways underlying these important traits.

We identified outlier regions for both traits that are atypically differentiated for what would be expected in this population and our null distribution model. The most divergent windows and TEs for antler and body traits are widely dispersed throughout the WTD genome. While effect sizes cannot be estimated from pooled samples, the sampling suggests we have maximized the additive variance of this population and these outlier windows are likely associated with the traits. The GO analysis, and more broadly a collection of small putative islands of divergence (Fig. 3), is consistent with a polygenic trait controlled by many genes: this interpretation is also supported by the literature on body size and ornaments in mammals [72–74], including red deer [39].

Genomic architecture of antlers and body size

Phenotypic association of SNPs has traditionally focussed on non-synonymous variants, but it is evident that non-coding regions impact phenotypic variation [20], often with clear relationships to promoters and enhancer regions [75–77]. While some outliers are surely false positives [78, 79], cattle GWAS for similar traits typically identify 100s to 1000s of QTL [80, 81]. Our data analysis – including the GO pathway analyses that incorporated TE insertions – suggest that it is the cumulative effects from these variants on biological pathways underlying antler and body trait variation in white-tailed deer.

Antlers are the only completely regenerable organ found in mammals [82], a unique process that involves simultaneous exploitation of oncogenic pathways and tumor suppressor genes, and the rapid recruitment of synaptic and blood vesicles [74]. Antlers might also be an honest signal of male quality [83]. A study of differential RNA expression in sika deer antlers identified numerous gene clusters, and we detected an overlap in gene pathways associated with cell death, cell wall formation, and protein metabolism (Fig. 4 [84]). One notable outlier in our study also detected by Ba et al. [84] is *IGF1R* that is responsible for antler mesenchymal cell proliferation and is a critical gene for antler development [85–87]. More broadly, the most divergent windows were often near genes linked to cancer and other biological processes related to development (Table 1), consistent with our understanding of antler growth and function. *LGALS9*, for example is a gene that produces Galactin-9, a well-studied protein expressed on tumour cells [88], and *DMBT1* is a tumour suppressor gene involved in bone cancer [89] that like *IGF1R* has been previously identified as key velvet antler peptide [74]. *MTMR2* (Myotubularin Related Protein 2) has been indicated as a candidate gene for litter size in pelibuey sheep [90] and plays a role in spermatogenesis [91]. Likewise, the qPCR validated *SRP54* gene has been

linked to bone marrow failure syndromes and skeletal abnormalities [92]. Thus, we report testable antler-genic links to bone formation, oncogenic pathways, and sexual selection detected from a pooled-GWAS approach.

Using analogous human tissue and cell expression, most of these reported genes appear to have low tissue specificity (*IGF1R*, *MTMR2*, *SRP54*) or are most highly expressed in adaptive immune tissues (*LGALS9*) or in the intestine (*DMBT1*). The specificity for the single cell type expression of these genes all fall into the categories of neuronal cells and glial cells (*IGF1R*, *MTMR2*), glandular epithelial cells (*LGALS9*, *DMBT1*), and germ cells (*MTMR2*, *SRP54*) [93].

The comparison of extreme body size phenotype sequences also revealed an array of genes of interest, but causal inferences are more challenging (see [94]) as traits of this nature, for example human height, involve thousands of SNPs and a multitude of biological pathways [95]. Despite the relatively high heritability of body size in this population [36], there were fewer windows meeting our threshold and no significant GO pathways. Similarly, Taye et al. [96] and Deng et al. [97] detected divergent genes, but no pathways associated with body size.

Considering a role for transposable elements

Traditionally, masking transposable elements reduces misalignments due to the repetitive nature of their sequences and current limitations with mapping software [98], 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 [21]. We found 275 highly divergent TE insertions in genic regions from our antler analysis and 104 from the body size analysis. The insertion of these highly variable sequences has the potential to impact gene function and studies are now starting to emerge that both validate insertions [99, 100] and show evidence for positive selection [61, 62]. Our GO pathway analysis that included TE insertions did identify pathways in antler, but not body size, and is an analytical consideration we wanted to highlight to the broader community. Validating the TEs and characterizing their influence on neighbouring genes is also warranted given the high frequency differences between pools.

Implications for studying quantitative traits in natural populations

Antler and body size are traits that are sexually selected and linked to reproductive success in WTD [35, 101];

they are also traits desired by hunters, managers, and farmers (e.g. antlers for velvet or large body for meat). We have detected highly divergent genomic regions, and thousands genome-wide variants with allele frequency differences between the phenotypic extremes for these traits. Additional lines of evidence, specifically RNA-seq studies (see [84]), should look at the regions identified here, but also consider interactions across the entire genomic, epigenomic, and transcriptional landscape.

We validated two out of three QTL allele frequencies via qPCR and showed an effect in a mixed model on phenotypes for WTD. While we applied the added stringency of accounting for overdispersion in the chi-square test, our inability to validate one outlier suggests a small effect size (e.g. [36]) and thus requires increased sample size (see [5]), or is a true type I error. Alternatively, epistatic interactions might also be at play [102], but detection requires whole-genome sequencing of many hundreds of individuals thereby defeating the purpose of pooled sequencing. Here, expanded individual genotyping might still reveal the predicted effect, and it is encouraging that approaches for low coverage pooled data are emerging [103].

The unrelated nature of males in our database [36] and null population analyses would support these patterns being inherent to the sampling strategy; however, replicating this approach in other populations would serve as further validation as we anticipate at least a portion of the outlier windows and genes to be shared across the species (e.g. [104]). More broadly, polygenic traits are difficult to characterize under ideal situations, and so the approach presented here is particularly promising for discrete traits, rare alleles and moderate-to-large effect loci (e.g. [17, 18, 25]). As we continue to identify and validate candidate QTLs, it is conceivable that a gene panel could be developed in white-tailed deer and assist in breeding and management programs or assess the effects of artificial selection by trophy hunting; however, until a reasonable amount of phenotypic variation can be attributed to specific QTL, a gene targeted approach for management and breeding of white-tailed deer will remain a challenge.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08333-x>.

Additional file 1.

Additional file 2.

Additional file 3.

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 assistants and technicians who collected samples and made phenotypic measurements throughout the duration of this project. We thank seven anonymous reviewers for their extensive comments and analytical suggestions on this manuscript.

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/graduate_theses). The raw qPCR genotype data table will be uploaded to Dryad.

Authors' contributions

S.J.A., S.D.C., and A.B.A.S. designed the study; S.D.C., 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. The author(s) read and approved the final manuscript.

Funding

This work was funded by the CFI-JELF (36905; A.B.A.S.), Compute Canada Resources for Research Groups (GME-665-01; A.B.A.S.), and NSERC Discovery Grants (A.B.A.S.: RGPIN-2017-03934).

Declarations

Ethics approval and consent to participate

Ethics approval was obtained for this study for the use of animal tissue through Trent University and Université Laval, and did not include the use of live animals.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Environmental and Life Sciences Graduate Program, Trent University, Peterborough K9J 7B8, Canada. ²Département de biologie, Centre d'études nordiques & Chaire de recherche industrielle CRSNG en aménagement intégré des ressources de l'île d'Anticosti, Université Laval, Québec, QC G1V 0A6, Canada. ³Forensics Department, Trent University, Peterborough K9J 7B8, Canada.

Received: 29 October 2021 Accepted: 24 January 2022

Published online: 12 February 2022

References

1. Stinchcombe JR, Hoekstra HE. Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity*. 2008;100(2):58–170 <https://doi.org/10.1038/sj.hdy.6800937>.
2. Kardos M, Shafer ABA. The Peril of Gene-Targeted Conservation. *Trends Ecol Evol*. 2018;33(11):827–39. <https://doi.org/10.1016/j.tree.2018.08.011>.
3. Barghi N, Hermisson J, Schlötterer C. Polygenic adaptation: a unifying framework to understand positive selection. *Nat Rev Genet*. 2020;1–13. <https://doi.org/10.1038/s41576-020-0250-z>.
4. Bünger L, Lewis RM, Rothschild MF, Blasco A, Renne U, Simm G. Relationships between quantitative and reproductive fitness traits in animals. *Philos Trans Royal Soc B Biol Sci*. 2005;360(1459):1489–502. <https://doi.org/10.1098/rstb.2005.1679>.

5. Kardos M, Husby A, McFarlane SE, Qvarnstrom A, Ellegren H. Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Mol Ecol Resour.* 2016;16(3):727–41. <https://doi.org/10.1111/1755-0998.12498>.
6. Barnett IJ, Lee S, Lin X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet Epidemiol.* 2013;37(2):142–51. <https://doi.org/10.1002/gepi.21699>.
7. Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat Genet.* 2012;44(8):886–9. <https://doi.org/10.1038/ng.2344>.
8. Gurwitz D, McLeod HL. Genome-wide studies in pharmacogenomics: harnessing the power of extreme phenotypes. *Pharmacogenomics.* 2013;14(4):337–9. <https://doi.org/10.2217/pgs.13.35>.
9. Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet Epidemiol.* 2011;35(8):790–9. <https://doi.org/10.1002/gepi.20628>.
10. Perez-Gracia JP, Ruiz-Ilundain MG, Garcia-Ribas I, Carrasco EM. The role of extreme phenotype selection studies in the identification of clinically relevant genotypes in cancer research. *Cancer.* 2002;95(7):1605–10. <https://doi.org/10.1002/cncr.10877>
11. Caballero A, Tenesa A, Keightley PD. The Nature of Genetic Variation for Complex Traits Revealed by GWAS and Regional Heritability Mapping Analyses. *Genetics.* 2015;201(4):1601–13. <https://doi.org/10.1534/genetics.115.177220>.
12. Schielzeth H, Husby A. Challenges and Prospects in Genome-Wide Quantitative Trait Loci Mapping of Standing Genetic Variation in Natural Populations. *Ann N Y Acad Sci.* 2014;1320:35–57.
13. Anand S, Mangano E, Barizzone N, Bordoni R, Sorosina M, Clarelli F, et al. Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering. *Sci Rep.* 2016;6(1):33735. <https://doi.org/10.1038/srep33735>.
14. Schlötterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nat Rev Genet.* 2014;15(11):749–63. <https://doi.org/10.1038/nrg3803>.
15. Endler L, Betancourt AJ, Nolte V, Schlötterer C. Reconciling differences in pool-GWAS between populations: A case study of female abdominal pigmentation in *Drosophila melanogaster*. *Genetics.* 2016;202(2):843–55. <https://doi.org/10.1534/genetics.115.183376>.
16. Kardos M, Luikart G, Bunch R, Dewey S, Edwards W, McWilliam S, et al. Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Mol Ecol.* 2015;24(22):5616–32. <https://doi.org/10.1111/mec.13415>.
17. Micheletti SJ, Narum SR. Utility of pooled sequencing for association mapping in nonmodel organisms. *Mol Ecol Resour.* 2018;18(4):825–37. <https://doi.org/10.1111/1755-0998.12784>.
18. Neethiraj R, Hornett EA, Hill JA, Wheat CW. Investigating the genomic basis of discrete phenotypes using a Pool-Seq-only approach: New insights into the genetics underlying colour variation in diverse taxa. *Mol Ecol.* 2017;26(19):4990–5002. <https://doi.org/10.1111/mec.14205>.
19. Kofler R, Gómez-Sánchez D, Schlötterer C. PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. *Mol Biol Evol.* 2016;33(10):2759–64. <https://doi.org/10.1093/molbev/msw137>.
20. Watanabe K, Stringer S, Frei O, Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet.* 2019;51(9):1339–48. <https://doi.org/10.1038/s41588-019-0481-0>.
21. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammeill M, et al. Ten things you should know about transposable elements. *Genome Biol.* 2018;19(1):199. <https://doi.org/10.1186/s13059-018-1577-z>.
22. Adrion JR, Begun DJ, Hahn MW. Patterns of transposable element variation and clinality in *Drosophila*. *Mol Ecol.* 2019;28(6):1523–36. <https://doi.org/10.1111/mec.14961>.
23. Mohamed A, Porto-Neto L, Reverter A, Kijas J. Evaluation of pooled whole genome sequencing (Pool-Seq) to recover known GWAS signals (gene effects). Proceedings of the 22nd Conference of the Association for the Advancement of Animal Breeding and Genetics (AAABG), Townsville, Queensland, Australia, 2–5 July 2017; 2017. p. 453–6.
24. Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R. Measuring genetic differentiation from pool-seq data. *Genetics.* 2018;210(1):315–30. <https://doi.org/10.1534/genetics.118.300900>.
25. Inbar S, Cohen P, Yahav T, Privman E. Comparative study of population genomic approaches for mapping colony-level traits. *PLoS Comput Biol.* 2020;16(3):e1007653. <https://doi.org/10.1371/journal.pcbi.1007653>.
26. Bastide H, Betancourt A, Nolte V, Tobler R, Stöbe P, Futschik A, Schlötterer C, Wittkopp P. A Genome-Wide Fine-Scale Map of Natural Pigmentation Variation in *Drosophila melanogaster*. *PLoS Genet.* 2013;9(6):e1003534. <https://doi.org/10.1371/journal.pgen.1003534>.
27. Porto-Neto LR, Reverter A, Prayaga KC, Chan EKF, Johnston DJ, Hawken RJ, et al. The Genetic Architecture of Climatic Adaptation of Tropical Cattle. *PLoS ONE.* 2014;9(11):e113284. <https://doi.org/10.1371/journal.pone.0113284>.
28. Moghadam HK, Poissant J, Fotherby H, Haidle L, Ferguson MM, Danzmann RG. Quantitative trait loci for body weight, condition factor and age at sexual maturation in Arctic charr (*Salvelinus alpinus*): Comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). *Mol Genet Genomics.* 2007;277(6):647–61. <https://doi.org/10.1007/s00438-007-0215-3>.
29. White MJ, Yaspan BL, Veatch OJ, Goddard P, Risse-Adams OS, Contreras MG. Strategies for Pathway Analysis Using GWAS and WGS Data. *Curr Protoc Hum Genet.* 2019;100(1):e79. <https://doi.org/10.1002/cphg.79>.
30. Holmans P. Statistical Methods for Pathway Analysis of Genome-Wide Data for Association with Complex Genetic Traits. *Adv Genet.* 2010;72. <https://doi.org/10.1016/B978-0-12-380862-2.00007-2>
31. Fridley BL, Biernacka JM. Gene set analysis of SNP data: Benefits, challenges, and future directions. *Eur J Hum Genet.* 2011. <https://doi.org/10.1038/ejhg.2011.57>.
32. Hewitt DG. Biology and management of white-tailed deer. CRC Press; 2011. Retrieved from <https://www.crcpress.com/Biology-and-Management-of-White-tailed-Deer/Hewitt/p/book/9781439806517>
33. DeYoung RW, Demarais S, Gee KL, Honeycutt RL, Hellickson MW, Gonzales RA. Molecular Evaluation of the White-tailed Deer (*Odocoileus virginianus*) Mating System. *J Mammal.* 2009;90(4):946–53. <https://doi.org/10.1644/08-MAMM-A-227.1>.
34. Jones PD, Strickland BK, Demarais S, Wang G, Dacus CM. Nutrition and ontogeny influence weapon development in a long-lived mammal. *Can J Zool.* 2018;96(9):955–62. <https://doi.org/10.1139/cjz-2017-0345>.
35. Newbolt CH, Acker PK, Neuman TJ, Hoffmann SJ, Ditchkoff SS, Steury TD. Factors Influencing Reproductive Success in Male White-Tailed Deer. *J Wildlife Manag.* 2016;81(2):206–17. <https://doi.org/10.1002/jwmg.21191>.
36. Jamieson A, Anderson SJ, Fuller J, Côté SD, Northrup JM, Shafer ABA. Heritability estimates of antler and body traits in white-tailed deer (*Odocoileus virginianus*) from genomic-relatedness matrices. *J Hered.* 2020. <https://doi.org/10.1093/jhered/esaa023>.
37. Michel ES, Demarais S, Strickland BK, Smith T, Dacus CM. Antler characteristics are highly heritable but influenced by maternal factors. *J Wildlife Manag.* 2016;80(8):1420–6. <https://doi.org/10.1002/jwmg.21138>.
38. Williams JD, Krueger WF, Harmel DH. Heritabilities for antler characteristics and body weight in yearling white-tailed deer. *Heredity.* 1994;73(1):78–83. <https://doi.org/10.1038/hdy.1994.101>.
39. Peters L, Huisman J, Kruuk LEB, Pemberton JM, Johnston SE. Genomic analysis reveals a polygenic architecture of antler morphology in wild red deer (*Cervus elaphus*). *Mol Ecol.* 2021;00:1–18. <https://doi.org/10.1111/MEC.16314>.
40. 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>
41. Simard MA, Coulson T, Gingras A, Côté SD. Influence of density and climate on the population dynamics of a large herbivore under harsh environmental conditions. *J Wildlife Manag.* 2010;74:1671–85.
42. Fuller J, Ferchaud AL, Laporte M, Le Luyer J, Davis TB, Côté SD, et al. Absence of founder effect and evidence for adaptive divergence in a recently introduced insular population of white-tailed deer (*Odocoileus virginianus*). *Mol Ecol.* 2020;29(1):86–104. <https://doi.org/10.1111/mec.15317>.
43. Taft HR, Roff DA. Do bottlenecks increase additive genetic variance? *Conservation Genetics.* 2012;13(2):333–42. <https://doi.org/10.1007/s10592-011-0285-y>.

44. Hamlin KL, Pac DF, Sime CA, DeSimone RM, Dusek GL. Evaluating the accuracy of ages obtained by two methods for Montana ungulates. *J Wildlife Manag.* 2000;64:441–9.
45. Simard M-A, Huot J, de Bellefeuille S, Côté SD. Influences of habitat composition, plant phenology, and population density on autumn indices of body condition in a northern white-tailed deer population. *Wildlife Monographs.* 2014;187:1–28.
46. Bundy RM, Robel RJ, Kemp KE. Whole Body Weights Estimated from Morphological Measurements of White-Tailed Deer. *Trans Kansas Acad Sci (1903-).* 1991;94(3/4):95–100. <https://doi.org/10.2307/3627856>.
47. Nilsen EB, Solberg EJ. Patterns of hunting mortality in Norwegian moose (*Alces alces*) populations. *Eur J Wildlife Res.* 2006;52:153–63.
48. Solberg EJ, Loison A, Gaillard J-M, Heim M. Lasting effects of conditions at birth on moose body mass. *Ecography.* 2004;27:677–87.
49. Smit AFA, Hubley R. RepeatModeler Open-1.0. 2008-2015. 2019. retrieved from <http://www.repeatmasker.org>
50. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015. 2019. Retrieved from <http://www.repeatmasker.org>
51. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics.* 2011;12(1):491. <https://doi.org/10.1186/1471-2105-12-491>.
52. Laine VN, Gossmann TI, Schachtschneider KM, Garroway CJ, Madsen O, Verhoeven KJF, et al. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat Commun.* 2016;7(1):10474. <https://doi.org/10.1038/ncomms10474>.
53. Korf I. Gene finding in novel genomes. *BMC Bioinformatics.* 2004;5(1):59. <https://doi.org/10.1186/1471-2105-5-59>.
54. Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36(Web Server):W5–W9. <https://doi.org/10.1093/nar/gkn201>
55. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics.* 2014;btu170.
56. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. Retrieved from <http://arxiv.org/abs/1303.3997>
57. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
58. Li J, Dani JA, Le W. The role of transcription factor Pitx3 in dopamine neuron development and Parkinson's disease. *Curr Topics Med Chem.* 2009;9(10):855–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19754401>
59. Broad Institute (2020). Picard Tools – By Broad Institute. Github. <http://broadinstitute.github.io/picard/>
60. Van der Auwera GA, Carneiro MO, Hartl C, ... DePristo MA. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43(1110). <https://doi.org/10.1002/0471250953.B1110543>.
61. Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, et al. PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE.* 2011a;6(1):e15925. <https://doi.org/10.1371/journal.pone.0015925>.
62. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics.* 2011b;27(24):3435–6. <https://doi.org/10.1093/bioinformatics/btr589>.
63. Kurland S, Wheat CW, Mancera M, de la PC, Kutschera VE, Hill J, Andersson A, ... Laikre L. Exploring a Pool-seq-only approach for gaining population genomic insights in nonmodel species. *Ecol Evol.* 2019;9(19):11448–63. <https://doi.org/10.1002/ECE3.5646>
64. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995;57(1):289–300.
65. Spitzer K, Pelizzola M, Futschik A. Modifying the Chi-square and the CMH test for population genetic inference: Adapting to overdispersion. *Ann Appl Stat.* 2020;14(1). <https://doi.org/10.1214/19-AOAS1301>.
66. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA.* 2015;6(1) <https://doi.org/10.1186/s13100-015-0041-9>.
67. Kofler R, Schlötterer C. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics.* 2012;28(15):2084–5. <https://doi.org/10.1093/bioinformatics/bts315>.
68. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE.* 2011;6(7):e21800. <https://doi.org/10.1371/journal.pone.0021800>.
69. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44–57. <https://doi.org/10.1038/nprot.2008.211>.
70. Guirao-Rico S, González J. Benchmarking the performance of Pool-seq SNP callers using simulated and real sequencing data. *Mol Ecol Resour.* 2021;1755-0998:13343. <https://doi.org/10.1111/1755-0998.13343>.
71. Haworth SE, Nituch L, Northrup JM, Shafer ABA. Characterizing the demographic history and prion protein variation to infer susceptibility to chronic wasting disease in a naïve population of white-tailed deer (*Odocoileus virginianus*). *Evol Appl.* 2021;eva.13214. <https://doi.org/10.1111/eva.13214>
72. Bouwman AC, Garrick DJ, Reecy J, Van Tassell CP. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet.* 2018;50:362–7. <https://doi.org/10.1038/s41588-018-0056-5>.
73. Visscher PM, Macgregor S, Benyamin B, et al. Genome partitioning of genetic variation for height from 11,214 sib-ling pairs. *Am J Human Genet.* 2007;81:1104–10.
74. Wang Y, Zhang C, Wang N, Li Z, Heller R, Liu R, ... Qiu Q. Genetic basis of ruminant headgear and rapid antler regeneration. *Science.* 2019;364(6446):eaav6335. <https://doi.org/10.1126/science.aav6335>
75. Foote AD, Vijay N, Ávila-Arcos MC, Baird RW, Durban JW, Fumagalli M, et al. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat Commun.* 2016;7(1):11693. <https://doi.org/10.1038/ncomms11693>.
76. Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet.* 2004;5(5):389–96. <https://doi.org/10.1038/nrg1327>.
77. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet.* 2015;24(R1):R102–10. <https://doi.org/10.1093/hmg/ddv259>.
78. Booker TR, Yeaman S, Whitlock MC. Global adaptation complicates the interpretation of genome scans for local adaptation. *Evolution Letters.* 2021;5(1):4–15. <https://doi.org/10.1002/evl3.208>.
79. Whitlock MC, Lotterhos KE. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F ST. *Am Nat.* 2015;186(5):524–36. <https://doi.org/10.1086/682949>.
80. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, ... Da Y. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics.* 2011;12(1):408. <https://doi.org/10.1186/1471-2164-12-408>
81. Jiang J, Ma L, Prakapenka D, VanRaden PM, Cole JB, Da, Y. A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. *Front Genet.* 2019;10:412. <https://doi.org/10.3389/fgene.2019.00412>
82. Li C, Zhao H, Liu Z, McMahon C. Deer antler – A novel model for studying organ regeneration in mammals. *Int J Biochem Cell Biol.* 2014;56:111–22. <https://doi.org/10.1016/j.biocel.2014.07.007>.
83. Vanpé C, Gaillard JM, Kjellander P, Mysterud A, Magnien P, Delorme D, et al. Antler size provides an honest signal of male phenotypic quality in roe deer. *Am Nat.* 2007;169(4):481–93. <https://doi.org/10.1086/512046>.
84. Ba H, Wang D, Yau TO, Shang Y, Li C. Transcriptomic analysis of different tissue layers in antler growth center in Sika Deer (*Cervus nippon*). *BMC Genomics.* 2019;20(1):173. <https://doi.org/10.1186/s12864-019-5560-1>.
85. Elliott JL, Oldham JM, Ambler GR, Bass JJ, Spencer GS, Hodgkinson SC, et al. Presence of insulin-like growth factor-I receptors and absence of growth hormone receptors in the antler tip. *Endocrinology.* 1992;130(5):2513–20.
86. Sadighi M, Haines SR, Skottner A, Harris AJ, Suttie JM. Effects of insulin-like growth factor-I (IGF-I) and IGF- II on the growth of antler cells in vitro. *J Endocrinol.* 1994;143(3):461–9.
87. Suttie JM, Gluckman PD, Butler JH, Fennessy PF, Corson ID, Laas FJ. Insulin-like growth factor 1 (IGF-1) antler-stimulating hormone? *Endocrinology.* 1985;116(2):846–8.
88. Heusschen R, Griffioen AW, Thijssen VL. Galectin-9 in tumor biology: A jack of multiple trades. *Biochim Biophys Acta - Rev Cancer.* Elsevier. 2013. <https://doi.org/10.1016/j.bbcan.2013.04.006>.

89. Mollenhauer J, Wiemann S, Scheurlen K, Poustka K. DMBT1, a new member of the SRCR superfamily, on chromosome 10q25.3-26.1 is deleted in malignant brain tumours. *Nat Genet.* 1997;17(1):32–9. <https://doi.org/10.1038/NG0997-32>.
90. Hernández-Montiel W, Martínez-Núñez MA, Ramón-Ugalde JP, Román-Ponce SI, Calderón-Chagoya R, Zamora-Bustillos R. Genome-wide association study reveals candidate genes for litter size traits in pelibuey sheep. *Animals.* 2020;10(3):434. <https://doi.org/10.3390/ani10030434>.
91. Mruk DD, Cheng CY. The myotubularin family of lipid phosphatases in disease and in spermatogenesis. *Biochemical Journal.* NIH Public Access. 2011. <https://doi.org/10.1042/BJ20101267>
92. Carapito R, Konantz M, Paillard C, Miao Z, Pichot A, Leduc MS, et al. Mutations in signal recognition particle SRP54 cause syndromic neutropenia with Shwachman-Diamond-like features. *J Clin Invest.* 2017;127(11):4090–103. <https://doi.org/10.1172/JCI92876>.
93. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein Sci.* 2018;27(1):233. <https://doi.org/10.1002/PRO.3307>.
94. Pavlidis P, Jensen JD, Stephan W, Stamatakis A. A Critical Assessment of Storytelling: Gene Ontology Categories and the Importance of Validating Genomic Scans. *Mol Biol Evol.* 2012;29(10):3237–48. <https://doi.org/10.1093/molbev/mss136>.
95. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, et al. Rare and low-frequency coding variants alter human adult height. *Nature.* 2017;542(7640):186–90. <https://doi.org/10.1038/nature21039>.
96. Taye M, Yoon J, Dessie T, Cho S, Oh SJ, Lee HK, et al. Deciphering signature of selection affecting beef quality traits in Angus cattle. *Genes Genomics.* 2018;40(1):63–75. <https://doi.org/10.1007/s13258-017-0610-z>.
97. Deng MT, Zhu F, Yang YZ, Yang FX, Hao JP, Chen SR, et al. Genome-wide association study reveals novel loci associated with body size and carcass yields in Pekin ducks 06 Biological Sciences 0604 Genetics. *BMC Genomics.* 2019;20(1):1–13. <https://doi.org/10.1186/s12864-018-5379-1>.
98. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet.* 2012;13(1):36–46. <https://doi.org/10.1038/nrg3117>.
99. Lerat E, Goubert C, Guirao-Rico S, Merenciano M, Dufour A, Vieira C, et al. Population-specific dynamics and selection patterns of transposable element insertions in European natural populations. *Mol Ecol.* 2019a;28(6):1506–22. <https://doi.org/10.1111/mec.14963>.
100. Lerat E, Goubert C, Guirao-Rico S, Merenciano M, Dufour A, Vieira C, et al. Population-specific dynamics and selection patterns of transposable element insertions in European natural populations. *Mol Ecol.* 2019b;28(6):1506–22. <https://doi.org/10.1111/mec.14963>.
101. Morina DL, Demarais S, Strickland BK, Larson JE. While males fight, females choose: male phenotypic quality informs female mate choice in mammals. *Anim Behav.* 2018;138:69–74. <https://doi.org/10.1016/J.ANBEHAV.2018.02.004>.
102. Knief U, Bossu CM, Saino N, et al. Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nat Ecol Evol.* 2019;3:570–6.
103. Tilk S, Bergland A, Goodman A, Schmidt P, Petrov D, Greenblum S. Accurate allele frequencies from ultra-low coverage Pool-seq samples in evolve-and-resequence experiments. *G3: Genes, Genomes, Genetics.* 2019;9(12):4159–68. <https://doi.org/10.1534/g3.119.400755>.
104. Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, et al. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat Commun.* 2016;7(1):1–10. <https://doi.org/10.1038/ncomms13195>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

