


# Development of a genotype-by-sequencing immunogenetic assay as exemplified by screening for variation in red fox with and without endemic rabies exposure

Michael E. Donaldson<sup>1</sup>  | Yessica Rico<sup>2</sup> | Karsten Hueffer<sup>3</sup> | Halie M. Rando<sup>4</sup> | Anna V. Kukekova<sup>4</sup> | Christopher J. Kyle<sup>1,5</sup>

<sup>1</sup>Environmental and Life Sciences Graduate Program, Trent University, Peterborough, ON, Canada

<sup>2</sup>CONACYT, Instituto de Ecología A.C., Centro Regional del Bajío, Pátzcuaro Michoacán, Mexico

<sup>3</sup>Department of Veterinary Medicine, University of Alaska Fairbanks, Fairbanks, AK, USA

<sup>4</sup>Department of Animal Sciences, College of ACES, University of Illinois at Urbana-Champaign, Urbana, IL, USA

<sup>5</sup>Forensic Science Department, Trent University, Peterborough, ON, Canada

## Correspondence

Christopher J. Kyle, Forensic Science Department, Trent University, Peterborough, ON, Canada.

Email: christopherkyle@trentu.ca

## Funding information

Natural Sciences and Engineering Research Council of Canada (NSERC); National Science Foundation Division of Graduate Education Integrative Graduate Education and Research Traineeship, Grant/Award Number: NSF DGE IGERT-1069157; National Institutes of Health, Grant/Award Number: GM120782

## Abstract

Pathogens are recognized as major drivers of local adaptation in wildlife systems. By determining which gene variants are favored in local interactions among populations with and without disease, spatially explicit adaptive responses to pathogens can be elucidated. Much of our current understanding of host responses to disease comes from a small number of genes associated with an immune response. High-throughput sequencing (HTS) technologies, such as genotype-by-sequencing (GBS), facilitate expanded explorations of genomic variation among populations. Hybridization-based GBS techniques can be leveraged in systems not well characterized for specific variants associated with disease outcome to “capture” specific genes and regulatory regions known to influence expression and disease outcome. We developed a multiplexed, sequence capture assay for red foxes to simultaneously assess ~300-kbp of genomic sequence from 116 adaptive, intrinsic, and innate immunity genes of predicted adaptive significance and their putative upstream regulatory regions along with 23 neutral microsatellite regions to control for demographic effects. The assay was applied to 45 fox DNA samples from Alaska, where three arctic rabies strains are geographically restricted and endemic to coastal tundra regions, yet absent from the boreal interior. The assay provided 61.5% on-target enrichment with relatively even sequence coverage across all targeted loci and samples (mean = 50×), which allowed us to elucidate genetic variation across introns, exons, and potential regulatory regions (4,819 SNPs). Challenges remained in accurately describing microsatellite variation using this technique; however, longer-read HTS technologies should overcome these issues. We used these data to conduct preliminary analyses and detected genetic structure in a subset of red fox immune-related genes between regions with and without endemic arctic rabies. This assay provides a template to assess immunogenetic variation in wildlife disease systems.

## KEYWORDS

arctic rabies virus, immunogenomics, local adaptation, red fox, sequence capture, wildlife disease

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Populations are exposed to spatially heterogeneous selective pressures that often lead to localized patterns of adaptation, where local individuals are expected to have higher fitness than those from other populations from different environments (Thompson, 2005). In this context, characterizing patterns of local adaptation illustrates the spatial and temporal interactions of species with their environment, describes processes that generate and maintain biodiversity, and identifies evolutionary forces shaping populations (Carlson, Cunningham, & Westley, 2016; Harrison, Pavlova, Telonis-Scott, & Sunnucks, 2014).

Pathogens can exert strong selective pressures on populations and are recognized as major drivers of adaptive divergence (Fumagalli et al., 2011). Clarifying adaptive patterns becomes ever more important as disease dynamics are rapidly altered by climatic changes and anthropogenic influences, with pathogens colonizing previously inaccessible environments, as evidenced by several emerging diseases (Gallana, Ryser-Degiorgis, Wahli, & Segner, 2013; Kim et al., 2014; Smith et al., 2012). The spatial patterns of local adaptation to selective pressure from disease can be discovered by assessing which gene variants are favored in local interactions among populations (Acevedo-Whitehouse & Cunningham, 2006; Beldomenico & Begon, 2016; Bonneaud, Balenger, Zhang, Edwards, & Hill, 2012; Eizaguirre, Lenz, Kalbe, & Milinski, 2012; Hansen, Olivieri, Waller, & Nielsen, 2012; Hess, Campbell, Close, Docker, & Narum, 2013; Kyle et al., 2014; Orsini, Andrew, & Eizaguirre, 2013; Rico, Morris-Pocock, Zigouris, Nocera, & Kyle, 2015; Schoville et al., 2012). These data are critical to understanding the capacity for natural populations to adapt to changing selective pressures and revealing spatially explicit adaptive responses to altered disease dynamics.

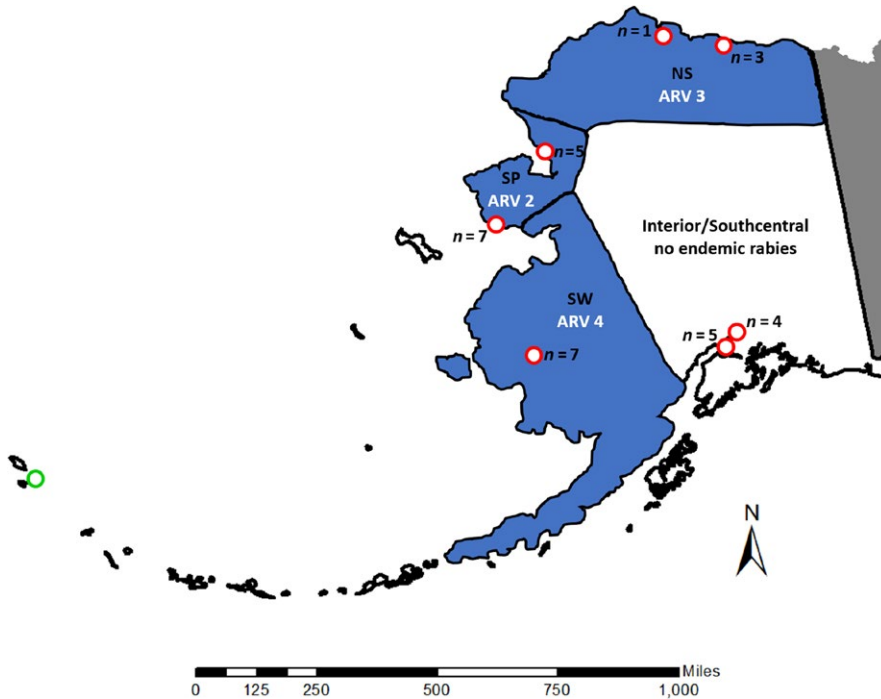
Host responses to pathogens are largely influenced by immune genes, where genetic diversity influences population-level resistance to disease via pathogen-mediated directional or balancing selection (Eizaguirre et al., 2012; Rico et al., 2015). Studies of wild populations generally focus on adaptive immunity, assessed via a small number of major histocompatibility complex (MHC) loci, due to their role in pathogen recognition and pathogen susceptibility (Acevedo-Whitehouse & Cunningham, 2006; Eizaguirre et al., 2012; Kyle et al., 2014). A growing number of studies also examine key receptor genes associated with innate immunity (e.g., toll-like receptors and interleukins) to clarify patterns of resistance to emerging infectious diseases (e.g., Bonneaud et al., 2012). The aforementioned loci, however, provide insight into only a small fraction of the genes associated with mounting an immune response, disease resistance, and adaptation. Overall, a more holistic assessment of immunogenetic variation is required to understand the capacity for adaptation to disease (Harrison et al., 2014; Morris, Wright, Grueber, Hogg, & Belov, 2015; Ng et al., 2009).

High-throughput sequencing (HTS) technologies allow for expanded explorations of genomic variation associated with local adaptation; however, for population-level assessments of genetic variation in wildlife with larger genomes, full-genome processing is generally not yet feasible (Ekblom & Wolf, 2014). For most non-model organisms, large chip arrays of single-nucleotide polymorphisms (SNP)

are not available, and those are maybe prone to ascertainment bias when applied to new populations (Albrechtsen, Nielsen, & Nielsen, 2010; Lachance & Tishkoff, 2013). As an alternative, genotype-by-sequencing (GBS) assays can be employed to obtain genomic subsets of population variation. GBS largely falls into three categories: restriction enzyme-, amplicon-, and hybridization-based models, each coming with important considerations for implementation (Jones & Good, 2016). For example, restriction site associated DNA (RAD) markers are well-suited for detecting neutral sequence variation across the genome to develop models of genetic population structure (Catchen et al., 2017), but there are some controversies as to their applicability in detecting patterns of local adaptation (Lowry et al., 2017). In more targeted GBS approaches, amplicon-based and target hybridization-based methods have been shown to be highly effective in enriching for subsets of genomic sequence. Amplicon-based methods rely on the ability to PCR amplify large numbers of targeted, overlapping short-sequences, using multiple sets of PCR primers in a single PCR (Samorodnitsky et al., 2015). Target capture utilizes overlapping biotinylated DNA or RNA probes, which bind to complementary targeted regions of DNA that are then selectively pulled down by magnetic streptavidin beads to enrich for target DNA (Chilamakuri et al., 2014; Ng et al., 2009). Both techniques, when coupled with HTS, enable the identification of genetic variation across specific loci in the genome.

Targeted sequence capture approaches have only recently been applied to wildlife systems. For example, Morris et al. (2015) used genome-level information from ten Tasmanian devils to identify SNPs within the “immunome,” consisting of the coding and regulatory regions of 167 immune genes. Using this information, they developed an amplicon-based assay for nine immune genes with nonsynonymous SNPs, and then, genotypes were generated for 220 Tasmanian devils in the context of a transmissible facial cancer decimating the species. Similarly, Schweizer, Robinson et al. (2016) employed an RNA-bait version of target capture to enrich for 1,040 candidate genes and their regulatory regions, as well as 5,000 1-kbp nongenic neutral regions, to determine genetic variation in 107 gray wolves from diverse ecotypes. In light of recent technological advances, multiple GBS-based strategies have emerged with the potential to study nonmodel wildlife species under various selective pressures.

Rabies viruses cause fatal encephalopathies in mammals with strains adapted to infect different primary hosts (Jackson & Wunner, 2007). In Alaska, arctic rabies (AR) is typically restricted to arctic coastal areas following arctic fox (*Vulpes lagopus*) distributions, where AR variants have discrete spatial distributions (Kuzmin, Hughes, Botvinkin, Gribencha, & Rupprecht, 2008; Nadin-Davis, Sheen, & Wandeler, 2012). Distinct phylogeographic patterns have been observed between AR variants (Kuzmin et al., 2008; Nadin-Davis et al., 2012), suggesting variable disease resistance may influence the spread of rabies. Three AR variants occur in Alaska that are isolated to regions with tundra red foxes (*Vulpes vulpes*) and arctic foxes, with nonoverlapping, temporally stable distributions, although there are no data regarding the relative virulence of these strains. The AR variants occur on the North Slope (NS; AR variant 3), Seward Peninsula (SP; AR variant 2), and southwestern (SW; AR variant 4) regions of Alaska (Figure 1).



**FIGURE 1** Schematic of Alaska arctic fox and red fox samples analyzed using immunogenetic profiling. Approximate arctic rabies variant (ARV) distribution was modified from Goldsmith et al. (2016). Arctic rabies and arctic rabies-free zones are indicated by blue and white background colors, respectively. Arctic fox and red fox sample locations are denoted by an open-green circle or open-red circles, respectively. Red fox sample size ( $n$ ) for each location used in the genetic structure analyses is indicated. SW, Southwest; SP, Seward Peninsula; NS, North Slope

Conversely, in interior and southcentral (SC) Alaska, only boreal red foxes are found, and AR is not endemic (Goldsmith et al., 2016). Red fox, a highly susceptible AR host, has been associated with AR spreading from Arctic regions to southern Canada (Kuzmin et al., 2008). In the Canadian Arctic, one AR variant predominates, including in regions where arctic foxes are absent and red fox are present. Therefore, it is unclear why rabies has not spread via red fox into central Alaska given the patterns observed across Canada (Mørk & Prestrud, 2004; Nadin-Davis et al., 2012). In red fox, recent experiments detected weak genetic structure in Alaska between tundra and boreal regions consistent with spatial distributions of AR variant presence and gene flow between those populations (Goldsmith et al., 2016). Questions remain, as to how DNA sequence diversity at functionally important immune-related loci for red fox is spatially structured in these regions, and if geographic patterns of diversity are correlated with AR strain distribution.

Herein, we describe the development of an immunogenetic assay designed to test the hypothesis that selective pressure from disease alters the spatial distribution of genetic variants among populations. Our goals were to (1) develop an assay capable of elucidating genetic variation at red fox immunogenic loci and (2) use the resulting SNPs to apply preliminary tests for genetic structure and  $F_{ST}$  outliers. We predicted that even with low sample sizes, we would be able to detect specific genetic variants based on AR presence/absence because AR exerts a strong selective pressure on fox populations. To develop the assay, we annotated immune-related red fox gene sequences and then applied high-throughput targeted sequence capture to enrich for 300-kbp of preselected genomic regions (per individual) that could be used to identify SNPs in a wide range of key immune genes and their regulatory regions, which can influence the expression of immune system genes and disease outcome (Fraser, 2013). To test how robust our

assay was, we assessed whether 1 ng DNA/sample could be used for target capture-based GBS to evaluate the potential of this technique for noninvasive samples with less DNA (such as from hair or feces), and whether the red fox hybridization probes could also be used to enrich for similar targets from an arctic fox DNA sample for future AR research. To address our hypothesis and elucidate patterns of local adaptation, we required an understanding of the underlying patterns of gene flow and genetic drift that may influence the presence/absence of genes beyond selective pressures. As such, we also included 23 neutral microsatellite loci and amelogenin (for sex determination) in this assay to provide insight into the demographic processes of the populations under investigation, thus providing an “all-in-one” test to generate data for our research aims. This study provides a template for future experiments aimed at identifying immunogenetic variants and genetic patterns of local adaptation to disease.

## 2 | METHODS

### 2.1 | Sample collection, DNA extraction, and quantification

We obtained red fox tissue ( $n = 33$ ) or extracted DNA samples from rabies-positive red fox ( $n = 8$ ) as described by Goldsmith et al. (2016) from a variety of organizations and individual trappers. The rabies-negative red fox tissue samples have voucher specimens in the University of Alaska Museum of the North, and accession numbers are provided, where applicable (Table S1). Rabies-positive red fox tissue was confirmed at the Alaska government health laboratories by serology (ELISA) and then at the Centres for Disease Control by sequencing. The arctic fox DNA sample ( $n = 1$ ) was obtained from Terry Spraker (Colorado State University, USA). We dissolved all

tissue samples in 1× lysis buffer (4 mol/L urea, 0.2 mol/L NaCl, 0.5% n-lauroyl sarcosine, 10 mmol/L 1,2-cyclohexanediaminetetraacetic acid, 0.1 mmol/L Tris-HCl pH 8.0) containing 600 U/ml proteinase K (time and temperature) at 56°C for two hours and extracted DNA using either the automated 96-well MagneSil Blood Genomic Max Yield System (Promega) or the DNeasy Blood and Tissue Kit (Qiagen). We quantified DNA extractions using the Quant-iTPicoGreen dsDNA Assay Kit (ThermoFisher Scientific). These DNA extractions yielded starting material for 45 target-enriched DNA libraries, which included 31 tundra red fox (11 SW, 15 SP, 5 NS), 10 boreal red fox (SC), additional replicates from the DNA dilution series (1, 10, 100 ng each), and one arctic fox (Figure 1; Table S1).

## 2.2 | Identification of red fox immune genes, neutral markers, and probe development

We compiled a list of candidate immune genes (Table S2) using those listed in The Dog Innate & Adaptive Immune Responses RT<sup>2</sup> Profiler PCR Array (Qiagen) and in the overview of key mediators of innate and adaptive immunity, development, and signaling (Knight, 2013). Using this candidate gene list, we downloaded genomic sequences (covering an entire locus), individual exon sequences, and protein sequences from the CanFam3.1 dog (*Canis lupus familiaris*) genome assembly (Hoepfner et al., 2014) using Ensembl release version 79 (Cunningham et al., 2015).

We used the dog genomic sequences to perform BLASTN (Altschul, Gish, Miller, Myers, & Lipman, 1990) against the *V. vulpes* draft genome (Kukekova et al., in review) to identify segments of the red fox genome assembly (NCBI BioProject PRJNA378561) with sequence similarity. The red fox genome assembly consists of 2,495,544,672 assembled bp and 676,878 scaffolds (N50 = 11.5 Mb). We extracted red fox immune-gene sequences (2,229,152 bp) from the red fox genome assembly that included upstream and downstream sequences ( $\pm 1,500$  bp). We annotated the red fox immune-gene structures (introns and exons) using dog exon sequences by performing BLASTN (Altschul et al., 1990) against the red fox DNA sequences with a word-size parameter set to 7 to include short sequence hits. We input the red fox immune-gene sequences and dog peptide sequences into FGENESH+ (Solovyev, 2007) on the SOFTBERRY online portal (<http://www.softberry.com/berry.phtml>) to resolve cases where BLASTN did not properly annotate the red fox intron/exon boundaries, which led to incomplete open reading frames (start codon to stop codon). We performed BLASTP (Altschul et al., 1990) against the NCBI refseq\_protin database to verify all red fox protein-coding sequences. The red fox gene annotations and predicted red fox immune-related protein sequences are available in Appendices S1 and S2, respectively.

We also generated a list of canine and red fox microsatellite markers from previous studies (An et al., 2010; Fredholm & Winterø, 1995; Mellers et al., 1997; Ostrander, Sprague, & Rine, 1993). We downloaded the red fox microsatellite region from NCBI when available. Otherwise, we extracted the genomic sequences from the CanFam3.1 dog assembly and used BLASTN to identify and extract regions of putative sequence similarity in the red fox genome. This

yielded another set of red fox DNA sequence targets (Table S3) that were included in the custom NimbleGen SeqCap EZ probe design (Roche).

We added 100-bp “padding” to each of the red fox DNA sequence targets (primary targets) to increase the efficiency of sequence capture. The custom NimbleGen SeqCap EZ probes were produced using the red fox DNA sequence targets. We only had limited access to the red fox genome assembly (Kukekova et al., in review); therefore, the probe set was compared to the dog genome reference sequence to test probe specificity. We finalized a “relaxed” probe design that allowed up to 20 close matches to the reference dog genome; however, 91.5% of the probes had only one match to the reference dog genome sequence, and 95.8% had five or fewer matches (not including zero) to the reference dog genome sequence, indicating a low likelihood for “off-target” sequence capture.

## 2.3 | Library preparation, sequence capture, and high-throughput sequencing

We prepared DNA libraries using the Kapa HTP Lib Prep Kit (Roche). Forty-two DNA libraries were prepared using 1 µg DNA, and an additional three DNA libraries were prepared for the dilution series using 1, 10, and 100 ng DNA from sample KH1354. Special considerations were required during library preparation based on the quantity of DNA put into the reaction (Table S4). Additional modifications to the KAPA DNA library protocol included the following: (i) TruSeq HT Dual-Index Adapters (Integrated DNA Technologies) resuspended in Nuclease Free Duplex Buffer (Integrated DNA technologies) were used instead of the SeqCap Adapter Kits A and B (Roche) during adapter ligation, and (ii) Illumina P5 and Illumina P7 primers (Integrated DNA technologies) were used instead of the Pre LM-PCR Oligos 1 & 2 (Roche) during Pre-Capture LM-PCR. We assessed library quality by ethidium bromide-stained gel electrophoresis using a 2% E-Gel (ThermoFisher Scientific).

Prior to target capture, we measured the concentration of the 45 DNA library preparations using a NanoDrop 8000 spectrophotometer (ThermoFisher Scientific). We pooled these DNA libraries in equimolar amounts to a final mass of 1 µg DNA. We performed target capture using the designed DNA oligos contained in the NimbleGen SeqCap EZ Developer Library (Roche) according to the manufacturer’s recommended protocol outlined in the SeqCap EZ Library SR User Guide v 5.0 (Roche) with the following modifications: (i) 1 µl of the xGen Universal Blocking Oligo TS HT-i5 (Integrated DNA Technologies) and 1 µl xGen Universal Blocking Oligo TS HT-i7 (Integrated DNA Technologies) were used instead of the NimbleGen Multiplex Hybridization Enhancing Oligo Pool (Roche), and (ii) NimbleGen SeqCap EZ Developer Reagent (Roche) was used instead of NimbleGen COT Human DNA (Roche) during hybridization sample preparation where the hybridization was carried out at 47°C for 72 hr. We assessed target-enriched DNA quality using a bioanalyzer (Agilent Technologies), and we performed HTS on a MiSeq run using 2 × 150-bp reads (Illumina).

## 2.4 | Sequence alignment and variant annotation

We used the `bwa-mem` command in the `BURROWS-WHEELER ALIGNER` v0.7.12 (BWA; Li, 2013) to align the paired-end reads to the red fox immune-gene sequences. We compiled sequence alignment metrics using `SAMTOOLS` v1.2 (Li et al., 2009). We used the `GENOME ANALYSIS TOOLKIT` v3.5 (GATK; McKenna et al., 2010) for base quality score recalibration, INDEL realignment, duplicate removal, depth of coverage calculations, SNP/INDEL discovery, and genotyping across all samples using standard hard filtering parameters or variant quality score recalibration according to GATK best practices recommendations (DePristo et al., 2011; Van der Auwera et al., 2013).

## 2.5 | SNP/INDEL filtering and analyses

For SNP/INDEL analyses, we removed the arctic fox sample library (KH1527), the three “dilution” sample libraries (KH-1354-1, KH-1354-10, KH-1354-100), and one sample library with <10 $\times$  coverage (KH989). Further, we removed the eight rabies-positive samples because we did not have enough samples from any one area (3 SP, 1 NS, 4 SW) to provide the sample size required for the preliminary downstream analyses presented in our study. We include the sequencing results from these rabies-positive samples as a future resource for researchers that may be able to use these data. Therefore, SNP/INDEL data from the 32 remaining sample libraries (Figure 1; Table S1) were used in the preliminary downstream analyses to help inform future experiments. We generated separate SNP and INDEL subdatasets containing genetic variants for those 32 red fox using GATK.

We filtered the SNP subdataset further using `VCFTOOLS` v0.1.14 (Danecek et al., 2011) to select for bi-allelic SNPs with a maximum missing genotype of 10% and a minor allele frequency of 5%. We reformatted the resulting variant call format file (.vcf) using `PGDSPIDER` v2.0.9.2 (Lischer & Excoffier, 2012) for the `STRUCTURE` v2.3.4 (Pritchard, Stephens, & Donnelly, 2000) and  $F_{ST}$  outlier analyses.

We used the tandem repeat finder `TRF` v4.09b (Benson, 1999) to annotate the INDEL subdataset corresponding to microsatellites and amelogenin and then `VCFTOOLS` to extract the genotypes of each individual for the microsatellite loci and amelogenin. We then used the python script “`getgenosfromvcf`” (De Wit et al., 2012) to extract microsatellite genotypes with a Phred quality score cutoff of 20, which yields genotypes with a 99% probability of being true. We created a filtered INDEL subdataset containing genetic variants for 15 microsatellite regions by removing ambiguous microsatellite calls with multiple sequence alignments. We also tested for significant departures from Hardy-Weinberg equilibrium (HWE) for the filtered INDEL subdataset using a probability test in `GENEPOP` v4.2 (Rousset, 2008) with a correction ( $p = 1.67E-03$ ) to reject the null hypothesis that loci are in HWE.

$F_{ST}$ -based outlier detection methods have high false-positive error rates when identifying SNPs under directional or balancing selection, and limited sensitivity in detecting SNPs under weak selection (Narum & Hess, 2011). Therefore, we used three different outlier detection programs to mitigate the number of false positives in the  $F_{ST}$  outlier

subdataset we used for preliminary analyses. We used `BAYESCAN` v2.1 (Foll & Gaggiotti, 2008), `LOSITAN` (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008; Beaumont & Nichols, 1996), and `OUTFLANK` (Whitlock & Lotterhos, 2015) to identify  $F_{ST}$  outliers putatively under selection from the filtered SNP subdataset. The three outlier detection programs do not rely on a set of “presumed” neutral loci to generate an empirical null distribution of  $F_{ST}$ . Rather, they simulate a null distribution of  $F_{ST}$  for the sample sizes observed in the dataset and identify departures from neutrality using different analytical approaches (Antao et al., 2008; Beaumont & Nichols, 1996; Foll & Gaggiotti, 2008; Whitlock & Lotterhos, 2015). We ran `BAYESCAN` using 1:10 prior odds for the neutral model and included 20 pilot runs consisting of 5,000 iterations each, followed by 550,000 iterations with a burn-in of 50,000 iterations. We ran `LOSITAN` for 1,000,000 iterations within a 99.5% confidence interval with the “Neutral mean  $F_{ST}$ ” and “Force mean  $F_{ST}$ ” options enabled under the infinite alleles model (IAM) and stepwise mutation model (SMM). We ran `OUTFLANK` using the default settings (LeftTrimFraction = 0.05, RightTrimFraction = 0.05, Hmin = 0.1) with NumberOfSamples = 2. We used false discovery rate (FDR) values of 0.05 for all  $F_{ST}$  outlier analyses. We extracted a  $F_{ST}$  outlier subdataset containing the 15 common  $F_{ST}$  outliers identified by the `LOSITAN-AIM`, `LOSITAN-SMM`, and `OUTFLANK` analyses using `VCFTOOLS`.

We performed principal component analysis (PCA) on the filtered SNP subdataset, INDEL subdataset, and  $F_{ST}$  outlier subdataset using `ADEGENET` v2.0.0 (Jombart & Ahmed, 2011). We obtained the required “genlight” objects for the `ADEGENET` analysis using a combination of `VCFTOOLS` and `PLINK` v1.07 (Purcell et al., 2007) to reformat the .vcf files to `PLINK` formatted files (.raw). We also tested for genetic structure using the  $F_{ST}$  outlier subdataset in `STRUCTURE`. We used the `STRAUTO` v1.0 script (Chhatre & Emerson, 2017) to run `STRUCTURE` over multiple processors at the same time. We ran `STRUCTURE` with a burn-in length of 50,000 followed by 200,000 iterations for  $K = 1$  through 8, and each run was performed 20 times. We calculated the  $\Delta K$  statistic (Evanno, Regnaut, & Goudet, 2005) to help determine the number of inferred genetic clusters using `STRUCTURE HARVESTER WEB` v0.6.94 (Earl & vonHoldt, 2011). We used the LargeKGreedy (10,000 repeats) algorithm in `CLUMPP` v1.1.2 (Jakobsson & Rosenberg, 2007) to combine the SNP results from the multiple `STRUCTURE` runs, and we visualized those results using `DISTRUCT` v1.1 (Rosenberg, 2004).

## 3 | RESULTS

### 3.1 | Identification of red fox immune genes, neutral markers, and probe development

We extracted regions of genomic sequence corresponding to 116 immune genes and their upstream regulatory regions (Table S2), 23 microsatellite regions, and a portion of intron 1 in the amelogenin sex-determining marker (Table S3) from the red fox genome sequence assembly (Kukekova et al., unpublished data). Using the red fox immune-gene sequences, we annotated the immunity gene intron/exon boundaries (Appendix S1). We validated the protein-coding annotations by translating all 116 nucleotide coding sequences to

proteins (Appendix S2) and performing BLASTp (Table S2). We generated full-length coding sequence information for 109 immunity-related genes (excluding: CD86, DLA-12, DLA-88, IL18, MAPK1, MAPK8, and MX1). We successfully extracted 1,500-bp of upstream genomic sequence representing putative regulatory regions for 93 genes and partial (<1,500-bp) regulatory region sequences for an additional 15 genes. Our red fox immune-gene sequence database lacked the upstream sequence necessary to target the regulatory regions for CD28, CD86, DLA-12, HLA-DPB1, IKBKG, IL18, MAPK1, and MX1 (Table S2). Those regulatory regions were either missing from, or spanned, multiple scaffolds in the red fox genome assembly and could not be included in our immune-gene sequence database. Using these annotations, the custom NimbleGen SeqCap EZ Developer Library probe set was synthesized.

### 3.2 | Library preparation, high-throughput sequencing, and sequence alignment

We obtained over 12.4 million paired-end reads from the HTS MiSeq run (Table S5). After the samples were demultiplexed, there was an average of 274,526 total reads per sample library and on average, 93.5% of the reads mapped to the primary target regions. We applied filters to remove mapped reads identified by the GATK as having low mapping quality (MAPQ < 20; mean = 13.4% per sample), being PCR duplicates (mean = 4.3% per sample), or having secondary alignments (mean = 0.3% per sample). This removed an average of 18.0% of the mapped reads from our analysis, yielding an average of 212,135 total reads pass filter per sample (Table 1; Fig. S1; Table S5). The average on-target enrichment (defined as the proportion of mapped reads hitting primary targets, out of the total reads obtained from sequencing) was 61.5%.

The sequence capture was successful in enriching for neutral microsatellite and functional immune-gene regions. Therefore, we considered two variables when deciding whether to proceed with preliminary downstream microsatellite and SNP analyses: sample size and sequencing depth of coverage. Current recommendations for the minimum depth of coverage to accurately determine genotypes and call SNPs vary. While Nielsen, Paul, Albrechtsen, and Song (2011) suggested >20× coverage to detect heterozygotes, ~40× coverage has also been reported for a 95% SNP detection sensitivity (Meynert, Ansari, FitzPatrick, & Taylor, 2014). We calculated depth of coverage on a per-sample basis (Fig. S2a; Table S6) and on a per-locus basis (Fig. S2b; Table S7) for the primary target regions corresponding to the

regulatory regions and exons for the immune-gene panel, the microsatellite markers, and amelogenin intron 1. The sample libraries and loci had an average depth of coverage of 49.8× (min = 10.4×, max = 93.6×) and 50.2× (min = 15.3× and max = 247.5×), respectively, which exceed suggested minimum standards. In future experiments, sequencing depth could be increased by increasing the number of MiSeq runs or by using a different platform (e.g., HiSeq 2500) to produce more reads. We decided to proceed with preliminary downstream analyses of the microsatellite and SNP markers to test for genetic structure and evidence of local adaptation.

### 3.3 | INDEL detection and microsatellite analysis

Of the 32 libraries processed for downstream analyses, we detected a single bi-allelic INDEL for the amelogenin sex-determining region and detected 25 males and seven females (Table S1). There was a skewed sex ratio and a peak in coverage for the amelogenin locus. Using independent field information (Table S1), we were only able to validate 15/21 of these calls; therefore, we cannot rule out the possibility that there are possible amelogenin paralogs in the red fox genome. In analyzing this filtered INDEL .vcf file, the microsatellite subdataset had a coverage of ~48× and 427/480 (~89%) of the possible genotypes were called with a 99% probability of being true (based on Phred scores). Our dataset partially overlapped the Goldsmith et al. (2016) dataset in having 21 of the same red foxes and six of the same loci (AHT121, AHTh171, CPH9, CPH15, REN105L03, REN247M23). We estimated concordance between the two studies of ~46.7% and noted a large difference in the percent of per-locus homozygote genotypes between our dataset (1.6%) and the previously published dataset (28.1%; Goldsmith et al., 2016). A PCA plot based on the microsatellite dataset did not reveal genetic structure between red fox from coastal tundra regions where arctic rabies is present and the boreal southcentral regions where rabies is absent (Fig. S3a).

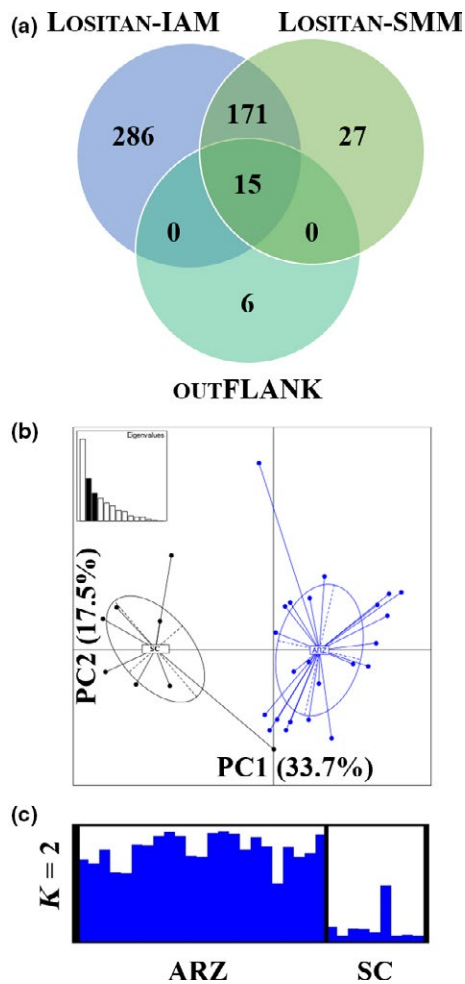
### 3.4 | SNP detection and analysis

We used the mapped reads from 32 libraries to generate a hard-filtered SNP dataset corresponding to the exon regions (932 SNPs), intron regions (2,829 SNPs), and upstream regulatory regions (1,058 SNPs). We used the entire SNP dataset (4,819 SNPs) to conduct PCA and then to detect  $F_{ST}$  outliers. PCA revealed a lack of genetic structure between red fox from coastal tundra regions where arctic

**TABLE 1** Sequence alignment summary statistics reveal high numbers of mapped reads pass GATK filters for each sample

	Reads	Mapped reads (%)	Mapped reads filtered (total) (%)	Mapped reads filtered (duplicates) (%)	Mapped reads filtered (mapping quality) (%)	Mapped reads filtered (not primary alignment) (%)	Mapped reads passing filter
Median	253,908	93.4	16.9	3.2	13.6	0.2	199,258
Mean	274,526	93.5	18.0	4.3	13.4	0.3	212,135
Minimum	98,466	90.5	15.0	2.3	11.3	0.2	47,235
Maximum	502,068	95.1	47.0	32.6	15.3	0.5	388,871

rabies is present and the boreal southcentral regions where rabies is absent (Fig. S3b). We identified  $F_{ST}$  outliers under directional selection using *LOSITAN* and *OUTFLANK* (FDR < 0.05; Figure 2a; Table S8), but zero  $F_{ST}$  outliers were identified using *BAYESCAN* (FDR < 0.05; data not shown). Furthermore, 1,195 SNPs were identified as candidates under balancing selection by both mutation models in *LOSITAN* (Table S8). There were 15  $F_{ST}$  outliers predicted by both mutation models in *LOSITAN* and *OUTFLANK* (Table 2). We used the genotypes from these 15  $F_{ST}$  outliers to visualize genetic structure by PCA and *STRUCTURE* (Figure 2b,c). Preliminary plots revealed genetic structure of red fox between coastal tundra regions where arctic rabies is present compared to red fox from the boreal southcentral regions where rabies is absent.



**FIGURE 2**  $F_{ST}$  outlier tests identify putative signatures of genetic structure in immune-related loci between red fox populations in arctic rabies zones and the arctic rabies-free zone. (a) Concordance between the results of  $F_{ST}$  outlier tests is visualized by the Venn diagram. Genetic structure was visualized by (b) principal component analysis and (c) *STRUCTURE* plots for  $K = 2$ , using the 15  $F_{ST}$  outlier SNPs identified by all three tests in (a). The percentage of variation for each principal component axis and a scatter plot of eigenvalues (inset) are included in the principal component analysis. IAM, infinite alleles model; SMM, stepwise mutation model; ARZ, arctic rabies zones (blue); SC, Southcentral (arctic rabies-free zone; black)

## 4 | DISCUSSION

In this study, we annotated red fox immune genes and developed a GBS target capture protocol that can be used to elucidate spatial patterns of genetic variation at neutral and functional loci of red fox. For the functional loci, we targeted genes associated with an innate, intrinsic, or adaptive immune response, as well as their regulatory segments that have been implicated in variable responses to disease exposure. We developed a system that accurately and evenly captured the targeted loci with sufficient coverage (mean = 50 $\times$ ; Fig. S2, Tables S5 and S6) for variant detection. Without a complete red fox genome assembly, we acknowledge the possibility of excess heterozygotes for a small proportion of our SNPs could be due to paralogous loci; however, our probe design tests suggested only one hit for most probes (91.5%). We used these data to perform a preliminary assessment of  $F_{ST}$  outliers. The vast majority of outliers were found to be under balancing selection, with a smaller subset under directional selection (Tables 2 and S8). Despite moderate sample sizes, our preliminary analyses found notable differences in the frequencies of the  $F_{ST}$  outliers in regions with and without AR. This assay provides a means to elucidate genetic variation from a large portion of the immunome with even coverage across samples and loci while alleviating some of the ascertainment bias of other genotype-by-sequencing approaches.

### 4.1 | Development of the red fox target capture genotyping-by-sequencing assay

Arctic rabies variants in Alaska are restricted to the coastal tundra regions coinhabited by arctic fox and red fox. Conversely, boreal interior regions inhabited by red fox, but not arctic fox, are devoid of endemic AR (Kuzmin et al., 2008; Nadin-Davis et al., 2012). Goldsmith et al. (2016) recently sequenced microsatellite markers from Alaska red fox. They observed high levels of admixture within populations and genetic structure between the coastal tundra regions and the boreal southcentral regions, consistent with the geographic distribution of AR. However, their study was limited to neutral marker analyses, and genetic diversity in arctic fox and red fox at functional loci that may be influenced by selective pressure from AR is unknown and the motivation to develop this GBS immunogenetic assay.

As the cost of HTS continues to decline, strategies have emerged to gain insight into the genetic diversity of regions of the genome that encode proteins, which are presumably functionally relevant for local adaptation of wildlife species. For example, to assess the genetic basis for adaptation of arctic foxes to a cold climate, a comparison of transcriptome sequences from two captive arctic fox and one red fox identified two fat metabolism genes under positive selection in the arctic fox transcriptome (Kumar, Kutschera, Nilsson, & Janke, 2015). However, isolating quality RNA from wildlife samples is problematic given the many environmental and individual variables associated with RNA expression, and the difficulty in obtaining fresh samples from a broad geographic range in regions that are difficult to access. Recently, GBS assays have emerged as a cost-effective alternative to whole-genome sequencing that aims to detect a genomic basis for population

**TABLE 2** Directional outliers detected by LOSITAN and OUTFLANK analyses (FDR < 0.05)

Gene	Immune-gene sequence name	Position	Location	Gene description
C3	Fox_ENSCAFG00000018625	1,558	Exon	Complement C3
C3	Fox_ENSCAFG00000018625	18,625	Intron	Complement C3
C3	Fox_ENSCAFG00000018625	2,190	Intron	Complement C3
C3	Fox_ENSCAFG00000018625	8,750	Intron	Complement C3
CRP	Fox_ENSCAFG00000011787	2,855	Exon	C-reactive protein precursor
DLA-DMA	Fox_ENSCAFG00000000848	2,510	Intron	Major histocompatibility complex, class II, DM alpha isoform X1
IL10	Fox_ENSCAFG00000011443	2,198	Intron	Interleukin-10
IL23R	Fox_ENSCAFG00000018542	40,066	Intron	Interleukin-23 receptor
IL23R	Fox_ENSCAFG00000018542	43,528	Intron	Interleukin-23 receptor
IL23R	Fox_ENSCAFG00000018542	52,788	Intron	Interleukin-23 receptor
ITGAM	Fox_ENSCAFG00000016881	10,880	Intron	Integrin alpha-M isoform X1
ITGAM	Fox_ENSCAFG00000016881	53,170	Intron	Integrin alpha-M isoform X1
NLRP3	Fox_ENSCAFG00000010686	44,328	Intron	NACHT, LRR and PYD domains-containing protein 3 isoform X1
NLRP3	Fox_ENSCAFG00000010686	44,349	Intron	NACHT, LRR and PYD domains-containing protein 3 isoform X1
TLR7	Fox_ENSCAFG00000011698	24,027	Exon	Toll-like receptor 7 isoform X1

variation and provide genetic signatures of local adaptation. For example, RAD sequencing found extremely low genetic variation among populations of distinct subspecies of Island fox (*Urocyon littoralis*) in southern California, an observation that can be used to guide management decisions (Funk et al., 2016). While useful in studies involving nonmodel species, RAD sequencing is not restricted to detecting  $F_{ST}$  outliers in protein-coding regions of the genome. Therefore, targeted GBS approaches using either amplicon- or hybridization-based workflows have been employed to identify phenotype-altering mutations in both regulatory and protein-coding regions.

While amplicon-based methods may require more straightforward sample preparation and have the ability to utilize smaller DNA inputs, hybridization-based approaches have performed slightly better in sequencing complexity and uniformity with respect to target enrichment (Samorodnitsky et al., 2015). The decision to use either approach may be a matter of user preference, and both have been successful in wildlife applications. Using the amplicon-based approach, Morris et al. (2015) found low levels of polymorphism in the Tasmanian devil. However, it is noteworthy that their amplicon design relied on the whole-genome sequence comparison of ten Tasmanian devils, which may be cost-prohibitive as a model for other wildlife investigations. Moreover, sample bias may have affected their choice in amplicon gene targets. Target capture has been used to identify functional variation in gray wolf ecotypes in North America (Schweizer, Robinson et al., 2016); however, the gene targets were first identified by supporting information from an established Affymetrix v2 Canine SNP array that included ~42K SNPs (Schweizer, Vonholdt et al., 2016). In this study, we designed a custom target capture assay for red fox to screen for

genetic variation in functional immune genes (Table S2), including their regulatory regions, and in neutral microsatellite regions (Table S3). The immune-gene annotations and their corresponding protein sequences (Appendices S1 and S2) helped validate a portion of the red fox genome assembly draft and are a resource for future immunome studies in red fox and related fox species. Our goal was to use these probes to screen for genomic signatures of local adaptation between tundra and boreal red fox from different AR regions in Alaska.

#### 4.2 | Target capture is successful with low copy number wildlife DNA and across species

For GBS, Samorodnitsky et al. (2015) suggested higher amounts of input DNA required for hybridization-based Nimblegen SeqCap (1 µg), relative to amplicon-based Illumina AmpliSeq (50 ng), is a limiting factor. The difference in the amount of template required for these respective methods is of concern for wildlife studies, which can include low-quality or low-template DNA; however, our results indicated the NimbleGen SeqCap worked equally well for a dilution series of sample KH1354 (1, 10, 100 ng, and 1 µg of input DNA). While the estimated target enrichment varied from 60.1% to 66.7% (Table S5), the 1 ng input DNA sample had the highest mean target coverage (66.6×; Table S6) of the dilution series. Coincidentally, during preparation of this manuscript, NimbleGen released new guidelines (SeqCap EZ Library SR v5.0), which now recommend 100 ng of input DNA. Illumina also now offers a low input library preparation kit (10 ng). Our results are encouraging for future experiments where samples may contain low copy number DNA (e.g., from hair/fur traps), as 1 ng of input DNA has the



potential to generate GBS data using a single MiSeq run; however, optimization of such experiments may require more intense investigations.

We estimated a target enrichment of 61.5% and 59.9% (Table S5) and a mean target coverage of 50.1 $\times$  and 37.0 $\times$  (Table S6) for red fox and arctic fox, respectively. Similarly, cross-species exon target capture has been successful for a broad range of cichlids (Ilves & López-Fernández, 2014). Red fox and arctic fox recently diverged from one another ~3.2 mya during the late Pliocene (Kumar et al., 2015), and these data suggest that our existing probe set can be used in future studies aimed at addressing similar questions regarding genetic variation in arctic fox or related fox species.

### 4.3 | Analysis of microsatellites

Genotyping microsatellite regions using PCR-based techniques can be difficult because the polymerase can “slip” at STRs, leading to amplicons that differ in length. Additionally, STR-containing reads from HTS data can be difficult to accurately map to a reference genome due to mismatch/INDEL penalties associated with STR expansion and contraction (Fungtammasan et al., 2015). Therefore, we manually inspected the bi-allelic INDELS of our 23 microsatellite targets for poor alignments, which generated a filtered microsatellite dataset containing 15 loci. Microsatellite genotypes from dinucleotide STRs have been accurately called with a 90% success rate given a 17 $\times$  depth of coverage (Fungtammasan et al., 2015), and our filtered subset had an average depth of coverage of 48 $\times$ . We did not detect genetic structure in this microsatellite dataset using STRUCTURE or ADEGENET PCA plots (Fig. S3a). These results did not fully align with genotypes from the same samples generated using traditional microsatellite amplification and profiling using a genetic analyzer by Goldsmith et al. (2016). Our observed heterozygosity is also high compared to reported estimates of expected heterozygosity in Polish red fox (72%, Mullins et al., 2014) and that of other carnivores, including the American badger (81%; Rico et al., 2016) and Canadian black bear (55%–81%; Pelletier et al., 2017). Possible causes of the conflicting results include amplification bias and allelic dropout in their dataset that is suggested by the large difference in observed homozygotes (1.6% vs. 28.1%) or challenges in our dataset in aligning sequences to generate accurate genotypes and hence the larger number of heterozygotes. We did not detect significant departures from HWE for any of the microsatellite loci used in our analyses; clearly, there are challenges associated with genotyping microsatellite regions using HTS. De Barba et al. (2017) recently performed high-throughput microsatellite genotyping of black bear samples and reported improved genotyping success compared to traditional methods. Alternatively, including several hundred, independent, random genomic regions containing SNPs to determine background genomic variation, could mitigate some of the technical and bioinformatic limitations we encountered.

### 4.4 | Target capture analysis of immunogenetic diversity—considerations

We used LOSITAN, BAYESCAN, and OUTFLANK to detect candidate outlier SNPs putatively under selection. While LOSITAN and OUTFLANK

detected  $F_{ST}$  outliers (Figure 2a; Tables 2 and S8), BAYESCAN did not. Narum and Hess (2011) reported lower type I (false positive) and type II (false negative) error rates for LOSITAN and BAYESCAN compared to ARLEQUIN, but false positives for candidate SNPs under directional and balancing selection are abundant with all three approaches. Further, these methods have limited sensitivity in detecting SNPs under weak selection (Narum & Hess, 2011). We also detected 1,195 SNPs under balancing selection dispersed through exons, introns, and regulatory regions of the red fox immunome (Table S8). Signatures of balancing selection in the context of response to pathogens have been found in human immune-related genes, including those associated with the major histocompatibility complex (MHC; Andrés et al., 2009), and in the MHC of wolverine (*Gulo gulo*; Rico et al., 2015). MHC genes are highly polymorphic and play an important role in the adaptive immune response to pathogens. Therefore, our findings are consistent with previous studies in finding candidate immune-related loci under balancing selection. While these results are promising, they should be viewed with caution as we concede that there is the potential for bias in the outlier analyses because the initial dataset was not a random distribution of genome-wide loci that would include both neutral and adaptive alleles. We also acknowledge that our tests for directional and balancing selection likely contain biases associated with linked loci because all of the SNPs cannot be considered as independent markers.

Outlier SNPs in exons under directional selection that were detected by both LOSITAN and OUTFLANK included C3, CRP, and TLR7 (Table 2). C3 encodes an activator in the complement system, which is involved in innate and adaptive immune responses and functions to lyse microorganisms, promote phagocytosis, trigger inflammation and aids in immune clearance (www.genecards.org). CRP recognizes foreign pathogens and promotes their elimination (www.genecards.org). TLR7 is a toll-like receptor that recognizes single-stranded RNA and activates the innate immune system (www.genecards.org). The candidate SNP in TLR7 may be an intriguing target for future studies because AR is a single-stranded RNA virus. The preliminary interpretation of our data suggests selective pressure at the molecular level on three immune genes in red fox. Overall, the SNP-based analysis found no evidence of population structure among red fox using all SNPs (Fig. S3b), but indicated population structure based on  $F_{ST}$  outliers only, where red fox from coastal tundra regions were distinct from those in the boreal Southcentral regions (Figure 2b,c). To support these preliminary findings, future studies are required using larger sample sizes from Alaska red fox that compare rabies-negative and rabies-positive individuals from the same population and they could help reveal a genetic link to rabies susceptibility/resistance.

## 5 | CONCLUSION

Understanding the capacity for local adaptation to disease in wild populations requires expanded genomic assessments of the genetic responses to these selective pressures. Herein, we developed an immunogenetic assay that bridges between genetic and full-genome research and has the potential to generate empirical data that set the

basis for predictive models to enhance our ability to anticipate epizootic disease spread and impacts under different climatic scenarios.

## ACKNOWLEDGMENTS

This research was funded by discovery grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to C.J.K., a National Science Foundation Division of Graduate Education Integrative Graduate Education and Research Traineeship grant (NSF DGE IGERT-1069157) to H.R., and a National Institutes of Health grant (GM120782) to A.V.K. We thank Caroline Drago, Katie Dogantzis, and Katelyn Jackson at Trent University for technical assistance; Matt Harnden from the Natural Resources DNA Profiling and Forensics Centre at Trent University for technical assistance; Mena Farag from Roche for helping revise the multiplexed target capture protocol; Tammy Craig, Matthew Smith, and Nick Downey from Integrated DNA Technologies for trouble-shooting library preparation; Aaron Goldman, Graham Cromar, Roger Shi, and Siwei Chen from The Clinical Genomics Centre at Mount Sinai Hospital in Toronto for performing the DNA library preparation and Illumina sequencing; Dr. Barry Saville for providing the computer capable of analyzing next-generation sequence data and Aaron Shafer for critical comments of our manuscript prior to submission; and two anonymous reviewers for critique leading to an improved version of our manuscript.

## CONFLICT OF INTEREST

None declared.

## AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: MED CJK KH. Red Fox tissue was assembled and provided by: KH. Performed the experiments: MED. Analyzed the data: MED YR. Contributed reagents/materials/analysis tools: CJK. Provided early-access to red fox genome sequence: HMR AVK. Wrote the manuscript: MED CJK. Revised the manuscript: MED CJK KH HMR AVK.

## DATA ACCESSIBILITY

High-throughput sequence data are available through the NCBI Sequence Read Archive (accession number SRP119314). The following information is available on the Dryad Digital Repository (<https://doi.org/10.5061/dryad.f81c5>): the subset of red fox immune-gene sequences (2,229,152 bp; .fasta) that were used to design the sequence capture assay, the corresponding coordinates (.bed) used to design the overlapping NimbleGen SeqCap EZ probes, and the resulting SNP variant call (.vcf) information.

## ORCID

Michael E. Donaldson  <http://orcid.org/0000-0002-6779-1455>

## REFERENCES

- Acevedo-Whitehouse, K., & Cunningham, A. A. (2006). Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution*, 21, 433–438. <https://doi.org/10.1016/j.tree.2006.05.010>
- Albrechtsen, A., Nielsen, F. C., & Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution*, 27, 2534–2547. <https://doi.org/10.1093/molbev/msq148>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- An, J., Bechet, A., Berggren, Å. S. A., Brown, S. K., Bruford, M. W., Cai, Q., ... Zhong, J. (2010). Permanent genetic resources added to molecular ecology resources database 1 October 2009–30 November 2009. *Molecular Ecology Resources*, 10, 404–408. <https://doi.org/10.1111/j.1755-0998.2009.02827.x>
- Andrés, A. M., Hubisz, M. J., Indap, A., Torgerson, D. G., Degenhardt, J. D., Boyko, A. R., ... Nielsen, R. (2009). Targets of balancing selection in the human genome. *Molecular Biology and Evolution*, 26, 2755–2764. <https://doi.org/10.1093/molbev/msp190>
- Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, 9, 1–5. <https://doi.org/10.1186/1471-2105-9-323>
- Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, 263, 1619–1626. <https://doi.org/10.1098/rspb.1996.0237>
- Beldomenico, P. M., & Begon, M. (2016). Disease spread, susceptibility and infection intensity: Vicious circles? *Trends in Ecology & Evolution*, 25, 21–27. <https://doi.org/10.1016/j.tree.2009.06.015>
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, 27, 573–580. <https://doi.org/10.1093/nar/27.2.573>
- Bonneaud, C., Balenger, S. L., Zhang, J., Edwards, S. V., & Hill, G. E. (2012). Innate immunity and the evolution of resistance to an emerging infectious disease in a wild bird. *Molecular Ecology*, 21, 2628–2639. <https://doi.org/10.1111/j.1365-294X.2012.05551.x>
- Carlson, S. M., Cunningham, C. J., & Westley, P. A. H. (2016). Evolutionary rescue in a changing world. *Trends in Ecology & Evolution*, 29, 521–530. <https://doi.org/10.1016/j.tree.2014.06.005>
- Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, 17, 362–365. <https://doi.org/10.1111/men.2017.17.issue-3>
- Chhatre, V. E., & Emerson, K. J. (2017). StrAuto: Automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*, 18, 192. <https://doi.org/10.1186/s12859-017-1593-0>
- Chilamakuri, C. S. R., Lorenz, S., Madoui, M.-A., Vodák, D., Sun, J., Hovig, E., ... Meza-Zepeda, L. A. (2014). Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics*, 15, 1–14. <https://doi.org/10.1186/1471-2164-15-449>
- Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., ... Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Research*, 43, D662–D669. <https://doi.org/10.1093/nar/gku1010>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- De Barba, M., Miquel, C., Lobréaux, S., Quenette, P. Y., Swenson, J. E., & Taberlet, P. (2017). High-throughput microsatellite genotyping in ecology: Improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, 17, 492–507. <https://doi.org/10.1111/men.2017.17.issue-3>

- De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., ... Palumbi, S. R. (2012). The simple fool's guide to population genomics via RNA-Seq: An introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources*, 12, 1058–1067. <https://doi.org/10.1111/1755-0998.12003>
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43, 491–498. <https://doi.org/10.1038/ng.806>
- Earl, D. A., & vonHoldt, B. M. (2011). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Eizaguirre, C., Lenz, T. L., Kalbe, M., & Milinski, M. (2012). Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nature Communications*, 3, 621. <https://doi.org/10.1038/ncomms1632>
- Eklom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, 7, 1026–1042. <https://doi.org/10.1111/eva.2014.7.issue-9>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620. <https://doi.org/10.1111/mec.2005.14.issue-8>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Fraser, H. B. (2013). Gene expression drives local adaptation in humans. *Genome Research*, 23, 1089–1096. <https://doi.org/10.1101/gr.152710.112>
- Fredholm, M., & Winterø, A. K. (1995). Variation of short tandem repeats within and between species belonging to the Canidae family. *Mammalian Genome*, 6, 11–18. <https://doi.org/10.1007/BF00350887>
- Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Pattini, L., & Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, 7, e1002355. <https://doi.org/10.1371/journal.pgen.1002355>
- Fungtammasan, A., Ananda, G., Hile, S. E., Su, M. S., Sun, C., Harris, R., ... Makova, K. D. (2015). Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Research*, 25, 736–749. <https://doi.org/10.1101/gr.185892.114>
- Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Silleet, T. S., ... Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: Genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). *Molecular Ecology*, 25, 2176–2194. <https://doi.org/10.1111/mec.2016.25.issue-10>
- Gallana, M., Ryser-Degiorgis, M.-P., Wahli, T., & Segner, H. (2013). Climate change and infectious diseases of wildlife: Altered interactions between pathogens, vectors and hosts. *Current Zoology*, 59, 427–437. <https://doi.org/10.1093/czoolo/59.3.427>
- Goldsmith, E. W., Renshaw, B., Clement, C. J., Himschoot, E. A., Hundertmark, K. J., & Hueffer, K. (2016). Population structure of two rabies hosts relative to the known distribution of rabies virus variants in Alaska. *Molecular Ecology*, 25, 675–688. <https://doi.org/10.1111/mec.13509>
- Hansen, M. M., Olivieri, I., Waller, D. M., & Nielsen, E. E. (2012). Monitoring adaptive genetic responses to environmental change. *Molecular Ecology*, 21, 1311–1329. <https://doi.org/10.1111/j.1365-294X.2011.05463.x>
- Harrisson, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, 7, 1008–1025. <https://doi.org/10.1111/eva.2014.7.issue-9>
- Hess, J. E., Campbell, N. R., Close, D. A., Docker, M. F., & Narum, S. R. (2013). Population genomics of Pacific lamprey: Adaptive variation in a highly dispersive species. *Molecular Ecology*, 22, 2898–2916. <https://doi.org/10.1111/mec.12150>
- Hoepfner, M. P., Lundquist, A., Pirun, M., Meadows, J. R., Zamani, N., Johnson, J., ... Grabherr, M. G. (2014). An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts (BP Chadwick, Ed.). *PLoS One*, 9, e91172. <https://doi.org/10.1371/journal.pone.0091172>
- Ivles, K. L., & López-Fernández, H. (2014). A targeted next-generation sequencing toolkit for exon-based cichlid phylogenomics. *Molecular Ecology Resources*, 14, 802–811. <https://doi.org/10.1111/men.2014.14.issue-4>
- Jackson, A. C., & Wunner, W. H. (2007). *Rabies: Scientific basis of the disease and its management*. Burlington, MA: Academic Press.
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, 25, 185–202. <https://doi.org/10.1111/mec.13304>
- Kim, B. I., Blanton, J. D., Gilbert, A., Castrodale, L., Hueffer, K., Slate, D., & Rupprecht, C. E. (2014). A conceptual model for the impact of climate change on fox rabies in Alaska, 1980–2010. *Zoonoses and Public Health*, 61, 72–80. <https://doi.org/10.1111/zph.2014.61.issue-1>
- Knight, J. C. (2013). Genomic modulators of the immune response. *Trends in Genetics*, 29, 74–83. <https://doi.org/10.1016/j.tig.2012.10.006>
- Kumar, V., Kutschera, V. E., Nilsson, M. A., & Janke, A. (2015). Genetic signatures of adaptation revealed from transcriptome sequencing of Arctic and red foxes. *BMC Genomics*, 16, 585. <https://doi.org/10.1186/s12864-015-1724-9>
- Kuzmin, I. V., Hughes, G. J., Botvinkin, A. D., Gribencha, S. G., & Rupprecht, C. E. (2008). Arctic and Arctic-like rabies viruses: Distribution, phylogeny and evolutionary history. *Epidemiology and Infection*, 136, 509–519. <https://doi.org/10.1017/S095026880700903X>
- Kyle, C. J., Rico, Y., Castillo, S., Srithayakumar, V., Cullingham, C. I., White, B. N., & Pond, B. A. (2014). Spatial patterns of neutral and functional genetic variations reveal patterns of local adaptation in raccoon (*Procyon lotor*) populations exposed to raccoon rabies. *Molecular Ecology*, 23, 2287–2298. <https://doi.org/10.1111/mec.12726>
- Lachance, J., & Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 35, 780–786. <https://doi.org/10.1002/bies.201300014>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1303, 1–3.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28, 298–299. <https://doi.org/10.1093/bioinformatics/btr642>
- 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, 142–152. <https://doi.org/10.1111/men.2017.17.issue-2>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A

- MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Mellersh, C. S., Langston, A. A., Acland, G. M., Fleming, M. A., Ray, K., Wiegand, N. A., ... Ostrander, E. A. (1997). A linkage map of the canine genome. *Genomics*, 46, 326–336. <https://doi.org/10.1006/geno.1997.5098>
- Meynert, A. M., Ansari, M., FitzPatrick, D. R., & Taylor, M. S. (2014). Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics*, 15, 1–11. <https://doi.org/10.1186/1471-2105-15-247>
- Mørk, T., & Prestrud, P. (2004). Arctic rabies – A review. *Acta Veterinaria Scandinavica*, 45, 1. <https://doi.org/10.1186/1751-0147-45-1>
- Morris, K. M., Wright, B., Grueber, C. E., Hogg, C., & Belov, K. (2015). Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (*Sarcophilus harrisii*). *Molecular Ecology*, 24, 3860–3872. <https://doi.org/10.1111/mec.2015.24.issue-15>
- Mullins, J., McDevitt, A. D., Kowalczyk, R., Ruczyńska, I., Górny, M., & Wójcik, J. M. (2014). The influence of habitat structure on genetic differentiation in red fox populations in north-eastern Poland. *Acta Theriologica*, 59, 367–376. <https://doi.org/10.1007/s13364-014-0180-2>
- Nadin-Davis, S. A., Sheen, M., & Wandeler, A. I. (2012). Recent emergence of the Arctic rabies virus lineage. *Virus Research*, 163, 352–362. <https://doi.org/10.1016/j.virusres.2011.10.026>
- Narum, S. R., & Hess, J. O. N. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11, 184–194. <https://doi.org/10.1111/j.1755-0998.2011.02987.x>
- Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., ... Shendure, J. (2009). Targeted capture and massively parallel sequencing of twelve human exomes. *Nature*, 461, 272–276. <https://doi.org/10.1038/nature08250>
- Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, 12, 443–451. <https://doi.org/10.1038/nrg2986>
- Orsini, L., Andrew, R., & Eizaguirre, C. (2013). Evolutionary ecological genomics. *Molecular Ecology*, 22, 527–531. <https://doi.org/10.1111/mec.12177>
- Ostrander, E. A., Sprague, G. F., & Rine, J. (1993). Identification and characterization of dinucleotide repeat (CA)<sub>n</sub> markers for genetic mapping in dog. *Genomics*, 16, 207–213. <https://doi.org/10.1006/geno.1993.1160>
- Pelletier, A., Obbard, M. E., Harnden, M., McConnell, S., Howe, E. J., Burrows, F. G., ... Kyle, C. J. (2017). Determining causes of genetic isolation in a large carnivore (*Ursus americanus*) population to direct contemporary conservation measures. *PLoS One*, 12, e0172319. <https://doi.org/10.1371/journal.pone.0172319>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559–575. <https://doi.org/10.1086/519795>
- Rico, Y., Ethier, D. M., Davy, C. M., Sayers, J., Weir, R. D., Swanson, B. J., ... Kyle, C. J. (2016). Spatial patterns of immunogenetic and neutral variation underscore the conservation value of small, isolated American badger populations. *Evolutionary Applications*, 9, 1271–1284. <https://doi.org/10.1111/eva.2016.9.issue-10>
- Rico, Y., Morris-Pocock, J., Zigorris, J., Nocera, J. J., & Kyle, C. J. (2015). Lack of spatial immunogenetic structure among wolverine (*Gulo gulo*) populations suggestive of broad scale balancing selection. *PLoS One*, 10, e0140170. <https://doi.org/10.1371/journal.pone.0140170>
- Rosenberg, N. A. (2004). distruct: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8, 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Samorodnitsky, E., Jewell, B. M., Hagopian, R., Miya, J., Wing, M. R., Lyon, E., ... Roychowdhury, S. (2015). Evaluation of hybridization capture versus amplicon-based methods for whole-exome sequencing. *Human Mutation*, 36, 903–914. <https://doi.org/10.1002/humu.2015.36.issue-9>
- Schoville, S. D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012). Adaptive genetic variation on the landscape: Methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, 43, 23–43. <https://doi.org/10.1146/annurev-ecolsys-110411-160248>
- Schweizer, R. M., Robinson, J., Harrigan, R., Silva, P., Galverni, M., Musiani, M., ... Wayne, R. K. (2016). Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in gray wolves. *Molecular Ecology*, 25, 357–379. <https://doi.org/10.1111/mec.13467>
- Schweizer, R. M., Vonholdt, B. M., Harrigan, R., Knowles, J. C., Musiani, M., Coltman, D., ... Wayne, R. K. (2016). Genetic subdivision and candidate genes under selection in North American gray wolves. *Molecular Ecology*, 25, 380–402. <https://doi.org/10.1111/mec.13364>
- Smith, A. L., Hewitt, N., Klenk, N., Bazely, D. R., Yan, N., Wood, S., ... Lipsig-Mummé, C. (2012). Effects of climate change on the distribution of invasive alien species in Canada: A knowledge synthesis of range change projections in a warming world. *Environmental Reviews*, 20, 1–16. <https://doi.org/10.1139/a11-020>
- Solovyev, V. V. (2007). Statistical approaches in Eukaryotic gene prediction. In D. J. Balding, M. Bishop & C. Cannings (Eds.), *Handbook of statistical genetics* (3rd ed., pp. 97–159). Chichester, UK: Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470061619.ch4>
- Thompson, J. N. (2005). *The geographic mosaic of coevolution*. Chicago, IL: University of Chicago Press.
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43, 11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>
- 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 FST. *The American Naturalist*, 186, S24–S36. <https://doi.org/10.1086/682949>

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Donaldson ME, Rico Y, Hueffer K, Rando HM, Kukekova AV, Kyle CJ. Development of a genotype-by-sequencing immunogenetic assay as exemplified by screening for variation in red fox with and without endemic rabies exposure. *Ecol Evol*. 2018;8:572–583. <https://doi.org/10.1002/ece3.3583>