Exploring End-User Accessible Microsatellite and Microhaplotype Nanopore Sequencing with the Oxford Nanopore's MinION™ Device

A Thesis Submitted to the Committee on Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in the Faculty of Arts and Science

TRENT UNIVERSITY
Peterborough, Ontario, Canada

Environmental and Life Sciences M.Sc. Graduate Program
April 2025

ABSTRACT

Exploring End-User Accessible Microsatellite and Microhaplotype Nanopore Sequencing with the Oxford Nanopore's MinION™ Device

Ella Clarke

Microsatellites and microhaplotypes are genetic markers that, through DNA amplicon sequencing, generate genotypes to distinguish between individuals from wildlife populations. Here, these markers were sequenced in caribou (*Rangifer tarandus*) specimens using Oxford Nanopore's MinION DNA sequencer for the first time. Microsatellite loci previously sequenced with an Illumina MiSeq were compared to MinION Mk1B sequencing data for the same samples/loci, revealing highly consistent microsatellite characterization across platforms. Additionally, a novel panel of caribou microhaplotype loci was developed and sequenced on the MinION Mk1B and Illumina MiSeq. Microhaplotype characterization of the same samples revealed that ambiguous read distributions for the top 3 reads per locus is a key challenge, particularly for the MinION, that hinders concordant haplotype calls across platforms. Potential reasons for this ambiguity include duplicated gene regions and PCR errors. Removing suspected duplicated gene regions and reducing the number of PCR cycles during target DNA amplification may mitigate this problem.

Keywords: Conservation Genomics, DNA Profiling, Nanopore Sequencing, MinION™, Genetics, Microsatellites, Microhaplotypes, Real-time Sequencing, Cost-effective, Wildlife Monitoring.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my co supervisors, Paul Wilson and Micheline Manseau, for their ongoing support, guidance, and encouragement throughout my graduate studies. I would also like to thank Derek Smith for generously sharing his nanopore sequencing expertise, allowing me to establish a new approach to caribou DNA sequencing at Trent University. Their insights greatly helped shape the direction of my thesis and their mentorship has been invaluable to the successful completion of this research. A big thank you to my colleagues and lab mates for their camaraderie, collaboration, and for fostering a positive working environment. Thank you, Bridget Redquest, Keta Patel, Austin Thompson, Sonesinh Keobouasone, Peng Liu, Julia Geue and Lisette Delgado-Aquije for your contributions to this project, which include but are not limited to, lab guidance, bioinformatics expertise, technical support, and constructive feedback to develop my thesis project. I extend my gratitude to the wildlife conservation community and the individuals who provided valuable field samples and assistance in data collection, without which this research would not have been possible. Special thanks to the faculty and staff in the biology department at Trent University for providing me with the resources and opportunities to pursue my research. I would like to acknowledge Trent University, National Wildlife Research Centre at Carleton University, Environment and Climate Change Canada, and Genome Canada for making this research possible by contributing caribou samples, data analyses, and financial support. Finally, thank you to my family and friends who encouraged me in the pursuit of my goals. I would like to thank my brother Manny for inspiring my interest in genetics and routinely checking in with me, asking how the caribou lab is going.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

DNA Profiling to Monitor Wildlife

Genetic markers offer unparalleled precision and efficiency in wildlife monitoring, revolutionizing our ability to track and study species diversity, population dynamics, and ecological interactions (Allendorf et al., 2010; Carroll et al., 2018; Schwartz et al., 2007). With continued advancements in human and wildlife genomic approaches, the ability to perform DNA profiling (identify individuals based on unique genetic markers) is greatly enhanced for many species (Allendorf et al., 2010; Ogden et al., 2009; Von Thaden et al., 2020). In the context of wildlife monitoring, analyzing the DNA profiles of individuals within a population provides insights into various wildlife population parameters such as genetic variation, population size, structure, connectivity, relatedness, genomic erosion, and adaptive potential (Carroll et al., 2018; Hohenlohe et al., 2020; Jones & Manseau, 2022; McFarlane et al., 2021; Schwartz et al., 2007).

By comparing the DNA profiles of individuals within a population, genetic variation (the number of different alleles and the frequency of each allele in a population) can be assessed (Deyoung & Honeycutt, 2005). High genetic variation indicates a diverse and healthy population, while low genetic variation may suggest inbreeding or a reduced ability to adapt to environmental changes (Giglio et al., 2016). DNA profiling also helps identify patterns of relatedness within a population (Jones & Manseau, 2022; Pemberton, 2008). Through contrasting DNA profiles, familial relationships are inferred, such as parent-offspring or sibling connections (Baetscher et al., 2018; Fournier et al., 2024; Jones et al., 2023; McFarlane et al., 2021). Measuring relatedness is vital for understanding breeding dynamics, dispersal patterns, and gene flow within and between populations (Arif et al., 2011; De Woody, 2005). Moreover, DNA

profiling identifies genetically distinct groups within a population based on differences at polymorphic loci (Arif et al., 2011; Ball et al., 2010). Overall, insight obtained from DNA profiles is crucial for conservation efforts as it helps determine the status of specific populations and guides the development of management strategies to maintain genetic diversity and minimize the risk of extinction (Hohenlohe et al., 2020).

Caribou (*Rangifer tarandus*) is a vulnerable species in Canada that benefits from DNA profiling in a monitoring context (www.EcoGenomicsCanada.ca). Caribou face a multitude of challenges, including habitat alteration, high predation risk via apparent competition, resource exploitation, and the effects of climate change, which collectively threaten the species' long-term survival (DeCesare et al., 2013; Ehlers et al., 2016; Environment and Climate Change Canada, 2023; Frenette et al., 2020). With continued population decline, informed conservation efforts to protect and manage at-risk *R. tarandus* populations are critical (Festa-Bianchet et al., 2011; Frenette et al., 2020; Mumma et al., 2018). Effective conservation and management strategies rely on accurate population assessments and individual identification, demanding robust and scalable genetic profiling methods (Hohenlohe et al., 2020).

DNA Profiling Markers

Individual caribou can be identified by analyzing genetic markers following DNA extraction (Ball et al., 2007). Two types of genetic markers compatible with generating genotypes for individual-identification are microsatellites (Arif et al., 2011; Carroll et al., 2018) and microhaplotypes (Oldoni et al., 2018; Lou et al., 2021). These markers are regions of the genome that exhibit high variability between individuals due to differences in allele sizes, allele sequences, and mutation rates, all contributing to high levels of polymorphism and heterozygosity (Hölzl-Müller et al., 2021; Osborne et al., 2022; Voskoboinik et al., 2018). The number and type of alleles present at each marker are determined through size separation or

genomic sequencing to create a unique genetic profile for each individual (De la Puente et al., 2020; Tytgat et al., 2022).

Microsatellite genotyping involves the analysis of short tandem repeats (STRs or microsatellites), which are repeating patterns of DNA typically composed of 1-6 base pairs (Ellegren, 2004; Rohilla et al., 2020). Traditionally, microsatellite genotyping has been conducted by PCR-amplifying target microsatellite regions followed by size separation, based on variable numbers of repeats, using capillary gel electrophoresis also called genotyping from capillary sizing (GCS) (Carratto et al., 2022; Klütsch et al., 2016; McFarlane et al., 2020; Vieira et al., 2016). However, there has been a recent shift to microsatellite amplicon sequencing to genotype these genetic markers, capturing the full microsatellite sequences and the variation they possess (e.g. SNPs) (Liu et al., 2024; Suez et al., 2015; Zhan et al., 2017). This approach also involves PCR amplification of target microsatellite regions, but these amplicons are then sequenced on a NGS sequencing platform (Carratto et al., 2022). The resulting high coverage amplicon sequences provide increased discriminatory power and facilitate more standardized scoring methods compared to GCS (De Barba et al., 2016; Liu et al., 2024).

Microsatellite markers are highly polymorphic and have been widely used in molecular fingerprinting for individual identification as the specific pattern, length, and number of microsatellites vary between individuals (Arif et al., 2011; De Barba et al., 2016). Microsatellite sequencing has numerous applications in wildlife conservation. It can be used to characterize genetic diversity within and between populations, identify individuals or populations, infer parentage, determine levels of inbreeding or outbreeding, and assess the impact of environmental changes or habitat fragmentation on genetic connectivity (Ball et al., 2010; Carroll et al., 2018; Flasko et al., 2017; Hohenlohe et al., 2020; Jones et al., 2010; Pemberton, 2008; Thaden et al.,

2020). Microsatellite data can identify individuals in wildlife populations based on extensive allelic diversity at different loci and enables comparisons of genetic differentiation among threatened populations (Hohenlohe et al., 2020; Arif et al., 2011). Microsatellite sequencing is also applicable to non-invasively collected DNA (e.g., fecal samples)—which are often the types of samples obtained from species of conservation concern due to the elusive nature of the species and limited population size (Ball et al., 2007; Carroll et al., 2018; Taberlet et al., 1997).

Single nucleotide polymorphisms (SNPs) are valuable complementary markers to microsatellites, offering the absence of stutter artefacts and simplified scoring (De la Puente et al., 2020; Kraus et al., 2014; Oldoni et al., 2018). However, due to the limited number of alleles per locus, typically biallelic resulting from a low mutation rate, individual SNPs lack high discriminating power (Carroll et al., 2018; De la Puente et al., 2020). Many SNPs are required to achieve the same or higher discriminating power than microsatellites, thus capturing numerous SNPs per target locus is desirable to confidently identify individuals while reducing the amount of DNA required and minimizing sequencing costs (Kraus et al., 2014). Microhaplotypes are multi-SNP genetic markers, providing an alternative to large panels of single SNPs (Oldoni et al., 2018; Van der Gaag et al., 2018).

Microhaplotyping involves sequencing specific regions of the genome that contain sets of closely linked SNPs known as microhaplotypes (Kidd et al., 2014; Oldoni et al., 2018). These SNPs are often in strong linkage disequilibrium, so they tend to be inherited together (Huang et al., 2022; Kidd et al., 2014). Since microhaplotypes contain multiple closely linked SNPs, more information, e.g. allelic diversity, can be obtained per locus compared to single SNP markers (Oldoni et al., 2018; Van der Gaag et al., 2018). These markers have shown to be useful in forensic genetics for inferring individual identification, biogeographic ancestry, and mixture

deconvolution (De la Puente et al., 2020; Huang et al., 2022; Kidd et al., 2014; Oldoni et al., 2018). Based on the specific combination of microhaplotypes and alleles present, individual genetic profiles can be created, providing a highly informative and discriminating tool for identification and relatedness analysis (Oldoni et al., 2018; Van der Gaag et al., 2018).

Illumina MiSeq DNA Profiling

Sequencing platforms like the lllumina MiSeq have frequently been used in conservation genomics research due to Illumina's compatibility with microsatellites (De Barba et al., 2016; Zhan et al., 2017) and microhaplotypes (Baetscher et al., 2018; Geue et al., 2024). Because it can perform high-throughput genotyping, the Illumina MiSeq is desirable when sequencing large numbers of DNA samples and genetic markers—which is often the case in long-term wildlife monitoring (www.EcoGenomicsCanada.ca; Zhan et al., 2017). This popular sequencing platform is a NGS technology that takes on a sequencing-by-synthesis approach (Quail et al., 2012). First, DNA libraries are prepared by combining fragments of DNA samples with adapter sequences which allows them to bind to a flow cell (Illumina, 2022). Fluorescently labelled nucleotides are then added to the flow cell and bind to clonally amplified DNA templates (Illumina, 2022; Quail et al., 2012). Fluorescent signals emitted upon nucleotide binding are analyzed to identify specific nucleotides (basecalling) (Illumina, 2022). Finally, data analysis is conducted. This approach usually takes between one to three days before results are generated (Illumina, 2022). In addition to the long wait periods involved in this sequencing process, current microsatellite and microhaplotype sequencing technologies have sizable startup costs and are expensive to run (Quail et al., 2012). These large-sized sequencing systems are also difficult to move from one location to another and involve laborious experimental procedures (Athanasopoulou et al., 2021). Ultimately, the expensive, specialized, and immobile nature of Illumina MiSeq (and similar

sequencing platforms) hinders widespread acquisition of genomic data needed for DNA profiling since these platforms are not generally accessible to most laboratories.

MinION Nanopore DNA Profiling

Rapidly developing nanopore sequencing technology may be the key to access and disseminate DNA profiling data relevant to wildlife management more easily. The MinION sequencer by Oxford Nanopore Technologies is a cost-effective, portable device that provides genomic data in real-time (Oxford Nanopore Technologies, 2022). Following a brief library preparation, this handheld MinION device contains a flow cell where DNA samples can be directly loaded. The flow cell contains 2,048 nanopores (small pore-forming proteins) embedded in an electro-resistant membrane. To bring native DNA strands into contact with the nanopore, a motor protein will bind to a DNA strand and direct it to an unoccupied nanopore. The motor protein further acts as a helicase, unwinding double-stranded DNA. To drive DNA strands through the nanopores, a constant voltage is applied to the membrane. This voltage also creates an ionic current within the nanopores; as a DNA strand passes through a nanopore, ionic current within the nanopore is disrupted by nucleotide bases occupying the pore space, creating an electric signal called a squiggle. The squiggle is then recorded by a sensor chip and is decoded using base-calling algorithms to generate the DNA sequence (Lu et al., 2016; Oxford Nanopore Technologies, 2022; Oxford Nanopore Technologies, n.d.; Wang et al., 2021). The MinION has a theoretical maximum output of 50 Gb for one sequencing run using an R10.4.1 flow cell, although its actual throughput depends on various factors such as the number of unblocked and active nanopores, DNA translocation speed through the nanopore, fragment lengths, and sequencing run time (Jain et al., 2015; Lu et al., 2016; Tytgat et al., 2022; Wang et al., 2021).

Unlike the Illumina MiSeq and most NGS platforms, the MinION generates unrestricted read lengths, enables rapid experimental procedure, and simplifies post-processing analysis (Oxford Nanopore Technologies, 2022; Athanasopoulou et al., 2021; Lu et al., 2016). Since the ionic current applied across the synthetic membrane drives the DNA molecule through the nanopore until it has reached the other side of the pore, the resulting read length is not limited by any fixed length constraints. This allows for the generation of long reads, often ranging from thousands to tens of thousands of bases in length (Athanasopoulou et al., 2021; Lu et al., 2016). The length of the reads generated by nanopore sequencing simply depends on the length of the DNA molecule being sequenced. The generation of long read lengths expedites library preparation workflow by eliminating the need for fragmentation steps, and simplifies genome assembly (Athanasopoulou et al., 2021). Additionally, the MinION sequencer enables rapid experimental procedure by eliminating the need for Polymerase Chain Reaction (PCR) amplification in some cases. However, when working with samples containing low quantities and quality of DNA or to perform amplicon sequencing, PCR is recommended to amplify target loci to increase read throughput and accuracy (Athanasopoulou et al., 2021). Finally, since DNA strands are sequenced as they pass through the nanopore, analyses can be conducted during the sequencing run (Jain et al., 2016). Conveniently, the MinION can plug directly into a computer as sequencing takes place to begin analysis right away.

Despite its advantages, the MinION's ability to generate DNA profiles based on microsatellite and microhaplotype data from wildlife animal populations has yet to be thoroughly assessed. Although, the MinION sequencer has already demonstrated utility in human forensic DNA profiling, showing promise for animal model applications. For example, when sequencing a panel of autosomal microsatellite loci for human identification, Hall et al. (2022) detected more

extensive genetic variability in microsatellite loci with the MinION in some cases compared to other NGS platforms. The MinION device has also been effective at sequencing highly polymorphic microhaplotypes, facilitating DNA mixture interpretation in forensic cases (Oldoni et al., 2019; Voskoboinik et al., 2018). Successful nanopore sequencing of highly repetitive tandem repeats and closely linked SNPs points to an opportunity to apply this technology to characterize short-read microsatellites and microhaplotypes in mammals within a conservation context (Cornelis et al., 2017; Tytgat et al., 2022).

Methods Overview

Here, using *R. tarandus* as a model species, the ability of the MinION Mk1B sequencer to perform accurate microsatellite and microhaplotype sequencing was assessed by comparing it to a leading sequencing platform: the Illumina MiSeq. Target microsatellite and microhaplotype regions were PCR amplified, and aliquots of the PCR products were used as input for the ONT Ligation Sequencing Amplicons protocol. Each DNA sample was indexed with barcodes from the Native Barcoding Kit 24 V14. The final DNA libraries were sequenced using the MinKNOW™ Software, which controls ONT sequencing devices. MinKNOW™ executed simplex live basecalling and demultiplexed FASTQ files according to each barcoded *R. tarandus* sample. The FASTQ files for each individual sample were then genotyped using Seq2Sat (https://github.com/ecogenomicscanada/Seq2Sat), an open-access genotyping pipeline (Liu et al., 2024). Microsatellite genotype data obtained from the microsatellite panel MinION run was compared to previously collected data for the same samples and loci on an Illumina MiSeq. Finally, the ability of the MinION to capture variation within microhaplotype regions of the *R. tarandus* genome was assessed by sequencing a novel set of microhaplotypes on a MinION Mk1B and an Illumina MiSeq system.

Specific Aims

The specific aims of this study were carried out to assess the applicability of the MinION Mk1B nanopore sequencing platform to perform routine profiling of individual caribou through microsatellite and microhaplotype sequencing.

Aim 1: Sequence a panel of previously optimized microsatellite and sex chromosome-specific locus amplicons on the MinION Mk1B to compare to the Illumina MiSeq platform.

Aim 2: Design primer sets for microhaplotype loci of interest in genic regions of the caribou genome.

Aim 3: Test primers and optimize microhaplotype multiplex PCR amplification.

Aim 4: Sequence a panel of novel microhaplotype markers using the MinION Mk1B and Illumina MiSeq platform.

Aim 5: Assess the ability of the MinION Mk1B to accurately sequence short reads using available bioinformatics tools.

Aim 6: Provide a cost breakdown of short-read amplicon sequencing with the MinION Mk1B compared to that of the Illumina MiSeq.

These aims will help inform which strategies should be implemented or changed to enhance the efficiency and accuracy of Canadian-wide caribou profiling. Cost-effectiveness remains a key consideration to facilitate sustainable population monitoring.

CHAPTER 2

Microsatellite Genotyping with the MinION Nanopore Sequencer: Proof-of-Concept Applied to Caribou Monitoring in Canada

Abstract

Profiling microsatellite loci has been an important tool to monitor wildlife, including caribou (*Rangifer tarandus*) populations in Canada. In this study, the functionality of a MinION Mk1B nanopore sequencer, developed by Oxford Nanopore Technologies (ONT), was applied to generating sequence-based microsatellites for profiling. The MinION Mk1B sequencer, including a startup kit, has historically cost ~$2714 CAD compared to $157,749 CAD for an Illumina MiSeq which is commonly used to characterize microsatellite within caribou (www.EcoGenomicsCanada.ca). Here, eight microsatellite loci and sex-specific loci were sequenced from the DNA of 19 caribou with a MinION Mk1B device using ONT's latest V14 chemistry and an R10.4.1 flow cell. These same samples and genetic regions had been sequenced and characterized using the Illumina MiSeq platform. Raw sequencing data generated from both sequencers were processed through the Seq2Sat/SatAnalyzer software before manually scoring target microsatellite amplicons. Nanopore-based microsatellite genotypes were 98.68% consistent with those generated by Illumina MiSeq for the same samples and loci, and 100% consistent for sex identification. Our findings suggest the feasibility of employing the MinION Mk1B for microsatellite profiling in wildlife genomic monitoring efforts, offering a cost-effective alternative when sequencing 96 samples or fewer in one sequencing run ($24.04/sample) with comparable performance to the Illumina MiSeq platform ($30.99/sample). When sequencing 384 samples on the MiSeq, where the MinION is currently limited to 96 samples, the Illumina platform is more cost-effective ($13.72/sample). Additional nanopore

barcodes are needed (>96) to lower per-sample and per-locus costs for MinION sequencing. Overall, despite the limited numbers of samples that can be run on a MinION compared to an Illumina MiSeq, the cost point demonstrates the potential of nanopore sequencing technology to be a viable alternative for microsatellite genomic profiling.

Introduction

Microsatellites are conventional molecular markers used in a conservation context (De Barba et al., 2016; Carroll et al., 2018; Fournier et al., 2024; Hettinga et al., 2012; Jones et al., 2023; McFarlane et al., 2018; Thaden et al., 2020). Microsatellites, also known as simple sequence repeats (SSRs) or short tandem repeats (STRs), are short, repetitive sequences of DNA that consist of units of one to five or six base pairs in length that are repeated in tandem (e.g., AAGAAGAAGAAG) (Arif et al., 2011; Ellegren, 2004). Due to their highly polymorphic nature, microsatellites are a powerful genetic marker for individual-specific DNA profiling and population monitoring (Arif et al., 2011; Carroll et al., 2018; Thaden et al., 2020). Microsatellites are co-dominant and display variability in the number, length, and pattern of repeats among individual animals (Arif et al., 2011; De Barba et al., 2016). This variation in microsatellite regions serves as a basis for assessing genetic diversity, population dynamics, relatedness, and evolutionary relationships in wildlife populations (Arif et al., 2011; Carroll et al., 2018; De Barba et al., 2016; Jones et al., 2010; McFarlane et al., 2021; Pemberton, 2008; Thompson et al., 2019).

Microsatellite loci are one of the genetic markers used to monitor at-risk caribou (*Rangifer tarandus*) in Canada, contributing to an extensive legacy database of genetic data available at www.EcoGenomicsCanada.ca. By leveraging microsatellite genotyping techniques in the *R. tarandus* genome, unique genetic profiles are derived which are useful for accurately

identifying individuals (Flasko et al., 2017), characterizing population structure (Ball et al., 2010), reconstructing evolutionary histories (Klütsch et al., 2016), and estimating population trends (Hettinga et al., 2012). This information can inform management and planning (Schmidt et al., 2024) by tracking temporal changes in such population metrics across various time spans (Schwartz et al., 2007).

Traditionally, microsatellites have been profiled using capillary-based size separation of amplified microsatellites (Ball et al., 2010; Guichoux et al., 2011; Klütsch et al., 2016; McFarlane et al., 2020; Vieira et al., 2016). The drawbacks of this technique involve low throughput, limited capacity for sample and microsatellite marker number, and differences in genotype data across sequencing platforms and facilities due to difficulties with automating allele scoring: all of which increase genotyping errors (De Barba et al., 2016; Guichoux et al., 2011). An alternative approach is to develop consensus genotypes by employing high-throughput sequencing of microsatellite genotypes from amplicon data and using bioinformatic analysis pipelines (Carroll et al., 2018; Darby et al., 2016; De Barba et al., 2016; Liu et al., 2024; Suez et al., 2015). Implementing a high-throughput sequencing microsatellite genotyping strategy has shown to decrease costs while improving microsatellite genotype accuracy, particularly for noninvasively collected DNA samples (i.e. fecal or tissue samples) (Carroll et al., 2018; De Barba et al., 2016). This approach also allows a greater number of samples and loci to be examined at once, further reducing cost. As a result, high-throughput sequencing platforms such as the Illumina MiSeq are commonly used for microsatellite profiling in many contexts (Darby et al., 2016; De Barba et al., 2016; Liu et al., 2024; Van Neste et al., 2014).

The Illumina MiSeq is a leading NGS technology that uses a sequencing-by-synthesis approach to sequence short-reads (Quail et al., 2012). However, the substantial capital

investment and costs per sequencing run associated with Illumina and other massively parallel sequencing (MPS) platforms remain barriers to widespread implementation for routine use (Hebert et al., 2024; Tytgat et al., 2022). The MinION nanopore sequencer, developed by Oxford Nanopore Technologies (ONT), may be a more cost-effective approach for microsatellite sequencing due to its substantially lower capital investment (~$2714 CAD) compared to the Illumina MiSeq (~$157,749 CAD) and can have lower costs per sequencing run, in addition to having no annual service fees (Hebert et al., 2024; Rang et al., 2018; Tytgat et al., 2022; Van der Reis et al., 2022). The MinION is a real-time DNA sequencer that stands out for its portability, native DNA strand sequencing capabilities, and its potential to be deployed beyond core facilities due to its low cost and simple library preparation workflows (Athanasopoulou et al., 2021; Hebert et al., 2024; Lu et al., 2016; Oxford Nanopore Technologies, 2022; Rang et al., 2018; Van der Reis et al., 2022).

MinION nanopore sequencing operates by passing DNA strands through tiny biological pores embedded on an electro-resistant membrane and detects variations in electrical conductivity at pinch points where the pore constricts (Oxford Nanopore Technologies, 2022; Wang et al., 2021). As the DNA strand moves through the pore, the changes in electrical signals correspond to different DNA bases, enabling real-time sequencing of the genetic material (Lu et al., 2016; Oxford Nanopore Technologies, 2022; Wang et al., 2021). Despite being considered a long-read third generation sequencing technology, the MinION can also sequence short amplicons (<500 bp long), enabling its application to microsatellite region genotyping (Oxford Nanopore Technologies, 2022; Tytgat et al., 2020). This is supported by forensic investigations of human STR regions that have demonstrated the MinION's ability to correctly genotype

microsatellite loci, including differentiating iso-alleles which are the same length but vary in nucleotide sequence (Tytgat et al., 2022).

Although the Illumina MiSeq has historically outperformed the MinION with respect to read accuracy, continued improvements in the MinION's sequencing fidelity are making it a viable alternative for routine microsatellite genotyping (Dokos et al., 2022; Rang et al., 2018; Tytgat et al., 2022). Coupled with improved base-calling software, ONT's latest version 14 chemistry (V14) for nanopore sequencing can generate 99.6% read accuracy for simplex reads ($Q20^{+}$) (Hebert et al., 2024; Van Dijk et al., 2023). Additionally, the newest R10.4.1 flow cells have notably improved data quality for homopolymer regions—which has previously been a pitfall of nanopore sequencing (Tytgat et al., 2020). Notable advancements in the new V14 chemistry and flow cells include reducing DNA translocation speed through the nanopore and switching to dual-constriction point nanopores so nucleotide signals are interpreted twice (Hebert et al., 2024; Van Dijk et al., 2023; Van der Verren et al. 2020). In recent nanopore sequencing experiments, Van der Reis et al. (2022) demonstrated that the MinION is well-suited for short-read DNA metabarcoding and offers swift access to sequencing outcomes.

For wildlife monitoring applications, accessible and widely deployable sequencing technology is vital to cost-effectively inform time-sensitive conservation and management efforts (Carroll et al., 2018; Hohenlohe et al., 2020; Thaden et al., 2020). Here, 8 microsatellites loci were profiled in *R. tarandus* tissue-extracted DNA samples on a MinION Mk1B sequencer as a beta-test of the nanopore platform. The genotypes at these 8 microsatellite loci were compared to genotypes of the same samples at the same loci sequenced with an Illumina MiSeq system. This enabled a direct comparison between microsatellite genotyping data across platforms and aims to validate the applicability of the MinION to accurately genotype *R. tarandus* microsatellite loci,

potentially providing a cost-effective alternative for microsatellite genotyping. Sex identification for each sample was also assessed across platforms by sequencing sex chromosome-specific locus amplicons designed by Ball et al. (2007).

Materials & Methods

MinION Mk1B DNA Sample Collection & Preparation

The results presented in this paper were derived from genomic DNA extracted from 19 *R. tarandus* tissue samples collected from Ontario caribou. DNA extracts were kept frozen since their initial collection and extraction. DNA concentrations for each sample were measured using a Qubit 4 Fluorometer (Invitrogen) with the High Sensitivity dsDNA Assay kit and diluted to 5 ng/µL, then 250 pg/µL. DNA samples at a concentration of 250 pg/µL were kept at −20°C to be used for two multiplex PCR amplification reactions.

MinION Mk1B Multiplex PCR

All multiplex PCR reactions conducted for MinION sequencing were followed according to the optimized primer pools and ratios developed in-house for Illumina microsatellite sequencing (not yet published). For the first multiplex PCR reaction (M1), each DNA sample was subjected to an 8-plex PCR amplification to amplify the 8 target microsatellite loci (Table A1). Samples were amplified using standard DNA Oligonucleotides primer pairs (Integrated DNA Technologies Inc, 25 nm scale) for the microsatellite targets. The 8-plex PCR reaction was performed in a total reaction volume of 13 µL, containing 4 µL of 250 pg/µL input DNA and 9 µL of cocktail which consisted of 2x Type-It PCR Master Mix (Qiagen), UltraPure DNase/RNase-Free Distilled Water (Invitrogen; Thermo Fisher Scientific), 0.2 mM BSA (Sigma Aldrich), and 8 microsatellite primer sets. Potential contamination was surveyed by including a no-DNA PCR blank. The following thermocycling regime for PCR amplification was followed:

95°C for 5 minutes; 30 cycles of 95°C for 30 sec, 63°C for 90 sec, 72°C for 30 sec; 68°C for 10 min.

For the second multiplex PCR reaction (M2), sex chromosome-specific locus amplicons (*Zfx*/*Zfy*) were amplified in each caribou DNA sample for sex identification (Table A2). Primers targeting prion protein gene (PRNP) amplicons were also included in this PCR reaction, although they were not assessed in this study. The LGL335 primer amplifies regions of the zinc finger gene intron on the X-chromosome and Y-chromosome while the SDP730 primer was designed specifically for caribou XY system profiling (Ball et al., 2007). This PCR amplification was performed in a total reaction volume of 11 µL, containing 4 µL of 250 pg/µL input DNA and 7 µL of cocktail which consisted of 2x Type-It PCR Master Mix (Qiagen), UltraPure DNase/RNase-Free Distilled Water (Invitrogen; Thermo Fisher Scientific), 0.2 mM BSA (Sigma Aldrich), and the standard LGL335 and SDP730 primers as well as forward and reverse PRNP primers (IDT, 25 nm scale). Potential contamination was surveyed by including a no-DNA PCR amplification. The same thermal cycling program was followed as in the 8-plex microsatellite PCR reaction.

After PCR amplification, equal volumes (6 µL) of PCR products from both multiplexes were pooled according to samples in a new plate. The resulting plate comprised 19 distinct DNA samples, containing 8 amplified microsatellite loci and the *Zfx*/*y* target amplicons. Pooled PCR products for each sample were screened on a QIAxcel Advanced System (Qiagen) with the DNA Screening Kit to validate PCR product amplification (Figure A1). DNA samples were then quantified using the Qubit 4 Fluorometer (Invitrogen) with the High Sensitivity dsDNA Assay kit before diluting each sample to a standard concentration (Table A3). The standard concentration was calculated using the NEBioCalculator based on an average amplicon length of

250 bp. Each sample containing amplified PCR products was diluted to 11.5 µL of 3 ng/µL, which corresponds to approximately 200 fmol of each sample. Although a post-PCR AMPure XP bead (Beckman Coulter) cleanup was not conducted here, the PCR multiplexes amplified sufficient DNA to add a post-PCR cleanup in future MinION Mk1B microsatellite sequencing runs. All samples were kept frozen leading up to library preparation. Aliquots of each sample from the M1/M2 pooled plate were used as input for DNA library preparation.

MinION Mk1B Library Preparation

The ONT Ligation Sequencing Amplicons protocol coupled with the Native Barcoding Kit 24 V14 (SQK-NBD114-24) was used to sequence microsatellite and sex targets on the MinION Mk1B sequencer. The procedure for the Ligation Sequencing Amplicons protocol was adhered to with some modifications. Library preparation involved preparing the DNA ends for adapter attachment, ligating native barcodes to the DNA ends, ligating sequencing adapters to the DNA ends, and priming the flow cell. During the end-prep steps, a diluted DNA control sample (DCS) was added to each sample. During the native barcode ligation steps, each DNA sample received a unique barcode, allowing all samples to be sequenced simultaneously on the same flow cell. Once DNA samples were barcoded, they were pooled in equal volumes. Lastly, the DNA library concentration was quantified using a Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher). The final prepared DNA library consisted of 12 µL containing 29.22 fmol of DNA and was stored short-term at 4°C in an Eppendorf DNA LoBind tube before sequencing. Although the Native Barcoding Amplicons protocol suggests loading 10-20 fmol of the final prepared library onto the flow cell, ONT recommends loading up to 50 fmol to achieve a pore occupancy of over 95% during experiments requiring extensive data output (Oxford Nanopore

Technologies, 2022). As a result, loading the flow cell with 29.22 fmol of DNA library was deemed appropriate.

MinION Mk1B Amplicon Sequencing

Here, 8 microsatellites and *Zfx*/*Zfy* amplicons to determine putative sex were sequenced from 19 unique caribou tissue DNA samples. A FLO-MIN114 flow cell (R.10.4.1) was primed and set up for sequencing as per the manufacturer's instructions, using Thermo Fisher's Invitrogen™ UltraPure™ Bovine Serum Albumin (50mg/mL). After an initial flow cell check, which revealed 1337 functional pores, the final prepared library was loaded onto the flow cell. The latest version of the MinKNOW™ software for Windows 10 (V23.07.15) was downloaded onto an M18 R1 Alienware computer that meets the MinION Mk1B IT requirements. Default MinKNOW™ parameters were selected while setting up the sequencing run with some modifications to perform live base-calling during sequencing. The SQK-NBD114-24 kit was selected to reflect the unique barcodes that each sample was indexed with during library preparation. Super-accurate basecalling (SUP) (400 bps) was conducted through the Guppy basecaller (V7.1.4) to optimize the number and accuracy of reads sequenced. The minimum read length was set to 20 bp to reduce the loss of short reads and maximize data yield. Finally, read filtering was set to a minimum Phred quality score of 10 (Q10) and read splitting was enabled. Phred quality scores represent error probabilities of basecalls (Ewing & Green., 1998).

The run limit was 72 hours; however, sequencing was stopped after 43 hours when fewer than 50 pores were available. SUP data output was set to POD5 files and FASTQ files for each barcode. Only FASTQ files that passed the Q10 threshold (FASTQ_pass) were used for downstream genotyping analysis. All FASTQ_pass basecalling data was demultiplexed according to ONT barcodes in MinKNOW™. Post-run barcode trimming in the FASTQ_pass

read files was performed, filtering out mid-read barcodes and reads that did not have an in-tact barcode on both ends. Unclassified data from the FASTQ_pass files was not included in downstream analysis. The NanoPlot tool was used to summarize Oxford Nanopore sequencing data (De Coster & Rademakers, 2023).

MinION Mk1B Microsatellite Genotyping and Sex Identification

All FASTQ_pass files with the same barcode designation were concatenated into one larger FASTQ file through Linux (WSL) using Ubuntu V22.04.3 LTS and were renamed to their corresponding 'sampleID'.fastq.gz instead of the barcode number. The concatenated FASTQ file for each sample was genotyped using the open-access Seq2Sat genotype scoring software (https://github.com/ecogenomicscanada/Seq2Sat) (Liu et al., 2024). Seq2Sat aligns microsatellite and sex allele reads from each sample to the forward and reverse complementary reverse primer sequences provided in an uploaded loci file, allowing some mismatches (Liu et al., 2024). To access this output in the SatAnalyzer web-interface, a sample file (Table A4), microsatellite loci file (Table A5), and sex loci file Table A6) specific to this experiment were uploaded. Genotypes at each locus were autoscored based on allelic peaks, the reads ratio of primary alleles, read variant sizes, and the presence of single nucleotide polymorphisms (SNPs) as outlined by Liu et al. (2024). Then, genotypes were reviewed and manually scored in accordance with scoring rules developed in-house for routine microsatellite profiling. During manual genotype scoring, the morphology of allelic peaks was thoroughly assessed to differentiate between homozygous, heterozygous, and inconclusive microsatellite genotypes. Genotypes were recorded as inconclusive when it was unclear whether an individual is heterozygous or homozygous for a given locus even after manual scoring based on the read ratio of the two most abundant alleles, or when the number of reads for either primary allele was below 20 (Liu et al., 2024).

Illumina Microsatellite Data Generation

Data from a previous Illumina MiSeq sequencing run was recovered and used for comparison with the MinION Mk1B amplicon sequencing experiment outlined above, focusing on the same samples and target loci. This Illumina MiSeq run consisted of 384 caribou samples and 23 loci—including the 8 microsatellite and *Zfx*/*Zfy* loci sequenced with the MinION Mk1B. The Illumina MiSeq workflow followed the 16S Metagenomic Sequencing Library Preparation protocol, performing multiplex PCRs, clean ups, and index reactions.

The initial PCR amplification used the Qiagen Type-It Microsatellite kit (Qiagen) with 1 ng of input DNA per sample, and the same thermal cycling program as the MinION Mk1B multiplex PCR was followed. Target regions were amplified with IDT's standard DNA oligonucleotides with the Illumina overhang adaptor sequence added to the locus-specific sequence (Forward overhang: 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, Reverse overhang: 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG). PCR products were pooled and cleaned with AMPure beads (Beckman Coulter). A subsequent index reaction with Type-It Master Mix, Illumina DNA/RNA UD Indexes Set A-D, 2.5 µL of cleaned PCR product, and ultrapure water. The index PCR thermocycling regime was followed: 95°C for 3 minutes; 8 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, and a final extension of 72°C for 5 min.

Indexed products were pooled, and DNA libraries were cleaned before being screened for correct amplification with the Qiaxcel System (Qiagen) using the DNA Screening Kit. DNA libraries were quantified with the High Sensitivity dsDNA Assay kit on a Qubit 4 Fluorometer (Invitrogen) and were diluted to 4 nM before further dilutions, denaturation steps, and pooling. The Illumina 500 Cycle v2 Kit was used to load DNA libraries onto the Illumina MiSeq with a

2x231 bp setting. After sequencing, FASTQ files were generated and indexed reads were demultiplexed. FASTQ files underwent Seq2Sat genotyping analysis which was visualized through the SatAnalyzer interface.

The same microsatellite loci file and sex loci file from the MinION Mk1B experiment were used to retrieve target loci data, and an Illumina-specific sample file filtered out samples that were not sequenced on both platforms (Table A7). Manual microsatellite genotype editing was carried out based on genotype scoring rules that were developed from analyzing large legacy datasets over time. Genotypes for the target microsatellite and sex identification based on the *Zfx*/*Zfy* amplicons were then compared to data obtained from the MinION Mk1B sequencing run. This enabled a direct comparison of inferred genotypes for each sample and locus as well as sex identification across the MinION Mk1B and Illumina MiSeq sequencing platforms.

Cost Comparison

To assess cost-efficiency based on the platform used, an approximate average cost of microsatellite sequencing for sequencing different numbers of samples on a MinION Mk1B vs an Illumina MiSeq was estimated.

Results and Discussion

Sequencing Summary

The MinION Mk1B sequencing run generated 6.15 M raw reads and 1.15 Gb of bases with an estimated N50 read length of 174 after filtering. Most reads were captured within the first 20 hours of sequencing. The mean amplicon read length was 187.8 bp and the median amplicon read length was 165 bp. MinION Mk1B average read quality relative to amplicon read lengths is shown (Figure 1.1). Most reads were <250 bp, which is consistent with the target microsatellite amplicon lengths. The raw reads that exceeded the length of target amplicon

fragments were likely concatemers which are filtered out before assessing amplicon sequencing results. Read qualities ranged from Q7 to slightly above Q40 on average. The mean read quality was Q13.7 and the median read quality was Q14.8. Most reads had a mean basecall quality between Q10 and Q20. There were 1131.7 Mb (97.8%) of reads above the Q10 quality score cutoff and 513.8 Mb (47.9%) of reads surpassed a quality score of Q15. Throughout the sequencing run, the distribution of basecall quality over time and the distribution of read lengths over time remained consistent.

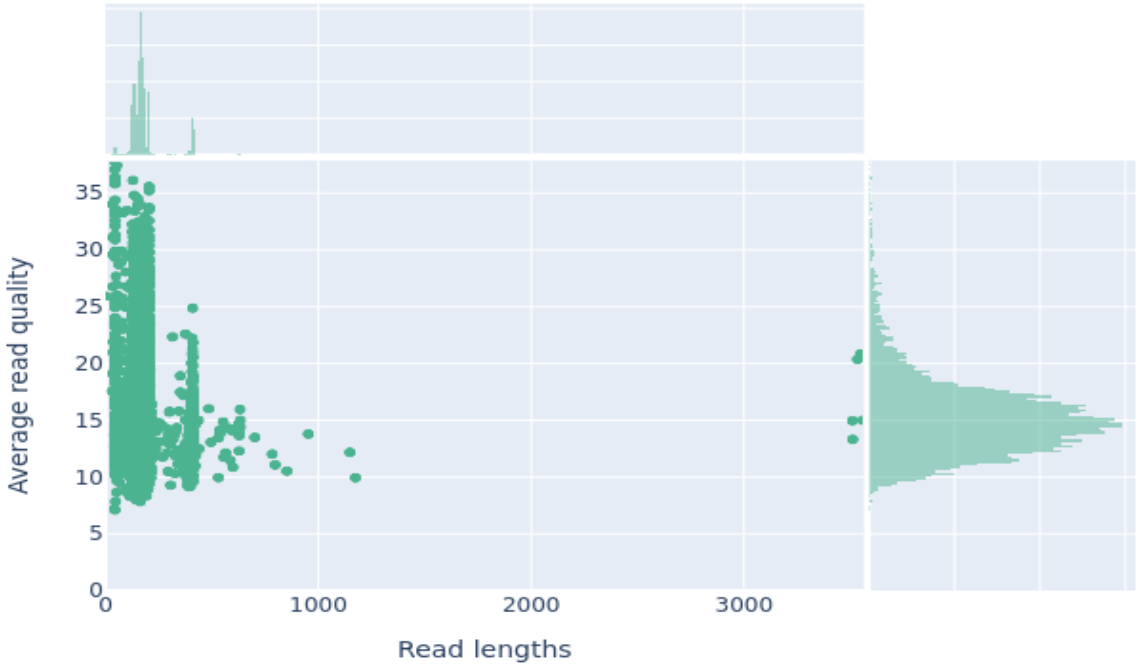


Figure 1.1 NanoPlot scatter plot of MinION Mk1B read lengths versus average read quality (Phred scores) generated from sequencing of microsatellite and sex amplicons in 19 caribou samples.

The Illumina MiSeq sequencing run generated a total of 1.32 M raw reads that pass the minimum quality score of 20 for the 19 caribou samples sequenced. The average quality scores of Illumina MiSeq raw reads before and after filtering for each sample, depicted by a different colour, are recorded (Figure 1.2). A quality score was assigned for each position in the read, calculated as the average quality score at that position across all reads from a specific sample.

The first sequencing cycle represents the first position in the read for a given sample (originally derived from Illumina's Sequencing by Synthesis approach). Up to the ~250th sequencing cycle is shown, reflecting the target microsatellite amplicon length of <250 bp.

Illumina MiSeq sequencing generated quality scores for two reads (read1 and read2) due to its paired-end sequencing method. The Illumina MiSeq raw read quality scores for read 1 and read 2 before and after filtering exhibited relatively high-quality scores (~35) at the starting positions of the reads. As the read position/sequencing cycle number increased, the quality scores followed a moderate downward curve. At the farthest read positions, quality scores for read 1 after filtering were ~Q25 for most samples. The quality scores for read 2 after filtering typically ranged from ~Q15 to ~Q20.

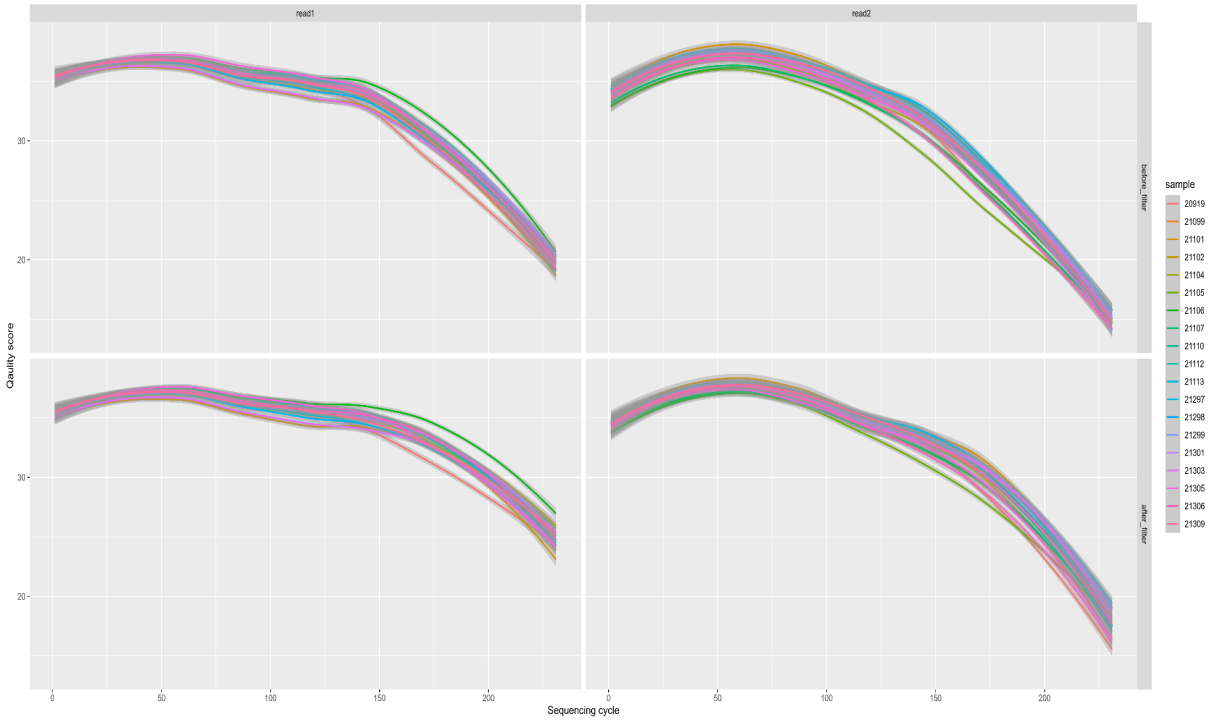


Figure 1.2 Illumina MiSeq raw read quality (Phred) scores before filtering (top) and after filtering (bottom) for all caribou samples represented by various colours. Sequencing cycle refers to the position in a read; one sequencing cycle represents one base pair. Quality scores at each read position are based on the average scores across all reads from a specific sample. Read1 (left) and Read2 (right) reflect the paired end reads generated from Illumina sequencing.

The total number of raw reads, clean reads, and assigned reads for each caribou sample based on Seq2Sat/SatAnalyzer data output is shown according to the sequencing platform used (Figure 1.3, Figure 1.4). Raw reads refer to the FASTQ file data that had not undergone any read quality filtering. Clean reads consisted of all reads that met the quality score threshold and assigned reads were clean reads that aligned to any of the target loci. Only clean and assigned reads that passed the quality score threshold of 10 for nanopore sequencing data and 20 for Illumina data were included. A quality score threshold of Q20+ for Illumina reads was selected for this experiment as it is the default minimum quality score filter applied to routine microsatellite genotyping conducted in-lab. Since the MinION typically generates a lower throughput and depth of coverage than other MPS instruments, like the Illumina MiSeq, a less stringent quality score threshold captures more target reads, increasing confidence in genotype calls while maintaining sufficient read quality (Voskoboinik et al., 2018). For both datasets, the majority of raw reads were retained as clean reads after filtering, and a small proportion of clean reads were aligned to target sequences to become assigned reads. Generating similar proportions of raw, clean, and target reads per sample from both sequencers suggests that the quality filtering and alignment steps of this workflow is comparable for the MinION Mk1B and Illumina MiSeq amplicon data with the quality score thresholds applied.

An important difference between the MinION Mk1B and Illumina MiSeq dataset was that additional caribou samples and microsatellite loci were sequenced during the Illumina run from which microsatellite data was obtained. Read depths shown for the Illumina MiSeq were subset to only include the same samples and loci sequenced during the MinION Mk1B experiment. Because the total number of samples and loci sequenced during the Illumina MiSeq run was greater than the number of samples and loci sequenced in the same MinION Mk1B

microsatellite sequencing experiment, the distribution of reads across samples and target loci is more spread out (Fumagalli, 2013). Therefore, it is not feasible to directly compare read depths per sample and locus between sequencing platforms. Despite this, read depths are shown for each sequencer to assess if sufficient depth was achieved for accurate microsatellite profiling and if reads were evenly distributed across loci, regardless of the number of samples and loci targeted during multiplex PCR and amplicon sequencing.

For the MinION sequencing run, the mean number of raw reads per sample was 320,551 raw reads, ranging from 188,020 to 488,612 raw reads per sample (Figure 1.3). All but one sample had more than 200,000 raw reads. The mean number of clean reads per sample was 318,970, ranging from 187,449 to 486,799 clean reads. The proportion of clean reads to raw reads, based on the mean number of clean reads and raw reads per sample, was 0.995. On average, each sample had 48,268 assigned reads, ranging from 27,377 to 73,911 assigned reads. The proportion of assigned reads to clean reads, based on the mean number of assigned reads and clean reads per sample, was 0.151. The proportion of assigned reads to raw reads, based on the mean number of assigned reads and raw reads per sample was also 0.150.

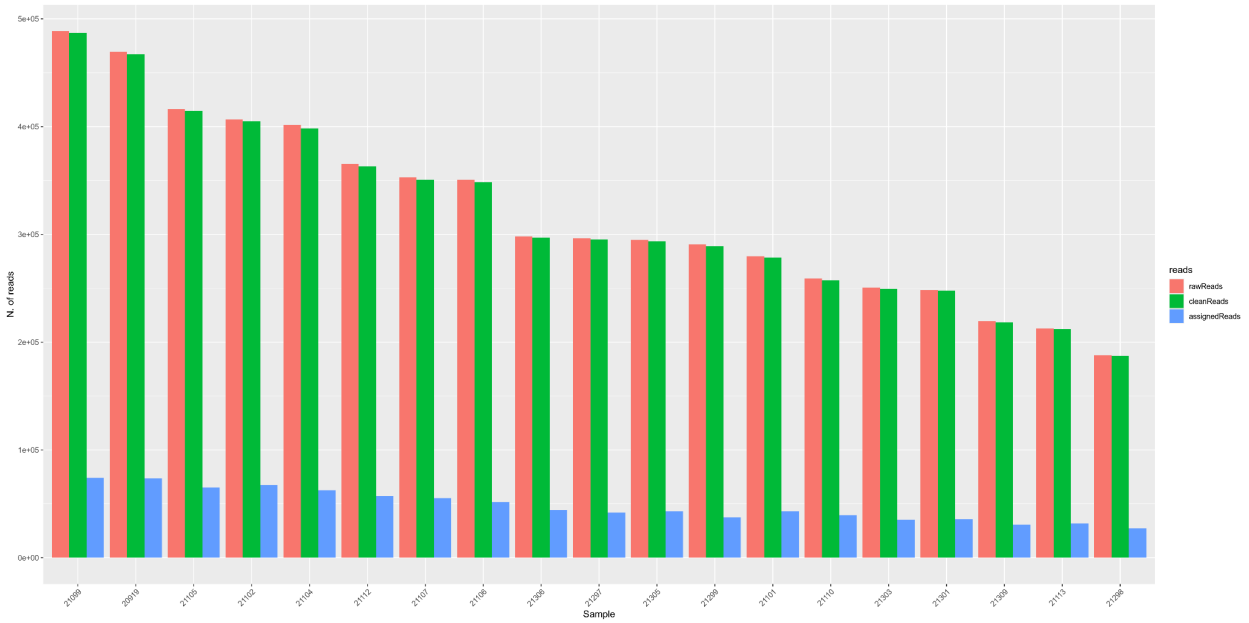


Figure 1.3 Total number of rawReads (raw reads) in red, cleanReads (clean reads) in green, and assignedReads (assigned reads) in blue per caribou sample that passed the quality score threshold generated from microsatellite sequencing on the MinION Mk1B platform. An average of 320,551 raw reads, 318,970 clean reads, and 48,268 assigned reads were obtained per sample.

The Illumina sequencing run generated a mean of 69,607 raw reads per sample, ranging from 38,544 to 98,576 raw reads (Figure 1.4). The mean number of clean reads per sample was 64,282, ranging from 35,914 to 92,344 clean reads. The proportion of clean reads to raw reads, based on the mean number of clean reads and raw reads per sample, was 0.924. There was a mean of 9,339 assigned reads per sample, ranging from 3.316 to 16,730 assigned reads. The proportion of assigned reads to clean reads, based on the mean number of assigned reads and clean reads per sample, was 0.145. The proportion of assigned reads to raw reads, based on the mean number of assigned reads and raw reads per sample, was 0.134.

Both sequencing platforms resulted in a much smaller number of assigned reads relative to the number of raw reads for all caribou samples. PCR conditions and primer specificity are known to impact off-target amplification (Brownie, 1997; Eriksson et al., 2020; Shum & Paul, 2009). If there was a lot off-target amplification, it would account for many raw reads that got

filtered out when aligning raw reads to the target microsatellite amplicons. Since the microsatellite data from both sequencers was generated using the same primers, number of PCR cycles, and highly similar multiplex PCR reactions, it is not surprising that the MinION Illumina MiSeq demonstrated similar proportions of assigned reads to raw reads. To increase these proportions, reducing the number of PCR cycles used to amplify target regions or adding more post-PCR AMPure XP bead (Beckman Coulter) cleanup steps in the microsatellite library preparation workflow could reduce the frequency and quantity of these sequence errors (Tytgat et al., 2020). This would be especially beneficial if applying this protocol to non-invasively collected fecal DNA samples, which are often used to genotype microsatellites in caribou (www.EcoGenomicsCanada.ca), since they are more susceptible to nontarget binding that can interfere with genotyping success (Eriksson et al., 2020). Ultimately, as long as sufficient assigned reads are generated from each platform, the low ratio of assigned reads to raw reads should not impede microsatellite calling.

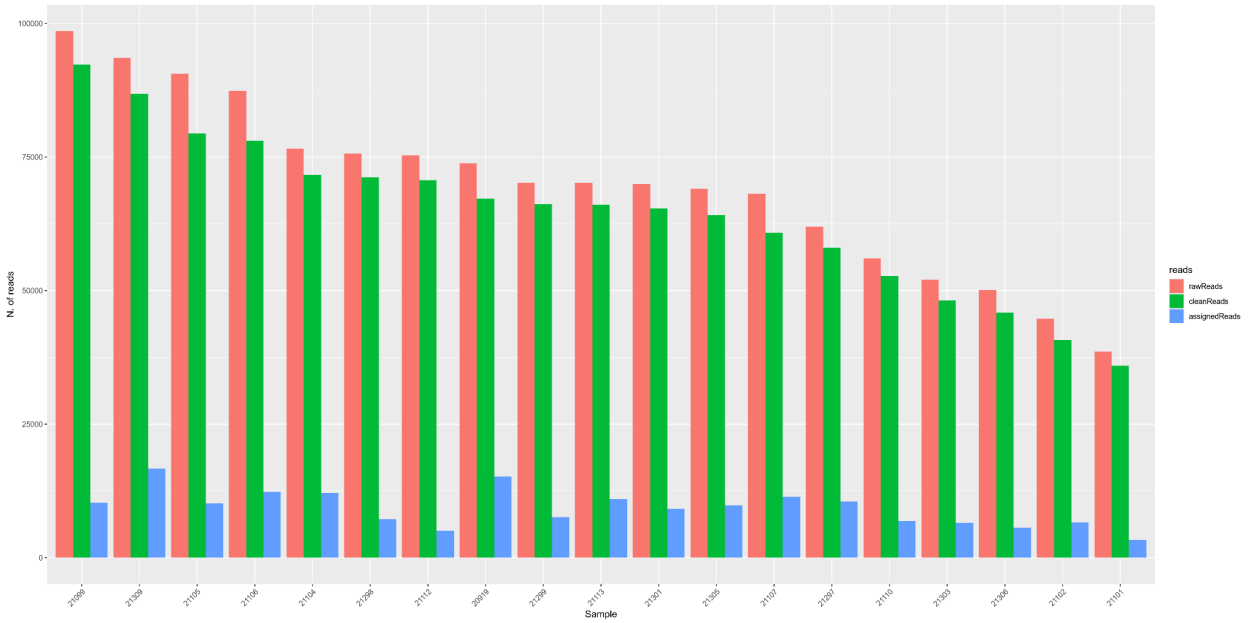


Figure 1.4 Total number of rawReads (raw reads) in red, cleanReads (clean reads) in green, and assignedReads (assigned reads) in blue per caribou sample that passed the quality score threshold generated from microsatellite sequencing on the Illumina MiSeq platform. An average of 69,607 raw reads, 64,282 clean reads, and 9,339 assigned reads were obtained per sample.

The relative read distribution of 8 target microsatellite loci and the *Zfx*/*Zfy* sex loci were assessed by depicting the mean percentages of number of target amplicon reads per sample to the total number of assigned reads per sample obtained from MinION Mk1B sequencing and Illumina MiSeq sequencing (Figure 1.5). The relative read distribution for each microsatellite locus sequenced with the MinION Mk1B and the Illumina MiSeq was intended to show which loci accounted for an even/uneven distribution of target microsatellite reads—it was not intended for a direct comparison across sequencing platforms.

On average, in the MinION Mk1B dataset 9.88% of the total assigned reads per sample were IGF reads, 2.62% were MAP2C reads, 12.50% were NVHRT48 reads, 16.93% were OHEQ reads, 11.47% were RT24 reads, 24.51% were RT27 reads, 10.81% were RT6 reads, 7.75% were RT7 reads, and 3.53% were *Zfx*/*Zfy* reads. The RT27 microsatellite locus accounted for the highest percentage of total assigned reads per sample on average (24.51%), whereas the

MAP2C locus accounted for the lowest percentage of assigned reads per sample on average (2.62%) with MinION Mk1B sequencing. For the Illumina MiSeq amplicon sequencing run, on average, 11.33% of the total assigned reads per sample were IGF reads, 4.53% were MAP2C reads, 6.52% were NVHRT48 reads, 3.05% were OHEQ reads, 8.96% were RT24 reads, 26.33% were RT27 reads, 4.89% were RT6 reads, 14.11% were RT7 reads, and 20.28% were *Zfx*/*Zfy* reads in the Illumina dataset. The RT27 microsatellite locus accounted for the highest percentage of the mean total assigned reads per sample (26.33%), and the OHEQ locus accounted for the lowest percentage of the mean total assigned reads per sample (3.05%) with Illumina MiSeq sequencing.

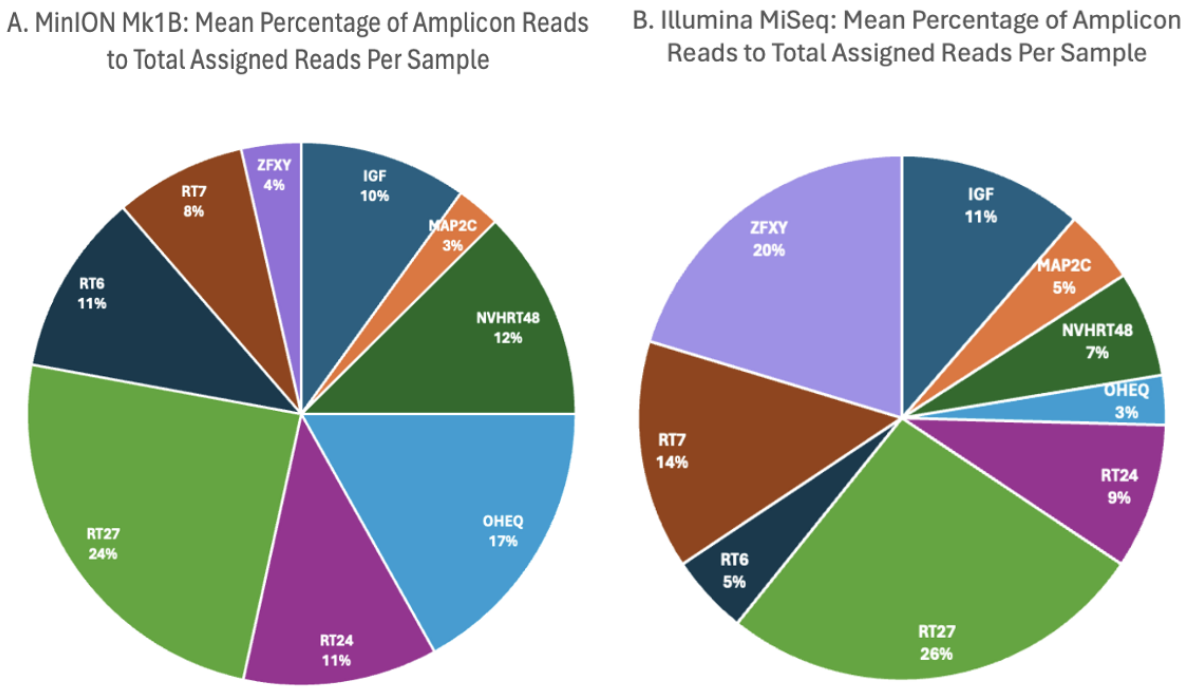


Figure 1.5 Percentages of mean number of reads for each target locus per sample relative to the mean total number of assigned reads per sample generated from MinION Mk1B amplicon sequencing (A) and Illumina MiSeq amplicon sequencing (B).

Loci that made up 9-13% of the mean assigned reads per sample were considered evenly distributed. For the MinION Mk1B dataset, 4 loci had an even distribution pattern: RT24, NVHRT48, IGF, and RT6. The RT27 and OHEQ loci had a disproportionately high percentage

of mean assigned reads per locus while the MAP2C, RT7 and the *Zfx*/*Zfy* locus had a disproportionately low percentage of mean assigned reads per sample. For the Illumina MiSeq dataset, only 2 loci were evenly distributed: RT24 and IGF. The RT27, RT7 and *Zfx*/*Zfy* locus had a disproportionately high percentage of mean assigned reads per locus while the OHEQ, NVHRT48, MAP2C, and RT6 locus had a disproportionately low percentage of mean assigned reads per sample.

The variance in mean number of target amplicon reads relative to the total assigned reads per sample across sequencing platforms is likely in part due to the difference in sample and locus number in the DNA library sequenced by the MinION Mk1B and Illumina MiSeq. Although, some of the variation in proportion of target amplicon reads to total assigned reads per sample could be attributed to uneven amplification of the target amplicons during PCR target enrichment steps (Peng et al., 2015). Having some variation in PCR product concentration is unavoidable (Hebert et al., 2024). This could result in final read counts for target amplicons not reflecting the true abundance of these sequences in the original DNA (Peng et al., 2015). In previous nanopore sequencing experiments with the MinION sequencer, highly variable depths of coverage for target loci have been observed, even when equimolar amounts of PCR amplicon products are pooled (Whitford et al., 2022). Whitford et al. (2022) also noted greater variance in depth of coverage between amplicons within samples compared to the variance in depth of coverage between different samples for the same amplicons, indicating read depth variability is highly locus dependent.

Sex Identification

Sex identification was determined for each caribou sample sequenced based on the ratio of *Zfy* (Y) to *Zfx* (X) reads generated from the MinION and Illumina sequencing. There were no

inconclusive sex identifications for any sample sequenced on either platform. All samples with a Y/X ratio of zero were categorized as female. Caribou samples identified as female based on Seq2Sat and SatAnalyzer output are shown according to the sequencing platform used (Figure 1.6). Every sample identified as female had at least 450 X reads across both datasets. All samples with a non-zero Y/X ratio were categorized as male. Caribou samples identified as male are shown according to the sequencing platform used (Figure 1.7). The Y/X ratios ranged from 0.46 to 1.07 for the MinION Mk1B dataset, with a mean Y/X ratio of 0.70 for each male. The Illumina MiSeq male Y/X ratios ranged from 1.70 to 2.71, with a mean Y/X ratio of 2.25 for each male. MinION Mk1B sequencing often resulted in a higher number of X reads than Y reads for most male samples, whereas Illumina MiSeq sequencing often generated more Y reads than X reads for most male samples. Out of the 19 caribou samples sequenced, 14 were identified as male and 5 were identified as female. There was 100% sex identification congruence for all samples sequenced on a MinION Mk1B and an Illumina MiSeq platform. Congruent sex identification for all samples sequenced with both instruments further supports the MinION Mk1B's comparability to the Illumina MiSeq for sex identification, which is an integral part of DNA profiling.

Figure 1.6 Number of reads for the *Zfx* (X) and *Zfy* (Y) alleles in caribou samples with a female sex identification sequenced on a MinION Mk1B (A) and an Illumina MiSeq (B) sequencer. Individuals with a Y/X ratio of zero are categorized as female based on Seq2Sat and SatAnalyzer data analysis. Each sample with a female sex identification has a separate bar graph depicting the number of reads for the two sex loci. The caribou sample ID is located at the top of each graph above the Y/X ratio. The X chromosome-specific allele is shown in orange. The Y chromosome-specific allele is shown in blue.

Figure 1.7 Number of reads for the *Zfx* (X) and *Zfy* (Y) alleles in caribou samples with a male sex identification sequenced on a MinION Mk1B (A) and an Illumina MiSeq (B) sequencer. Individuals with a non-zero Y/X ratio are categorized as male based on Seq2Sat and SatAnalyzer data analysis. Each sample with a male sex identification has a separate bar graph depicting the number of reads for the two sex loci. The caribou sample ID is located at the top of each graph above the Y/X ratio. The X chromosome-specific allele is shown in orange. The Y chromosome-specific allele is shown in blue.

Microsatellite Genotyping

After sequencing 19 caribou samples on a MinION Mk1B and an Illumina MiSeq system, sufficient read depth was attained for 301 out of the total 304 microsatellite alleles sequenced. In this experiment, all primary microsatellite allele genotypes with a read depth of 20 or greater were scorable, regardless of the sequencing platform used. For each microsatellite locus, two primary alleles (allele 1 and allele 2) comprised the genotype. Allele 1 and allele 2 genotypes called for the 8 microsatellite loci in each sample sequenced on a MinION Mk1B (Table A8) and an Illumina MiSeq (Table A9) are recorded. Inconclusive allele genotypes were denoted as "-99". Only primary alleles with sufficient read depth and Phred quality scores were considered the true genotype at a given locus.

Examples of the Seq2Sat/SatAnalyzer interface output figures are shown for a standard tetranucleotide microsatellite locus (OHEQ), a dinucleotide microsatellite with a tetranucleotide insert (AC)AAAT(AC) (IGF), and a standard dinucleotide microsatellite (RT27) using data from caribou samples 21102, 21099 and 21106 respectively. SatAnalyzer output figures included a bar graph of the number of reads for all allele sizes detected per microsatellite locus, as well as a genotype summary table that displays the marker of interest, microsatellite repeat unit, microsatellite repeat array (MRA) base, MRA size, allele size, number of reads for each allele, forward flanking region, MRA region, and reverse flanking region. Note that some loci did not show forward or reverse flanking regions in cases where the boundary between the flanking region and MRA region is unclear (Liu et al., 2024). Any microsatellite reads that did not contain the primary alleles are recognized as sequencing errors/PCR artefacts and do not reflect allelic diversity.

SatAnalyzer genotyping analysis of both sequencing platform's OHEQ microsatellite data characterized sample 21102 as homozygous (Figure 1.8). The assigned allele 1 variant contained the MRA region (TATC)10 according to both sequencing platforms. The MinION-based OHEQ allele 1 was 108 bp with 11,337 reads. The Illumina-based OHEQ allele 1 was also 108 bp with 244 reads. In addition to the primary allele, MinION Mk1B sequencing revealed 2 sequencing artefacts for the OHEQ microsatellite locus whereas no sequencing artefacts were obtained for OHEQ in the Illumina data for this caribou sample.

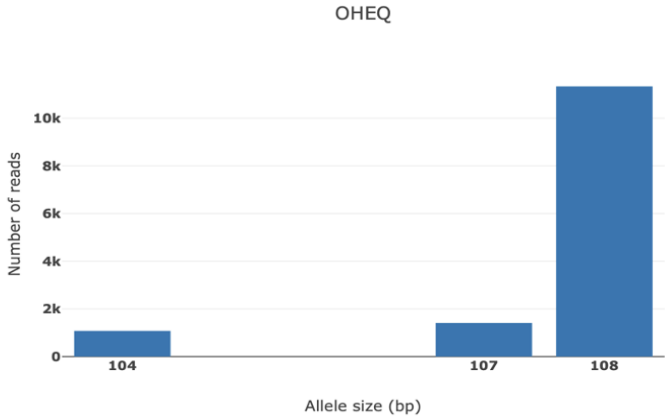
Genotyping through SatAnalyzer characterized sample 21099 as homozygous for the IGF locus analysis, which is consistent across sequencing platforms (Figure 1.9). In both datasets, the called IGF genotype had an allele size of 81 bp and contained the MRA region (AC)5AAAT(AC)8. This IGF allele had 5,181 MinION reads and 1,258 Illumina reads. MinION sequencing also resulted in three sequencing artifacts for the IGF microsatellite locus in sample 21099, while Illumina sequencing resulted in 1 sequencing artifact.

SatAnalyzer genotyping analysis profiled sample 21106 as heterozygous for the RT27 locus according to both sequencing platforms (Figure 1.10). The MinION and Illumina RT27 allele 1 contained the MRA region GA(CA)1GATA(CA)17 and RT27 allele 2 contained the MRA region GA(CA)1GATA(CA)19 for sample 21106. RT27 allele 1 and allele 2 generated from both sequencing platforms were 103 bp and 107 bp, respectively. RT27 allele 1 had 3,333 MinION reads and 882 Illumina reads. RT27 allele 2 had 2,274 MinION reads and 456 Illumina reads. Genotyping analysis captured 5 RT27 sequencing artifacts from MinION sequencing and 4 RT27 artifacts from Illumina sequencing. For the RT27 locus within sample 21106, the Illumina MiSeq morphology exhibited two clear peaks, one allele at 103 bp and the other at 107 bp, resulting in the called genotype. In comparison, the MinION Mk1B morphology showed a

peak at 103 bp, followed by a stutter product at 105 bp, and then a second peak at 107 bp. It is not uncommon for this dinucleotide microsatellite locus to have a stutter product directly after the main target allele peak, before the second target allele peak. These locus-specific patterns in morphology have been determined based on previous ABI work with these microsatellite loci used for caribou monitoring (www.EcoGenomicsCanada.ca).

MinION Mk1B sequencing generated more sequencing artefacts for the target microsatellite loci, sometimes double the number of artefacts detected in the Illumina MiSeq dataset for the same sample or microsatellite locus. This is consistent with the higher error rate observed for nanopore sequencing than Illumina-based approaches (Tytgat et al., 2022; Voskoboinik et al., 2018). Additionally, the use of DNA polymerase during PCR target enrichment prior to amplicon sequencing introduces sequencing errors known as artefacts or more specifically, stutter artefacts (Daniels et al., 1998; Peng et al., 2015). Stutter artefacts arise from strand slippage of DNA polymerase during PCR extension stages, which often results in an additional or deleted microsatellite repeat unit in the newly formed DNA strand (Daniels et al., 1998; Hölzl-Müller et al., 2021; Walsh et al., 1996). Despite sequencing artifacts occurring more frequently in the MinION Mk1B microsatellite dataset in this study, this did not impact accurate genotype calling capabilities.

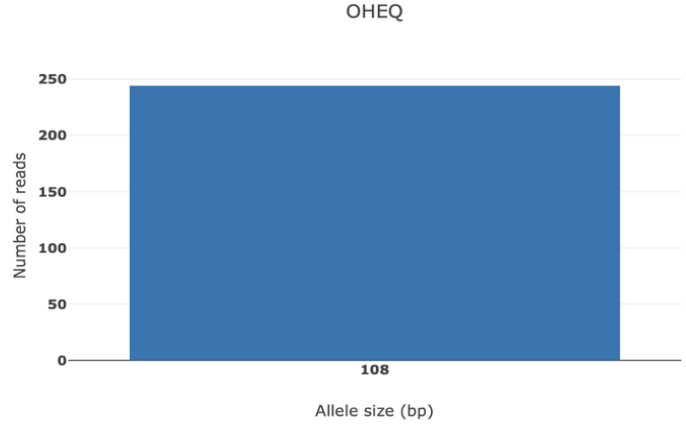
A. Sample 21102 MinION Data



SNPs/artifacts are highlighted in red. N. of Reads are in red are the warnings

Marker	Repeat unit	MRA base	MRA size	Allele size	N. of Reads	Forward flanking region	MRA	Reverse flanking region
Reference	TATC	(TATC)	40	108	N.A.	GGGGTCAAAGACTAAATATTAGCAGGAGTCAG	(TATC)10	ATCATCATCTATTAATATCTATCATCTACCTAATTT
OHEQ	TATC	(TATC)	36	104	1073	GGGGTCAAAGACTAAATATTAGCAGGAGTCAG	(TATC)9	ATCATCATCTATTAATATCTATCATCTACCTAATTT
OHEQ	TATC	ATC(TATC)	39	107	1406	GGGTCAAAGACTAAATATTAGCAGGAGTCAGT	ATC(TATC)9	ATCATCATCTATTAATATCTATCATCTACCTAATTT
OHEQ	TATC	(TATC)	40	108	11337	GGGGTCAAAGACTAAATATTAGCAGGAGTCAG	(TATC)10	ATCATCATCTATTAATATCTATCATCTACCTAATTT

B. Sample 21102 Illumina Data

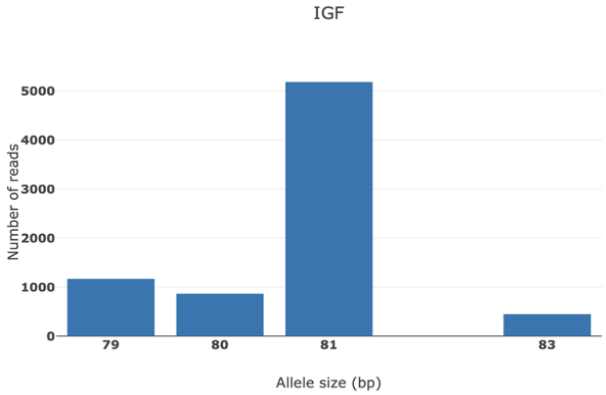


SNPs/artifacts are highlighted in red. N. of Reads are in red are the warnings

Marker	Repeat unit	MRA base	MRA size	Allele size	N. of Reads	Forward flanking region	MRA	Reverse flanking region
Reference	TATC	(TATC)	40	108	N.A.	GGGGTCAAAGACTAAATATTAGCAGGAGTCAG	(TATC)10	ATCATCATCTATTAATATCTATCATCTACCTAATTT
OHEQ	TATC	(TATC)	40	108	244	GGGGTCAAAGACTAAATATTAGCAGGAGTCAG	(TATC)10	ATCATCATCTATTAATATCTATCATCTACCTAATTT

Figure 1.8 SatAnalyzer bar graph for the OHEQ tetranucleotide microsatellite locus in sample 21102 of the MinION Mk1B dataset (A) and Illumina MiSeq dataset (B). Each bar represents the number of reads for a unique microsatellite allele size (bp) detected. Bars with more than one colour indicate that multiple distinct microsatellite reads were detected that have the same allele size. The summary genotype table displays the target microsatellite marker (column 1), microsatellite repeat units (column 2), microsatellite repeat array (MRA) (column 3), MRA size (column 4), allele size (column 5), number of reads for each allele present (column 6), forward flanking region (column 7), MRA region (column 8), and the reverse flanking region (column 9). Blue text represents information about the reference sequence (row 1). The primary allele size (108 bp) for the OHEQ locus is highlighted in green. The individual is homozygous for the OHEQ locus.

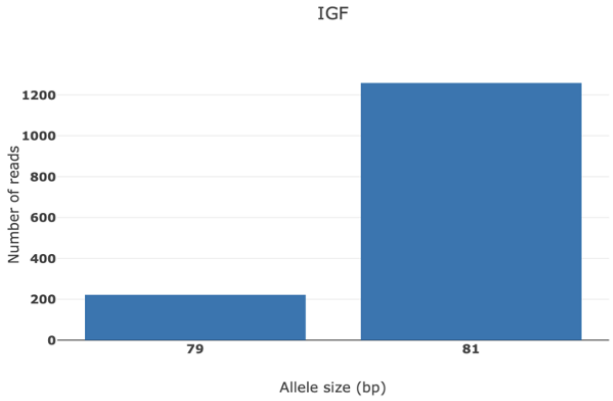
A. Sample 21099 MinION Data



SNPs/artifacts are highlighted in red. N. of Reads are in red are the warnings

Marker	Repeat unit	MRA base	MRA size	Allele size	N. of Reads	Forward flanking region	MRA	Reverse flanking region
Reference	AC	(AC)AAAT(AC)	30	81	N.A.	GTGTTATTTAGAATACACAAAAAATGGGGGAAAGAAAATGCACTCACGTGC	(AC)5AAAT(AC)8	
IGF	AC	(AC)AAAT(AC)	28	79	1167	GTGTTATTTAGAATACACAAAAAATGGGGGAAAGAAAATGCACTCACGTGC	(AC)5AAAT(AC)7	
IGF	AC	C(AC)AAAT(AC)	29	80	866	GTGTTATTTAGAATACACAAAAAATGGGGGAAAGAAAATGCACTCACGTGCA	C(AC)4AAAT(AC)8	
IGF	AC	(AC)AAAT(AC)	30	81	5181	GTGTTATTTAGAATACACAAAAAATGGGGGAAAGAAAATGCACTCACGTGC	(AC)5AAAT(AC)8	
IGF	AC	(AC)AAAT(AC)	32	83	449	GTGTTATTTAGAATACACAAAAAATGGGGGAAAGAAAATGCACTCACGTGC	(AC)5AAAT(AC)9	

B. Sample 21099 Illumina Data

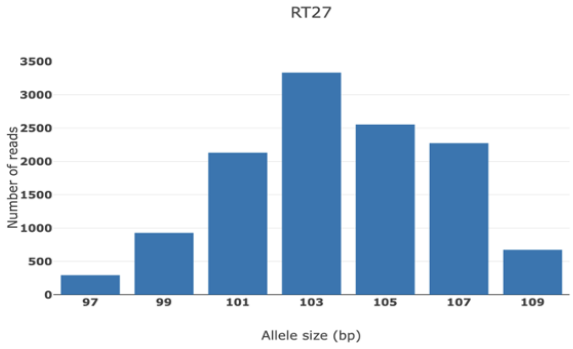


SNPs/artifacts are highlighted in red. N. of Reads are in red are the warnings

Marker	Repeat unit	MRA base	MRA size	Allele size	N. of Reads	Forward flanking region	MRA	Reverse flanking region
Reference	AC	(AC)AAAT(AC)	30	81	N.A.	GTGTTATTTAGAATACACAAAAAATGGGGGAAAGAAAATGCACTCACGTGC	(AC)5AAAT(AC)8	
IGF	AC	(AC)AAAT(AC)	28	79	222	GTGTTATTTAGAATACACAAAAAATGGGGGAAAGAAAATGCACTCACGTGC	(AC)5AAAT(AC)7	
IGF	AC	(AC)AAAT(AC)	30	81	1258	GTGTTATTTAGAATACACAAAAAATGGGGGAAAGAAAATGCACTCACGTGC	(AC)5AAAT(AC)8	

Figure 1.9 SatAnalyzer bar graph for the IGF microsatellite locus in sample 21099 of the MinION Mk1B dataset (A) and Illumina MiSeq dataset (B). Each bar represents the number of reads for a unique microsatellite allele size (bp) detected. The summary genotype table displays the target microsatellite marker (column 1), microsatellite repeat units (column 2), microsatellite repeat array (MRA) (column 3), MRA size (column 4), allele size (column 5), number of reads for each allele present (column 6), forward flanking region (column 7), MRA region (column 8), and the reverse flanking region (column 9). Blue text represents information about the reference sequence (row 1). The MRA base reference sequence for the IGF locus displays a dinucleotide microsatellite with a tetranucleotide insert. The primary allele size (81 bp) is highlighted in green. This individual is homozygous for the IGF locus.

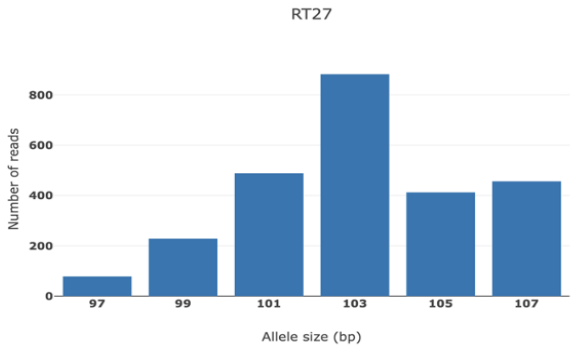
A. Sample 21106 MinION Data



SNPs/artifacts are highlighted in red. N. of Reads are in red are the warnings

Marker	Repeat unit	MRA base	MRA size	Allele size	N. of Reads	Forward flanking region	MRA	Reverse flanking region
Reference	CA	GA(CA)GATA(CA)	42	103	N.A.	T	GA(CA)1GATA(CA)17	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	36	97	292	T	GA(CA)1GATA(CA)14	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	38	99	928	T	GA(CA)1GATA(CA)15	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	40	101	2131	T	GA(CA)1GATA(CA)16	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	42	103	3333	T	GA(CA)1GATA(CA)17	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	44	105	2554	T	GA(CA)1GATA(CA)18	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	46	107	2274	T	GA(CA)1GATA(CA)19	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	48	109	673	T	GA(CA)1GATA(CA)20	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC

B. Sample 21106 Illumina Data



SNPs/artifacts are highlighted in red. N. of Reads are in red are the warnings

Marker	Repeat unit	MRA base	MRA size	Allele size	N. of Reads	Forward flanking region	MRA	Reverse flanking region
Reference	CA	GA(CA)GATA(CA)	42	103	N.A.	T	GA(CA)1GATA(CA)17	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	36	97	78	T	GA(CA)1GATA(CA)14	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	38	99	228	T	GA(CA)1GATA(CA)15	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	40	101	488	T	GA(CA)1GATA(CA)16	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	42	103	882	T	GA(CA)1GATA(CA)17	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	44	105	412	T	GA(CA)1GATA(CA)18	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC
RT27	CA	GA(CA)GATA(CA)	46	107	456	T	GA(CA)1GATA(CA)19	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC

Figure 1.10 SatAnalyzer bar graph for the RT27 microsatellite locus in sample 21106 of the MinION Mk1B dataset (A) and Illumina MiSeq dataset (B). Each bar represents the number of reads for a unique microsatellite allele size (bp) detected. The genotype summary table displays the target microsatellite marker (column 1), microsatellite repeat units (column 2), microsatellite repeat array (MRA) (column 3), MRA size (column 4), allele size (column 5), number of reads for each allele present (column 6), forward flanking region (column 7), MRA region (column 8), and the reverse flanking region (column 9). Blue text represents information about the reference sequence (row 1). The two primary allele sizes (103 bp and 107 bp) are highlighted in green. This individual is heterozygous for the RT27 locus.

In most cases, microsatellite allele reads detected in the Illumina MiSeq dataset were also detected in the MinION Mk1B dataset for the same sample and locus. Out of the 304 total

primary microsatellite alleles sequenced and genotyped, 300 alleles were assigned the same genotype in the MinION Mk1B and Illumina MiSeq dataset, resulting in a genotyping congruence rate of 98.68% (Figure 1.11). The MinION Mk1B dataset contained 2 inconclusive allele genotypes, accounting for 0.65% of the total number of alleles sequenced. The Illumina MiSeq dataset also contained 2 inconclusive allele genotypes, accounting for 0.65% of alleles sequenced. No alleles had differing assigned genotypes between platforms excluding inconclusive allele genotypes, resulting in a genotyping non-congruence rate of 0% across sequencing platforms. Overall, the cross-platform consistency suggests that the MinION Mk1B sequencer provides reproducible microsatellite sequencing data to that of the standard Illumina MiSeq.

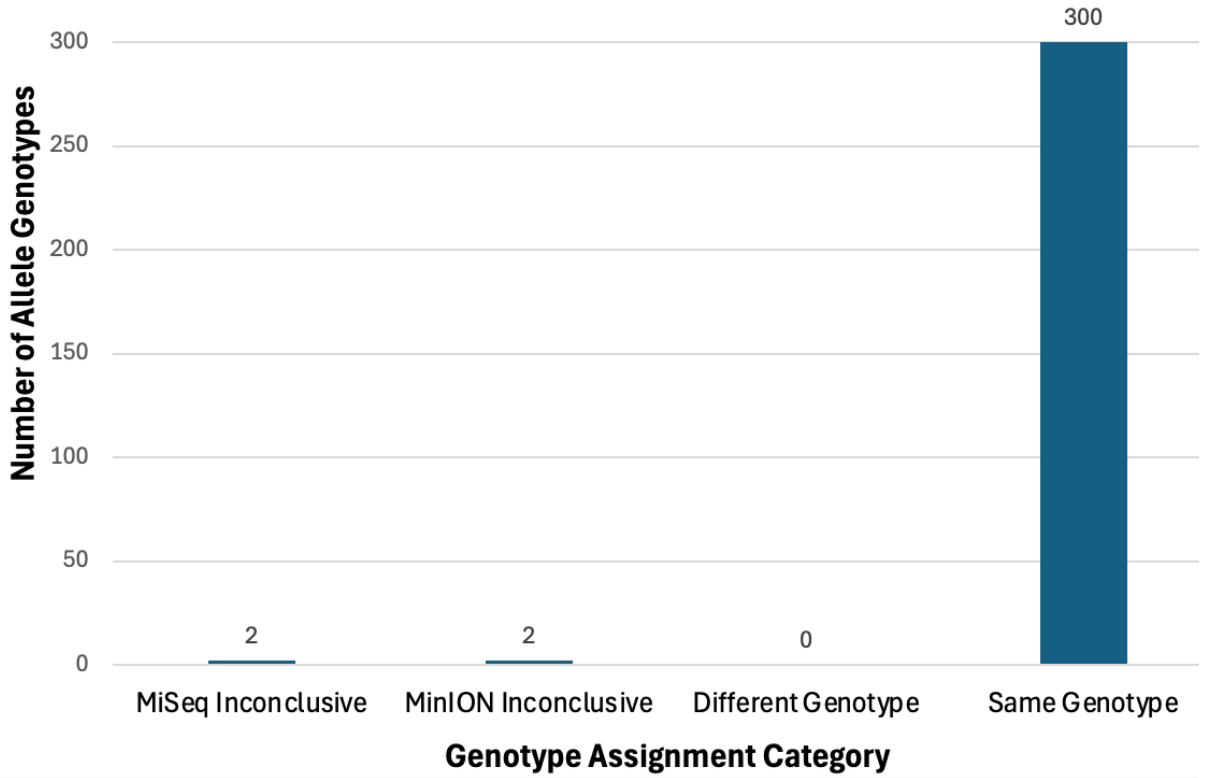


Figure 1.11 Number of microsatellite alleles in each genotype assignment category for the 304 alleles sequenced on the MinION Mk1B and Illumina MiSeq sequencer. Allele 1 and allele 2 genotypes for 8 microsatellites in 19 caribou samples are compared across sequencing platforms. If the same genotype was assigned for the same microsatellite allele and sample across platforms, the genotype assignment is categorized as "Same Genotype". If a different genotype was assigned for a specific microsatellite allele and sample across platforms, the genotype assignment is categorized as "Different Genotype". Inconclusive allele genotypes are categorized as "MiSeq Inconclusive" or "MinION Inconclusive" based on what platform had the inconclusive genotype. Inconclusive genotypes are not included in the "Same Genotype" or "Different Genotype" categories count.

The number of same, different, and inconclusive allele genotypes across sequencing platforms are shown according to microsatellite allele (Figure 1.12). There were 16 primary alleles (allele 1 and allele 2 for 8 microsatellite loci) sequenced within 19 caribou samples. Twelve out of the 16 microsatellite alleles sequenced in each sample were identical across sequencing platforms for every sample. There were no samples in which both IGF alleles, NVHRT48 alleles, RT24 alleles, RT27 alleles, RT6 alleles or RT7 alleles had different genotypes based on the sequencing platform used. Both MAP2C alleles each had one sample

containing a MinION inconclusive allele genotype and both OHEQ alleles each had one sample containing a MiSeq inconclusive allele genotype.

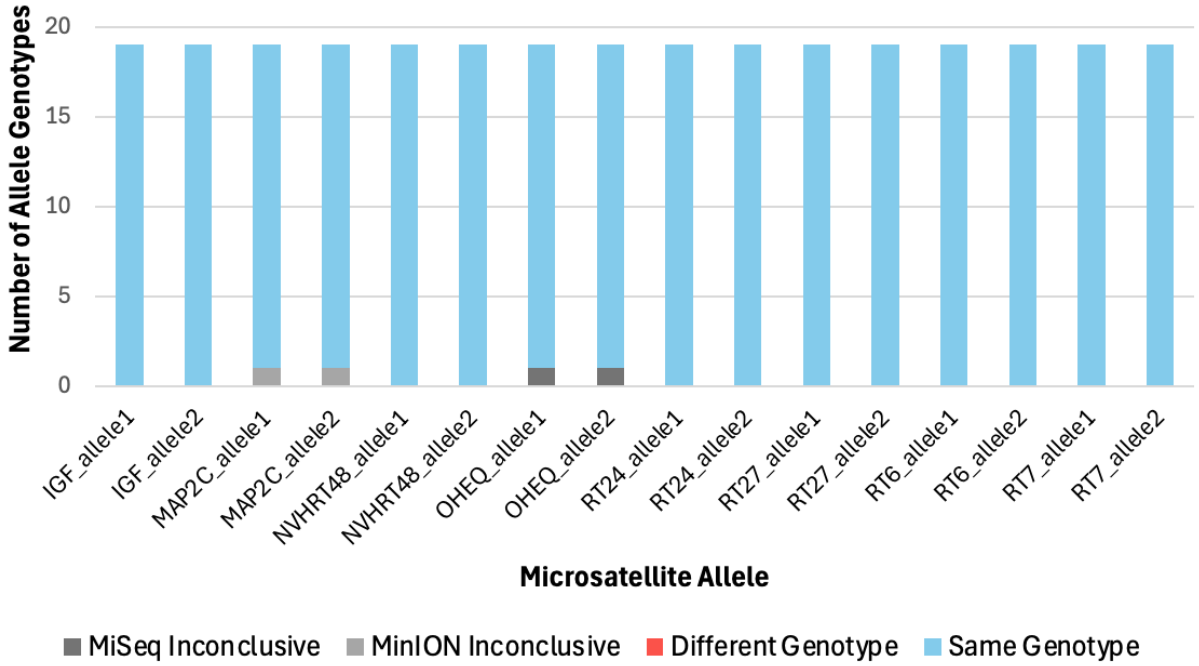


Figure 1.12 Number of alleles in each genotype assignment category according to microsatellite allele for 19 caribou samples sequenced on a MinION Mk1B and an Illumina MiSeq sequencer. The two primary allele genotypes (allele 1 and allele 2) for 8 microsatellite loci are categorized as "Same Genotype", "Different Genotype", "MiSeq Inconclusive" or "MinION Inconclusive". If the same genotype was assigned for the same microsatellite allele and sample across platforms, the genotype assignment is categorized as "Same Genotype". If a different genotype was assigned for a specific microsatellite allele and sample across platforms, the genotype assignment is categorized as "Different Genotype". Inconclusive allele genotypes are categorized as "MiSeq Inconclusive" or "MinION Inconclusive" based on what platform had the inconclusive genotype. Inconclusive genotypes are not included in the "Same Genotype" or "Different Genotype" categories count.

The number of same, different, and inconclusive microsatellite allele genotypes across sequencing platforms are shown according to caribou sample ID (Figure 1.13). Seventeen out of 19 caribou samples (89.47%) were assigned the same genotypes for every allele sequenced with no inconclusive genotypes: 20919, 21099, 21102, 21104, 21105, 21106, 21107, 21110, 21112, 21113, 21297, 21299, 21301, 21303, 21305, 21306, and 21309. The remaining samples (21101 and 21298) contained two inconclusive alleles each. In the Illumina MiSeq dataset, two alleles

were assigned as MiSeq inconclusive for sample 21101. In the MinION dataset, there were two MinION inconclusive alleles occurring in sample 21298 data.

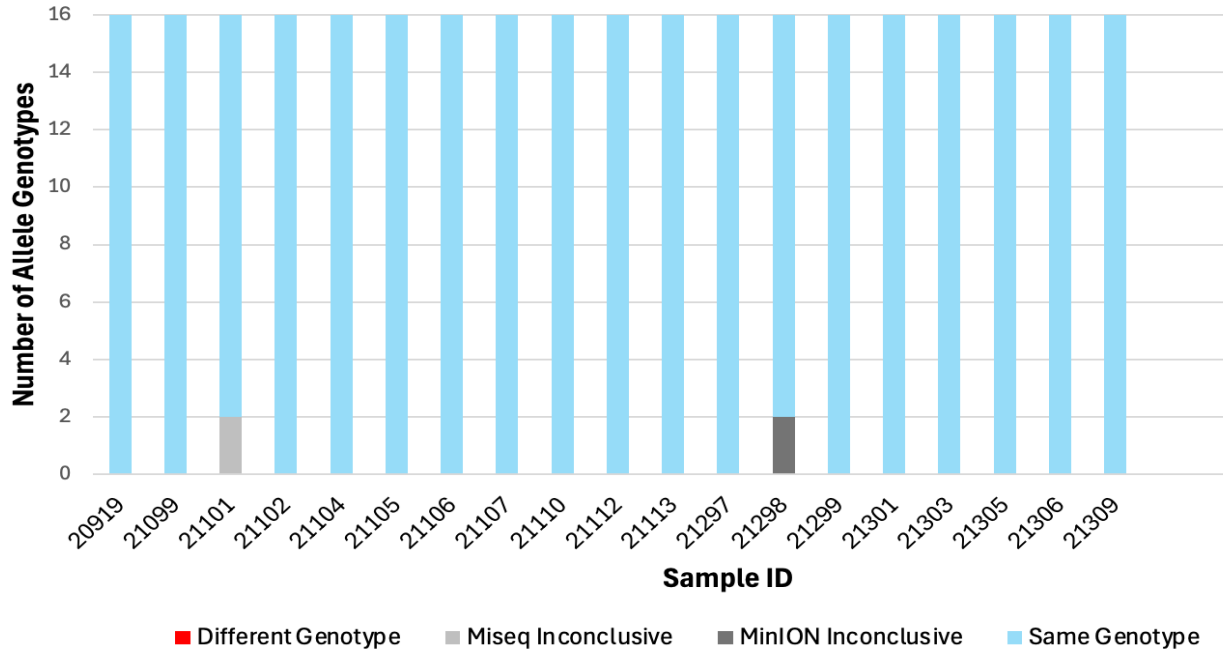


Figure 1.13 Number of microsatellite alleles in each genotype assignment category according to caribou sample ID. The two primary allele genotypes for 8 microsatellite loci sequenced on a MinION Mk1B and an Illumina MiSeq sequencer are categorized as "Same Genotype", "Different Genotype", "MiSeq Inconclusive" or "MinION Inconclusive". If the same genotype was assigned for the same microsatellite allele and sample across platforms, the genotype assignment is categorized as "Same Genotype". If a different genotype was assigned for a specific microsatellite allele and sample across platforms, the genotype assignment is categorized as "Different Genotype". Inconclusive allele genotypes are categorized as "MiSeq Inconclusive" or "MinION Inconclusive" based on what platform had the inconclusive genotype. Inconclusive genotypes are not included in the "Same Genotype" or "Different Genotype" categories count.

The two inconclusive allele genotypes in the MinION Mk1B dataset were MAP2C allele 1 and MAP2C allele 2 in sample 21298. Neither allele had 20 or more reads, making this microsatellite locus not scorable for sample 21298. As a result, its genotype was recorded as -99. The two inconclusive allele genotypes in the Illumina MiSeq dataset were OHEQ allele 1 and OHEQ allele 2 in sample 21101. Only 22 reads were obtained for OHEQ allele 1 while 8 reads were obtained for OHEQ allele 2. Although allele 1 had slightly more than 20 reads, the number of reads was still quite low and there were not enough reads for allele 2 to confidently call a

genotype for this locus. None of the inconclusive genotypes observed in either dataset qualified for genotype scoring because of the low read depth for one or both potential primary alleles.

A limitation of this study is the manual scoring of genotypes. Although microsatellite genotypes were initially scored automatically in Seq2Sat, genotypes were manually assessed and occasionally overridden based on scoring rules developed for each microsatellite/sex locus. Because microsatellite genotypes were manually assigned following automatic scoring, there is potential for incorrect or inconsistent genotype calls. Accurately calling genotypes is especially challenging in cases where the two potential primary alleles have similar length, making it difficult to identify the true allele length(s) based on the observed allele distribution patterns (Suez et al., 2015). To manually score genotypes as consistently as possible, two individuals independently applied the same set of genotype scoring rules that were developed to determine the genotype for each target locus in all 19 caribou samples sequenced on a MinION Mk1B and an Illumina MiSeq. Since all microsatellite genotypes with a minimum read depth 20 reads and a quality score of Q10+ for MinION reads or Q20+ for Illumina reads were scorable and the same across sequencing platforms, this suggests that MinION Mk1B-derived microsatellite genotypes can be called on par with the well-established Illumina MiSeq using these parameters. Therefore, if the appropriate genotype scoring rules are applied, microsatellite genotypes should be consistent whether a MinION Mk1B or an Illumina MiSeq instrument was used to sequence the target loci within each caribou sample.

Cost Breakdown

Cost breakdowns are provided for several scenarios involving microsatellite amplicon sequencing with the MinION Mk1B and Illumina MiSeq for various sample size sets, including the approach applied in this study (Table A10). The cost breakdown for microsatellite amplicon

sequencing performed here included 11 target loci: 8 microsatellite loci, the LGL335 and SDP730 (*Zfx*/*Zfy*) loci, and the PRNP locus. Additional cost breakdowns provided included approximate expenses incurred for sequencing a full microhaplotype loci panel developed in-house, which consists of 23 microsatellites plus the *Zfx*/*Zfy* loci and the PRNP locus, in 96 samples on a MinION Mk1B (Table A11), 384 samples on a MinION Mk1B (Table A11, repeated 4X), as well as 96 samples on an Illumina MiSeq (Table A12) and 384 samples on an Illumina MiSeq platform (Table A13). All prices were approximate and were shown in Canadian Dollars ($CAD). The cost of a MinION Mk1B and Illumina MiSeq sequencer was not included in the cost breakdown. However, a MinION Mk1B instrument (including startup kit) costs ~$2714 CAD and an Illumina MiSeq costs $157,749 CAD according to manufacturer list prices.

For MinION Mk1B sequencing, costs were separated into MinION-specific expenses and additional expenses. MinION-specific expenses included an R10.4.1 flow cell, ONT Native Barcoding Kit, and 3rd party NEB reagents. It was assumed that flow cells are not reused and that the recommended reagent amounts as indicated in the ONT Ligation Sequencing Amplicons protocol are used. Additional expenses included 2x Type-It PCR Master Mix (Qiagen) and IDT's standard DNA Oligonucleotides primer pairs for PCR amplification of the target loci. The amount of 2x Type-It PCR Master Mix (Qiagen) and IDT standard DNA Oligonucleotide primers were totalled for all multiplex PCR amplifications involved in executing a sequencing run. Note that the estimated total volume of primers across both PCR multiplexes included a primer set for the PRNP locus (Geue et al., 2024), despite this locus having not been examined in this study. The cost of Invitrogen's Qubit dsDNA High Sensitivity dsDNA Assay Kit reagents was also recorded as an additional expense. Qubit reagents were needed to quantify each sample

to confirm each sample had 200 fmol of DNA before starting library preparation and they were also needed to quantify the final DNA library.

The estimated total submission cost for sequencing 11 loci in 19 samples on the MinION Mk1B is $1,255.92, with a cost per sample and cost per locus of approximately $66.10 and $6.01, respectively (Table 1.1). The number of samples comprising a single DNA library will be maximized to 96 samples moving forward to reduce the total submission cost and cost per sample/locus when sequencing the full panel of microsatellites. This would reduce the cost to $24.04 per sample and $0.92 per locus on the MinION Mk1B for 26 target loci. To sequence 384 samples, the 96 sample MinION Mk1B run must be repeated four times. This does not change the cost per sample or cost per locus, however, the total submission cost increases from $2,308.24 for 96 samples to $9,232.96 for 384 samples.

Currently, the MinION flow cell is the most expensive when sequencing the maximum number of samples (96) at a time, adding $9.97 per sample. The Native Barcoding Kit 96 v14 (SQK-NBD114.96) is the second largest expense per sample. Unfortunately, nothing can be done by users to change these set costs. However, there is an opportunity to reduce the NEB 3rd party reagents costs which are also a big contributor to MinION Mk1B sequencing expenses. Previous experiments have demonstrated the feasibility of halving the NEBNext Ultra II End Repair/dA-tailing Module (NEB, cat # E7546L) and NEB Blunt/TA Ligase Master Mix (NEB, cat # M0367L) reagent volumes consumed during library preparation steps (van der Reis et al., 2022). By halving these NEB end prep and native barcode ligation reagent volumes, the total submission cost for 96 samples on the MinION Mk1B would be reduced from $2,308.24 to $2,033.21 and the cost per sample would be $21.18 instead of $24.04. Additionally, rinsing and reusing flow cells is an option to reduce sequencing costs. Rinsing the flow cell with a wash kit

and sequencing a second sample set of 96 individuals would lower the total submission cost to $1,840.91 per sequencing run (96 samples), at $19.18 per sample. If halving NEB end prep and native barcode ligation reagent volumes and reusing flow cells, sequencing costs could potentially be lowered to $16.31 per sample for 96 samples. This strategy would require optimization to ensure adequate reads are generated per sample and locus for each run. Finally, it is also advantageous to purchase NEB 3rd party reagents as well as flow cells in bulk. This adds to upfront costs but is cheaper in the long run when routinely performing metabarcoding experiments (van der Reis et al., 2022).

Illumina MiSeq expenses included the Illumina 500 Cycle V2 Kit, 2x Type-It PCR Master Mix (Qiagen), IDT's standard DNA Oligonucleotide primer pairs for all PCR amplifications, AMPure XP beads (Beckman Coulter), Illumina barcoding indexes, and Qubit High Sensitivity DsDNA Assay Kit (Invitrogen) reagents. Typically, 384 caribou samples, with each sample having undergone 4 multiplex PCR amplification reactions, are pooled and sequenced in-lab to lower Illumina MiSeq sequencing costs. The cost of these 4 PCR multiplexes accounted for enough 2x Type-It PCR Master Mix (Qiagen) and IDT standard DNA Oligonucleotide primers to amplify the full panel of 23 microsatellite loci, the LGL335 and SDP730 (*Zfx*/*Zfy*) loci, and the PRNP locus. As a result, the cost per locus estimate was based on sequencing 26 target loci in either 96 or 384 samples on the Illumina MiSeq.

The estimated total submission cost for sequencing 26 target loci in 96 samples and 384 samples on the Illumina MiSeq is $2,975.32 and $5,268.28, respectively (Table 1.1). The approximate cost per sample is $30.99 when sequencing 96 samples in one run, while the cost per sample is $13.72 when sequencing 384 samples in one run. This corresponds to a per locus cost of $1.19 for a 96-sample set and $0.53 for a 384-sample set. The largest expense for

sequencing microsatellite amplicons on an Illumina MiSeq system is the cost of the Illumina 500 Cycle V2 Kit. The Qiagen 2x Type-It PCR Master Mix is the second largest expense and the indexes for all PCR reactions are the third largest expense for each Illumina sequencing run. Although the price of the Illumina 500 Cycle V2 Kit and PCR indexes are set by the manufacturers, investigating ways to further reduce the amount of Qiagen 2x Type-It PCR Master Mix would lower the total submission cost and cost per sample for a sequencing run.

Table 1.1 Estimated total submission cost, average cost per sample, and average cost per locus for sequencing a variable number of samples and microsatellite loci with a MinION Mk1B or an Illumina MiSeq. The cost estimate for this study is shown (11 target loci; 19 samples). A cost estimate is also shown for sequencing the full microsatellite panel, which includes 26 loci, in 96 or 384 samples with a MinION Mk1B and an Illumina MiSeq. Prices are shown in $CAD.

Study Design	Total Submission Cost	Cost Per Sample	Number of Loci	Cost Per Locus
19 Samples MinION Mk1B	$1,255.82	$66.10	11 (Subset of Microsatellite Panel)	$6.01
96 Samples MinION Mk1B	$2,308.24	$24.04	26 (Full Microsatellite Panel)	$0.92
384 Samples MinION Mk1B	$9,232.96	$24.04	26 (Full Microsatellite Panel)	$0.92
96 Samples Illumina MiSeq	$2,975.32	$30.99	26 (Full Microsatellite Panel)	$1.19
384 Samples Illumina MiSeq	$5,268.28	$13.72	26 (Full Microsatellite Panel)	$0.53

When sequencing 96 samples/1 PCR plate in one run, it is more cost-effective to use the MinION Mk1B platform. If conducting Illumina MiSeq sequencing, the Illumina V2 500 Kit alone exceeds the total submission cost and cost per sample of that of the MinION Mk1B to sequence one plate of samples. Therefore, when the goal of an experiment is to sequence 96 samples or fewer in one sequencing run, the MinION Mk1B is cheaper. If the goal is to sequence more than 96 samples for a given experiment, whether it is 2, 3 or 4 PCR plates worth of

samples, the Illumina MiSeq sequencing is the better option financially to minimize sequencing costs. The greater the number of samples sequenced on an Illumina MiSeq that exceeds 96 samples, the more cost-effective Illumina sequencing becomes.

It is important to note that the price of sequencing platforms and costs associated with their use frequently change over time. Opportunities to further reduce costs of microsatellite profiling will continue to arise alongside technological advances, which may or may not include the MinION Mk1B and Illumina MiSeq platforms. However, the MinION Mk1B platform and workflow shown here could add value to the pre-existing Illumina MiSeq-based approach that is currently taken by Eco Genomics Canada (www.EcoGenomicsCanada.ca) to characterize microsatellite loci in caribou for routine population monitoring.

Conclusion

This experiment provides proof of concept that MinION Mk1B amplicon sequencing is a viable alternative to the widely used Illumina MiSeq for microsatellite genotyping, while reducing costs for sample sets of 96 or fewer unique caribou individuals. Next steps involve sequencing these microsatellite and sex target amplicons on a MinION Mk1B with an increased sample size of 96 caribou. This would greatly reduce the per sample cost compared to sequencing 24 or fewer samples per MinION Mk1B run. Moreover, although the current per sample cost for sequencing 384 samples on an Illumina MiSeq is lower than the estimated per sample cost of sequencing 96 samples on a MinION Mk1B, there are advantages to using the latter approach. For example, if a sequencing run fails, more time and money is lost from a 384 sample Illumina MiSeq run since the laboratory procedure takes less time to complete for a 96 sample MinION Mk1B run and its total submission cost is lower.

Additionally, increasing the MinION Mk1B sample size to 96 would provide the opportunity to add a subset of caribou fecal samples to the existing set of 19 caribou tissue samples. DNA extracted from caribou fecal samples are often used in lab since they are obtained via non-invasive sampling methods, which is preferable for sampling DNA from vulnerable or elusive species. If the MinION Mk1B can generate microsatellite genotypes from fecal DNA samples that are comparable to ones generated by the Illumina MiSeq used for caribou population monitoring, this would provide further support for the MinION's implementation in routine population genetics research.

CHAPTER 3

Genotyping a Novel Microhaplotype Panel for Caribou Monitoring in Canada with the MinION Nanopore Sequencer and Illumina MiSeq

Abstract

Microhaplotype profiling is becoming increasingly popular in the field of conservation genomics as this type of genetic marker offers high discriminating power that can be harnessed for individual identification and relatedness analysis in wildlife populations. Microhaplotypes, which are short DNA regions containing closely linked SNPs, provide more information per locus than individual SNPs and in theory should allow for more automated genotype scoring than sequencing microsatellite repeats, which have been typically used for these applications. This study presents the development and validation of a novel panel of 25 exonic microhaplotype genetic markers to aid population monitoring of caribou (*Rangifer tarandus*) across Canada. To evaluate the robustness and accuracy of this marker panel, 23 caribou tissue DNA samples were sequenced using two high-throughput platforms: the Oxford Nanopore MinION Mk1B and the Illumina MiSeq. Longer amplicon fragments (186-394 bp) were designed to be compatible with MinION Mk1B and Illumina MiSeq sequencing while maximizing the capture of synonymous and nonsynonymous changes. Comparative analyses of sequencing performance, including read depth/distribution, read accuracy, genotyping consistency, and cost-efficiency, were conducted across both platforms. Microhaplotype scoring for each sample and locus was conducted using the Seq2Sat V2 custom bioinformatics tool. Our findings demonstrate that ambiguous read distributions tend to negatively impact haplotype calling concordance across sequencing platforms. We recommend manually checking the read depth of the top 3 reads per locus and filtering out sample data at loci where the read depth of read 3 is very similar to the read depth of

read 2 (within a heterozygote) to mitigate false haplotype calls when using either sequencing platform. MinION Mk1B sequencing resulted in higher error rates and poorer read distribution compared to the Illumina MiSeq approach. Consequently, further research is needed to determine whether microhaplotype profiling in caribou using the MinION Mk1B platform is a viable strategy.

Introduction

The highly polymorphic and discriminating nature of microhaplotypes (Kidd et al., 2014) makes it a promising genetic marker for wildlife population monitoring. Microhaplotypes consist of two or more closely linked SNPs over a short region of DNA (typically <300 bp) that have multiple allelic combinations due to the inclusion of multiple SNPs in proximity (Bennett et al., 2019; Carratto et al., 2022; Oldoni et al., 2018). These characteristics allow for accurate assessments of individual identification, genetic diversity, and biogeographic ancestry, even for closely related species or populations with complex genetic histories (Carratto et al., 2022; Kidd et al., 2014; Oldoni et al., 2018; Xue et al., 2023).

Although microhaplotypes are a relatively newer type of genetic marker, there are advantages of using microhaplotypes over more traditional markers like single SNPs and microsatellite repeats. For example, microhaplotypes reveal more extensive levels of polymorphism relative to single SNPs since they are characterized by multiple alleles and contain more information per locus (De la Puente et al., 2020; Oldoni et al., 2018). Rather than only obtaining individual SNP changes, microhaplotypes capture added sequence variation spanning the entire amplicon length as well as all SNP variation present in the microhaplotype sequence. Additionally, microhaplotypes are preferable for downstream genotyping analysis compared to microsatellites because they circumvent common issues, such as stutter fragments,

that pose challenges to individual identification (Bennett et al., 2019; Carratto et al., 2022; De la Puente et al., 2020; Xue et al., 2023). Microhaplotypes are amenable to automated scoring while capturing extensive variability since SNP alleles can be assessed directly from the target haplotype sequences obtained via high-throughput DNA sequencing (Eriksson et al., 2020; Geue et al., 2024). Automated microhaplotype scoring also generates more standardized results, which is needed to bridge the gap between researchers and conservation managers to make use of conservation genomic inferences (Baetscher et al., 2018; Eriksson et al., 2020; Geue et al., 2024; Von Thaden et al., 2020).

In the field of conservation genomics, the ability to accurately identify individuals, characterize diversity, and establish relatedness is paramount to effectively monitor wildlife populations and inform conservation management decisions (Arif et al., 2011; Baetscher et al., 2018; Delomas et al., 2023; Hohenlohe et al., 2020). As such, developing microhaplotype panels to genotype wildlife DNA samples is gaining popularity. Recently, Delomas et al. (2023) selected a panel of highly polymorphic microhaplotype loci to facilitate monitoring of gray wolves (*Canis lupus*). This genotyping panel provided an accurate assessment of population parameters, including species differentiation, individual identification and relationship inference, contributing to more informed decisions made by wildlife managers (Delomas et al., 2023). Baetscher et al. (2018) also found that microhaplotype markers bolster relationship inference capabilities after targeting these multi-SNP regions in kelp rockfish (*Sebastes atrovirens*).

A similar microhaplotype-based approach could be implemented to enhance long-term caribou (*Rangifer tarandus*) monitoring strategies. With the improved accuracy and analytical power achieved through microhaplotype sequencing, a more efficient method of identifying individuals and determining kinship is available, as well as an opportunity to reduce genotyping

costs (Baetscher et al., 2018). An important consideration for conservation genomics research is cost-effectiveness as more affordable protocols help sustain longstanding projects and allow smaller laboratory facilities with limited resources/funding to contribute to population monitoring efforts. In addition to the reduced genotyping costs associated with microhaplotypes due to their high discriminating power, allowing fewer loci to be sequenced per individual to attain strong inferences, continued advancements in high-throughput sequencing technologies are resulting in more options to lower costs per individual (Baetscher et al., 2018).

To date, many wildlife population genetics studies involving the development of novel microhaplotype panels rely on Illumina sequencing technology. However, the MinION Mk1B nanopore DNA sequencer by Oxford Nanopore Technologies (ONT) may provide a comparable alternative. Since microhaplotypes can be much longer DNA regions than microsatellite repeats and individual SNPs, the long-read nature of nanopore sequencing might be better suited for this type of genetic marker compared to shorter-read sequencers like the Illumina MiSeq. Furthermore, the MinION Mk1B had a substantially lower start-up cost of ~$2714 CAD for a sequencer and start-up kit in contrast with the popular Illumina MiSeq system which cost ~$157,749 CAD. Validating these technologies' applicability to generate informative multi-allelic microhaplotype data would provide options to increase data accessibility and standardization, which are vital to effectively monitor wildlife populations and inform management strategies.

Here, a novel panel of 25 microhaplotype loci at coding regions of various genes were sequenced within 23 caribou tissue samples on a MinION Mk1B and an Illumina MiSeq instrument. Full haplotype sequences for each caribou sample and microhaplotype locus were assessed to validate the use of either sequencer to perform microhaplotyping and identify

haplotypic diversity. A cost estimate associated with each microhaplotyping approach was also provided. Sequencing the same microhaplotype loci within the same caribou samples on a MinION Mk1B and an Illumina MiSeq enabled a direct comparison of microhaplotype sequencing capabilities across sequencing platforms. Finally, recommendations for an optimal microhaplotype assay are delivered to supplement individual identification and kinship analysis strategies for caribou population monitoring over time.

Materials & Methods

Microhaplotype Panel Optimization

Thirty-five caribou chromosomes were surveyed at coding regions of genes using the annotated chromosome-level reference caribou genome by Taylor et al. (2019) in Geneious Prime (V2023.1.1). Over 60 genes were identified as having numerous non-synonymous (Ka) and synonymous (Ks) SNPs that could generate variable microhaplotypes across the caribou genome. Genic regions with large numbers of Ka and Ks changes were preferentially selected. Of the identified genes containing Ka and Ks changes, only exons with approximately 300-400 bp regions were included for optimization to make the regions compatible with the Illumina MiSeq platform as well as the MinION Mk1B platform (65 genes).

Standard DNA Oligonucleotides primers (Integrated DNA Technologies Inc, 25 nm scale) for nanopore sequencing were used to amplify the 65 target microhaplotype loci. These primers were tested for downstream sequencing through a temperature gradient PCR amplification and subsequent capillary electrophoresis using a QIAxcel Advanced System (Qiagen). Three single-plex PCR amplifications were performed for each primer pair: one using DNA extracted from brain tissue of an adult caribou (CTRL3A), a second using DNA extracted

from brain tissue of a caribou calf (CTRL3C), and a third being a no DNA control. The temperature gradient indicated that the optimal annealing temperature was 58°C.

This annealing temperature was then tested on the same microhaplotype loci with Illumina-specific primers, which were IDT's standard DNA oligonucleotides with the Illumina overhang adaptor sequence added to the locus-specific sequence (Forward overhang: 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, Reverse overhang: 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG). Primers were again single-plex PCR amplified in a CTRL3A, CTRL3C, and no DNA control. The annealing temperature of 58 °C was successful at amplifying many of the microhaplotype regions with Illumina-specific primers. For the Illumina PCR amplification, 30 cycles were selected since this number of cycles has shown success when microhaplotyping the PRNP locus in the work of Geue et al. (2024). However, since fainter bands were observed on the QIAxcel Advanced System (Qiagen) at 30 PCR cycles when using primers for nanopore sequencing, the number of PCR cycles was increased from 30 to 35 cycles prior to library preparation and nanopore sequencing. Primers that generated prominent bands from capillary electrophoresis were selected for the final microhaplotype panel to be sequenced on a MinION Mk1B and an Illumina MiSeq platform. This resulted in a novel panel of 25 microhaplotype loci across the caribou genome (Table B1).

DNA Sample Collection & Preparation

Caribou genomic DNA was extracted from 23 tissue samples collected across Canada. Two of the tissue samples, CTRL3A and CTRL3C acted as DNA controls as they are known to have high quality DNA. These two samples were used to confirm proper PCR amplification. All DNA extracts from tissue samples had been kept frozen since their extraction. DNA

concentrations of tissue samples were quantified using a Qubit 4 Fluorometer (Invitrogen) with the High Sensitivity dsDNA Assay kit and each sample was diluted to 250 pg/μL.

Four multiplex PCR amplifications (M1, M2, M3 and M4) were prepared separately for downstream MinION and Illumina sequencing. All MinION M1-M4 reactions comprised 1 ng of input DNA and cocktail which consisted of 2x Type-It PCR Master Mix (Qiagen), UltraPure DNase/RNase-Free Distilled Water (Invitrogen; Thermo Fisher Scientific), 0.2 mM BSA (Sigma Aldrich), and 0.2 μM standard primer pairs (IDT). Similarly, all Illumina M1-M4 reactions comprised 1 ng of input DNA and cocktail which consisted of 2x Type-It PCR Master Mix (Qiagen), UltraPure DNase/RNase-Free Distilled Water (Invitrogen; Thermo Fisher Scientific), 0.2 mM BSA (Sigma Aldrich), and 0.4 μM standard primer pairs (IDT). Both MinION and Illumina M1-M3 reactions contain 6 microhaplotype primer pairs each while M4 contains 7 microhaplotype primer pairs. The following thermocycling regime for PCR amplification was followed: 95 °C for 5 minutes; 30/35 cycles of 95°C for 30 sec, 58°C for 90 sec, 72°C for 30 sec; 68 °C for 10 min. Thirty-five cycles were run for the MinION PCR amplification while 30 cycles were run for the Illumina PCR amplification.

MinION M1-M4 PCR products for the same sample were pooled equivolumetrically (3 μL of each sample) in a new plate, and Illumina M1-M4 PCR products were also pooled equivolumetrically according to sample (6.25 μL of each sample) in a new plate. Both pooled plates were then size separated using capillary electrophoresis performed on a QIAxcel Advanced System (Qiagen) to ensure proper amplification of the microhaplotype target regions (Figure B1, Figure B2). Aliquots of each sample from the M1-M4 pooled plates were used as input for DNA library preparation. It is important to note that a PCR cleanup would typically be conducted prior to library preparation, however, too much DNA was lost from attempting this so

uncleaned DNA that was diluted to the standard concentration underwent downstream library preparation.

MinION Mk1B Library Preparation

DNA sample concentrations for the MinION M1-M4 pooled plate were re-quantified using the Qubit 4 Fluorometer (Invitrogen) and the High Sensitivity dsDNA Assay kit. Samples were diluted to 3.74 ng/μL in a total volume of 11.5 μL to generate 200 fmol of DNA per sample based on an estimated average amplicon length of 350 bp. Library preparation steps were followed according to the ONT Ligation Sequencing Amplicons protocol. Briefly, library preparation included DNA-end preparation for adapter attachment, native barcode ligation, sequencing adapter ligation, and flow cell priming/loading. During the end-prep steps, a diluted DNA control sample (DCS) was added to each sample. During the native barcode ligation steps, barcodes from the Native Barcoding Kit 24 V14 (SQK-NBD114-24) were used to index each unique DNA sample. After barcoding, all samples were pooled in equal volumes to make up the DNA library (4.42 ng/μL). The DNA library was prepared with Thermo Fisher's Invitrogen™ UltraPure™ Bovine Serum Albumin (50mg/mL) and loaded onto a FLO-MIN114 flow cell (R.10.4.1). The final prepared DNA library contained 27.83 fmol (6 ng total) of DNA in an Eppendorf DNA LoBind tube.

MinION Mk1B Sequencing

MinION Mk1B sequencing was performed in MinKNOW™ V23.07.15 for Windows 10 with Guppy V7.1.4 (GPU). An initial flow cell check revealed 1,529 functional pores. Default sequencing run parameters were selected during the sequencing run along with the FLO-MIN114/HD DNA Kit 14 and live (400 bps) Super-Accurate (SUP) 5kHz basecalling configuration. Sequencing stopped at 66 hours when only 63 available pores remained. Post-

sequencing, barcodes were trimmed and mid-read strand barcodes as well as reads lacking barcodes on both ends were removed. Data output included POD5 files and FASTQ files for each barcode and were renamed according to the caribou sample ID associated with each barcode number. FASTQ files with a minimum quality score of 8 (MinKNOW default for FASTQ_pass) were used for downstream genotyping analysis. FASTQ_fail files and unclassified data from the FASTQ_pass files were omitted. FASTQ_pass files for each barcode/sample were concatenated through Linux (WSL) using Ubuntu V20.04.6 LTS. The overall MinION sequencing run was summarized using NanoPlot tools (De Coster & Rademakers, 2023).

Illumina Library Preparation

Library preparation for Illumina sequencing was done according to the 16S Illumina Amplicon protocol with modifications using the Illumina DNA Prep Kit. The Illumina M1-M4 pooled plate was cleaned with AMPure XP beads (Beckman Coulter) before adding Illumina MiSeq indexes (IDT-ILMN Nextera DNA UD Indexes) to the DNA ends. Samples were transferred to a new PCR plate to undergo an additional PCR reaction adding Illumina indexes to each sample. The index addition PCR reaction included 5 uL of UltraPure DNase/RNase-Free Distilled Water (Invitrogen; Thermo Fisher Scientific), 12.5 μL of 2x Type-It PCR Master Mix (Qiagen), 5 μL of Illumina DNA UD Indexes, and 2.5 uL of amplified product to each well. The following thermocycling regime for PCR amplification was followed: 95 °C for 3 minutes; 8 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec; 72 °C for 5 min. After the index addition, PCR products from each sample were pooled in a new 1.5 mL tube (4 μL of each sample) and underwent a second AMPure XP bead (Beckman Coulter) clean-up. A Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific) was used to determine the concentration of the pooled DNA library which was 12.6 ng/μL. The DNA library was diluted to

5 ng/uL using ddH2O and then 4 nM. In a separate tube, 1 mL of 20 pM denatured library was prepared by combining the 4 nM pooled library with 0.2N NaOH in equal volumes and topping up with HT1. Finally, equal volumes of the 20 pM library and HT1 were combined along with a 12.5 pM PhiX control. After a short heat denaturation, the Illumina MiSeq V2 flow cell/cartridge was loaded according to manufacturer instructions.

Illumina MiSeq Sequencing

The GenerateFASTQ V3.0.1 module was used to conduct Illumina MiSeq amplicon sequencing. Once the cartridge was loaded into the machine, a submission sheet was uploaded to the MiSeq submission folder and on-screen instructions were followed to start the sequencing run. Default settings were selected for 251 by 251 reads. After 37 hours, sequencing data was available through Illumina's BaseSpace genomics cloud.

SNP Microhaplotyping

Microhaplotype reads from MinION and Illumina FASTQ_pass files were genotyped using the open-access genotyping software Seq2Sat V2 (https://github.com/ecogenomicscanada/Seq2Sat) (Liu et al., 2024). MinION FASTQ_pass files for each sample were kept in a separate directory than Illumina FASTQ_pass files. Based on the uploaded target microhaplotype loci file (Table B2), Seq2Sat was run for every FASTQ file in both directories through the command line in a remote server accessed with MobaXterm Personal Edition V23.6 for Windows. The loci file contained locus names, forward and reverse primer sequences, 5' and 3' end trimmed lengths, known SNPs positions, and full reference sequences for the 25 target microhaplotype loci. Seq2Sat V2 default parameters were selected for genotyping analysis. The default minimum quality score threshold was Q10 for MinION sequencing data and Q20 for Illumina sequencing data.

Microhaplotype loci were excluded from genotype analysis if neither sequencing platform generated any target amplicon reads for a locus or if a microhaplotype locus generated reads in less than 5 of the 23 caribou samples sequenced. Microhaplotype loci were also excluded if no reads corresponded to the correct target gene region. Following microhaplotype locus filtering, caribou samples that did not generate any target amplicon data or Seq2Sat output were classified as sample dropouts and removed prior to genotyping analysis, as well as samples that lacked target amplicon data for more than half of the microhaplotype loci. Target microhaplotype reads with a read depth of <20 were considered to have a low read depth and were also not genotyped. This is consistent with the threshold applied by Baetscher et al. (2018), retaining haplotypes with 20 reads or more per locus within an individual when targeting microhaplotype regions in kelp rockfish (*Sebastes atrovirens*). Similarly, Turchi et al. (2019) applied a minimum read depth threshold of 20x to genotype microhaplotype loci for forensic genetics applications.

Cost Comparison

A cost analysis of microhaplotype sequencing was recorded for the MinION Mk1B and Illumina MiSeq approach to assess cost-efficiency of the two sequencing technologies. The cost of a MinION Mk1B or an Illumina MiSeq sequencer is not included in the cost breakdown. Costs listed as MinION Mk1B sequencing expenses include 2x Type-It PCR Master Mix (Qiagen), IDT's standard DNA Oligonucleotide primers needed for PCR amplification of the target microhaplotype amplicons, Qubit dsDNA High Sensitivity Assay Kit reagents (Invitrogen) needed to quantify each sample and the final DNA library concentration, the ONT Native Barcoding Kit V14 (SQK-NBD114.96), NEB 3rd party reagents, and an R10.4.1 flow cell. It is assumed that flow cells are not reused and that the recommended reagent amounts as indicated in

the ONT Ligation Sequencing Amplicons protocol are used. Illumina MiSeq sequencing expenses include 2x Type-It PCR Master Mix (Qiagen) and standard DNA Oligonucleotide primers (IDT) needed for all PCR amplifications, AMPure XP Beads (Beckman Coulter), barcoding indexes for each unique sample (IDT-ILMN Nextera DNA UD Indexes), Qubit 4 (Invitrogen) High Sensitivity dsDNA Assay kit, reagents to quantify the final DNA library, and an Illumina V2 500 Kit.

Results and Discussion

Sequencing Summary

The MinION Mk1B sequencing run only included 23 caribou samples while the Illumina MiSeq targeted a 384-sample run for cost reduction purposes. As a result, the total number of reads and bases generated from Illumina MiSeq sequencing included the 23 caribou samples relevant to this study plus additional samples used for other experiments. The Illumina MiSeq sequencing run generated a yield of 8.33 M raw reads and 2.08 Gbp of bases spread over 384 caribou samples, with an error rate of 1.37+/- 0.05%. Of the total Illumina MiSeq raw reads, 86.78% passed the minimum quality score threshold of Q30. The MinION Mk1B sequencing run generated 13.69 M reads and 4.37 Gbp of bases across 23 caribou samples post filtering. The MinION Mk1B mean read length is 319.8 bp, with an estimated N50 read length of 329, and a mean read quality of Q13.5 after filtering (Figure B3). There were 4343.5 Mb (99.0%) MinION of reads above the Q10 quality score cutoff and 1449.5 Mb (36.2%) of reads had a quality score of >Q15.

Since the MinION Mk1B device has been shown to achieve a lower throughput and depth of coverage than the Illumina MiSeq, a less stringent quality score threshold was applied to filter MinION raw reads (Voskoboinik et al., 2018). This was done to preserve more target reads,

making it easier to distinguish between primary alleles and rare or false haplotypes, as well as increasing confidence in genotype calls (Voskoboinik et al., 2018). Quality score thresholds can be manually changed on the SmarTyper interface for Seq2Sat to balance read depth with read quality.

Sample & Locus Dropout

The following loci were excluded from downstream genotyping analysis due to no reads being generated from one or both sequencing platforms: 5_Fbxo25, 13_Fgb, 15_Stoml1, 26_Cip2a, and 27_Abt1. This is likely due to issues with PCR amplification and/or the primer sets. Because these microhaplotype loci did not perform on either sequencing platform for any sample despite the same samples generating microhaplotype data for other loci, this suggests that further primer optimization is required to make these markers amenable to proper amplification and downstream microhaplotype genotyping. The 3_Nudt12, 11_Or51g1, and 15_Serpina4 loci were also dropped since these markers generated reads within fewer than 4 out the 23 caribou samples on one or both sequencing platforms. The reads that were obtained had low read depth and were not successfully sequenced on both sequencing platforms. According to Voskoboinik et al. (2018), a true microhaplotype should be observed with a minimum frequency of at least 5/2N, where N is the number of individuals in the population sample. Since these loci were inconsistently sequenced across platforms in a small number of samples, these loci were eliminated from the microhaplotype panel. Finally, the microhaplotype loci 4_Pgap2, 17_Shb, 21_Dpm3, and 30_Htra3 were removed since none of their reads corresponded to the correct genomic region. Following locus dropout filtering, the remaining 13 microhaplotype loci consist of 1_Gimap6, 1_Rsl24d1, 2_Cd51, 2_Lexm, 2_Vcam1, 3_Slc34a1, 4_Abo, 11_Oaf, 18_Eme1, 18_Krt16, 18_Rnft1, 20_Scn11a, and 32_Gpm6a.

Of the 23 caribou DNA samples that were used to sequence target microhaplotypes, 5 samples dropped out from the Illumina MiSeq dataset (21.73%) and 1 sample dropped out from the MinION Mk1B dataset (4.34%). Three of the Illumina MiSeq sample dropouts (samples 20919, 21112, 21297) did not generate any target amplicon data. The other two dropouts (samples 21306 and 3A) did not generate amplicon data for 11 out of the 13 loci remaining after microhaplotype locus dropout filtering. The single MinION Mk1B sample dropout (sample 21301) was removed because no amplicon data was generated for 5 out the 13 final microhaplotype loci and 2 additional loci did not pass the minimum read depth threshold of 20 reads required for downstream genotyping.

The number of caribou samples that resulted in target microhaplotype reads for these 13 loci in the MinION Mk1B dataset ranges from 20 to 22 samples per locus, with a mean of 22 samples per locus out of 23 total samples sequenced. MinION target reads for 1_Gimap6, and 4_Abo were captured in the fewest number of samples (20 each). In the Illumina MiSeq dataset, 18 out of 23 total samples sequenced generated target microhaplotype reads at each locus.

Read Depth

Although read depths for each caribou sample and microhaplotype locus were recorded according to sequencing platform, read depths cannot be directly compared across sequencers because a different number of caribou samples were sequenced during the MinION Mk1B run as were sequenced during the Illumina MiSeq run. The Illumina MiSeq run included a total of 384 samples, only 23 of which are relevant to this study. The other samples and loci are not shown but they impacted the read distribution across the target caribou samples and microhaplotype loci examined here (Fumagalli, 2013). The purpose of showing read depth was to establish whether adequate read depth was achieved to confidently call microhaplotypes. Microhaplotypes with

insufficient reads were removed in accordance with the read depth threshold of 20 reads applied in the work of Baetscher et al. (2018), where a panel of microhaplotypes was described for a fish species lacking a reference genome.

The mean number of target microhaplotype reads per caribou sample is shown according to microhaplotype locus for the MinION Mk1B dataset (Figure 2.1) and Illumina MiSeq dataset (Figure 2.2). For each dataset, the mean number of reads for the 2 primary (highest) microhaplotype alleles for a given locus was compared to the mean number of total microhaplotype reads per sample for that locus. The mean read depth calculation only included microhaplotype loci with a conclusive zygosity and minimum read depth of 20 for the highest 2 sequences per sample.

MinION Mk1B amplicon sequencing resulted in a mean read depth of 413 reads per sample across all loci for the 2 highest microhaplotype alleles, with a minimum mean read depth of 43 reads per sample (2_Vcam1 locus) and a maximum mean read depth of 1,515 reads per sample (18_Krt16 locus). The mean total microhaplotype reads across all samples was 7,237 reads per microhaplotype locus, ranging from 975 to 13,549 total reads per locus. The 2_Abo locus had the lowest mean total microhaplotype reads per sample, while the 2_Lexm locus had the highest. The mean percentage of 2 highest microhaplotype reads to total reads per sample across all loci was 5.53%.

Illumina MiSeq amplicon sequencing generated a mean read depth of 407 reads for the 2 highest microhaplotype sequences per sample, ranging from 52 to 1,392 reads across all microhaplotype loci. The mean total microhaplotype reads across all samples for each microhaplotype locus was 1,186 reads, ranging from 110 to 3,570 total reads per locus. 1_Gimap6 had the lowest mean read depth of primary sequences per sample while 20_Scn11a

had the fewest total reads per locus. The 18_Krt16 locus generated the largest mean read depth for the 2 highest microhaplotype sequences per sample and mean total microhaplotype reads per sample out of all the microhaplotypes sequenced with the Illumina MiSeq. Of the mean 1,186 total target microhaplotype reads per sample, 33.66% of reads accounted for the 2 highest sequences per sample.

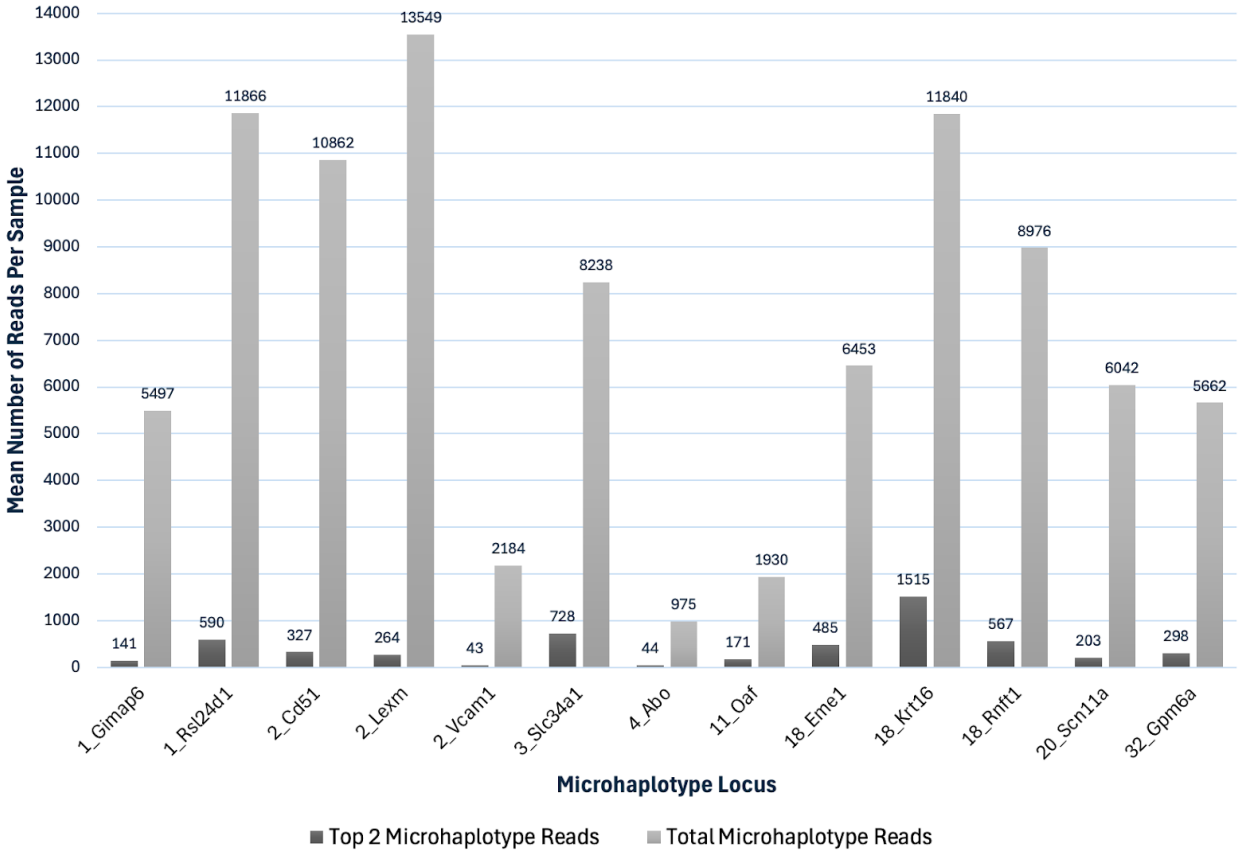


Figure 2.1 Mean number of target microhaplotype reads per caribou sample generated from MinION Mk1B amplicon sequencing according to microhaplotype locus. The mean number of reads for the 2 most abundant (top 2) microhaplotype alleles per sample (dark gray) is shown in comparison to the mean number of total microhaplotype reads per sample (light gray).

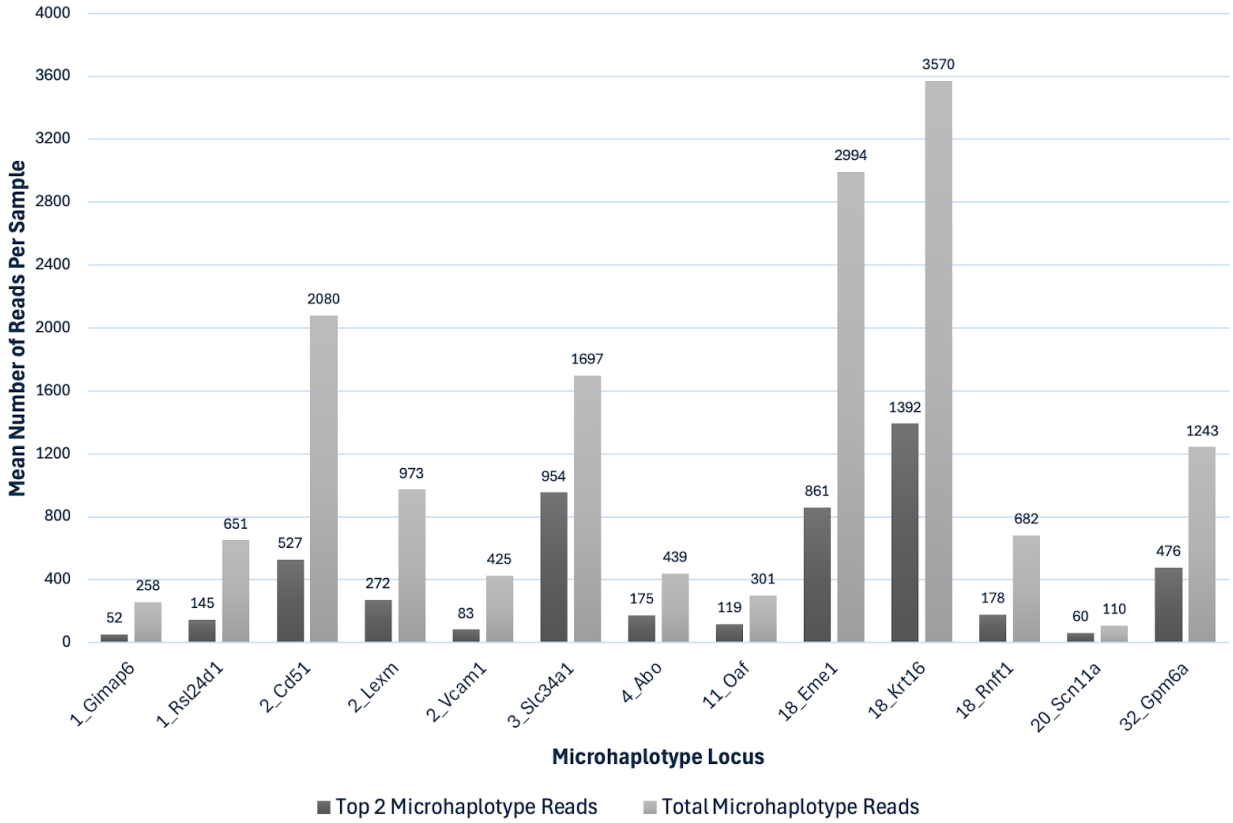


Figure 2.2 Mean number of target microhaplotype reads per caribou sample generated from Illumina MiSeq amplicon sequencing according to microhaplotype locus. The mean number of reads for the 2 most abundant (top 2) microhaplotype alleles per sample (dark gray) is shown in comparison to the mean number of total microhaplotype reads per sample (light gray).

The number of reads for the 2 highest sequences as well as the number of total reads captured for a given sample and locus were highly variable within and between sequencing platforms. Regardless of whether the MinION Mk1B or Illumina MiSeq was used, certain loci tended to generate more reads than others. For example, the 18_Krt16 locus generated a disproportionate number of 2 highest reads and total reads in both datasets. The substantial differences in read depth from locus to locus suggests the read distribution is largely locus dependent, which is consistent with the findings of Whitford et al. (2022), who also pooled PCR amplicon products in equimolar quantities and observed highly variable depth of coverage across target amplicons using the MinION platform. Despite this, sufficient read depth was achieved in most cases to confidently call microhaplotypes.

In future experiments, adding a second threshold may be beneficial to exclude false haplotypes or artifacts that exceed a read depth of 20. In some cases, false microhaplotypes can have high read depths resulting from sequencing errors or index switching (Baetscher et al., 2018). To mitigate this, Tytgat et al. (2022) proposed adding a requirement that haplotypes constitute a certain percentage of the most abundant haplotype for a given locus to limit the retainment of artifacts instead of true haplotypes. Similarly, Baetscher et al. (2018) required microhaplotypes to have a read depth ratio of >0.2 of the haplotype with the highest read depth. Since it can be difficult to differentiate between a false call/artifact and a rare or newly discovered haplotype variant, another option would be to set a minimum haplotype frequency across all samples sequenced to maintain less common but true microhaplotype calls (Voskoboinik et al., 2018). Increasing the index assignment stringency thresholds when basecalling MinION sequencing runs as well as removing unligated adapters could mitigate index switching (Eriksson et al., 2020; Esnault et al., 2022), helping to capture true microhaplotype reads for the correct caribou sample. The unique dual indices used for Illumina-based microhaplotype amplicon sequencing is known to reduce the incidence of index switching when working with highly multiplexed samples (Guenay-Greunke et al., 2021; Kircher et al., 2011). Integrating a MinION-compatible double-indexing strategy may further reduce index switching across samples sequenced on a nanopore-based platform. Ultimately, the goal is to find a balance between eliminating false calls and preventing true calls from being filtered out (Broquet & Petit, 2004; Voskoboinik et al., 2018).

Microhaplotyping

When calling haplotypes at microhaplotype loci, it is important to consider the distribution of reads beyond the 2 most abundant alleles per locus within each sample. If the read

depth of additional sequence reads captured are too close to the supposed primary alleles, confidence in the true haplotype for that locus is diminished. Ideally, microhaplotype loci would have high read depth for the two primary alleles and a low read depth for all other reads, which should be sequencing errors (Baetscher et al., 2018). The read depths for the 3 highest reads generated at each microhaplotype locus within all samples in this study were categorized as having interpretable or uninterpretable read distribution patterns prior to sample and locus filtering (Table B3).

Examples of good, interpretable read distributions are shown for a heterozygous haplotype call for the 18_Krt16 locus within sample 21102, and a homozygous haplotype call for the 3_Slc34a1 locus within sample 21104 (Figure 2.3). The MinION read distribution across the 3 highest reads generated for 18_Krt16 within sample 21102 was 2564 reads (read 1), 2051 reads (read 2), and 66 reads (read 3). The Illumina read distribution across the 3 highest reads generated for the same sample and locus was 1634 reads (read 1), 1400 reads (read 2), and 32 reads (read 3). In both cases, the 2 highest allele sequences represented the same haplotypes across sequencing platforms and there was a substantial drop off in read depth from the two primary haplotypes to the third read. Similarly, for the 3_Slc34a1 locus within sample 21104, there was a big difference in read depth of the primary alleles to the next most common read (read 2 for homozygotes) captured by MinION and Illumina amplicon sequencing. Since 3_Slc34a1 was homozygous in this sample, the read depth of read 1 reflects both primary sequences. In the MinION dataset, read 1 had 1206 reads and read 2 had 60 reads. In the Illumina dataset, read 1 had 1142 reads and read 2 had 8 reads. These loci and samples showed an expected read distribution pattern that supports the haplotype calls made.

A. Interpretable Read Distribution for a Heterozygous Haplotype Call

MinION Mk1B:

Marker	SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads
Reference	ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
18_Krt16	59(A|T)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G)	TGACGTGG	2564	0.555600	heter	59|0.533400	15.530900	16509
18_Krt16	59(A|A)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G)	AGACGTGG	2051	0.444400	heter		12.423500	16509
18_Krt16	59(A|T)72(C|T)	TT	66	NA	seq error		0.399800	16509

Illumina MiSeq:

Marker	SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads
Reference	ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
18_Krt16	59(A|A)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G)	AGACGTGG	1634	0.538600	heter	59|0.530400	38.106300	4288
18_Krt16	59(A|T)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G)	TGACGTGG	1400	0.461400	heter		32.649300	4288
18_Krt16	18(G|T)	T	32	NA	seq error		0.746300	4288

B. Interpretable Read Distribution for a Homozygous Haplotype Call

MinION Mk1B:

Marker	SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads
Reference	ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
3_Slc34a1	25(G|G)50(A|A)94(G|G)104(C|C)110(G|G)146(C|C)	GAGCGC	1206	0.952600	homo	0|0.553900	12.192900	9891
3_Slc34a1		indel	60	NA	seq error		0.606600	9891
3_Slc34a1		indel	57	NA	seq error		0.576300	9891

Illumina MiSeq:

Marker	SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads
Reference	ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
3_Slc34a1	25(G|G)50(A|A)94(G|G)104(C|C)110(G|G)146(C|C)	GAGCGC	1142	0.993000	homo	80|0.993700	72.738900	1570
3_Slc34a1	80(T|C)	C	8	NA	seq error		0.509600	1570
3_Slc34a1	142(T|C)	C	8	NA	seq error		0.509600	1570

Figure 2.3 Interpretable read distributions for the 3 highest amplicon reads resulting in a heterozygous genotype call for the 18_Krt16 microhaplotype locus within caribou sample 21102 (A) and a homozygous genotype call for the 3_Slc34a1 microhaplotype locus within caribou sample 21104 (B) according to sequencing platform used (MinION Mk1B or Illumina MiSeq). The haplotype, number of reads, haplotype ratio and zygosity is highlighted in green for the primary sequences per sample as determined by Seq2Sat analysis.

Examples of ambiguous read distributions are shown for the 18_Eme1 and 1_Rsl24d1 locus within caribou sample 21111 and 21303, respectively (Figure 2.4). The MinION read distribution across the 3 highest reads generated for 18_Eme1 within sample 21111 was 504 reads (read 1), 219 reads (read 2), and 113 reads (read 3). The Illumina read distribution across the 3 highest reads generated for the same sample and locus was 1222 reads (read 1), 448 reads (read 2), and 386 reads (read 3). For this sample and locus, both sequencers exhibited a high read depth for read 3, with little difference in read depth from read 2. Additionally, the read 1 depth for both platforms was more than double the depth for read 2, which reduced confidence that the haplotype was heterozygous.

For the 1_Rsl24d1 locus within sample 21303, the MinION read distribution across the 3 highest reads was 268 reads (read 1), 208 reads (read 2), and 203 reads (read 3). The Illumina read distribution for the same sample and locus comprised 112 reads (read 1), 110 reads (read 2), and 96 reads (read 3). Both sequencers demonstrated a high read depth for read 3, which was very close to the read depth for read 1 and read 2. This made it challenging to determine which two alleles represented the true haplotype (i.e. read 1 and read 2, read 2 and read 3, or read 1 and read 3). Examples such as these indicate that a removal of read depths less than 20% on the highest read and taking the highest two reads (Baetscher et al., 2018; Tytgat et al., 2022) may result in an erroneous genotype and demonstrates the need to factor in the third read depth in the scoring rule.

A. Uninterpretable Read Distribution

MinION Mk1B: Sample 21111

Marker	SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads
Reference	ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
18_Eme1	18(T|T)28(C|C)47(A|A)54(G|G)100(C|C)115(C|C)124(C|C)139(T|T)162(C|C)	TCAGCCCTC	504	0.697100	heter	162|0.515000	11.610200	4341
18_Eme1	18(T|T)28(C|C)47(A|A)54(G|G)100(C|C)115(C|C)124(C|C)139(T|T)162(C|T)	TCAGCCCTT	219	0.302900	heter		5.044900	4341
18_Eme1	54(G|T)162(C|T)	TT	113	NA	seq error		2.603100	4341

Illumina MiSeq: Sample 21111

Marker	SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads
Reference	ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
18_Eme1	18(T|T)28(C|C)47(A|A)54(G|G)100(C|C)115(C|C)124(C|C)139(T|T)162(C|C)	TCAGCCCTC	1222	0.731700	heter	162|0.566200	41.395700	2952
18_Eme1	18(T|T)28(C|C)47(A|A)54(G|T)100(C|C)115(C|C)124(C|C)139(T|T)162(C|T)	TCATCCCTT	448	0.268300	heter		15.176200	2952
18_Eme1	162(C|T)	T	386	NA	seq error		13.075900	2952

B. Uninterpretable Read Distribution

MinION Mk1B: Sample 21303

Marker	SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads
Reference	ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
1_Rsl24d1	21(A|A)31(T|T)75(A|A)102(A|A)109(A|A)121(C|C)223(C|C)231(T|T)	ATAAACCT	268	0.563000	heter	75|0.534500	3.315200	8084
1_Rsl24d1	21(A|G)31(T|T)75(A|G)102(A|G)109(A|G)121(C|C)223(C|C)231(T|A)	GTGGGCCA	208	0.437000	heter		2.573000	8084
1_Rsl24d1	21(A|G)75(A|G)98(C|T)	GGT	203	NA	seq error		2.511100	8084

Illumina MiSeq: Sample 21303

Marker	SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads
Reference	ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
1_Rsl24d1	21(A|G)31(T|T)75(A|G)102(A|G)109(A|G)121(C|C)223(C|C)231(T|A)	GTGGGCCA	112	0.504500	heter	231|0.627000	18.481800	606
1_Rsl24d1	21(A|A)31(T|T)75(A|A)102(A|A)109(A|A)121(C|C)223(C|C)231(T|T)	ATAAACCT	110	0.495500	heter		18.151800	606
1_Rsl24d1	21(A|G)75(A|G)98(C|T)	GGT	96	NA	seq error		15.841600	606

Figure 2.4 Uninterpretable read distributions for the 3 highest amplicon reads resulting in ambiguous genotype calls across the MinION Mk1B and Illumina MiSeq sequencing platforms for the 18_Eme1 microhaplotype locus within caribou sample 21111 (A), and the 1_Rsl24d1 microhaplotype locus within caribou sample 21303 (B). The haplotype, number of reads, haplotype ratio and zygosity is highlighted in green for the primary sequences per sample as determined by Seq2Sat analysis.

The total number of interpretable, ambiguous, and blank/missing (N/A) read distributions across the 3 highest amplicon reads per locus across all caribou samples are shown based on MinION Mk1B sequencing (Figure 2.5) and Illumina MiSeq sequencing (Figure 1.6). The Illumina dataset had a much higher prevalence of blank/missing haplotypes. Excluding blank haplotypes from both datasets, the Illumina platform showed a higher incidence of interpretable read distributions across the top 3 amplicon reads for all samples and loci (Illumina: 61.98% vs MinION: 55.29%). Conversely, the MinION dataset had a greater incidence of uninterpretable read distributions (Illumina: 38.02% vs MinION: 44.71%).

Ambiguous read distribution patterns were seen across all microhaplotype loci, though it was more prevalent of an issue with specific loci, where unclear read distributions were generated from both sequencers within many samples. However, in some cases, one sequencing platform showed read distribution ambiguity while the same sample and locus had a clear read distribution on the other platform. Although ambiguous read distributions were an issue for the MinION Mk1B and Illumina MiSeq platform, it appeared to be more prominent with nanopore sequencing across samples and loci given the higher incidence of uninterpretable read distributions.

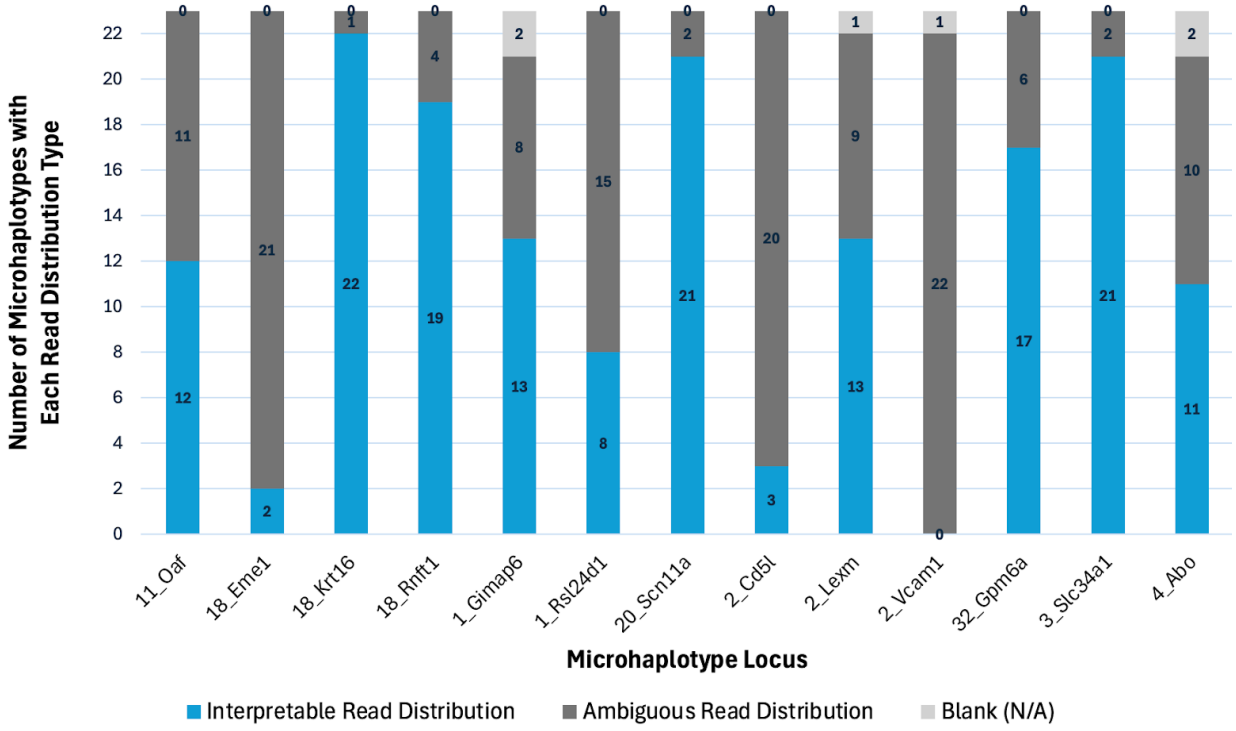


Figure 2.5 Number of interpretable read distributions (blue), ambiguous read distributions (dark gray) and blank (N/A) read distributions (light gray) for the highest 3 amplicon reads according to microhaplotype locus across all 23 caribou samples sequenced with a MinION Mk1B device.

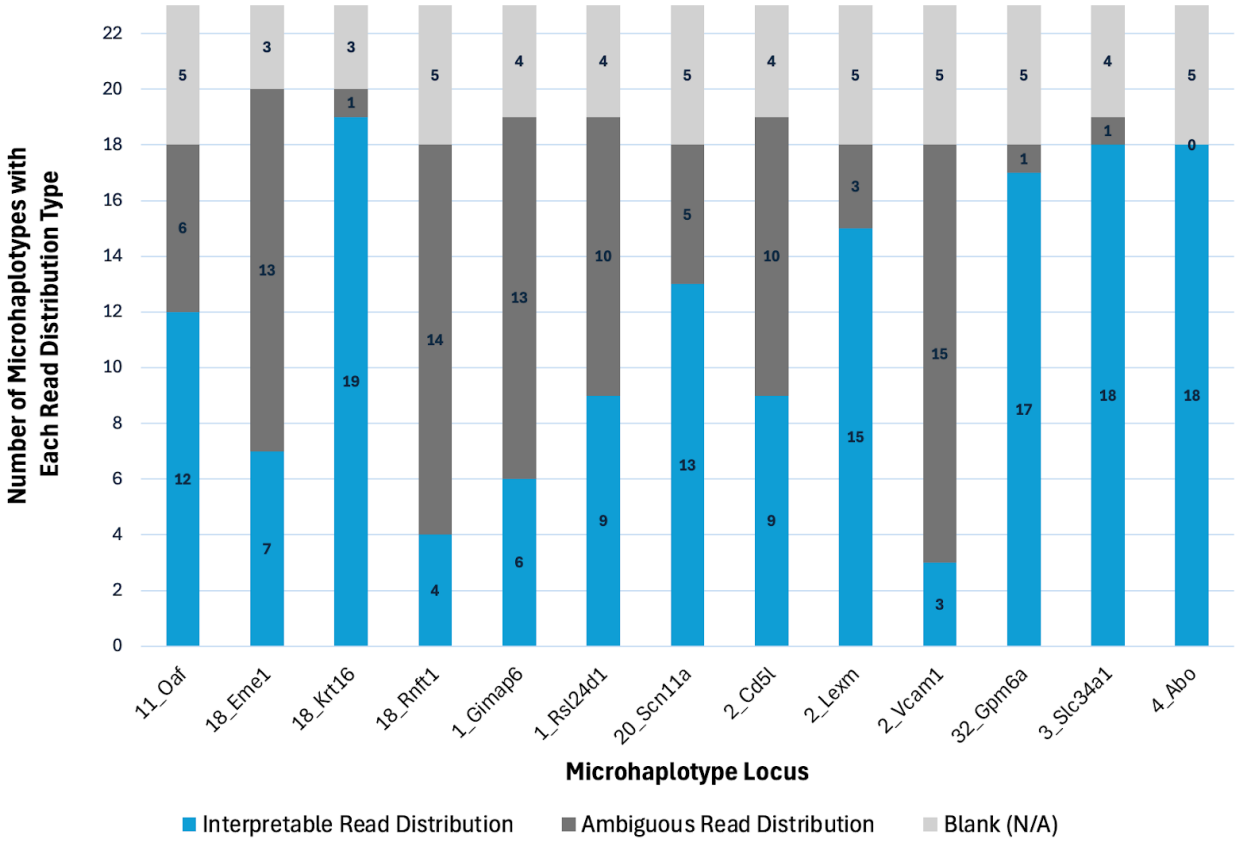


Figure 2.6 Number of interpretable read distributions (blue), ambiguous read distributions (dark gray) and blank (N/A) read distributions (light gray) for the highest 3 amplicon reads according to microhaplotype locus across all 23 caribou samples sequenced with an Illumina MiSeq device.

Having a high read depth for sequences beyond the top 2 reads significantly reduces confidence that these additional sequences are sequencing errors. Instances of poor read distribution observed at a higher frequency or intensity for some microhaplotype loci may result from gene duplications (Guan et al., 2020). If there was a duplication of the gene where the target microhaplotype is located, they may both be detected under one target amplicon due to having very similar DNA sequences. This can result in a disproportionately high depth of coverage for a locus since more than one gene is mapping to the target region, hindering accurate microhaplotyping. This phenomenon was also observed in the work of Baetscher et al. (2018).

An example of a microhaplotype locus that is unlikely to be a duplicated region (1_Gimap6) was compared to a microhaplotype locus that is likely a duplicated region

(1_Rsl24d1) (Figure 2.7, Figure 2.8). The depth of coverage for the 1_Gimap6 genome region was around average (~30X) for the caribou samples examined, which suggested only the target region is captured. Conversely, the 1_Rsl24d1 region was likely duplicated since the coverage was more than double in some cases, indicating multiple gene regions are contributing to the depth of coverage. All caribou samples shown in Figure 2.8 showed a genomic duplication but to different degrees. After exploring the depth of coverage of the 13 microhaplotype loci within caribou samples that have undergone whole genome sequencing in Integrated Genomics Viewer (IGV), 1_Rsl24d1, 2_Cd51, 18_Eme1 and 18_Rnft1 appeared to be likely duplicated regions. These same microhaplotype loci frequently exhibited high read depths for all top 3 reads, sometimes even the top 4 reads per sample.

The suspected duplicated gene regions likely reflect recent duplications. The accumulation of variation and mutations over time makes it more difficult to map fragments to duplicated target regions. Since the coverage is so high across many samples at these loci, this means fragments are mapping well to those regions, suggesting that little variation and mutations have accumulated due to a recent duplication. Duplicated genomic regions are not uncommon in mammals (Assis & Bachtrog, 2015), however we recommend removing these loci from the microhaplotype marker panel for future sequencing runs. Alternatively, target amplicons could

be lengthened to differentiate between duplicated regions containing microhaplotypes.

Figure 2.7 Image from Integrated Genomics Viewer (IGV) exemplifying a microhaplotype locus (1_Gimap6) on scaffold 1 (chromosome 1) of the caribou genome that is unlikely to be a duplicated genomic region. The target locus is outlined in red. Each row represents a unique caribou sample. Data was obtained from whole genome sequencing at approximately 30X depth of coverage.

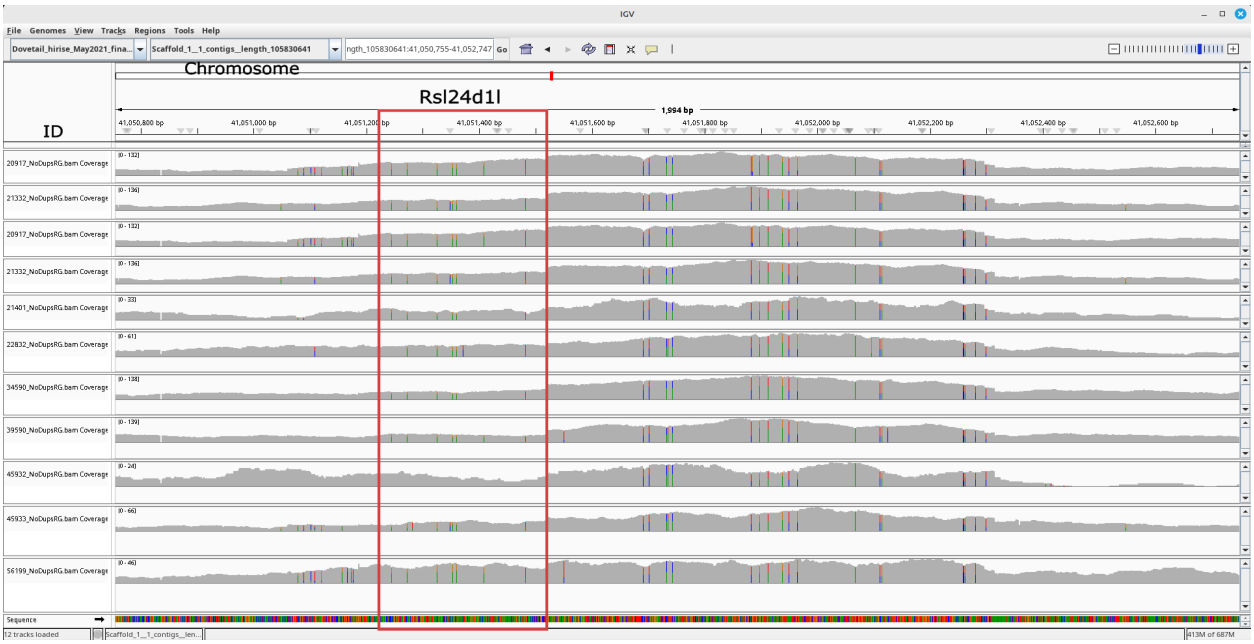


Figure 2.8 Image from Integrated Genomics Viewer (IGV) exemplifying a microhaplotype locus (1_Rsl24d1) on scaffold 1 (chromosome 1) of the caribou genome that is likely a duplicated genomic region. The target locus is outlined in red. Each row represents a unique caribou sample. Data was obtained from whole genome sequencing at approximately 30X depth of coverage.

Another explanation for poor read distribution may be sample contamination, more specifically index switching (Baetscher et al., 2018). Index switching, also known as index-hopping, can occur in multiplexed sequencing and involves indexes being assigned to the wrong sample, resulting in incorrectly assigned reads after sample and locus demultiplexing (Eriksson et al., 2020). To screen for this, the loci all sample data was run through Allelematch in R Studio (Galpern et al., 2012). Any caribou samples that clustered to the wrong sample from the other sequencing platform were excluded from microhaplotyping.

All 23 caribou samples sequenced with a MinION Mk1B and Illumina MiSeq were analyzed with the Allelematch R package based on 13 microhaplotype loci. Proper sample clustering was observed across sequencing platforms for samples 21099, 21101, 21102, 21104, 21105, 21106, 21110, 21111, 21113, 21303, 21305, 21307, 21309, and 3C. However, caribou samples 20919, 21107, 21112, 21297, 21298, 21299, 21301, 21306, and 3A exhibited poor

clustering between sequencing platforms. Poor clustering appears to be largely due to low or no target reads at many loci for these samples. For example, samples 20919, 21112, 21297 did not generate any target microhaplotype data in the Illumina MiSeq dataset. If no or too few reads were generated from one or both sequencers, a comparison of target microhaplotype data cannot be made for those samples across sequencing platforms.

Following Seq2Sat analysis of raw FASTQ files, microhaplotype alleles captured for all target loci were visible in an HTML output file for each caribou sample. Information provided includes the microhaplotype locus name, amplicon positions of any SNPs, combination of SNPs detected (haplotype), number of sequence reads, ratio of the number of top 1 reads X100 relative to the product of top 1 and top 2 reads (haplotype ratio), zygosity, SNP variant ratio, total target reads, reads percentage of unique amplicon reads to total target reads, amplicon read length (bp), and full microhaplotype sequence for all target reads captured. Within the full microhaplotype sequences shown in the HTML files, target SNPs, new SNPs, and sequence artefacts (errors) were highlighted along the amplicon.

Example Seq2Sat microhaplotype output figures/tables are shown for the microhaplotype locus 18_Krt16 in caribou sample 21099 according to both sequencing platforms used (Figure 2.9, Figure 2.10). The 2 highest microhaplotype reads called for the 18_Krt16 locus in caribou sample 21099 were the same across sequencing platforms. The MinION Mk1B and Illumina MiSeq captured the same SNP positions, haplotype, zygosity, amplicon length and SNPs. One of the primary microhaplotypes called consisted of the following SNPs at the specified amplicon positions: 59(A|A)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G). This haplotype was recorded as AGACGTGG. The other primary microhaplotype called consisted of the following SNPs/amplicon positions:

59(A|T)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G). This haplotype was recorded as TGACGTGG. Both primary microhaplotypes were 164 bp in length. The 18_Krt16 microhaplotype locus was deemed heterozygous according to the haplotype ratio and variant ratio for sample 21099.

A. MinION 18_Krt16 called haplotypes **B. Illumina 18_Krt16 called haplotypes**

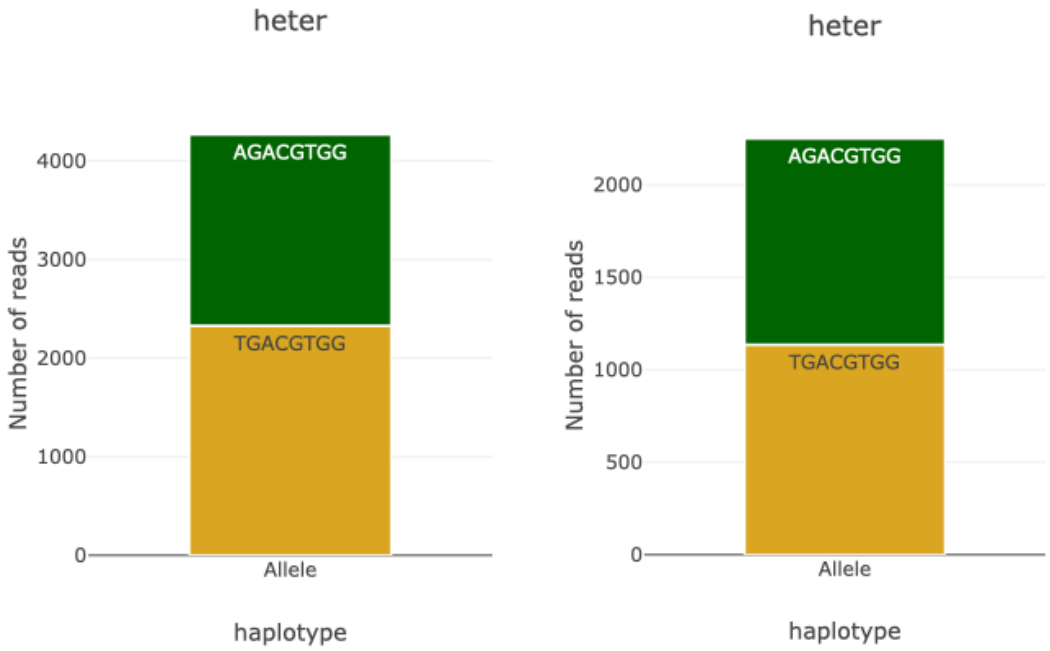


Figure 2.9 Seq2Sat output figure showing the number of reads for the 2 highest microhaplotype alleles for the 18_Krt16 locus sequenced from caribou sample 21099 with a MinION Mk1B (A) and an Illumina MiSeq (B). The combination of nucleotides for each microhaplotype (represented in green or yellow) illustrates the combination of single nucleotide variants detected throughout the amplicon sequence.

A. MinION 18_Krt16 microhaplotype reads

Marker	SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads	Length
Reference	ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	164
18_Krt16	59(A|T)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G)	TGACGTGG	2330	0.545900	heter	59|0.515500	15.457100	15074	164
18_Krt16	59(A|A)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G)	AGACGTGG	1938	0.454100	heter		12.856600	15074	164

Sequence
TGGCATCCCTTCCTAGGAGATGCTTGCCTTGCAGAATCAGACTGGTGGGGATGTCAGCGAGGAGATGGATGCCGCCCCCAGCGTGGACCTGAGCCACATCCTGAATGAGATGGGTGACCAGTATGAGCAGATAGCAGACAGGAACTGCAGGGATGCCGAGGCCT
TCGCATCCCTTCCTAGGAGATGCTTGCCTTGCAGAATCAGACTGGTGGGGATGTCAGCGTGGAGATGGATGCCCCCCCAGCGTGGACCTGAGCCACATCCTGAATGAGATGGGTGACCAGTATGAGCAGATAGCAGACAGGAACTGCAGGGATGCCGAGGCCT
TCGCATCCCTTCCTAGGAGATGCTTGCCTTGCAGAATCAGACTGGTGGGGATGTCAGCGAGGAGATGGATGCCCCCCCAGCGTGGACCTGAGCCACATCCTGAATGAGATGGGTGACCAGTATGAGCAGATAGCAGACAGGAACTGCAGGGATGCCGAGGCCT

B. Illumina 18_Krt16 microhaplotype reads

SNV	Haplotype	N. of Reads	Haplotype Ratio	Zygosity	Variant Ratio	Reads(%)	Total Reads	Length
ref	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	164
59(A|T)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G)	TGACGTGG	1136	0.504900	heter	59|0.542800	37.993300	2990	164
59(A|A)73(G|G)95(A|A)112(C|C)140(G|G)145(T|T)146(G|G)154(G|G)	AGACGTGG	1114	0.495100	heter		37.257500	2990	164

Sequence
TCGCATCCCTTCCTAGGATGCTTGCCTTGCAGAATCAGACTGGTGGGGATGTCAGCGAGGAGATGGATGCCGCCCCCCAGCGTGGACCTGAGCCACATCCTGAATGAGATGCGTGACCAGTATGAGCAGATAGCAGACAGGAACTGCAGGGATGCCGAGGCCT
TCGCATCCCTTCCTAGGAGATGCTTGCCTTGCAGAATCAGACTGGTGGGGATGTCAGCGTGGAGATGGATGCCGCCCCCCAGCGTGGACCTGAGCCACATCCTGAATGAGATGCGTGACCAGTATGAGCAGATAGCAGACAGGAACTGCAGGGATGCCGAGGCCT
TCGCATCCCTTCCTAGGAGATGCTTGCCTTGCAGAATCAGACTGGTGGGGATGTCAGCGAGGAGATGGATGCCGCCCCCCAGCGTGGACCTGAGCCACATCCTGAATGAGATGCGTGACCAGTATGAGCAGATAGCAGACAGGAACTGCAGGGATGCCGAGGCCT

Figure 2.10 Seq2Sat output tables showing the 2 highest 18_Krt16 reads captured across caribou samples from MinION Mk1B amplicon sequencing (A) or Illumina MiSeq amplicon sequencing (B). Each unique microhaplotype sequence is compared to the 18_Krt16 reference data. Seq2Sat output tables include the microhaplotype locus name ("Marker" column), amplicon positions of single nucleotide polymorphisms ("SNP" column), combination of SNPs detected ("Haplotype" column), number of sequence reads ("N. of Reads" column), ratio of the number of top 1 reads X100 relative to the product of top 1 read and top 2 read ("Haplotype Ratio" column), heterozygosity/homozygosity/inconclusive due to sequence error ("Zygosity" column), variant ratio of top 1 reads with a particular nucleotide variant compared to top 2 reads with an alternate nucleotide variant for each SNP ("Variant Ratio" column), percentage of number of reads to total target reads ("Reads %" column), total target reads ("Total Reads" column), amplicon read length ("Length" column), and full microhaplotype sequence ("Sequence" column). The 2 called microhaplotypes according to each sequencing platform for caribou sample 51655 have their Haplotype, N. of Reads (number of reads), Haplotype Ratio, and Zygosity highlighted in green. For the full 18_Krt16 sequences, target SNPs are highlighted in red and sequence artifacts (errors) are highlighted in gray.

For samples that clustered appropriately across sequencing platforms, the concordance of primary microhaplotype alleles was compared (Figure 2.11). Microhaplotype allele concordance was indicated by the same primary allele being identified in the MinION Mk1B and Illumina MiSeq dataset for the same caribou sample and microhaplotype locus. If no data was obtained for one or both primary microhaplotype alleles for a specific locus within a sample, that allele was denoted as "blank". Primary alleles with fewer than 20 reads were also categorized as blanks. Blank data was separated based on the dataset it came from (MinION Mk1B or Illumina MiSeq) as well as if there was blank data for the same locus and individual in both datasets.

Based on all 13 target loci sequenced within 14 caribou samples, 75.82% of all microhaplotype alleles were concordant between the MinION Mk1B and Illumina MiSeq datasets. An average of 10.44% of microhaplotype alleles were discordant across all samples and loci. MinION Mk1B sequencing resulted in 10.99% of target alleles being blank, either failing to be sequenced or having too low of a read depth (<20 reads). Conversely, 2.75% of target alleles sequenced with the Illumina MiSeq failed to amplify. No microhaplotype sequences failed to amplify for the same sample and locus between sequencing platforms. The microhaplotype loci ranked from most to least concordant alleles across sequencing platforms after read depth filtering are as follows; 32_Gpm6a (100.00% concordant), 18_Krt16 (100.00% concordant), 2_Lexm (96.43% concordant), 3_Slc34a1 (96.43% concordant), 1_Rsl24d1 (92.86% concordant), 18_Rnft1 (85.71% concordant), 18_Eme1 (78.57% concordant), 20_Scn11a (78.57% concordant), 11_Oaf (75.00% concordant), 2_Vcam1 (60.71% concordant), 2_Cd51 (50.00% concordant), 4_Abo (42.86% concordant) and 1_Gimap6 (28.57% concordant).

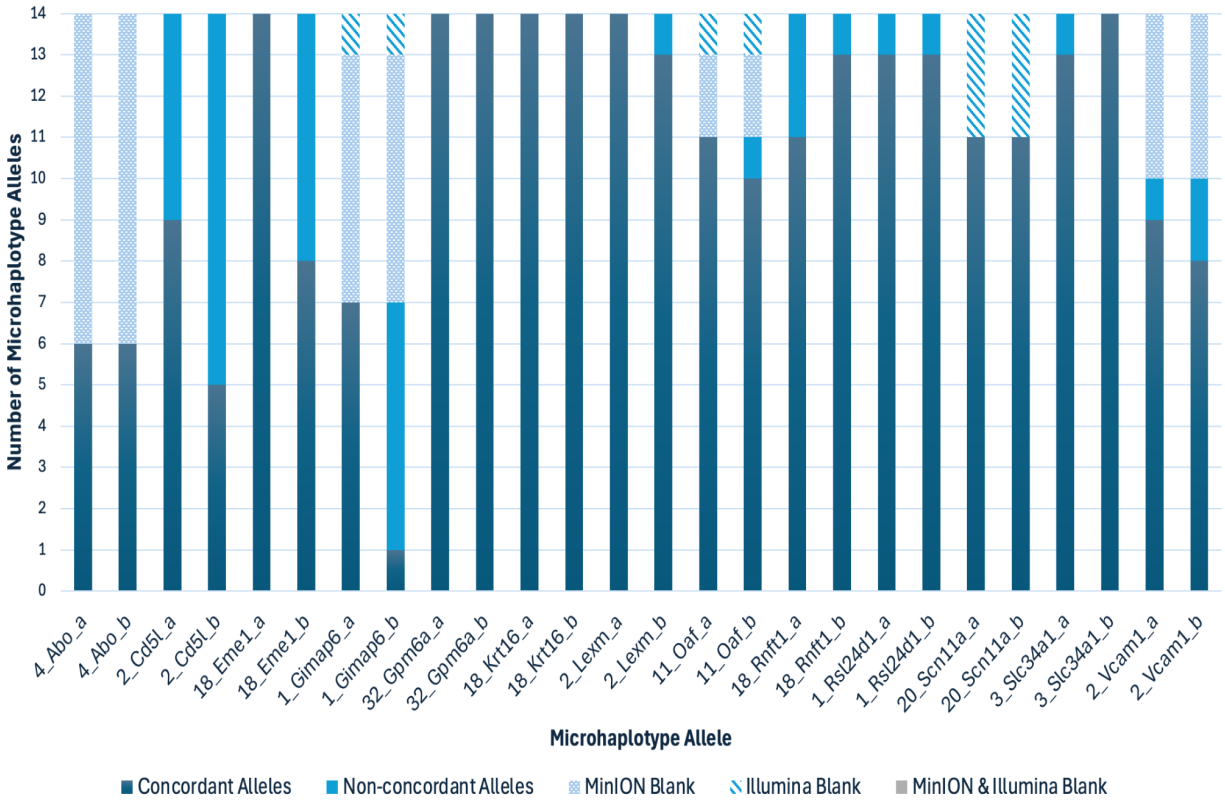


Figure 2.11 Microhaplotype allele concordance across sequencing platforms (MinION Mk1B and Illumina MiSeq) for the two primary alleles assigned at 13 loci within 14 caribou samples. The two primary microhaplotype alleles assigned for each locus are represented by an "_a" or "_b" following the locus name.

Discordance between haplotype calls for the same caribou samples and microhaplotype loci sequenced with the MinION Mk1B and Illumina MiSeq is likely due to ambiguous read distributions, e.g. high reads at the third most common sequence. Of the 9 microhaplotype loci with at least one discordant allele across platforms, the 4 loci hypothesized to be duplicated gene regions accounted for 68.42% of the discordant alleles (1_Rsl24d1, 2_Cd51, 18_Eme1, and 18_Rnft1). Excluding the 1_Rsl24d1, 2_Cd51, 18_Eme1, and 18_Rnft1 loci, the discordance of all microhaplotype alleles between datasets lowered from 10.44% to 4.76% discordant. These four loci generated particularly ambiguous read distributions for the top 3 reads per locus per sample. This suggests that having clear read distributions, where the depth of read 3 is substantially lower than the depth of read 1 and read 2 for a heterozygote, or where the depth of

read 2 is substantially lower than the depth of read 1 for a homozygote, is crucial to consistently score microhaplotypes. However, the caribou samples assessed in this study should be resequenced and scored again to confirm consistency in the observed microhaplotype concordance/discordance and to further explore the impact of read distribution across the top 3 reads per locus on microhaplotype scoring.

1_Gimap6 and 2_Vcam1 also had relatively high rates of discordance. Although these loci were not suspected gene duplications, over 25% of their corresponding alleles were missing from either dataset. This could indicate issues with the primers or initial PCR amplification steps, which lowers confidence in the called haplotypes and could negatively impact haplotype concordance. The microhaplotype loci 2_Lexm, 3_Slc34a1 and 11_Oaf only had one discordant sequence each, so these instances of haplotype discordance are more likely to be anomalies.

Sequencing Errors

At a dataset level, the mean error rate (%) along each position of a microhaplotype amplicon is shown according to microhaplotype locus, as seen in Geue et al. (2024). Briefly, the number of reads with a sequencing error was divided by the total reads to generate an error rate percentage at each amplicon position. Only samples that generated amplicon error rates from Seq2Sat analysis were included in the mean error rate calculations for each microhaplotype locus. Mean error rate comprised sequence errors but not insertions/deletions (indels). If any indels were present in a microhaplotype read, Seq2Sat did not calculate error rate. Examples of microhaplotype error rate figures are shown for microhaplotypes 18_Eme1, 18_Krt16, and 32_Gpm6a. Error rate figures for all other microhaplotype loci are shown in Appendix B (Figure B4-Figure B12).

The 18_Eme1 microhaplotype locus was a 169 base pair amplicon. The mean error rate across each position in the 18_Eme1 amplicons was 0.44% based on the MinION Mk1B dataset compared to 0.30% in the Illumina MiSeq dataset (Figure 2.12). The minimum mean error rate at each position in the 18_Eme1 amplicons was 0.06% and 0.01% for the MinION and Illumina MiSeq, respectively. The maximum mean error rate at each 18_Eme1 amplicon position was 3.50% for the MinION Mk1B and 6.23% for the Illumina MiSeq.

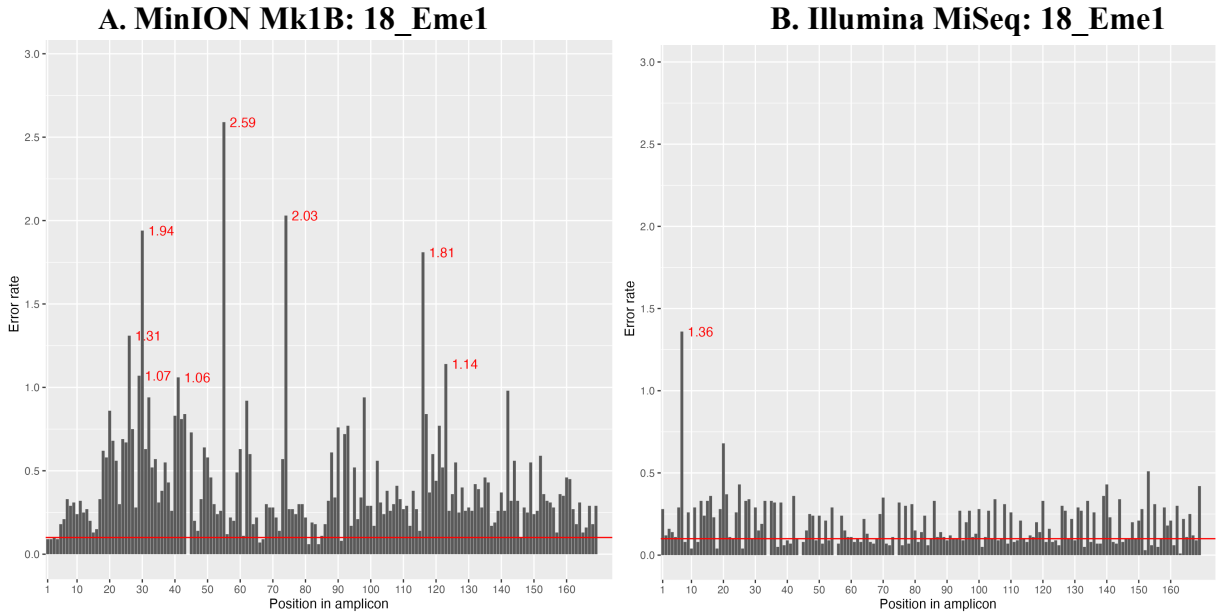


Figure 2.12 Mean error rate (%) at each position along the 18_Eme1 microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

The microhaplotype locus 18_Krt16 was a 164 base pair amplicon. Based on the MinION Mk1B dataset, the mean error rate across each position in the 18_Krt16 amplicons was 0.37%, with a minimum mean error rate of 0.06% and a maximum mean error rate of 2.50% (Figure 2.13). The Illumina MiSeq dataset exhibited a mean error rate of 0.20% across each 18_Krt16 amplicon position, with a minimum mean error rate of 0.05% and a maximum mean error rate of 2.02%.

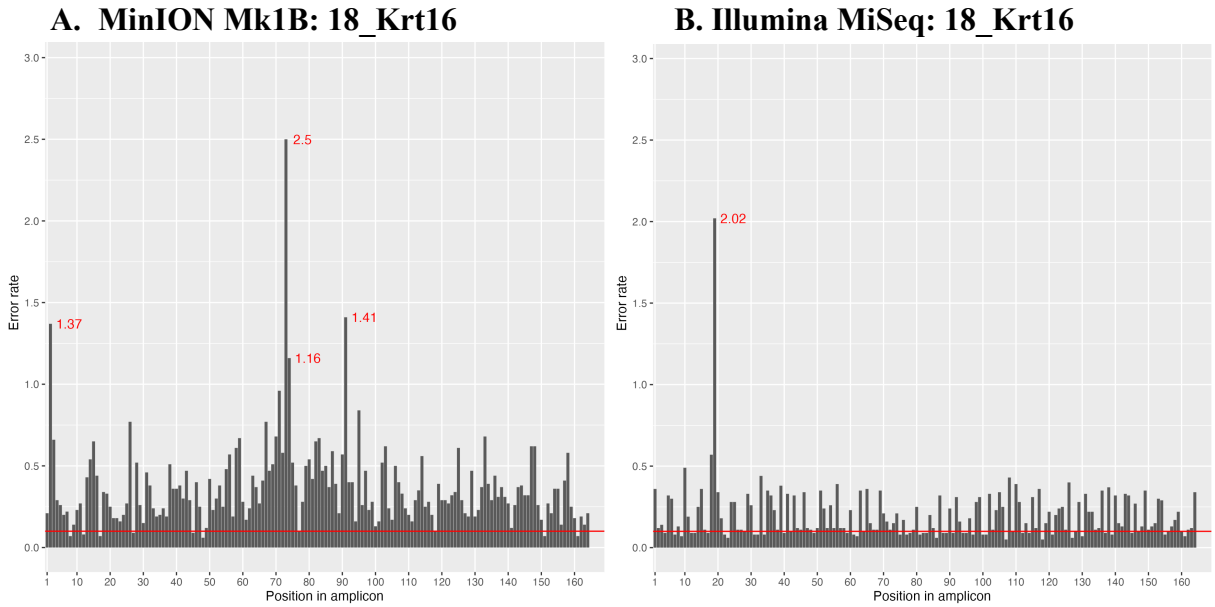


Figure 2.13 Mean error rate (%) at each position along the 18_Krt16 microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

The 32_Gpm6a microhaplotype amplicon was 223 base pairs in length. The mean error rate across all 32_Gpm6a amplicon positions was 0.34% for MinION Mk1B sequencing data, with a minimum mean error rate of 0.03% and a maximum mean error rate of 2.09% (Figure 2.14). For the 32_Gpm6a Illumina MiSeq sequencing data, the mean error rate across all amplicon positions was 0.19%, with a minimum mean error rate of 0.01% and a maximum mean error rate of 4.00%.

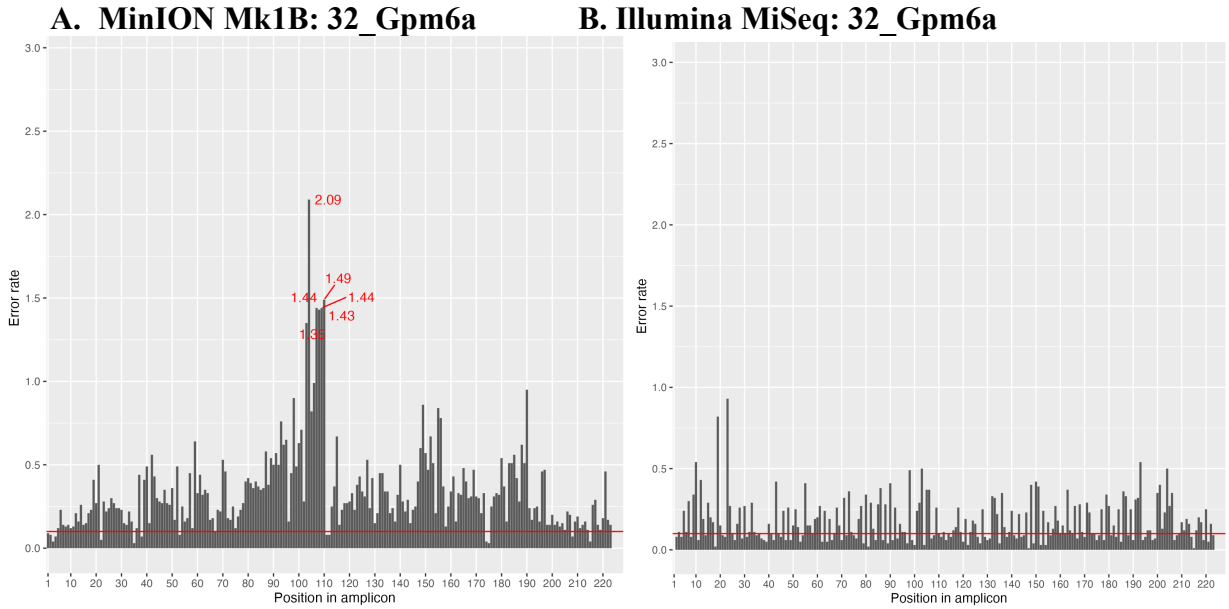


Figure 2.14 Mean error rate (%) at each position along the 32_Gpm6a microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

The MinION Mk1B dataset demonstrated a higher mean error rate for most amplicon positions in all target microhaplotypes. Nanopore sequencing is known to have a higher error rate than Illumina sequencing (Tytgat et al., 2022; Voskoboinik et al., 2018). However, the higher error rate observed from MinION Mk1B sequencing may be in part due to using a greater number of PCR cycles to amplify target microhaplotypes in the MinION Mk1B dataset compared to the Illumina MiSeq dataset (35 versus 30 cycles, respectively). More PCR cycles during initial amplification of target microhaplotypes results in more errors that can interfere with downstream analyses (Geue et al., 2024). Further optimization to reduce the number of PCR cycles while ensuring adequate PCR product is generated for each target amplicon is necessary for quality assurance purposes. Despite the high error rates observed here, confidence in microhaplotype calls can be restored when there is microhaplotype allele concordance across sequencing platforms for the same samples and loci. It is unlikely that the exact same sequence errors would be observed along ~350 bp amplicons for the primary microhaplotype alleles within a particular sample sequenced with two different instruments, especially at high read depth.

The presence/absence of indels in microhaplotype alleles generated from MinION Mk1B and Illumina MiSeq amplicon sequencing is shown across all caribou samples according to microhaplotype locus (Figure 2.15, Figure 2.16). Only the two primary microhaplotype alleles per locus within each sample were assessed for indels. Eleven microhaplotype loci did not generate indels in any of the 2 primary alleles for each sample in either dataset. These loci included 1_Rsl24d1, 2_Cd51, 2_Lexm, 2_Vcam1, 3_Slc34a1, 4_Abo, 11_Oaf, 18_Eme1, 18_Krt16, 20_Scn11a, and 32_Gpm6a.

In the MinION Mk1B dataset, two primary alleles for the 1_Gimap6 locus contained one or more indels and the rest did not. However, no 1_Gimap6 alleles sequenced with the Illumina MiSeq contained indels. The other microhaplotype locus with indels in the primary allele sequences was 18_Rnft1. All 18_Rnft1 primary alleles contained at least one indel for every sample sequenced in both datasets. For these two microhaplotype loci with alleles containing indels, the percentage of the 2 primary alleles per sample containing indels relative to the total number of primary alleles for all caribou samples according to locus was as follows: 1_Gimap6 (MinION 7.69%; Illumina 0.00%) and 18_Rnft1 (MinION 100%; Illumina 100%).

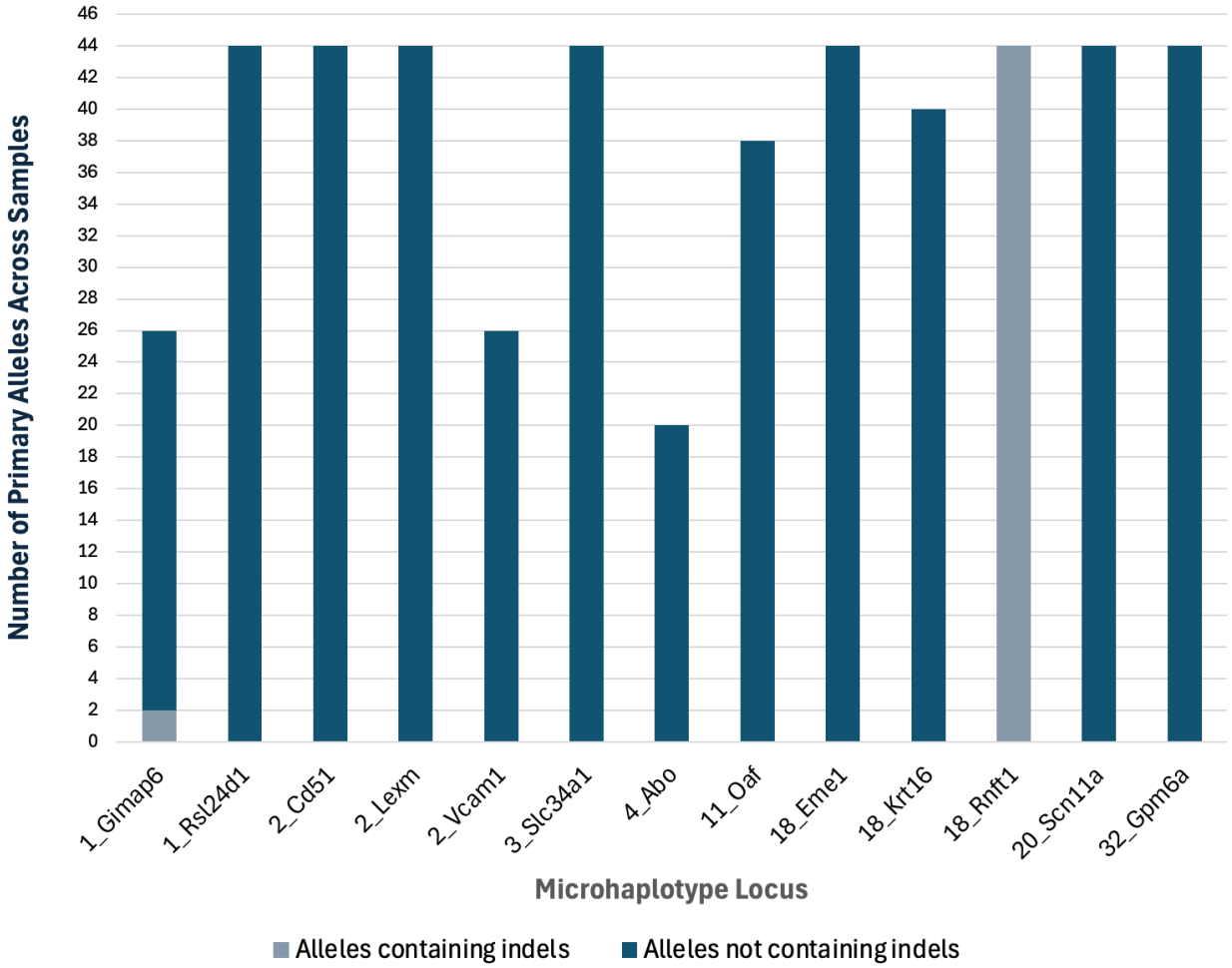


Figure 2.15 Presence of indels in the two primary microhaplotype alleles per sample according to microhaplotype locus for amplicon sequences generated with a MinION Mk1B. Microhaplotype alleles containing one or more indels are shown in gray. Microhaplotype alleles that do not contain any indels are shown in navy.

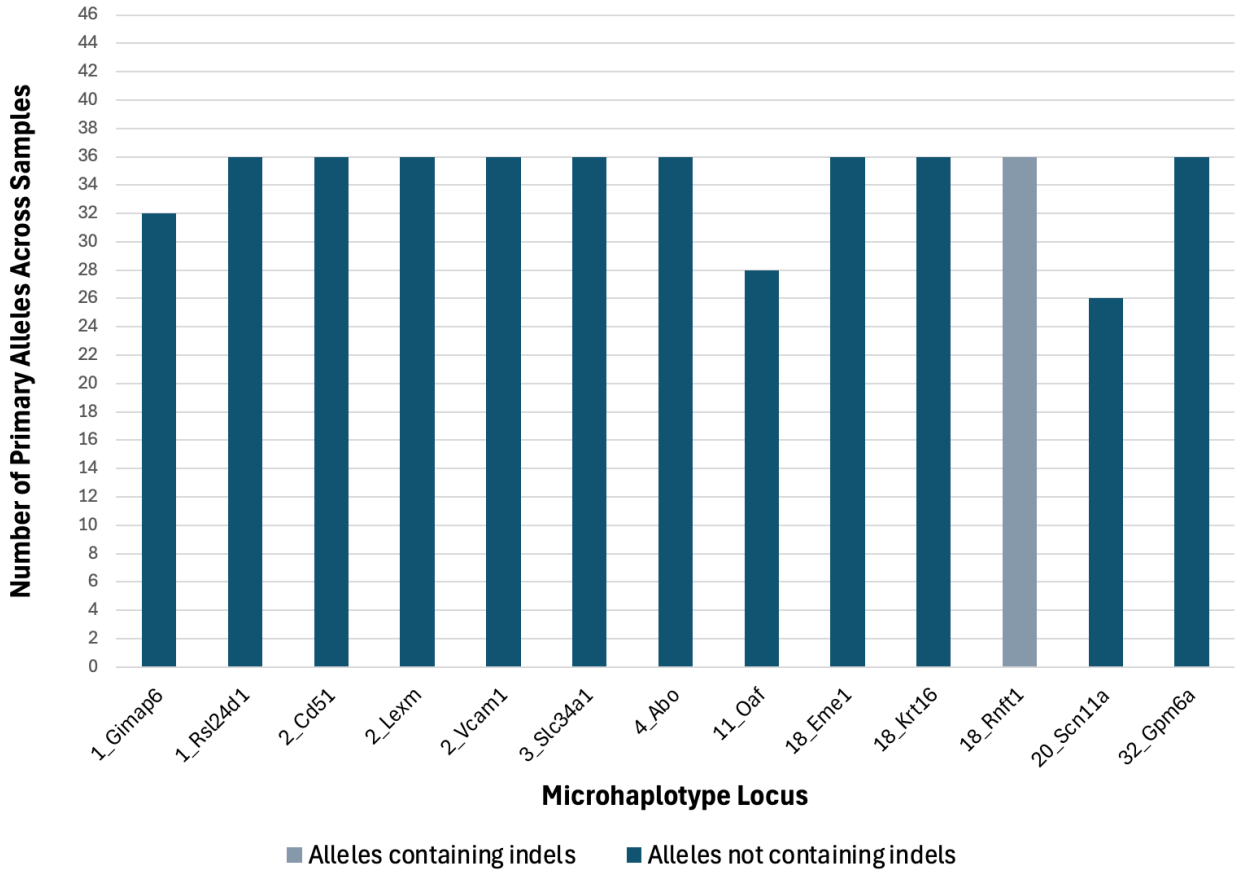


Figure 2.16 Presence of indels in the two primary microhaplotype alleles per sample according to microhaplotype locus for amplicon sequences generated with an Illumina MiSeq. Microhaplotype alleles containing one or more indels are shown in gray. Microhaplotype alleles that do not contain any indels are shown in navy.

The one instance where an indel was observed in the primary microhaplotype alleles for the 1_Gimap6 locus in the MinION Mk1B dataset was within caribou sample 21107. 1_Gimap6 was characterized as homozygous for this individual, however, the relatively low read depth (27 reads) and reads percentage (1.34%) suggests a likely sequencing error. This is further supported by this microhaplotype allele being identified as sequencing errors in other caribou samples at low read depths. Moreover, it is unlikely that this allele represents a true indel since it was much longer in length (bp) compared to the 1_Gimap6 reference sequence and had a weak alignment to the reference sequence. Instead of a true indel, this microhaplotype allele may be a PCR artefact originating from the initial PCR amplification for MinION sequencing, which could

explain why this supposed “indel” was not present for the same sample and locus in the Illumina dataset.

All primary alleles for the 18_Rnft1 locus contained indels in both datasets. Based on the read distribution pattern of 18_Rnft observed for many samples sequenced with the MinION Mk1B and Illumina MiSeq, as well as the exceptionally high coverage attained in this gene region from whole genome sequencing, it is hypothesized that 18_Rnft reads reflect a duplicate gene. As such, the “indels” recorded may not be true insertions or deletions and perhaps reflect reads captured from a highly similar gene elsewhere in the caribou genome.

Although unlikely, it is possible that the recorded indels were true indels. Genuine indels within target microhaplotype regions can be advantageous as they reveal additional variation, increasing the discriminating power per locus (Wendt et al., 2016). For example, Xue et al. (2023) developed a novel panel of microhaplotypes that contain SNPs and/or indels designed for complex kinship analysis in human forensics. A similar approach could be applied for relationship analysis in caribou. However, it is important to note that indels should be absent or minimal given the targeted exons in this microhaplotype locus set.

Cost Breakdown

The cost of a MinION Mk1B (~$2714 CAD) and Illumina MiSeq ($157,749 CAD) sequencer is not included in the cost breakdown. The costs of microhaplotype sequencing, including the total submission cost per sequencing run, cost per sample, and cost per microhaplotype locus, are shown according to the sequencer used, sample size, and number of target amplicons (Table 2.1). Comprehensive cost estimates for sequencing a panel of 25 microhaplotypes in 96 or 384 caribou samples with a MinION Mk1B or an Illumina MiSeq are shown in Appendix B (Table B4, Table B5, Table B6). The cost estimates were calculated based

on using the maximum number of samples for a MinION Mk1B (96 samples) and an Illumina MiSeq (384 samples) sequencing run to reduce cost per sample as much as possible. All prices are approximate and are shown in Canadian Dollars.

When sequencing 25 microhaplotype amplicons in 96 caribou samples with a MinION Mk1B device, the total submission cost is approximately $2,399.37. This corresponds to a per sample cost of $24.99 and a per locus cost of $1.00. To sequence 384 caribou samples with a MinION Mk1B sequencer, a 96-sample run would have to be repeated 4 times. This would not change the cost per locus or cost per sample, although the total submission cost for 384 samples on a MinION Mk1B would be approximately $9,597.48. In comparison, when sequencing the same panel of microhaplotypes in 96 caribou samples with an Illumina MiSeq, the estimated total submission cost is $3,058.53, with a cost per sample of $31.86 and a cost per locus of $1.27. Finally, for sample size of 384 caribou samples on the Illumina MiSeq platform, the total submission cost for sequencing a panel of 25 microhaplotype loci is $5,601.12, with a cost per sample of $14.59 and a cost per locus of $0.58.

Table 2.1 Estimate of total submission cost, average cost per sample, and average cost per locus according to the sequencing platform (MinION Mk1B or Illumina MiSeq) and sample size used for sequencing a panel of 25 microhaplotype amplicons. Prices are shown in $CAD.

Sequencing Platform	Sample Size	Total Submission Cost	Cost Per Sample	Number of Microhaplotype Loci	Cost Per Locus
MinION Mk1B	96 Samples	$2,399.37	$24.99	25	$1.00
MinION Mk1B	384 Samples	$9,597.48	$24.99	25	$1.00
Illumina MiSeq	96 Samples	$3,058.53	$31.86	25	$1.27
Illumina MiSeq	384 Samples	$5,601.12	$14.59	25	$0.58

Maximizing the number of samples per sequencing run is a huge factor in reducing overall costs. In experiments with a sample size of 384 samples, conducting Illumina MiSeq

sequencing is the more cost-effective option compared to MinION Mk1B sequencing, offering a lower total submission cost, cost per sample, and cost per locus. A larger sample size needs to be sequenced per run on the MinION Mk1B to be financially competitive with the Illumina MiSeq when dealing with more than 96 samples at a time. This could be accomplished with the ONT Ligation Sequencing DNA V14 Dual Barcoding Protocol, which includes the Native Barcoding Kit 24 V14 (SQK-NBD114.24) and PCR Barcoding Expansion pack (EXP-PBC096), allowing up to 2,304 samples (24 pools of 96 samples) to be sequenced in one run (Oxford Nanopore Technologies, 2024). This strategy involves an initial PCR amplification to add barcodes from the PCR Barcoding Expansion pack to each unique sample (up to 96) to be pooled. Then, a second PCR amplification adds a native barcode from the Native Barcoding Kit to each pool of samples. All pools are combined into one library before beginning library preparation steps.

In experiments with a sample size of 96 samples or fewer, the MinION Mk1B platform has a lower total submission cost, cost per sample, and cost per locus than the Illumina MiSeq system. However, routine microhaplotype profiling of individual caribou typically involves sequencing hundreds to thousands of samples as is conducted in our lab. Despite this, a benefit of conducting 4 sequencing runs of 96 samples on the MinION Mk1B instead of sequencing 384 samples in one run on the Illumina MiSeq is that the total submission cost is lower per MinION Mk1B sequencing run. Therefore, if a run were to fail, financial losses are mitigated. Additionally, for labs that do not have access to an Illumina MiSeq, the MinION Mk1B is a viable alternative for microhaplotype sequencing, regardless of sample size, and has a substantially lower capital investment.

Adopting multiplex PCR to amplify amplicons containing many SNPs and sequencing these regions with high throughput sequencing is recognized as a cost-efficient method of

microhaplotype genotyping (Eriksson et al., 2020). However, minimizing the number of microhaplotype loci sequenced in the same instrument run while maintaining statistical confidence is a key step towards more affordable data collection, thus increasing the utility of microhaplotypes (Baetscher et al., 2018). If the panel of 25 microhaplotype loci shown here is reduced to only include enough markers to identify individuals and correctly call microhaplotypes, costs would be diminished by minimizing the number of multiplex PCR amplifications required. Furthermore, having fewer genomic targets directs more sequencing power to allow a greater number of samples to be sequenced in one sequencing run, lowering per-sample costs (Baetscher et al., 2018).

Conclusion

Read distribution patterns should be considered when calling haplotypes, particularly from nanopore reads. The imbalance in read distribution across read 1, read 2 and read 3 observed here appears to heavily contribute to haplotype non-concordance across sequencing platforms when sequencing the same microhaplotype loci within the same caribou samples. Ambiguous read distributions tended to result in inconsistent haplotype calls with low confidence. These unclear read distributions are thought to arise from loci with gene duplications, or sequencing errors from initial PCR amplification. To improve clarity and consistency of haplotype calling, we recommend excluding microhaplotype loci that are suspected to be duplicated gene regions, as well as lowering the number of PCR cycles during amplification of the target DNA regions. Future experiments are needed to determine whether the incidence of ambiguous read distributions can be mitigated with these strategies, especially for the nanopore platform. This will provide a better understanding of whether the MinION

Mk1B is a viable option for sequencing microhaplotype loci and confidently calling haplotypes to monitor caribou populations in Canada.

CHAPTER 4
GENERAL DISCUSSION/CONCLUSION

The microsatellite chapter demonstrates that Oxford Nanopore's MinION Mk1B device can effectively sequence microsatellite markers captured by short amplicon regions from caribou tissue DNA samples, with comparable performance to the Illumina MiSeq technology. Both sequencers resulted in consistent microsatellite scoring across loci within the same caribou samples, validating the MinION's application to rapidly identify unique individuals using these markers, despite its long-read sequencing nature. The reproducible microsatellite scoring observed across sequencing platforms supports the cross-compatibility of this wildlife monitoring data, which is an important step towards sustaining long-term wildlife management projects (Schmidt et al., 2024).

Although microsatellites were characterized within tissue DNA samples in this study, DNA can be extracted from the mucosal lining of caribou fecal pellets (Ball et al., 2007; McFarlane et al., 2020) and sequenced on a high throughput sequencer to identify individuals (www.EcoGenomicsCanada.ca; Liu et al., 2024). Since fecal pellets are the predominant sample type that undergo genetic profiling under the current Canada-wide caribou monitoring framework (www.EcoGenomicsCanada.ca), testing the applicability of MinION Mk1B sequencing to profile microsatellites using fecal instead of tissue DNA samples is highly recommended. Non-invasively collected DNA samples, including fecal pellets, are widely used for wildlife monitoring research, particularly for elusive species (www.EcoGenomicsCanada.ca; Eriksson et al., 2020). Thus, validating MinION-based microsatellite sequencing with caribou fecal samples could provide an alternate population monitoring approach for research laboratories of varying sizes and resources to conduct on-site microsatellite genotyping analyses that rely on non-invasive sampling methods.

The microhaplotype chapter highlights the need to consider the read distribution beyond the two supposed primary alleles at each microhaplotype locus within each caribou sample to call individual haplotypes. When sequencing a novel panel of microhaplotype markers on a MinION Mk1B and an Illumina MiSeq platform, various read distribution patterns across the top 3 reads per locus were observed—some of which made it difficult to distinguish the true primary alleles as well as zygosity. This source of ambiguity can lead to inconsistent haplotype calls. We observed that achieving sufficient read depth with a clear read distribution across the top 3 allele reads is an important part of producing reliable microhaplotype calls. Although a clear read distribution across the top 3 reads does not guarantee that the resulting haplotype call is correct, it reduces the likelihood of that haplotype call being spurious (Baetscher et al., 2018).

We hypothesize that much of the read distribution ambiguity observed here is attributed to certain microhaplotype loci having gene duplications, as well as the high number of PCR cycles used to amplify target microhaplotype regions. The microhaplotype loci deemed to be duplicate gene regions can be excluded from future microhaplotyping, however, to establish whether the number of PCR cycles is directly related to the incidence of read distribution problems, we propose conducting a PCR-cycle optimization experiment. This would consist of amplifying the target microhaplotype loci with a reduced number of PCR cycles (20 cycles and 15 cycles) and performing the same amplicon sequencing and genotyping analysis to determine if fewer PCR cycles substantially mitigate the read distribution problem observed, particularly with nanopore sequencing.

Additional testing is needed to validate microhaplotype sequencing results on the MinION Mk1B and Illumina MiSeq platform. First, in an effort to recover some of the microhaplotype loci that dropped out, we recommend redesigning primers that did not map to the

appropriate genic region and reordering all primer sets, excluding primers targeting suspected duplicate gene regions. Second, the microhaplotype primers on a larger sample size, ideally 96 individuals as in the work of Baetscher et al. (2018), with duplicates for each caribou sample, would increase confidence in both sequencing platforms' ability to accurately call microhaplotypes from amplicon sequencing. Third, to conduct said amplicon sequencing, we suggest incorporating the optimal number of PCR cycles for target DNA amplification steps as determined from the previously described PCR-optimization experiment. Fourth, a greater number of microhaplotype loci will likely be required to achieve sufficient variability for accurate individual identification, and potentially relatedness analysis. In the current body of research, studies have shown larger microhaplotype panels to be sufficient for mixture detection and deconvolution in human forensic identifications (113 loci) (De la Puente et al., 2020), individual identification and relatedness analysis in gray wolves (321 loci) (Delomas et al., 2023) and resolving challenging relationship inferences in kelp rockfish (165 loci) (Baetscher et al., 2018). Before investing in additional microhaplotype loci for caribou monitoring, the target regions assessed in this thesis must be validated to determine if accurate and consistent microhaplotyping is feasible with the MinION Mk1B and/or Illumina MiSeq platform.

Although the MinION nanopore sequencer has shown some success in the microsatellite and microhaplotype studies presented here, cost remains a key factor in deciding whether to pursue this approach. Funding sustainability for long-term wildlife monitoring poses a considerable challenge, necessitating affordable data collection, analysis, and interpretation (Eriksson et al., 2020). When processing large number of samples to undergo genetic profiling, as is in the case of the EcoGenomics Canada research program (www.EcoGenomicsCanada.ca),

the Illumina MiSeq is more cost-effective, offering a lower cost per sample and cost per locus compared to the MinION Mk1B for microsatellite and microhaplotype characterization.

At present, a MinION Mk1B strategy may not be a worthwhile investment for routine microhaplotype profiling given its higher cost, sequencing error rate, and frequency of ambiguous read distributions observed. However, the MinION Mk1B could be more advantageous than the Illumina MiSeq for microsatellite profiling in caribou in some contexts, despite its higher per-sample and per-locus costs. Since the MinION Mk1B device is drastically cheaper than the Illumina MiSeq system, MinION Mk1B sequencing protocols could be conducted in more labs with the same net budget, as well as in remote or resource-limited labs. Decentralizing current microsatellite profiling efforts aimed at caribou population monitoring could expedite sample processing as well as reduce the supplement costs and inconveniences associated with shipping caribou samples across Canada to reach a laboratory facility possessing an Illumina MiSeq. Although, it is important to recognize that an approach relying on decentralized facilities would still require substantial infrastructure, such as proper laboratory equipment and trained individuals, which possesses a different set of challenges.

REFERENCES

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of Conservation Genetics. *Nature Reviews Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Arif, I. A., Khan, H. A., Bahkali, A. H., Al Homaidan, A. A., Al Farhan, A. H., Al Sadoon, M., & Shobrak, M. (2011). DNA marker technology for wildlife conservation. *Saudi Journal of Biological Sciences*, *18*(3), 219–225. https://doi.org/10.1016/j.sjbs.2011.03.002

Arifin, M. I., Staskevicius, A., Shim, S. Y., Huang, Y., Fenton, H., McLoughlin, P. D., Mitchell, G., Cullingham, C. I., & Gilch, S. (2020). Large-scale prion protein genotyping in Canadian caribou populations and potential impact on chronic wasting disease susceptibility. *Molecular Ecology*, *29*(20), 3830–3840. https://doi.org/10.1111/mec.15602

Assis, R., & Bachtrog, D. (2015). Rapid divergence and diversification of mammalian duplicate gene functions. *BMC Evolutionary Biology*, *15*(1), 138. https://doi.org/10.1186/s12862-015-0426-x

Athanasopoulou, K., Boti, M. A., Adamopoulos, P. G., Skourou, P. C., & Scorilas, A. (2021). Third-generation sequencing: The spearhead towards the radical transformation of Modern Genomics. *Life*, *12*(1), 30. https://doi.org/10.3390/life12010030

Baetscher, D. S., Clemento, A. J., Ng, T. C., Anderson, E. C., & Garza, J. C. (2018). Microhaplotypes provide increased power from short-read dna sequences for relationship inference. *Molecular Ecology Resources*, *18*(2), 296–305. https://doi.org/10.1111/1755-0998.12737

Ball, M. C., Finnegan, L., Manseau, M., & Wilson, P. (2010). Integrating multiple analytical approaches to spatially delineate and characterize genetic population structure: An application to boreal caribou (Rangifer tarandus caribou) in central Canada. *Conservation Genetics*, *11*(6), 2131–2143. https://doi.org/10.1007/s10592-010-0099-3

Ball, M. C., Pither, R., Manseau, M., Clark, J., Petersen, S. D., Kingston, S., Morrill, N., & Wilson, P. (2007). Characterization of target nuclear DNA from faeces reduces technical issues associated with the assumptions of low-quality and quantity template. *Conservation Genetics*, *8*(3), 577–586. https://doi.org/10.1007/s10592-006-9193-y

Bayerl, H., Kraus, R. H., Nowak, C., Foerster, D. W., Fickel, J., & Kuehn, R. (2017). Fast and cost-effective single nucleotide polymorphism (snp) detection in the absence of a reference genome using semideep next-generation random amplicon sequencing (ramseq). *Molecular Ecology Resources*, *18*(1), 107–117. https://doi.org/10.1111/1755-0998.12717

Bennett, L., Oldoni, F., Long, K., Cisana, S., Madella, K., Wootton, S., Chang, J., Hasegawa, R., Lagacé, R., Kidd, K. K., & Podini, D. (2019). Mixture deconvolution by massively parallel sequencing of microhaplotypes. *International Journal of Legal Medicine*, *133*(3), 719–729. https://doi.org/10.1007/s00414-019-02010-7

Bourgeois, S., Kaden, J., Senn, H., Bunnefeld, N., Jeffery, K. J., Akomo-Okoue, E. F., Ogden, R., & McEwing, R. (2019). Improving cost-efficiency of faecal genotyping: New tools for Elephant species. *PLOS ONE*, *14*(1). https://doi.org/10.1371/journal.pone.0210811

Broque, T., & Petit, E. (2004). Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology*, *13*(11), 3601–3608. https://doi.org/10.1111/j.1365-294x.2004.02352.x

Brownie, J. (1997). The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Research*, *25*(16), 3235–3241. https://doi.org/10.1093/nar/25.16.3235

Carratto, T. M., Moraes, V. M., Recalde, T. S., Oliveira, M. L., & Teixeira Mendes-Junior, C. (2022). Applications of massively parallel sequencing in Forensic Genetics. *Genetics and Molecular Biology*, *45*(3 suppl 1). https://doi.org/10.1590/1678-4685-gmb-2022-0077

Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications*, *11*(7), 1094–1119. https://doi.org/10.1111/eva.12600

Casanova-Adán, L., Mosquera-Miguel, A., González-Bao, J., Ambroa-Conde, A., Ruiz-Ramírez, J., Cabrejas-Olalla, A., González-Martín, E., Freire-Aradas, A., Rodríguez-López, A., Phillips, C., Lareu, M., & De la Puente, M. (2023). Adapting an established Ampliseq microhaplotype panel to Nanopore sequencing through direct PCR. *Forensic Science International: Genetics*, *67*, 102937. https://doi.org/10.1016/j.fsigen.2023.102937

Cornelis, S., Gansemans, Y., Deleye, L., Deforce, D., & Van Nieuwerburgh, F. (2017). Forensic SNP genotyping using Nanopore Minion Sequencing. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep41759

Daniels, J., Holmans, P., Williams, N., Turic, D., McGuffin, P., Plomin, R., & Owen, M. J. (1998). A simple method for analyzing microsatellite allele image patterns generated from DNA pools and its application to Allelic Association Studies. *The American Journal of Human Genetics*, *62*(5), 1189–1197. https://doi.org/10.1086/301816

Darby, B. J., Erickson, S. F., Hervey, S. D., & Ellis-Felege, S. N. (2016). Digital fragment analysis of short tandem repeats by high-throughput amplicon sequencing. *Ecology and Evolution*, *6*(13), 4502–4512. https://doi.org/10.1002/ece3.2221

De Barba, M., Miquel, C., Lobréaux, S., Quenette, P. Y., Swenson, J. E., & Taberlet, P. (2016). High-throughput microsatellite genotyping in ecology: Improved accuracy, efficiency,

standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, *17*(3), 492–507. https://doi.org/10.1111/1755-0998.12594

De Coster, W., & Rademakers, R. (2023). NanoPack2: Population-scale evaluation of long-read sequencing data. *Bioinformatics*, *39*(5). https://doi.org/10.1093/bioinformatics/btad311

De la Puente, M., Ruiz-Ramírez, J., Ambroa-Conde, A., Xavier, C., Amigo, J., Casares de Cal, M. Á., Gómez-Tato, A., Carracedo, Á., Parson, W., Phillips, C., & Lareu, M. V. (2020). Broadening the applicability of a custom multi-platform panel of microhaplotypes: Biogeographical ancestry inference and expanded reference data. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.581041

De Woody, J. A. (2005). Molecular approaches to the study of parentage, relatedness, and fitness: Practical applications for wild animals. *Journal of Wildlife Management*, *69*(4), 1400–1418. https://doi.org/10.2193/0022-541x(2005)69[1400:mattso]2.0.co;2

Delomas, T. A., Struthers, J., Hebdon, T., & Campbell, M. R. (2023). Development of a microhaplotype panel to inform management of Gray Wolves. *Conservation Genetics Resources*, *15*(1–2), 49–57. https://doi.org/10.1007/s12686-023-01301-x

DeCesare, N. J., Hebblewhite, M., Bradley, M., Hervieux, D., Neufeld, L., & Musiani, M. (2013). Linking habitat selection and predation risk to spatial variation in survival. *Journal of Animal Ecology*, *83*(2), 343–352. https://doi.org/10.1111/1365-2656.12144

Deyoung, R. W., & Honeycutt, R. L. (2005). The molecular toolbox: Genetic techniques in wildlife ecology and Management. *Journal of Wildlife Management*, *69*(4), 1362–1384. https://doi.org/10.2193/0022-541x(2005)69[1362:tmtgti]2.0.co;2

Dokos, R. S., Reid, S., & Jayasinghe, L. (2022). Update from Oxford Nanopore Technologies. Available online: https://nanoporetech.com/resource-centre/video/ncm22/update-from-oxford-nanopore (accessed on 3 December 2023).

EcoGenomics Canada. (n.d.). *Analytical and Collaborative Genomics for Wildlife Monitoring*. Retrieved January 5, 2024, from https://www.ecogenomicscanada.ca/

Ehlers, L. P., Johnson, C. J., & Seip, D. R. (2016). Evaluating the influence of anthropogenic landscape change on Wolf Distribution: Implications for woodland caribou. *Ecosphere*, *7*(12). https://doi.org/10.1002/ecs2.1600

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, *5*(6), 435–445. https://doi.org/10.1038/nrg1348

Environment and Climate Change Canada. (2023, February 1). *Caribou in Canada*. Canada.ca. Retrieved November 27, 2023, from https://www.canada.ca/en/environment-climate-change/services/species-risk-education-centre/caribou.html

Environment and Climate Change Canada. (2020, November 27). *Amended Recovery Strategy for the Woodland Caribou (Rangifer tarandus caribou), Boreal population, in Canada [Proposed] 2019*. Canada.ca. Retrieved November 14, 2022, from https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/recovery-strategies/woodland-caribou-boreal-2019.html

Eriksson, C. E., Ruprecht, J., & Levi, T. (2020). More affordable and effective noninvasive single nucleotide polymorphism genotyping using high-throughput amplicon sequencing. *Molecular Ecology Resources*, *20*(6), 1505–1516. https://doi.org/10.1111/1755-0998.13208

Esnault, G., Earley, B., Cormican, P., Waters, S. M., Lemon, K., Cosby, S. L., Lagan, P., Barry, T., Reddington, K., & McCabe, M. S. (2022). Assessment of Rapid Minion Nanopore DNA virus meta-genomics using calves experimentally infected with bovine herpes virus-1. *Viruses*, *14*(9), 1859. https://doi.org/10.3390/v14091859

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, *8*(3), 186–194. https://doi.org/10.1101/gr.8.3.186

Festa-Bianchet, M., Ray, J. C., Boutin, S., Cote, S. D., & Gunn, A. (2011). Conservation of caribou (*Rangifer tarandus*) in Canada: an uncertain future. *Canadian Journal of Zoology*, *89*(5), 419-434. http://dx.doi.org.proxy1.lib.trentu.ca/10.1139/Z11-025

Flasko, A., Manseau, M., Mastromonaco, G., Bradley, M., Neufeld, L., & Wilson, P. (2017). Fecal DNA, hormones, and pellet morphometrics as a noninvasive method to estimate age class: An application to wild populations of Central Mountain and boreal woodland caribou (*Rangifer tarandus caribou*). *Canadian Journal of Zoology*, *95*(5), 311–321. https://doi.org/10.1139/cjz-2016-0070

Fournier, C., Manseau, M., Redquest, B., Andrew, L., Kelly, A., Hervieux, D., Hegel, T., Pittoello, G., Trim, V., Brannen, D., & Wilson, P. (2024). Genetic network analysis uncovers spatial variation in diversity and connectivity of a species presenting a continuous distribution. *Global Ecology and Conservation*, *54*. https://doi.org/10.1016/j.gecco.2024.e03119

Frenette, J., Pelletier, F., & St-Laurent, M.-H. (2020). Linking habitat, predators and alternative prey to explain recruitment variations of an endangered Caribou Population. *Global Ecology and Conservation*, *22*. https://doi.org/10.1016/j.gecco.2020.e00920

Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLOS ONE*, *8*(11). https://doi.org/10.1371/journal.pone.0079667

Galpern, P., Manseau, M., Hettinga, P., Smith, K., & Wilson, P. (2012). Allelematch: An R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. *Molecular Ecology Resources*, *12*(4), 771–778. https://doi.org/10.1111/j.1755-0998.2012.03137.x

Geue, J. C., Liu, P., Keobouasone, S., Wilson, P., & Manseau, M. (2024). MhGeneS: An analytical pipeline to allow for robust microhaplotype genotyping. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.14027

Giglio, R. M., Ivy, J. A., Jones, L. C., & Latch, E. K. (2016). Evaluation of alternative management strategies for maintenance of genetic variation in wildlife populations. *Animal Conservation*, *19*(4), 380–390. https://doi.org/10.1111/acv.12254

Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T. J., Clayton, J. B., Johnson, T. J., Hunter, R., Knights, D., & Beckman, K. B. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in Microbiome Studies. *Nature Biotechnology*, *34*(9), 942–949. https://doi.org/10.1038/nbt.3601

Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, *36*(9), 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025

Guenay-Greunke, Y., Bohan, D. A., Traugott, M., & Wallinger, C. (2021). Handling of targeted amplicon sequencing data focusing on index hopping and demultiplexing using a nested metabarcoding approach in ecology. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-98018-4

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Leger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., & Petit, R. J. (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources*, *11*(4), 591–611. https://doi.org/10.1111/j.1755-0998.2011.03014.x

Hall, C. L., Kesharwani, R. K., Phillips, N. R., Planz, J. V., Sedlazeck, F. J., & Zascavage, R. R. (2022). Accurate profiling of forensic autosomal strs using the Oxford Nanopore Technologies Minion Device. *Forensic Science International: Genetics, 56*, 102629. https://doi.org/10.1016/j.fsigen.2021.102629

Hebert, P. D., Floyd, R., Jafarpour, S., & Prosser, S. W. (2024). Barcode 100k specimens: In a single Nanopore run. *Molecular Ecology Resources*, *25*(1). https://doi.org/10.1111/1755-0998.14028

Hettinga, P. N., Arnason, A. N., Manseau, M., Cross, D., Whaley, K., & Wilson, P. J. (2012). Estimating size and trend of the north interlake woodland caribou population using fecal-dna and capture–recapture models. *The Journal of Wildlife Management*, *76*(6), 1153–1164. https://doi.org/10.1002/jwmg.380

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2020). Population Genomics for Wildlife Conservation and Management. *Molecular Ecology*, *30*(1), 62–82. https://doi.org/10.1111/mec.15720

Hölzl-Müller, P., Bodner, M., Berger, B., & Parson, W. (2021). Exploring STR sequencing for forensic DNA intelligence databasing using the Austrian national DNA database as an example. *International Journal of Legal Medicine*, *135*(6), 2235–2246. https://doi.org/10.1007/s00414-021-02685-x

Huang, S., Sheng, M., Li, Z., Li, K., Chen, J., Wu, J., Wang, K., Shi, C., Ding, H., Zhou, H., Ma, L., Yang, J., Pu, Y., Yu, Y., Chen, F., & Chen, P. (2022). Inferring bio-geographical ancestry with 35 microhaplotypes. *Forensic Science International*, *341*, 111509. https://doi.org/10.1016/j.forsciint.2022.111509

Illumina. (2022). *NGS Workflow Steps*. Illumina sequencing workflow. Retrieved November 2, 2022, from https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html

Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B., & Akeson, M. (2015). Improved Data Analysis for the minion nanopore sequencer. *Nature Methods*, *12*(4), 351–356. https://doi.org/10.1038/nmeth.3290

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., … Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, *36*(4), 338–345. https://doi.org/10.1038/nbt.4060

Jones, A. G., Small, C. M., Paczolt, K. A., & Ratterman, N. L. (2010). A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, *10*(1), 6–30. https://doi.org/10.1111/j.1755-0998.2009.02778.x

Jones, T. B., & Manseau, M. (2022). Genetic networks in ecology: A guide to population, relatedness, and pedigree networks and their applications in Conservation Biology. *Biological Conservation*, *267*, 109466. https://doi.org/10.1016/j.biocon.2022.109466

Jones, T. B., Manseau, M., Merriell, B., Pittoello, G., Hervieux, D., & Wilson, P. J. (2023). Novel multilayer network analysis to assess variation in the spatial co-occurrences of close kin in wild caribou populations. *Global Ecology and Conservation*, *47*. https://doi.org/10.1016/j.gecco.2023.e02688

Karamitros, T., & Magiorkinis, G. (2018). Multiplexed targeted sequencing for Oxford Nanopore Minion: A detailed library preparation procedure. *Methods in Molecular Biology*, 1712, 43–51. https://doi.org/10.1007/978-1-4939-7514-3_4

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagacé, R., Chang, J., Wootton, S., Haigh, E., & Kidd, J. R. (2014). Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International: Genetics*, *12*, 215–224. https://doi.org/10.1016/j.fsigen.2014.06.014

Kircher, M., Sawyer, S., & Meyer, M. (2011). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1). https://doi.org/10.1093/nar/gkr771

Klütsch, C. F., Manseau, M., Trim, V., Polfus, J., & Wilson, P. J. (2016). The Eastern Migratory Caribou: The role of genetic introgression in Ecotype Evolution. *Royal Society Open Science*, *3*(2), 150469. https://doi.org/10.1098/rsos.150469

Kraus, R. H., vonHoldt, B., Cocchiararo, B., Harms, V., Bayerl, H., Kühn, R., Förster, D. W., Fickel, J., Roos, C., & Nowak, C. (2014). A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in europe based on noninvasively collected samples. *Molecular Ecology Resources*, *15*(2), 295–305. https://doi.org/10.1111/1755-0998.12307

Liu, P., Wilson, P., Redquest, B., Keobouasone, S., & Manseau, M. (2024). Seq2Sat and SatAnalyzer toolkit: Towards comprehensive microsatellite genotyping from sequencing data. *Molecular Ecology Resources*, *24*(3). https://doi.org/10.1111/1755-0998.13929

Lou, R. N., Jacobs, A., Wilder, A., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*, *30*(23). https://doi.org/10.22541/au.160689616.68843086/v4

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore Minion Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, *14*(5), 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

McFarlane, S., Manseau, M., Flasko, A., Horn, R. L., Arnason, N., Neufeld, L., Bradley, M., & Wilson, P. (2018). Genetic influences on male and female variance in reproductive success and implications for the recovery of severely endangered Mountain Caribou. *Global Ecology and Conservation*, *16*. https://doi.org/10.1016/j.gecco.2018.e00451

McFarlane, S., Manseau, M., Steenweg, R., Hervieux, D., Hegel, T., Slater, S., & Wilson, P. J. (2020). An assessment of sampling designs using SCR analyses to estimate abundance of boreal caribou. *Ecology and Evolution*, *10*(20), 11631–11642. https://doi.org/10.1002/ece3.6797

McFarlane, S., Manseau, M., & Wilson, P. J. (2021). Spatial familial networks to infer demographic structure of wild populations. *Ecology and Evolution*, *11*(9), 4507–4519. https://doi.org/10.1002/ece3.7345

Mumma, M. A., Gillingham, M. P., Parker, K. L., Johnson, C. J., & Watters, M. (2018). Predation risk for boreal woodland caribou in human-modified landscapes: Evidence of wolf spatial responses independent of apparent competition. *Biological Conservation*, *228*, 215–223. https://doi.org/10.1016/j.biocon.2018.09.015

Ogden, R., Dawnay, N., & McEwing, R. (2009). Wildlife DNA forensics—bridging the gap between conservation genetics and law enforcement. *Endangered Species Research*, *9*, 179–195. https://doi.org/10.3354/esr00144

Oh, K. P., Aldridge, C. L., Forbey, J. S., Dadabay, C. Y., & Oyler-McCance, S. J. (2019). Conservation genomics in the sagebrush sea: Population divergence, demographic history, and local adaptation in sage-grouse (Centrocerchus spp.). *Genome Biology and Evolution*, *11*, 2023–2034

Oldoni, F., Kidd, K. K., & Podini, D. (2018). Microhaplotypes in forensic genetics. Forensic Science International: *Genetics*, *38*, 54–69. https://doi.org/10.1016/j.fsigen.2018.09.009

Osborne, M. J., Caeiro-Dias, G., & Turner, T. F. (2022). Transitioning from microsatellites to snp-based microhaplotypes in genetic monitoring programmes: Lessons from paired data spanning 20 years. *Molecular Ecology*, *32*(2), 316–334. https://doi.org/10.1111/mec.16760

Oxford Nanopore Technologies. (2022). *Kit 14 sequencing and duplex basecalling*. Retrieved November 30, 2022, from https://community.nanoporetech.com/docs/plan/best_practice/kit-14-device-and-informatics/v/k14_s1019_v1_revk_29jun2022

Oxford Nanopore Technologies. (2022). *MinION*. Retrieved September 20, 2022, from https://nanoporetech.com/products/minion?utm_campaign=Digital+Gearbox+Google+adverts&utm_source=ppc&gclid=CjwKCAjwyaWZBhBGEiwACslQo6Z659gcdYc7AkTYkTE0dsN4DMcEaa0nKZFhTXHkgnKdfK56nk9htxoCDXgQAvD_BwE#tabs-0=mk1b

Oxford Nanopore Technologies. (2024). *Ligation sequencing DNA V14 - dual barcoding (SQK-NBD114.24 with EXP-PBC096) - minion - overview of the Protocol*. Retrieved April 30, 2024, from https://community.nanoporetech.com/docs/prepare/library_prep_protocols/ligation-sequencing-dual-barcoding-v14/v/dbc_9184_v114_revi_22mar2023/overview-of-the-protocol?devices=minion

Oxford Nanopore Technologies. (n.d.). *How basecalling works*. Retrieved January 3, 2025, from https://nanoporetech.com/platform/technology/basecalling#capturing-the-signal

Peng, Q., Vijaya Satya, R., Lewis, M., Randad, P., & Wang, Y. (2015). Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-1806-8

Pemberton, J. M. (2008). Wild pedigrees: The way forward. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1635), 613–621. https://doi.org/10.1098/rspb.2007.1531

Quail, M., Smith, M. E., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., & Gu, Y. (2012). A tale of three next generation sequencing platforms:

Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, *13*(1), 341. https://doi.org/10.1186/1471-2164-13-341

Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1462-9

Rohilla, K. J., Ovington, K. N., Pater, A. A., Barton, M., Henke, A. J., & Gagnon, K. T. (2020). Systematic microsatellite repeat expansion cloning and validation. *Human Genetics*, *139*(10), 1233–1246. https://doi.org/10.1007/s00439-020-02165-z

Schmidt, T. L., Thia, J. A., & Hoffmann, A. A. (2024). How can genomics help or Hinder Wildlife Conservation? *Annual Review of Animal Biosciences*, *12*(1), 45–68. https://doi.org/10.1146/annurev-animal-021022-051810

Schwartz, M., Luikart, G., & Waples R. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution*, *22*(1), 25–33. https://doi.org/10.1016/j.tree.2006.08.009

Shum, J., & Paul, N. (2009). Chemically modified primers for improved multiplex polymerase chain reaction. *Analytical Biochemistry*, *388*(2), 266–272. https://doi.org/10.1016/j.ab.2009.02.033

Suez, M., Behdenna, A., Brouillet, S., Graça, P., Higuet, D., & Achaz, G. (2015). MICNESS: Genotyping microsatellite loci from a collection of (NGS) reads. *Molecular Ecology Resources*, *16*(2), 524–533. https://doi.org/10.1111/1755-0998.12467

Taberlet, P., Camarra, J. -J., Griffin, S., Uhrès, E., Hanotte, O., Waits, L. P., Dubois-Paganon, C., Burke, T., & Bouvet, J. (1997). Noninvasive genetic tracking of the endangered pyrenean brown bear population. *Molecular Ecology*, *6*(9), 869–876. https://doi.org/10.1046/j.1365-294x.1997.00251.x

Taylor, R. S., Horn, R. L., Zhang, X., Golding, G. B., Manseau, M., & Wilson, P. J. (2019). The Caribou (*Rangifer tarandus*) genome. *Genes*, *10*(7), 540. https://doi.org/10.3390/genes10070540

Thaden, A., Nowak, C., Tiesmeyer, A., Reiners, T. E., Alves, P. C., Lyons, L. A., Mattucci, F., Randi, E., Cragnolini, M., Galián, J., Hegyeli, Z., Kitchener, A. C., Lambinet, C., Lucas, J. M., Mölich, T., Ramos, L., Schockert, V., & Cocchiararo, B. (2020). Applying genomic data in wildlife monitoring: Development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. *Molecular Ecology Resources*, *20*(3), 662–680. https://doi.org/10.1111/1755-0998.13136

Therkildsen, N. O., & Palumbi, S. R. (2016). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in

nonmodel species. *Molecular Ecology Resources*, *17*(2), 194–208. https://doi.org/10.1111/1755-0998.12593

Thompson, L. M., Klütsch, C. F., Manseau, M., & Wilson, P. J. (2019). Spatial differences in genetic diversity and northward migration suggest genetic erosion along the boreal caribou southern range limit and continued range retraction. *Ecology and Evolution*, *9*(12), 7030–7046. https://doi.org/10.1002/ece3.5269

Turchi, C., Melchionda, F., Pesaresi, M., & Tagliabracci, A. (2019). Evaluation of a microhaplotypes panel for forensic genetics using massive parallel sequencing technology. *Forensic Science International: Genetics*, *41*, 120–127. https://doi.org/10.1016/j.fsigen.2019.04.009

Tytgat, O., Gansemans, Y., Weymaere, J., Rubben, K., Deforce, D., & Van Nieuwerburgh, F. (2020). Nanopore sequencing of a forensic STR multiplex reveals loci suitable for single-contributor STR profiling. *Genes*, *11*(4), 381. https://doi.org/10.3390/genes11040381

Tytgat, O., Škevin, S., Deforce, D., & Van Nieuwerburgh, F. (2022). Nanopore sequencing of a forensic combined STR and SNP multiplex. *Forensic Science International: Genetics*, *56*, 102621. https://doi.org/10.1016/j.fsigen.2021.102621

Urban, L., Perlas, A., Francino, O., Martí-Carreras, J., Muga, B. A., Mwangi, J. W., Boykin Okalebo, L., Stanton, J. L., Black, A., Waipara, N., Fontsere, C., Eccles, D., Urel, H., Reska, T., Morales, H. E., Palmada-Flores, M., Marques-Bonet, T., Watsa, M., Libke, Z., … van Oosterhout, C. (2023). Real-time genomics for One Health. *Molecular Systems Biology*, *19*(8). https://doi.org/10.15252/msb.202311686

Van der Gaag, K. J., de Leeuw, R. H., Laros, J. F. J., den Dunnen, J. T., & de Knijff, P. (2018). Short hypervariable microhaplotypes: A novel set of very short high discriminating power loci without stutter artefacts. *Forensic Science International: Genetics*, *35*, 169–175. https://doi.org/10.1016/j.fsigen.2018.05.008

Van der Reis, A. L., Beckley, L. E., Olivar, M. P., & Jeffs, A. G. (2022). Nanopore short-read sequencing: A Quick, cost-effective and accurate method for dna metabarcoding. *Environmental DNA*, *5*(2), 282–296. https://doi.org/10.1002/edn3.374

Van der Verren, S. E., Van Gerven, N., Jonckheere, W., Hambley, R., Singh, P., Kilgour, J., Jordan, M., Wallace, E. J., Jayasinghe, L., & Remaut, H. (2020). A dual-constriction biological nanopore resolves homonucleotide sequences with high fidelity. *Nature Biotechnology*, *38*(12), 1415–1420. https://doi.org/10.1038/s41587-020-0570-8

Van Dijk, E. L., Naquin, D., Gorrichon, K., Jaszczyszyn, Y., Ouazahrou, R., Thermes, C., & Hernandez, C. (2023). Genomics in the long-read sequencing era. *Trends in Genetics*, *39*(9), 649–671. https://doi.org/10.1016/j.tig.2023.04.006

Van Neste, C., Vandewoestyne, M., Van Criekinge, W., Deforce, D., & Van Nieuwerburgh, F. (2014). My-forensic-loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing. *Forensic Science International: Genetics*, *9*, 1–8. https://doi.org/10.1016/j.fsigen.2013.10.012

Vieira, M. L., Santini, L., Diniz, A. L., & Munhoz, C. de. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, *39*(3), 312–328. https://doi.org/10.1590/1678-4685-gmb-2016-0027

Vincent, A. T., Derome, N., Boyle, B., Culley, A. I., & Charette, S. J. (2017). Next-generation sequencing (NGS) in the Microbiological World: How to make the most of your money. *Journal of Microbiological Methods*, *138*, 60–71. https://doi.org/10.1016/j.mimet.2016.02.016

Von Thaden, A., Nowak, C., Tiesmeyer, A., Reiners, T. E., Alves, P. C., Lyons, L. A., Mattucci, F., Randi, E., Cragnolini, M., Galián, J., Hegyeli, Z., Kitchener, A. C., Lambinet, C., Lucas, J. M., Mölich, T., Ramos, L., Schockert, V., & Cocchiararo, B. (2020). Applying genomic data in wildlife monitoring: Development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. *Molecular Ecology Resources*, *20*(3), 662–680. https://doi.org/10.1111/1755-0998.13136

Voskoboinik, L., Motro, U., & Darvasi, A. (2018). Facilitating complex DNA mixture interpretation by sequencing highly polymorphic haplotypes. *Forensic Science International: Genetics*, *35*, 136–140. https://doi.org/10.1016/j.fsigen.2018.05.001

Walsh, P. S., Fildes, N. J., & Reynolds, R. (1996). Sequence Analysis and characterization of stutter products at the tetranucleotide repeat locus VWA. *Nucleic Acids Research*, *24*(14), 2807–2812. https://doi.org/10.1093/nar/24.14.2807

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, Bioinformatics and Applications. *Nature Biotechnology*, *39*(11), 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Wendt, F. R., Warshauer, D. H., Zeng, X., Churchill, J. D., Novroski, N. M. M., Song, B., King, J. L., LaRue, B. L., & Budowle, B. (2016). Massively parallel sequencing of 68 insertion/deletion markers identifies novel microhaplotypes for utility in human identity testing. *Forensic Science International: Genetics*, *25*, 198–209. https://doi.org/10.1016/j.fsigen.2016.09.005

Whitford, W., Hawkins, V., Moodley, K. S., Grant, M. J., Lehnert, K., Snell, R. G., & Jacobsen, J. C. (2022). Proof of concept for multiplex amplicon sequencing for mutation identification using the Minion Nanopore sequencer. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-12613-7

Xue, J., Tan, M., Zhang, R., Chen, D., Liu, G., Zheng, Y., Wu, Q., Xiao, Y., Liao, M., Qu, S., & Liang, W. (2023). Evaluation of microhaplotype panels for complex kinship analysis

using massively parallel sequencing. *Forensic Science International: Genetics*, *65*, 102887. https://doi.org/10.1016/j.fsigen.2023.102887

Zhan, L., Paterson, I. G., Fraser, B. A., Watson, B., Bradbury, I. R., Nadukkalam Ravindran, P., Reznick, D., Beiko, R. G., & Bentzen, P. (2017). Megasat: Automated inference of microsatellite genotypes from sequence data. *Molecular Ecology Resources*, *17*(2), 247–256. https://doi.org/10.1111/1755-0998.12561

Zou, C., Karn, A., Reisch, B., Nguyen, A., Sun, Y., Bao, Y., Campbell, M. S., Church, D., Williams, S., Xu, X., Ledbetter, C. A., Patel, S., Fennell, A., Glaubitz, J. C., Clark, M., Ware, D., Londo, J. P., Sun, Q., & Cadle-Davidson, L. (2020). Haplotyping the Vitis collinear core genome with rhampseq improves marker transferability in a diverse genus. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-019-14280-1

APPENDIX A

Table A1 Microsatellite 8-plex PCR reaction for 19 caribou samples to be sequenced with a MinION Mk1B sequencer. Enough Master Mix for 40 reactions were prepared, each with a total reaction volume of 13 µL.

Reagent	**Stock**	**[ ] / rxn**	**µL / - µL rxn**	**Master Mix (µL)**
dd Water			0.59	23.73
Multiplex PCR Type-It Master Mix (2x)	2	1	6.5	260
BSA	3	0.2	0.87	34.67
Map2c-F	40	0.1	0.03	1.3
Map2c-R	40	0.1	0.03	1.3
Rt27-F	40	0.2	0.07	2.6
Rt27-R	40	0.2	0.07	2.6
Oheq-F	40	0.1	0.03	1.3
Oheq-R	40	0.1	0.03	1.3
IGF1-F	40	0.1	0.03	1.3
IGF1-R	40	0.1	0.03	1.3
NVHRT48-F	40	0.6	0.2	7.8
NVHRT48-R	40	0.6	0.2	7.8
Rt6-F	40	0.1	0.03	1.3
Rt6-R	40	0.1	0.03	1.3
RT7-F	40	0.1	0.03	1.3
Rt7-R	40	0.1	0.03	1.3
Rt24-F	40	0.3	0.1	3.9
Rt24-R	40	0.3	0.1	3.9
DNA template		**250pg/µL**	**4.00**	

TOTAL COCKTAIL VOLUME	**13.00**	**360**
	Volume of mm/tube	**9.00**

Table A2 *Zfx*/*Zfy* multiplex PCR reaction for 19 caribou samples to be sequenced with a MinION Mk1B sequencer. Enough Master Mix for 40 reactions were prepared, each with a total reaction volume of 11 µL. The only primers relevant to this experiment are LGL335 and SDP730 which correspond to caribou sex markers.

Reagent	Stock	[ ] / rxn	µL / - µL rxn	Master Mix (µL)
dd Water			0.22	8.67
Multiplex PCR Type-It Master Mix (2x)	2	1	5.50	220.00
BSA	3	0.2	0.73	29.33
PRNP-F	40	0.5	0.14	5.50
PRNP-R	40	0.5	0.14	5.50
LGL335	40	0.5	0.14	5.50
SDP730	40	0.5	0.14	5.50
DNA template		**250pg/µL**	**4.00**	
TOTAL COCKTAIL VOLUME			**11.00**	**280.00**
			Volume of mm/tube	**7.00**

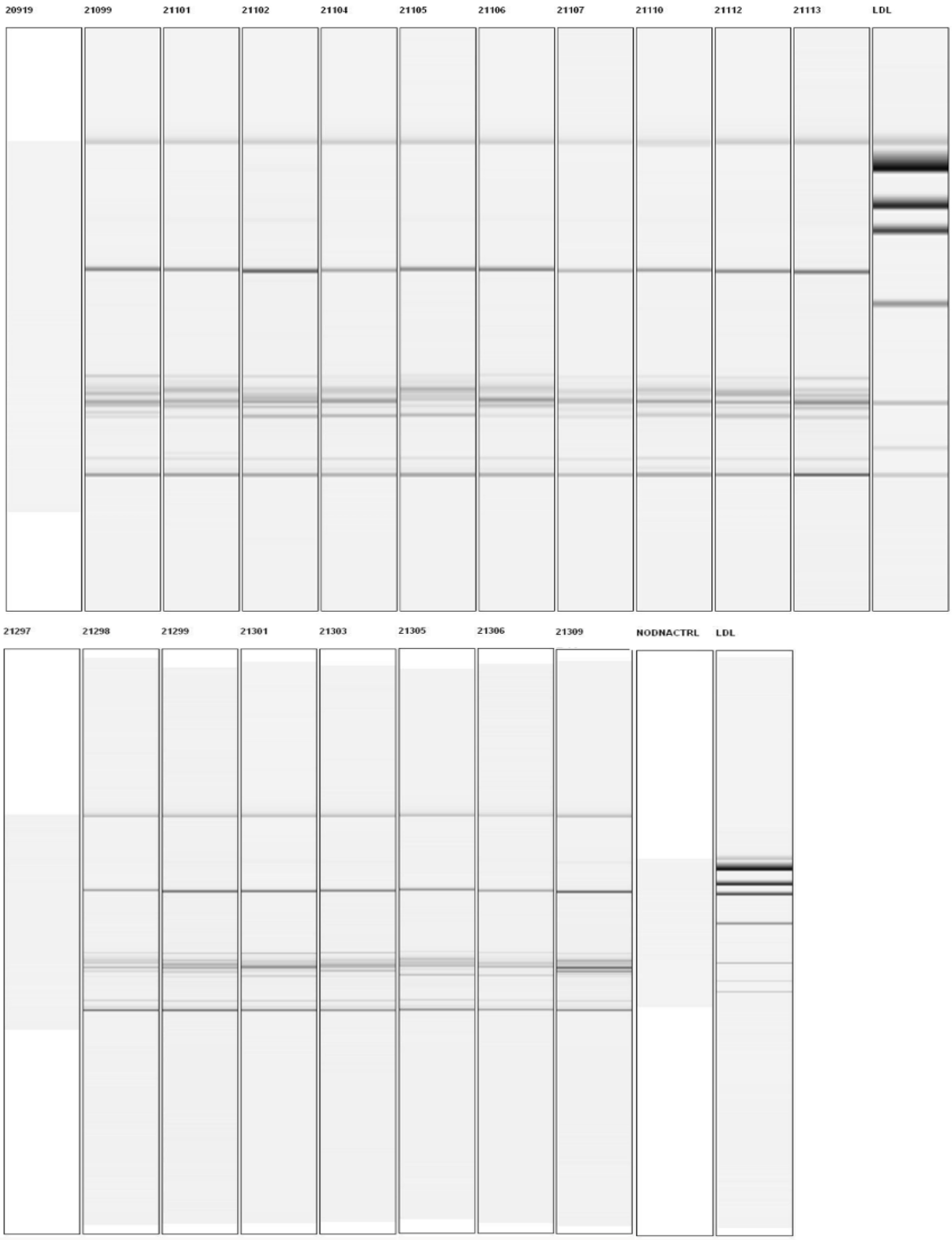


Figure A1 Capillary electrophoresis images of pooled PCR-amplified microsatellite loci and sex targets in each caribou sample ID generated from a QIAxcel Advanced System by Qiagen.

Table A3 DNA quantification of each sample after combining 6 µL of PCR products from the microsatellite multiplex with 6 µL of PCR products from the sex marker multiplex for each caribou sample ID. DNA concentrations were measured using a Qubit 4 Fluorometer (Invitrogen) with the High Sensitivity dsDNA Assay kit. DNA quantification was necessary to dilute each sample to 11.5 µL of 3 ng/µL to achieve the proper input DNA requirements (calculated using NEBioCalculator).

Sample ID	Initial DNA Concentration (ng/uL)	Initial DNA volume needed for 3 ng/µL dilution	Volume of H2O needed for final volume of 11.5 µL
20919	90.40	0.38	11.12
21099	95.40	0.36	11.14
21101	91.00	0.38	11.12
21102	108.00	0.32	11.18
21104	65.00	0.53	10.97
21105	81.00	0.43	11.07
21106	86.00	0.40	11.10
21107	78.60	0.44	11.06
21110	67.80	0.51	10.99
21112	82.60	0.42	11.08
21113	81.20	0.42	11.08
21297	65.60	0.53	10.97
21298	44.60	0.77	10.73
21299	74.00	0.47	11.03
21301	73.60	0.47	11.03
21303	60.00	0.58	10.93
21305	58.20	0.59	10.91
21306	66.00	0.52	10.98
21309	85.80	0.40	11.10

Table A4 Microsatellite loci information used in the input loci file for the Seq2Sat/SatAnalyzer software to perform microsatellite genotyping. The microsatellite loci file contains the locus names, forward primer sequences, reverse complementary reverse primer sequences, forward flanking regions, reverse flanking regions, microsatellite repeat units, number of repeats, and MRA regions. The loci file was saved as a tab separated .txt file.

Locus	Forward Primer Sequence	Reverse Complementary Reverse Primer Sequence	Forward Flanking Region	Reverse Flanking Region	Microsatellite Repeat Unit	Number of Repeats	MRA Region
MAP2C	ATTTACCAGACAGTTTAGTTTTGAGAC	TCCTGATGTTGGCAAGGACG	TTTAATCAGAAAGTAA	TCTAAGTGGTGTCAGACAGAATCTTCAAGACCTGCCTAATCAGATACAG	GT	19	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
NVHRT48	CCGTGAATCTTAACCAGGTCTCA	ACTGATGGGGAGTTTTCGGT	TGTGTATGTA	TATTGAGACAAAAAACTATTGTTTCTAAATGAAGCTGACCACATTTGGGTTCATGAAGACTTCAGGAGACAA	TG	18	TGTGTGTGTGTGTGTATGTGTGTGTGTGTGTGTGTGTG
OHEQ	CAGGAGCTGTATGTCAGGAACT	ACCAAAGAAGACCTTCACTGACA	GGGGTCAAAGACTAAATATTAGCAGGAGTCAG	ATCATCATCTATTAATATCTATCATCTACCTAATTT	TATC	10	TATCTATCTATCTATCTATCTATCTATCTATCTATC
RT24	TTTTTGGAGCAGTTTTCACTTTGC	TCTGCAATTTGTTTTTCTGACTTAGTA	AGTATTTGTTTAATATATCTCAACAATACACATTATTGTATACTTCTAAGCTCTTTCTCTCTTCCT	TCACACTCACTCT	AC	10	ACACACACACACACACACAC
RT27	ACACAGCAAAAGCATTTACTTGTG	CAACTGTTGGGTCTTTGGTTTTC	T	CCCTTCCAGAAGACCCTTTTGTCTTCTGCAATAAAGAGTCTAGGCTTGCTTGATTGGCTC	CA	21	GACAGATACACACACACACACACACACACACACACACA
RT6	GGCAACAAGACTGCAATTAGCT	GCCAGCCTCCAAGAATGAGT	CATCATTCTTCTTCACTTTACATGTTGACATTAACTGGGAACTCTGTCTAATGTCGGATTTTGAGACTGTTACC	AGCATGAGTAAGACTGCCCCT	TG	21	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
RT7	CTCCCTACATGCCACAGTCC	CCTCCCCTGGTCTCTGCTAA	TGATCTCTTTGCCCTGTTCTACTCTTCTTCTCATGTAGCTTCTAACAT	GTAATTTGCTTAGCTATTACTTTGGCTATGTTTATCTTCTGTTCCTTCTCT	AC|AT	22	ACACACACACACACACACACACACACACACACACACAT
IGF	GAGGGTATTGCTAGCCAGCTG	AGGTTCAAGTTATGCAGAAAAAATATG	GTGTTATTTAGAATACACAAAAATGGGGGAAAGAAAATGCACTCACGTGC	NA	AC	15	ACACACACACAAATACACACACACACACAC

Table A5 *Zfx*/*Zfy* loci information contained in the sex loci file for the Seq2Sat/SatAnalyzer software to assign putative sex for each caribou sample ID. The sex loci file contains the sex locus name, forward primer sequence, reverse complementary reverse primer sequence, X locus sequence, and Y locus sequence. The loci file was saved as a tab separated .txt file.

Sex Locus	ZFXY
Forward Primer Sequence	GGAAATCATTCATGAATATCAC
Reverse Complementary Reverse Primer Sequence	GTACTGTCTGGAATCAGGTCT
X Locus Sequence	TGAATTCTTAAAATTATATTTTTAAATTCAATACACAAAAACTCTATGTGG TCTAGCAGCTAAAATGCCATCACAACACCTTTAAGGATACATACTAGAGT TTCATCTGAGAGCTCACAAAGCACGCTGCGCTGTGGAACTCGTATGCCCTC ACCTGTTTG
Y Locus Sequence	TTAATTCTTAAAAGTACACAAAAACTGCATGTATTCTAACAACTAAAATG CCATCACACCTTTATGGAATATATACTGGAATTTCCTCTGAGAGCTCGCAA AGCATGCTGTGCTGTGGAACTCGTGTGCCCTCACCTGTTTG

Table A6 Sample file information for Seq2Sat/SatAnalyzer software to perform microsatellite genotyping and sex identification in each caribou sample based on MinION Mk1B sequencing data. The sample loci file contains a list of sample IDs and their corresponding FASTQ file names. Each sample has a unique FASTQ file. The fourth column is a placeholder column that is reserved for future use (not currently used by Seq2Sat). The loci file was saved as a tab separated .txt file.

Sample ID	Fastq File Name	Placeholder
20919	20919.fastq.gz	contrl
21099	21099.fastq.gz	contrl
21101	21101.fastq.gz	contrl
21102	21102.fastq.gz	contrl
21104	21104.fastq.gz	contrl
21105	21105.fastq.gz	contrl
21106	21106.fastq.gz	contrl
21107	21107.fastq.gz	contrl
21110	21110.fastq.gz	contrl
21112	21112.fastq.gz	contrl
21113	21113.fastq.gz	contrl
21297	21297.fastq.gz	contrl
21298	21298.fastq.gz	contrl
21299	21299.fastq.gz	contrl
21301	21301.fastq.gz	contrl
21303	21303.fastq.gz	contrl
21305	21305.fastq.gz	contrl
21306	21306.fastq.gz	contrl
21309	21309.fastq.gz	contrl

Table A7 Sample file information for Seq2Sat/SatAnalyzer software to perform microsatellite genotyping and sex identification in each caribou sample based on Illumina Mk1B sequencing data. The sample loci file contains a list of sample IDs and their corresponding FASTQ file names. Each sample has two unique FASTQ files, one for the first read and one for the second read data obtained from Illumina paired-end sequencing. The fourth column is a placeholder column that is reserved for future use (not currently used by Seq2Sat). The loci file was saved as a tab separated .txt file.

Sample ID	FASTQ File Read 1	FASTQ File Read 2	Placeholder
20919	20919_S302_L001_R1_001.fastq.gz	20919_S302_L001_R2_001.fastq.gz	contrl
21099	21099_S281_L001_R1_001.fastq.gz	21099_S281_L001_R2_001.fastq.gz	contrl
21101	21101_S282_L001_R1_001.fastq.gz	21101_S282_L001_R2_001.fastq.gz	contrl
21102	21102_S283_L001_R1_001.fastq.gz	21102_S283_L001_R2_001.fastq.gz	contrl
21104	21104_S284_L001_R1_001.fastq.gz	21104_S284_L001_R2_001.fastq.gz	contrl
21105	21105_S285_L001_R1_001.fastq.gz	21105_S285_L001_R2_001.fastq.gz	contrl
21106	21106_S286_L001_R1_001.fastq.gz	21106_S286_L001_R2_001.fastq.gz	contrl
21107	21107_S287_L001_R1_001.fastq.gz	21107_S287_L001_R2_001.fastq.gz	contrl
21110	21110_S290_L001_R1_001.fastq.gz	21110_S290_L001_R2_001.fastq.gz	contrl
21112	21112_S292_L001_R1_001.fastq.gz	21112_S292_L001_R2_001.fastq.gz	contrl
21113	21113_S293_L001_R1_001.fastq.gz	21113_S293_L001_R2_001.fastq.gz	contrl
21297	21297_S295_L001_R1_001.fastq.gz	21297_S295_L001_R2_001.fastq.gz	contrl
21298	21298_S296_L001_R1_001.fastq.gz	21298_S296_L001_R2_001.fastq.gz	contrl
21299	21299_S297_L001_R1_001.fastq.gz	21299_S297_L001_R2_001.fastq.gz	contrl
21301	21301_S298_L001_R1_001.fastq.gz	21301_S298_L001_R2_001.fastq.gz	contrl
21303	21303_S300_L001_R1_001.fastq.gz	21303_S300_L001_R2_001.fastq.gz	contrl
21305	21305_S303_L001_R1_001.fastq.gz	21305_S303_L001_R2_001.fastq.gz	contrl
21306	21306_S304_L001_R1_001.fastq.gz	21306_S304_L001_R2_001.fastq.gz	contrl
21309	21309_S306_L001_R1_001.fastq.gz	21309_S306_L001_R2_001.fastq.gz	contrl

Table A8 Microsatellite genotypes for the two primary alleles at each microsatellite locus (8) for 19 caribou samples sequenced with a MinION Mk1B sequencer. Inconclusive allele genotypes are recorded as -99.

Sample	IGF_allele1	IGF_allele2	MAP2C_allele1	MAP2C_allele2	NVHRT48_allele1	NVHRT48_allele2	OHEQ_allele1	OHEQ_allele2	RT24_allele1	RT24_allele2	RT27_allele1	RT27_allele2	RT6_allele1	RT6_allele2	RT7_allele1	RT7_allele2
20919	81	83	93	101	118	118	108	112	113	127	87	103	125	139	127	135
21099	81	81	93	107	118	118	112	143	105	111	93	107	135	141	135	135
21101	81	81	99	107	118	120	139	143	105	113	103	107	139	139	137	143
21102	81	81	93	107	118	118	108	108	113	119	87	87	129	129	127	137
21104	81	83	89	101	118	120	108	143	111	113	87	87	139	139	127	137
21105	81	81	89	101	120	126	108	139	113	119	87	87	139	139	137	143
21106	81	81	89	101	118	120	108	147	111	113	103	107	139	139	127	137
21107	81	81	89	93	118	118	135	139	111	117	87	99	139	139	127	127
21110	81	81	93	93	118	118	108	112	111	113	87	89	139	139	135	143
21112	81	81	89	93	118	118	108	139	113	127	87	89	139	139	135	135
21113	81	81	89	93	118	118	108	143	113	113	87	103	125	133	127	135
21297	81	81	93	101	118	118	108	135	111	113	103	103	127	127	135	135
21298	81	81	-99	-99	118	118	131	139	111	113	101	103	137	139	135	145
21299	81	81	93	93	118	118	127	131	113	131	99	103	127	129	127	135
21301	81	81	89	101	118	120	131	135	113	113	87	103	139	139	127	127
21303	81	81	93	101	118	126	108	127	113	115	103	103	127	139	127	135
21305	81	81	93	99	118	126	108	131	113	113	87	87	139	139	127	135
21306	81	81	99	99	118	118	131	131	113	113	87	87	125	127	135	139
21309	81	81	101	101	118	118	108	135	113	115	87	99	139	139	127	135

Table A9 Microsatellite genotypes for the two primary alleles at each microsatellite locus (8) for 19 caribou samples sequenced with an Illumina MiSeq sequencer. Inconclusive allele genotypes are recorded as -99.

Sample	IGF_allele1	IGF_allele2	MAP2C_allele1	MAP2C_allele2	NVHRT48_allele1	NVHRT48_allele2	OHEQ_allele1	OHEQ_allele2	RT24_allele1	RT24_allele2	RT27_allele1	RT27_allele2	RT6_allele1	RT6_allele2	RT7_allele1	RT7_allele2
20919	81	83	93	101	118	118	108	112	113	127	87	103	125	139	127	135
21099	81	81	93	107	118	118	112	143	105	111	93	107	135	141	135	135
21101	81	81	99	107	118	120	-99	-99	105	113	103	107	139	139	137	143
21102	81	81	93	107	118	118	108	108	113	119	87	87	129	129	127	137
21104	81	83	89	101	118	120	108	143	111	113	87	87	139	139	127	137
21105	81	81	89	101	120	126	108	139	113	119	87	87	139	139	137	143
21106	81	81	89	101	118	120	108	147	111	113	103	107	139	139	127	137
21107	81	81	89	93	118	118	135	139	111	117	87	99	139	139	127	127
21110	81	81	93	93	118	118	108	112	111	113	87	89	139	139	135	143
21112	81	81	89	93	118	118	108	139	113	127	87	89	139	139	135	135
21113	81	81	89	93	118	118	108	143	113	113	87	103	125	133	127	135
21297	81	81	93	101	118	118	108	135	111	113	103	103	127	127	135	135
21298	81	81	93	101	118	118	131	139	111	113	101	103	137	139	135	145
21299	81	81	93	93	118	118	127	131	113	131	99	103	127	129	127	135
21301	81	81	89	101	118	120	131	135	113	113	87	103	139	139	127	127
21303	81	81	93	101	118	126	108	127	113	115	103	103	127	139	127	135
21305	81	81	93	99	118	126	108	131	113	113	87	87	139	139	127	135
21306	81	81	99	99	118	118	131	131	113	113	87	87	125	127	135	139
21309	81	81	101	101	118	118	108	135	113	115	87	99	139	139	127	135

Table A10 Total cost of MinION-specific expenses and additional expenses incurred from sequencing 11 loci in 19 samples on a MinION Mk1B in this study. Prices are shown in $CAD.

MinION-Specific Expenses							
Item	**Amount Item Comes With**	**Cost of Item**	**Cost Per Amount**	**Amount Needed for 19 Samples**	**Cost for 19 Samples**	**Cost per Sample**	**Cost per Locus**
Flow Cell R10.4.1	1 Flow Cell	$956.84	$956.84	1 Flow Cell	$956.84	$50.36	$4.58
Qubit™ Reagents	500 Samples	$492.00	$0.98	1 DNA Library	$0.98	$0.05	Negligible
Native Barcoding Kit 24 v14 (SQK-NBD114.24)	188 Barcodes (6 sets of 24)	$959.95	$5.11	19 Barcodes	$97.02	$5.11	$0.46
NEBNext Ultra II End Repair / dA-tailing Module (NEB, cat # E7546L)	672 µL Buffer, 288 µL Enzyme Mix	$1,211.00	$1.80	33.25 µL Buffer, 14.25 µL Enzyme Mix	$59.92	$3.15	$0.29
NEB Blunt/TA Ligase Master Mix (NEB, cat # M0367L)	1,250 µL	$644.00	$0.52	95 µL	$48.94	$2.58	$0.23
Quick T4 DNA Ligase in NEBNext® Quick Ligation Module (NEB, cat # E6056L)	500 µL Ligase, 1000 µL Buffer	$2,116.00	$4.23	5 µL Ligase, 10 µL Buffer	$21.16	$1.11	$0.10
Subtotal					**$1,184.86**	**$62.36**	**$5.67**
Additional Expenses							
Item	**Amount Item Comes With**	**Cost of Item**	**Cost Per Amount**	**Amount Needed for 19 Samples**	**Cost for 19 Samples**	**Cost per Sample**	**Cost per Locus**
Qiagen Type-It PCR Master Mix 2x for All Reactions	25,000 µL	$2,697.82	$0.11	480 µL	$51.80	$2.73	$0.25
IDT Standard Primers for All Reactions	2,500 µL	$22.00	$0.01	64 µL	$0.56	$0.03	Negligible
Qubit™ Reagents	500 Samples	$492.00	$0.98	19 Samples	$18.70	$0.98	$0.09
Subtotal					**$71.06**	**$3.74**	**$0.34**
TOTAL					**$1,255.92**	**$66.10**	**$6.01**

Table A11 Cost estimate of MinION-specific expenses and additional expenses incurred when sequencing the full panel of microsatellites (26 loci) in 96 samples on a MinION Mk1B. Prices are shown in $CAD.

MinION-Specific Expenses							
Item	**Amount Item Comes With**	**Cost of Item**	**Cost Per Amount**	**Amount Needed for 96 Samples**	**Cost for 96 Samples**	**Cost per Sample**	**Cost per Locus**
Flow Cell R10.4.1	1 Flow Cell	$956.84	$956.84	1 Flow Cell	$956.84	$9.97	$0.38
Qubit™ Reagents	500 Samples	$492	$0.98	1 DNA Library	$0.98	$0.01	Negligible
Native Barcoding Kit 96 v14 (SQK-NBD114.96)	288 Barcodes (3 sets of 96)	$1,092.17	$3.79	96 Barcodes	$364.06	$3.79	$0.15
NEBNext Ultra II End Repair / dA-tailing Module (NEB, cat # E7546L)	672 µL Buffer, 288 µL Enzyme Mix	$1,211	$1.80	168 µL Buffer, 72 µL Enzyme Mix	$302.75	$3.15	$0.12
NEB Blunt/TA Ligase Master Mix (NEB, cat # M0367L)	1,250 µL	$644	$0.52	480 µL	$247.30	$2.58	$0.10
Quick T4 DNA Ligase in NEBNext® Quick Ligation Module (NEB, cat # E6056L)	500 µL Ligase, 1000 µL Buffer	$2,116	$4.23	5 µL Ligase, 10 µL Buffer	$21.16	$0.22	$0.01
Subtotal					**$1,893.09**	**$19.72**	**$0.76**
Additional Expenses							
Item	**Amount Item Comes With**	**Cost of Item**	**Cost Per Amount**	**Amount Needed for 96 Samples**	**Cost for 96 Samples**	**Cost per Sample**	**Cost per Locus**
Qiagen Type-It PCR Master Mix 2x for All Reactions	25,000 µL	$2,697.82	$0.11	2,915 µL	$314.57	$3.28	$0.13
IDT Standard Primers for All Reactions	2,500 µL	$22.00	$0.01	695.48 µL	$6.12	$0.06	Negligible
Qubit™ Reagents	500 Samples			96 Samples	$94.46	$0.98	$0.04

		$492.00	$0.98				
Subtotal					**$178.91**	**$1.86**	**$0.17**
TOTAL					**$2,308.24**	**$24.04**	**$0.92**

Table A12 Cost estimate of expenses incurred when sequencing the full panel of microsatellites (26 loci) in 96 samples on an Illumina MiSeq. Prices are shown in $CAD.

Illumina MiSeq Expenses							
Item	**Amount Item Comes With**	**Cost of Item**	**Cost Per Amount**	**Amount Needed for 96 Samples**	**Cost for 96 Samples**	**Cost per Sample**	**Cost per Locus**
Qiagen Type-It PCR Master Mix 2x for All Reactions	25,000 µL	$2,697.82	$0.11	4,015 µL	$433.27	$4.51	$0.17
IDT Standard Primers for All Reactions	2,500 µL	$22.00	$0.01	695.48µL	$6.12	$0.06	Negligible
AMPure Beads for All Reactions	60,000 µL	$2,031.00	$0.03	2,155 µL	$72.95	$0.76	$0.03
Indexes for All Reactions	1,152 Samples (3 sets of 384)	$3,012.00	$2.61	96 Samples	$251.00	$2.61	$0.10
Qubit™ Reagents	500 Samples	$492.00	$0.98	1 DNA Library	$0.98	$0.01	Negligible
Illumina V2 500 Kit	384 Samples	$2,211.00	$2,211.00	1 Kit	$2,211.00	$23.03	$0.89
TOTAL					**$2,975.32**	**$30.99**	**$1.19**

Table A13 Cost estimate of expenses incurred when sequencing the full panel of microsatellites (26 loci) in 384 samples on an Illumina MiSeq. Prices are shown in $CAD.

Illumina MiSeq Expenses							
Item	**Amount Item Comes With**	**Cost of Item**	**Cost Per Amount**	**Amount Needed for 384 Samples**	**Cost for 384 Samples**	**Cost per Sample**	**Cost per Locus**
Qiagen Type-It PCR Master Mix 2x for All Reactions	25,000 µL	$2,697.82	$0.11	16,060 µL	$1,733.08	$4.51	$0.17
IDT Standard Primers for All Reactions	2,500 µL	$22.00	$0.01	2,781µL	$24.47	$0.06	Negligible
AMPure Beads for All Reactions	60,000 µL	$2,031.00	$0.03	8,620 µL	$291.79	$0.76	$0.03
Indexes for All Reactions	1,152 Samples (3 sets of 384)	$3,012.00	$2.61	384 Samples	$1,004.00	$2.61	$0.10
Qubit™ Reagents	500 Samples	$492.00	$0.98	4 DNA Libraries	$3.94	$0.01	Negligible
Illumina V2 500 Kit	384 Samples	$2,211.00	$2,211.00	1 Kit	$2,211.00	$5.76	$0.22
TOTAL					**$5,268.28**	**$13.72**	**$0.53**

APPENDIX B

Table B1 Microhaplotype loci details for the optimized panel of 25 microhaplotypes sequenced with a MinION Mk1B and an Illumina MiSeq system. Column 1 represents the chromosome number followed by the gene name where each microhaplotype is located. Column 2 refers to the size of the microhaplotype region in base pairs (bp). Column 3 is the position of the microhaplotype region. Columns 4 and 5 provide the forward and reverse primer sequences, respectively.

CHR_GENE	SIZE (bp)	POSITION	FWD-PRIMER	REV-PRIMER
1_Gimap6	317	100,067,401	GAGAAGCCAAGCCTCCTGAG	ATGGGGCTCCCTTGCTGC
1_Rsl24d1	310	41,051,232	GAAGCGCAATCCTCACAAGG	CAGCTGCTTTCCTTTGCCTG
2_Cd5l	300	95,362,995	AACGGAGACACCAACTGCTC	CCTGAGCAGAATACGACCACA
2_Lexm	330	25,798,100	CCTGGCAAAAGTCACGGGC	AGACAGCCTCCTGGGGTTC
2_Vcam1	345	68,739,254	GCTTTCTCTAAGGATCCAGAGAT	CGCAGACAGACACAAGCACT
3_Nudt12	307	6,835,427	TCAATCACCGATACAAGAAAGCA	GACACTCTTCAGGTGGCACA
3_Slc34a1	234	65,180,443	AGACCCTGATCAGTCCCGTC	GGGCTAGAGTGTTCTGCCAA
4_Abo	288	2,378,887	GACGCCAAGTCCTCCTGGAA	TTTGCTGGCAAACACGGC
4_Pgap2	333	77,212,499	AATGGGAGCCTGGTCCGG	TCTCCACGAGGCTGAGGC
5_Fbxo25	308	87,897,110	AGACACTTAACACGTGGGGC	GCAGGATGTTGCTGAGCATG
11_Oaf	286	25,961,589	CGCCTTTTCTCCCAGGTGT	CCGGGCTAGGGATCCTCATC
11_Or51g1	236	45,483,005	TATTCTCCGTACTGTGCTGGC	TCTGCTGGGTCTTGATGCTT
13_Fgb	290	2,334,196	ACGAGACTGTTAGAATACTACGTC	GGACAGAGCCATCCGTGTAC
15_Serpina4	238	51,128,108	TGGTCATTTCCCTTTCAGAGCT	TAACAGACAAGCCAGCCCTG
15_Stoml1	299	29,443,785	GCCTCTCCTAACAGTTGGCC	GTGCTTCGCTAAAGTGCCC
17_Shb	258	53,939,203	CCAGTGTACTCCAAACCCCC	ACCCGTTATGATCGTCAGGC
18_Eme1	211	55,128,468	CAGGTCATTTACTCCTTTCAGGC	CTGAGGGCTGGCTTCAGTC
18_Krt16	204	50,908,493	GGGGTAAGTTGAGCACCTCC	TCCCACCTTGCTCAGAAACC
18_Rnft1	237	9,608,274	TGCTGAAAATCCAGGCTCCA	TGCTGCATAACAAGTTTGACACC
21_Dpm3	296	15,296,613	GATGGAAGAGCAGGCTGTCC	AATTAACAAGGGGCGGTGGG
26_Cip2a	186	45,844,748	CTGTTCAGTGACTCAGCTGC	TCCTTGAACAACTCCAGTGC
27_Abt1	393	19,331,235	CCCATCTCAGCGAGCATCTT	TGGGCTAGGAGGTGGAAGG
30_Htra3	236	2,384,767	TCAGAATTGCCCATGACCCC	GACGCCACAAAACCCCAAAA
32_Gpm6a	263	5,496,749	ATGGTGGCAGCAGAATCCTC	GCCCCACTCTTCTGTGAGAC
20_Scn11A	394	44,057,025	AAACCAAAGTCCAGGTAGCCC	CACGGCAACTCCAGCTTTAA

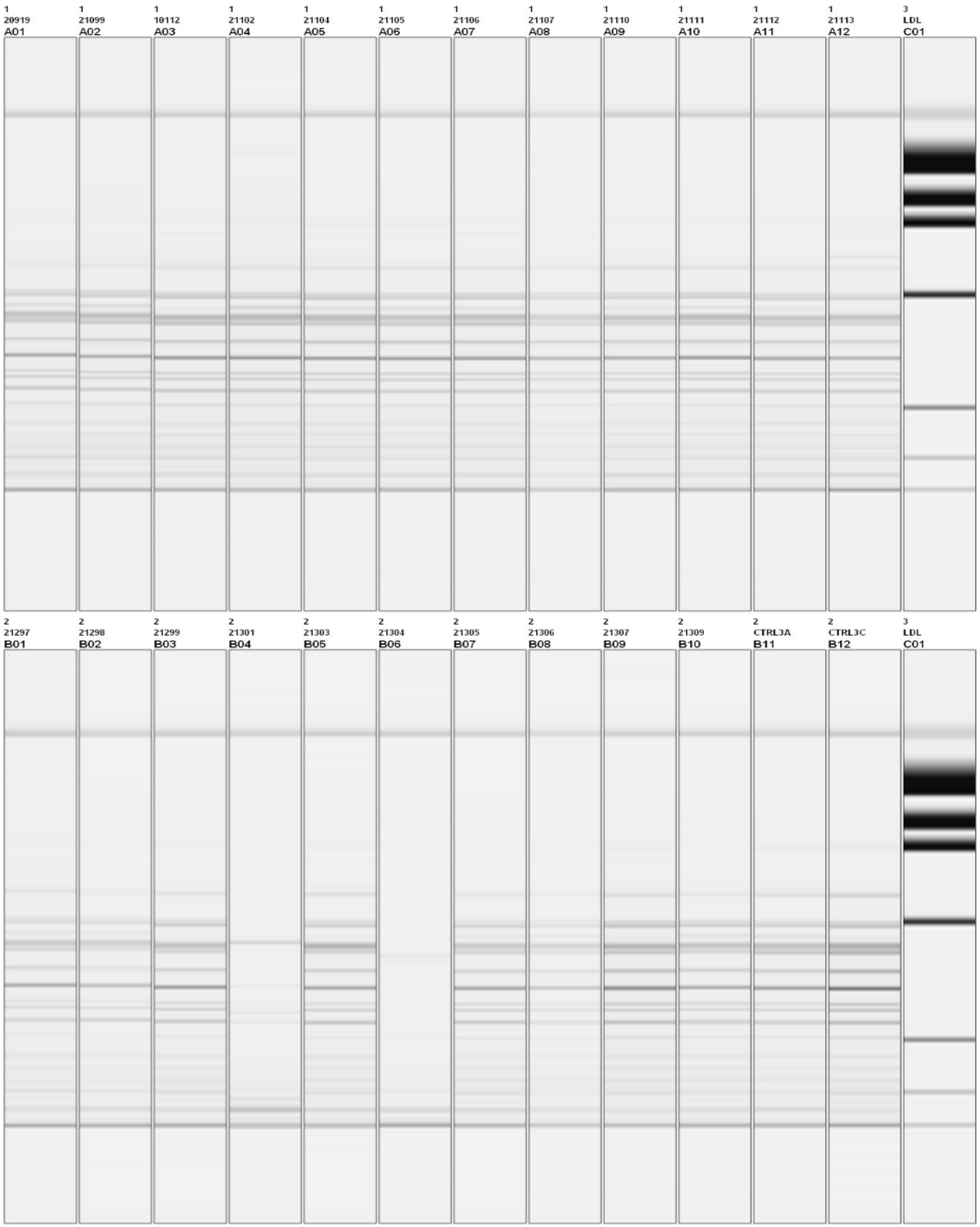


Figure B1 Capillary electrophoresis images of MinION Mk1B pooled PCR-amplified microhaplotype amplicon targets in each caribou sample ID generated from a QIAxcel Advanced System by Qiagen.

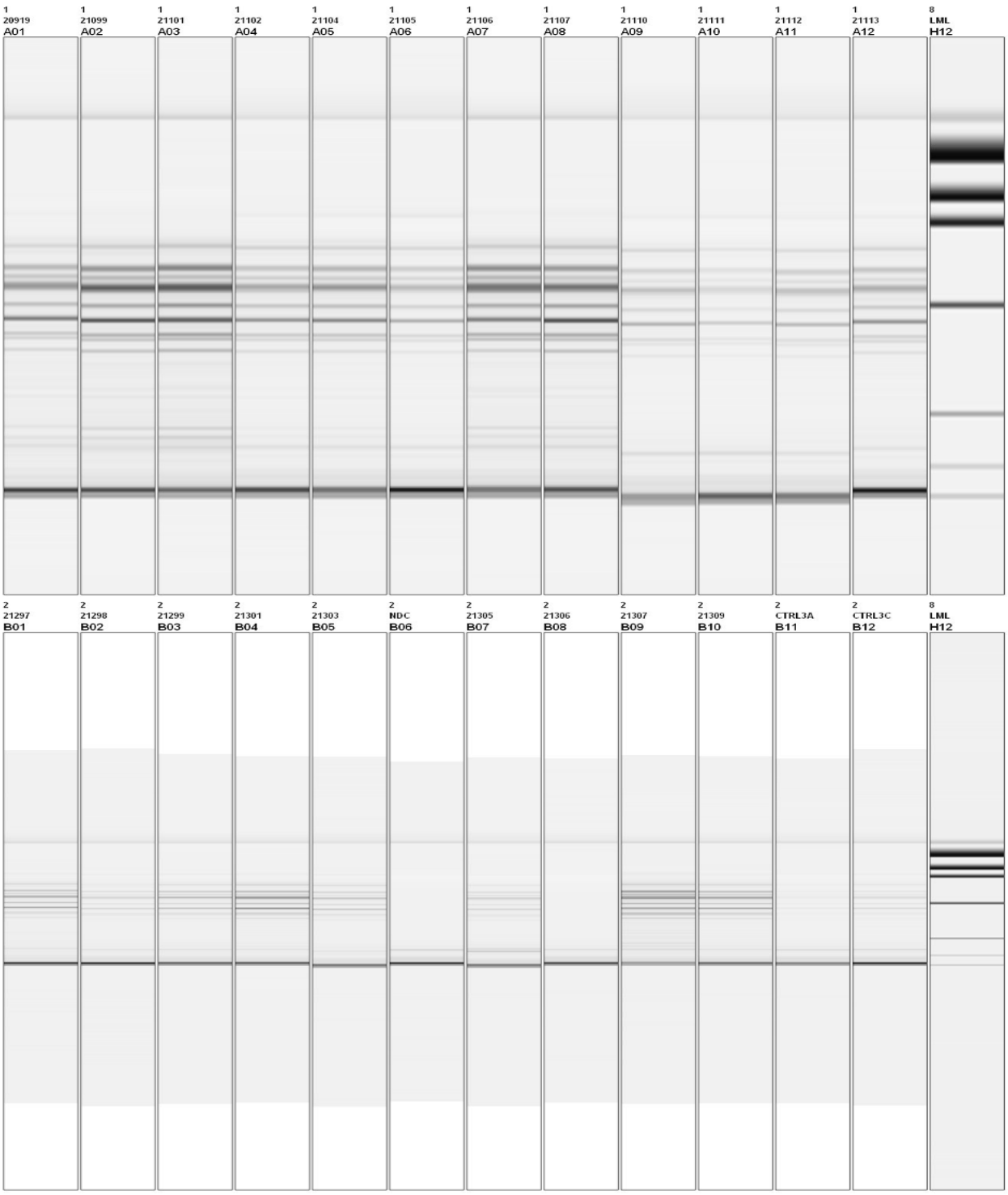


Figure B2 Capillary electrophoresis images of Illumina MiSeq pooled PCR-amplified microhaplotype amplicon targets in each caribou sample ID generated from a QIAxcel Advanced System by Qiagen.

Table B2 Microhaplotype loci information used to run Seq2Sat/SatAnalyzer genotyping software on raw FASTQ files generated from MinION Mk1B and Illumina MiSeq microhaplotype sequencing. Column 1 is the microhaplotype locus name. Columns 2 and 3 are the forward and reverse primer sequences, respectively. Columns 4 and 5 represent the trimmed length for the 5' and 3' end, respectively. Column 6 contains the position of known SNPs from the reference sequence. Column 7 provides the full reference sequence for each microhaplotype locus.

Locus Name	F-Primer Seq	R-Primer Seq	5' End Trimmed Length	3' End Trimmed Length	SNPs Positions	Reference Sequence
1_Gimap6	GAGAAGCCAAGCCTCCTGAG	ATGGGGCTCCCTTGCTGC	0	0	15|16|21|22|27|29|33|36|49|56|75|100|101|109|141|208|232|234|259	GCGGGCCTGTGCTCCCTTCTGAGGTGGAAGCCCTGCAGCATCCCGGCCACTGGAGGCTCGGCTTATAACTCGATGCAGATCAGCGTGGAATGTGGACCGGGCAACTCTGCGTCTCTTGGGCTGTGTCCTTGCCCGTTAAGAGGGCAGGCACGATGGCTTTTAGGGCCCAGCGCCATGTGGGGCAGGACAGGCTCTGGGCTGGGGAGGAGAAGGCAACTAGAAACGGGGGAGGGCCGCAAGGAGCTGCTCTTGATTCGCTCTTGAGAGAGGCCACCACCC
1_Rsl24d1	GAAGCGCAATCCTCACAAGG	CAGCTGCTTTCCTTTGCCTG	0	0	21|31|102|109|121|223	TCAGATGGACTAAAGCGTTCCAGAAAGCAGCTGGTAAAGAGCTCACGGTGGATAATTCATTTGAATTTGAAAAACATAGAAATGAACCTGTCAAATACCAGCAAGAACTATGGAATAAAACCATTGATGCAATGAAGAGAGTTGAAGAGATCAAACAGAAGCGACAAGCTAAATTTATAATGAACAGGTTGAAGAAAAATAAAGAACTACAGAAAGTTCAGGACGTTAAAGTGGTCAAGCAAAACATCCATCTTATCCGGGCCCCTCTTG
2_Cd5l	AACGGAGACACCAACTGCTC	CCTGAGCAGAATACGACCACA	0	0	5|21|42|56|75|86|90|96|105|111|127|143|156|157|166|175|177|188|192|193|196|198|201|202	TGGGCGACTGGAGGTGCTGCACAAGGGCACATGGGGCTCTGTCTGTGATGACGGCTGGGGGACAAAAGAGGAGCAGGTGGTGTGCCGGCAACTGGGCTGTAGAAGGCCAATCTTTGTACCTGCCAAATTGCGGAGAAAGTTTGCCCCTGGAGATGGGCGCATCTGGCTAGATGATGTTCATTGCAAAGGGCATGAGGAGTCCTTGGAGCAGTGTCAGCATAGGTCCTGGGGGTACCATGACTGCAACCACAAGGAAGA
2_Lexm	CCTGGCAAAAGTCACGGGC	AGACAGCCTCCTGGGGTTC	0	0	4|82|119|164|180	ACCCGTGGCCCCTACGACATCTTCTCTGGTGACCGAAGCAAGCCCCAGCCTTACGGACATTACTCTGTGCAGGTGTGTACGCAGACCGGCTGTGCCTTCCCCGGCAGGAAGACCAGCCTCGCTCATCTACAGCACCTTGTGTGCCAGGCTCTGCGGGGGGAGGTGGGGGAACAGGGGGCCGTTCCCATATCCTCCTCACTCCCCACTCCCAGCAAGGGAGGGATGAGGCTCTTCGTGCTAGATGGGGAGACTGATGCTAGCCATGCAGCTGGAAGAGGCAGAGCCGGAATTT
2_Vcam1	GCTTTCTCTAAGGATCCAGAGAT	CGCAGACAGACACAAGCACT	0	0	14|50|75|85|104|162|252|274|291	TCATTTGGGTAGCCCCCCAGAGGTCGGGAAGCCAGTCACAGTCACGTGTTCGGTGCCTGATGTTTACCCATTTGAGAGGCTGGAGATAGAGTTGTTCAAAGGGAACCGTTCCATGAAAATACAGGAATTTCTGGAGCCTTCAGAAAAAAAGGCCCAGGAAACAAAGAGCTTGGAAGTGACCTTCTCGCCTACTGATGAGGATATTGGGAAAGCTCTAGTTTGTCAAGCTACATTACACATTTATGATGATGACTCTCCACCCAAAGTGAGGAAAGCTACAAAAGAACTGGAAGTCTACAGTA
3_Nudt12	TCAATCACCGATACAAGAAAGCA	GACACTCTTCAGGTGGCACA	0	0	80|108|165|204|214|221|231	ATTGTAAATGAGTTTTTTCCTATTTTCCTCTTTTAAGGAAGAGATGTCTTCTGTAAAAAGAAGTCTGAATCAAGAAATAATATCCCAGTTTCACTCTTCAGCTGCAGACGGAGATATTTGCCAAGTTAACAGCAATACTCAGTCATTCTCCATCTCTTCTCAATGAAACTTCTGAAAATGGCTGGACTGCTTTAATGTATGCTGCGAGGAATGGGCACCCAGATGTTGTCCAATTTCTACTTGAGAAAGGGTAAACATTTTAGGCT
3_Slc34a1	AGACCCTGATCAGTCCCGTC	GGGCTAGAGTGTTCTGCCAA	0	0	25|50|94|104|110|146	CCCCAGTCTTGCACAGGATTCCAGGGACCTCAGCCTACGGCTTCCCCAGCATGGGCCCCGTGGCCCTCCCAGAGCGTGGCTGCCCCTACGGGGAGGTTGTGGAGCGCCACGACCCACTGCCTGCCAAGCTGGCCCTGGAGGATGAGCGGAAGCCAGGTGGGCCTGGGTTGGGGTGGCAGGGGTGTTGGCGGTCA
4_Abo	GACGCCAAGTCCTCCTGGAA	TTTGCTGGCAAACACGGC	0	0	25|32|50|92|103|108|111|240	GGACAGTCAGCCCGTGGTGGGCACCGGGGGGGACGGAGGCAGAGGGGTCCCACTGGGATGTGGGCGCCAGGGTCCCGGTGTGTCCGCCCTCCTGCCAGCTCACACGGCATCACGGACGGGCTTTCTCGCCCAGCAGCAGGACTGATGTCCTAACCCTGGCCCCTTGGCTGGCGCCCATGGTCTGGGAGGGGACCTTCAACATCGACATCCCGAATGTGCCGTTCCGGCTCCAGAGCGCCACCGTCGGGCTG
4_Pgap2	AATGGGAGCCTGGTCCGG	TCTCCACGAGGCTGAGGC	0	0	40|187|198|229|237|257|276	CCCCACTTCGCCCTGCTGGCCCGGTCACGGTCTGCAGCCCACTGTCGCCTTCCTCTTCTGTGTACTCGTCTCTGCTCTTCACTTCAAGGAGATTCGGGCCACATACTGCGGCGTGCCCATTACCTTCCGTCAGTGAGCTCAGCCATCGGCCGAGAGGTGCCCCCGCCGTACGGGCGGTGCTTCTACATCGGCCTGCACTCGGCGCCCCACTTCTTGGTGGCCTTCGCCTACTGGAGCCACTACTTCAGCTGTGCCTCCCGGGTCCGCACTAGTGCCCGCTTTCCCACCTCCGCTCCG
5_Fbxo25	AGACACTTAACACGTGGGGC	GCAGGATGTTGCTGAGCATG	0	0	18|50|128|180|233	GTGTCTGCTTTCCAGACCTGGCCCAGGTGCCACTGTCCCCGAGTCCCCAATGTGCAGTAACGCCAGGACACCCAGCGCACTCGGGGGCTAGGGTGTCGGTGTCTGGGGCCCCAGGTGCAGCCAGGGCCACCGCAGGGTGTCCTCTGGGAGGAGTGGGGAGGCAGCAGACGTCGCTGTGCCCGACGCGCTCCTGCGTTGCCGGTGGAGTGAGTGTGGGGCCCTCCCAGCAGGTGGACAGCGGCCTGACGCTCAGCGACCTCCCTGTGCACA
11_Oaf	CGCCTTTTCTCCCAGGTGT	CCGGGCTAGGGATCCTCATC	0	0	10|91|152|240	GGATGGCTCCGTGTTCGAGGCTCTCCCCAAGGCCTCGGAGCAGGTGTTGCTGCCTCGCTGCGGGCAGGTGGGGGACCGCGGGAAGCCCTGCTGTCTGCCGCTATGGTCTGAGCCTCGCCTGGTACCCCTGCATGCTCAAGTACTGCCACGGCCGCGACCGCCCAGCGCCCTACAAGTGCGGCATCCGCAGCTGCCAGAAGAACTATAGCTTCGACTTCTACGTGCCCCAGAAGCAGCTGTGTCTCTGG
11_Or51g1	TATTCTCCGTACTGTGCTGGC	TCTGCTGGGTCTTGATGCTT	0	0	5|13|58|61|67|68|108|153|154	CTTTGCCTCTGCTCGGGAGAGGCTCAAGGCCTTCAACACTTGTGTCTCCCACCTCCTGGCTATACTCCGCTTCTGTGTGCCTATACTAGGTCTGTCCATCGTACACAGGTTTGGGAGCCACACTTCACTCCTGGTGCACATCCTTATGGGCACCATCTCTGTGCTCTTCCCACCCTTGATGAATCCTGTTATTTAA
13_Fgb	ACGAGACTGTTAGAATACTACGTC	GGACAGAGCCATCCGTGTAC	0	0	6|8|13|24|26|51|56|82|166|168|235	TCAAGTGTTCCACAACACTCTTCATTGTATTTCTGCCTGGTCCCCTGCAGTTACATTTGGATGCCCGTGGTCATCGGCCCTATGACAAGAGGAGGGAAGAGGCTCCCAGCCTGAGACCTGTACCCCCTCCCATCAGCGGAGGTGGCTACCGGGCTCGTCCCCCCAAAGCAGCTCTGGGCCAGAAGAAAGTGGAGAGAAAGCCCCCTGATGCTGATGGCTGCCTGCACGCTGACCCGGACCTGGTGGG
15_Serpina4	TGGTCATTTCCCTTTCAGAGCT	TAACAGACAAGCCAGCCCTG	0	0	2|4|54|58|70|111|115|139|140	TCCGCAAGGCCATCCTGGAGGTGGGTGAAGTTGGCACCCAGGCTGCAGTGGTCACGGGGAGTTCTGTCACCTTTTGGCCCTGGGACAACCGCCAAGCCCTTTGGTTGAACCGGTTCTTCCTTGTGGTGATCTTTTCCACATATGCCCAGAGCATCCTCTTTCTGGGAAAAGTGGTCAACCCCATGAAACCATAGCCC
15_Stoml1	GCCTCTCCTAACAGTTGGCC	GTGCTTCGCTAAAGTGCCC	0	0	14|20|23|50|143|173|185	TCTAAGGATGGGGCTGTGCTATCCGTGGGGGCTGATGTCCAGTTCCGCATCTGGGACCCGGTACTGTCGGTGATGACGGTGAAGGACCTGAACACGGCCACACGCATGACCGCCC

						AGAACGCCATGACCAAGGCCCTGCTCAGGAGGCCCCTGCGAGAGATCCAGATGGAG ACACTCAAGATCAGCGACCAGCTCCTGGTAGGCAGCCCCGCACCGGGCAGAGTTCCG TGGCTGGCACGTGGGCCCAGCTCTGCCTGCCGG
17_Shb	CCAGTGTACT CCAAACCCCC	ACCCGTTATGA TCGTCAGGC	0	0	44|56|61|73|74|91|98|123|135	CAGAGTCTCCAAGCCTTCTGTGACTCCAGCACAGGGAAAGCCCCGAAGCCCAGTGCC CCCCGCCCCAGAGCAGCGCGCACACTCCAGATTACGGCACCTGGTCGGCCTCCAACG GTCAGAGTAAGTTTTGTGCCCGCCTCTGGTGTGGTGGTGACACTGGAGGGCAGGTAG GTGGGCGTGTGGGAAGAGCAGCTTCCTCCCTCATCAGGGACAGG
18_Eme1	CAGGTCATTT ACTCCTTTCA GGC	CTGAGGGCTGG CTTCAGTC	0	0	18|28|47|54|100|115|124|139|162	TTATAAGCAGTGTTTTTCTGAGGAAGAACGCCAGAACTTGCTCGCAAACTTACAGGTG CGCCGTGGGGAAGGCGTGACAGCCACCTCGCGCCGTGTTGGACCAGAGCTGTCCAGG CGTATCTACCTTCAGATGACAGCTTTGCAGCCAGATCTCTCCTTAGACAGTGCA
18_Krt16	GGGGTAAGTT GAGCACCTCC	TCCCACCTTGC TCAGAAACC	0	0	59|73|95|112|140|145|146|154	TCGCATCCCTTCCTAGGAGATGCTTGCCTTGCAGAATCAGACTGGTGGGGATGTCAGC GAGGAGATGGATGCCGCCCCCAGCGTGGACCTGAGCCACATCCTGAATGAGATGCGT GACCAGTATGAGCAGATAGCAGACAGGAACTGCAGGGATGCCGAGGCCT
18_Rnft1	TGCTGAAAAT CCAGGCTCCA	TGCTGCATAAC AAGTTTGACAC C	0	0	50|68|69|72|85|116|161|187	GAATACAAGGTCAGGTGTCCACAGCTGTACGCATGGATGTGTACATAGTCGCTTACG GAGCCACTCCCACAGCGAAGCCAGGCAACCTGATGATAGCGACAGGGAGTATGGAG ATCATGGTAGCAGCTCCTTCTCAGAATTCCGGTATCTCTTCAAGTGGTTGCAGAAGAG TCTTCCATATATTTTGATTCTGG
21_Dpm3	GATGGAAGA GCAGGCTGTC C	AATTAACAAGG GGCGGTGGG	0	0	37|74|91|122|239	AGGTGGGAATGGAGTTAGGGTGCTAGGAACACATCCCTTCGCAGGTCAAGTCAGTTC CAGGGTCCTGGATTTGGTTCTGCAGCTCGAGGGCGGCGTCCTCGCAGCCGTGAAAATT GGCCACGCGGTAGCCCACAGTGCCCAGGGCATAGCAGTCAGCGGACACCAGAAAGT AGATGGGCAGTGGCCACAGGACCTCCTGGCAAGGGGAGGGCGGCGCCAGGCCCAGG GCCGCCGGGTGGCGCCCAGGAGCACCAGCC
26_Cip2a	CTGTTCAGTG ACTCAGCTGC	TCCTTGAACAA CTCCAGTGC	0	0	17|47|79|82|97|99|100|107|116|12 7	GCGCCACGCGCTCACACAGGTGATATTTGAACAACCCCCGTCTGGCAGCGCCGTTCT GGGAAGCCGTTCTAAGTCTTTAGAGCCGACTGTGGCTCTCCTTCGCTGGTCACACCAG CCTCTGGATGGATCAGAAAACTGCTCTGTTTTAG
27_Abt1	CCCATCTCAG CGAGCATCTT	TGGGCTAGGAG GTGGAAGG	0	0	20|23|28|49|57|88|89|109|118|120| 124	GCTTTTGAGCGCCAGGTGCGTAGGCAGCGTCTGAGGGCTGAGGTTGCCCAGGCCAAG CGTGAGACTGACTTCTATCTTCGAAGTGTGGAGCGCGGACAGCGTTTCCTTGCGGCTG ATGGGGACTCTGCCCGCCCGAATGGTTCCTGGGCCTTTGCCCAGCGTCCTACTGAGCA GGAGCTGAGGGCCCGGAAGGCAGCTCGGCCAGGGGGACGTGAACGAGCTCGCCTGG CTAACGCTCAGGACCAGGCCCGCTCCAACCGAGGGCTTCTTGCCAAGATCTTTGGAG CTCCTACACCCTCAGAGAGCAGGGAGGACGCCTCACTGGTCAGGAACTCTTGAGGGT CAGAGAGGCCCC
30_Htra3	TCAGAATTGC CCATGACCCC	GACGCCACAA AACCCCAAAA	0	0	96|113|167|190	TTTCTTGGGTTGGATCGTTTGCTAGAGCAGGAAAGCTGTTTGCTCGAGAGATAACCAG TCTATTGTAAAGCATATGTCTCAGGAACAGCCTGATGAAAGAGGCGCGTCGGGCAAG GTGTGGGGGAAGGGCAGGGAGCCCACGGCCTCTGGTGCCCCTCCACCCGACCGCCAG CCCAGAAGCTCTTCAGACCCATCTT
32_Gpm6a	ATGGTGGCAG CAGAATCCTC	GCCCCACTCTT CTGTGAGAC	0	0	20|90|154|166|195|203	AGCCTCAGAAGATGATTGTTGGGAGCATCTCCTGGGCCCTGGAGCACAGTTGCCATG GTAACTCTGCTGCTCTGCCAAGACTGCTTCTCTGTGTGGGAGGAAAGAAGCTCCCCTG CCAGCAGCTACGTGGGATTCTGGCAGCTGGAGAGAACACCGAAATGGTGCACTGCTT CCCCAGATGGGAGTTCTGAAAGCCGCCTTCACAAGCCAATAGGCAACTCAG
20_Scn11A	AAACCAAAG TCCAGGTAGC CC	CACGGCAACTC CAGCTTTAA	0	0	14|64|142|302|315	TGGATCGTTTCCGCCGGGCTTTTTGTTTTGTTAATGCATACTCTTGAGCATTTTTTCAGG AAGCGGTGCAGGAGGCAACATTTATCAAAGCATAAAGAGGTGACCGAAGGCCCTGG TGGAGAGAGCAAAGATATCATTCCCCTGGTCACAGGGATAAGAAAGGGCCCAGAGA TCTGGGAGGAGTTTGGTGTACTAACTTCTGTACCAATGACCTTGCATGATCGGACTTG GTTGGCCCCACTTGCAGAGGAAGAAGATGATGCTGAATTCCCTGGTGAAGATAAGGC ACAGGCTGTCACACAACCTGAGGCTGGAAAACAGGTATGCAGGTTCATACACAGACT CAGAGATCGT

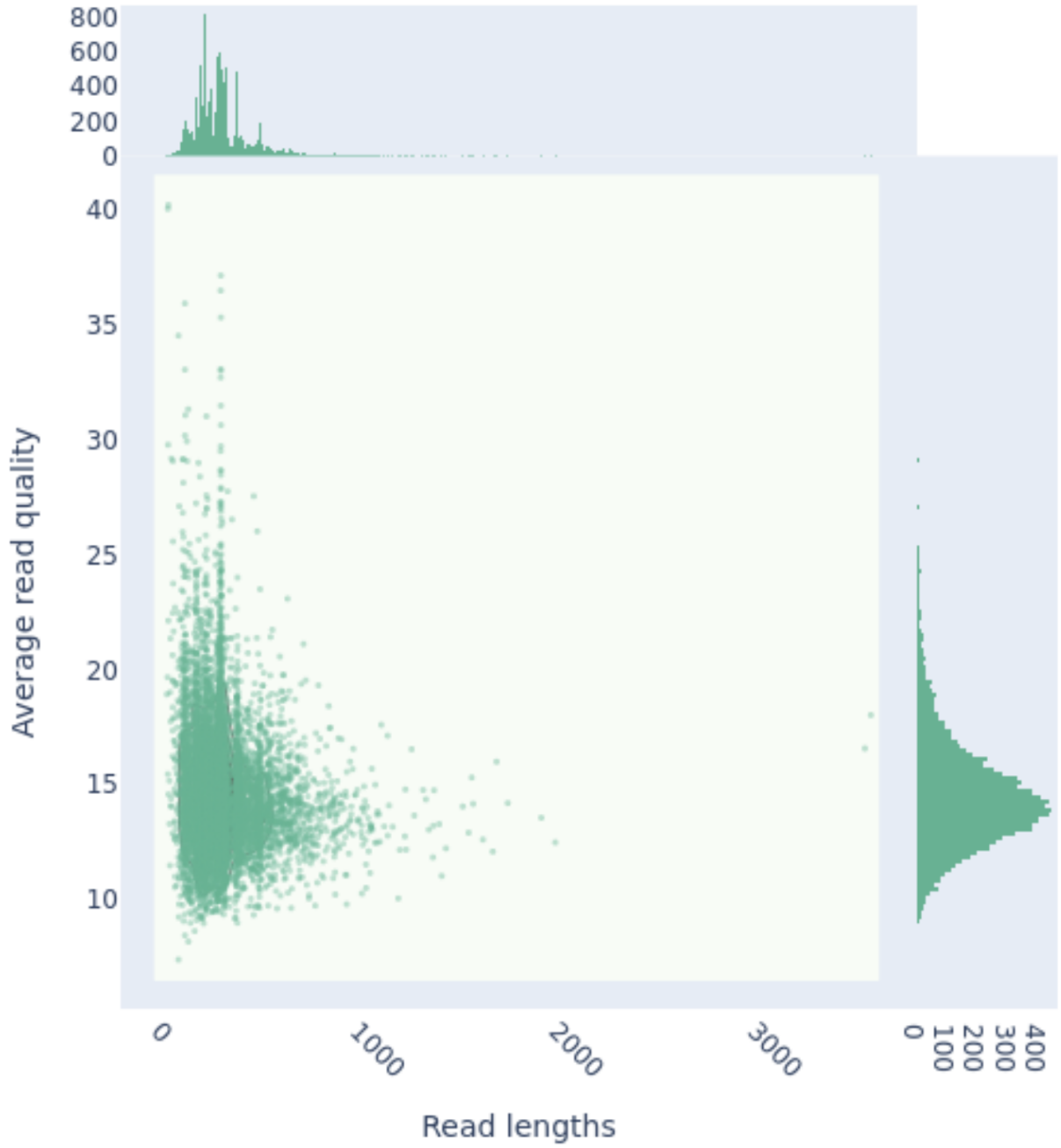


Figure B3 NanoPlot Kde scatter plot of MinION Mk1B read lengths versus average read quality generated from sequencing of 25 microhaplotype regions in 23 caribou DNA samples.

Table B3 Read depths of the top 3 highest reads for 13 microhaplotype loci within 23 caribou samples sequenced with a MinION Mk1B (left) and Illumina MiSeq (right) device. Top 3 reads with poor read distribution are highlighted in gray. Top 3 reads with target read distribution are highlighted in green. Reads <20 constitute low read depth. Zygosity at each locus is characterized as heterozygous (het), homozygous (homo), or inconclusive (NA). If no reads were obtained for a certain locus/sample, it is also denoted as NA. If a zygosity is highlighted in yellow, the zygosity assigned by Seq2Sat was manually overridden.

MinION Mk1B					
Caribou Sample	Microhaplotype Locus	Zygosity	Read Depth: Read 1	Read Depth: Read 2	Read Depth: Read 3
20919	11_Oaf	Het	72	51	42
	18_Eme1	Het	329	185	105
	18_Krt16	Het	990	814	161
	18_Rnft1	Het	215	195	186
	1_Gimap6	Homo	52	3	3
	1_Rsl24d1	Het	161	146	126
	20_Scn11a	Homo	62	5	4
	2_Cd5l	Het	243	76	15
	2_Lexm	Het	94	77	8
	2_Vcam1	Het	16	13	11
	32_Gpm6a	Het	79	73	30
	3_Slc34a1	Het	215	189	58
	4_Abo	Homo	33	3	3
21099	11_Oaf	Homo	457	12	7
	18_Eme1	Het	590	426	91
	18_Krt16	Het	2330	1938	55
	18_Rnft1	Het	515	459	435
	1_Gimap6	Het	125	111	32
	1_Rsl24d1	Het	962	405	385
	20_Scn11a	Het	134	114	6
	2_Cd5l	Het	534	197	148
	2_Lexm	Het	240	181	34
	2_Vcam1	Het	71	48	35
	32_Gpm6a	Het	521	466	18
	3_Slc34a1	Homo	1460	61	57
	4_Abo	Homo	109	5	5
21101	11_Oaf	Homo	571	14	13
	18_Eme1	Het	1041	299	293
	18_Krt16	Het	2348	1997	367
	18_Rnft1	Het	1108	695	589
	1_Gimap6	NA	26	3	3
	1_Rsl24d1	Homo	2693	175	74
	20_Scn11a	Homo	406	15	13

Illumina MiSeq					
Caribou Sample	Microhaplotype Locus	Zygosity	Read Depth: Read 1	Read Depth: Read 2	Read Depth: Read 3
20919	11_Oaf	NA	NA	NA	NA
	18_Eme1	NA	NA	NA	NA
	18_Krt16	NA	NA	NA	NA
	18_Rnft1	NA	NA	NA	NA
	1_Gimap6	NA	NA	NA	NA
	1_Rsl24d1	NA	NA	NA	NA
	20_Scn11a	NA	NA	NA	NA
	2_Cd5l	NA	NA	NA	NA
	2_Lexm	NA	NA	NA	NA
	2_Vcam1	NA	NA	NA	NA
	32_Gpm6a	NA	NA	NA	NA
	3_Slc34a1	NA	NA	NA	NA
	4_Abo	NA	NA	NA	NA
21099	11_Oaf	Homo	236	14	10
	18_Eme1	Het	826	532	118
	18_Krt16	Het	1136	1114	28
	18_Rnft1	Het	110	106	84
	1_Gimap6	Het	60	58	24
	1_Rsl24d1	Het	140	72	58
	20_Scn11a	Het	14	8	2
	2_Cd5l	Het	606	238	124
	2_Lexm	Het	220	200	58
	2_Vcam1	Het	132	56	52
	32_Gpm6a	Het	332	280	10
	3_Slc34a1	Homo	1080	6	6
	4_Abo	Homo	300	8	6
21101	11_Oaf	Homo	258	10	10
	18_Eme1	Het	846	272	250
	18_Krt16	Het	748	746	158
	18_Rnft1	Het	346	188	182
	1_Gimap6	Het	30	20	10
	1_Rsl24d1	Homo	186	4	4
	20_Scn11a	Homo	22	2	2

	2_Cd5l	Het	863	502	48
	2_Lexm	Het	361	284	136
	2_Vcam1	Het	62	56	36
	32_Gpm6a	Homo	940	27	27
	3_Slc34a1	Homo	2099	109	100
	4_Abo	Homo	52	3	2
21102	11_Oaf	Homo	630	22	17
	18_Eme1	Het	1015	495	290
	18_Krt16	Het	2564	2051	66
	18_Rnft1	Het	1328	1280	102
	1_Gimap6	Homo	320	17	11
	1_Rsl24d1	Het	839	684	413
	20_Scn11a	Homo	289	9	9
	2_Cd5l	Het	418	359	254
	2_Lexm	Het	254	192	51
	2_Vcam1	Het	77	51	29
	32_Gpm6a	Het	413	295	109
	3_Slc34a1	Het	763	641	146
	4_Abo	Het	26	15	4
21104	11_Oaf	Homo	264	6	5
	18_Eme1	Het	728	550	19
	18_Krt16	Homo	2989	117	76
	18_Rnft1	Het	470	467	370
	1_Gimap6	Homo	148	9	6
	1_Rsl24d1	Het	443	443	245
	20_Scn11a	Het	95	87	6
	2_Cd5l	Het	273	256	219
	2_Lexm	Homo	482	16	13
	2_Vcam1	Het	38	35	19
	32_Gpm6a	Homo	664	29	25
	3_Slc34a1	Homo	1206	60	57
	4_Abo	Homo	53	4	4
21105	11_Oaf	Het	145	59	56
	18_Eme1	Het	846	192	186
	18_Krt16	Het	2610	2573	78
	18_Rnft1	Het	979	887	107
	1_Gimap6	Homo	265	16	13
	1_Rsl24d1	Het	1618	433	228
	20_Scn11a	Homo	360	20	10
	2_Cd5l	Het	517	347	341
	2_Lexm	Het	284	226	99
	2_Vcam1	Het	52	32	21
	32_Gpm6a	Het	468	431	28
	3_Slc34a1	Het	694	577	148
	4_Abo	Homo	81	5	4

	2_Cd5l	Homo	600	64	22
	2_Lexm	Het	142	126	112
	2_Vcam1	Het	50	46	30
	32_Gpm6a	Homo	514	22	6
	3_Slc34a1	Homo	1136	12	6
	4_Abo	Homo	208	4	4
21102	11_Oaf	Homo	200	12	6
	18_Eme1	Het	1322	578	464
	18_Krt16	Het	1634	1400	32
	18_Rnft1	Het	270	232	18
	1_Gimap6	Homo	74	16	10
	1_Rsl24d1	Het	96	94	70
	20_Scn11a	Homo	60	4	2
	2_Cd5l	Het	364	330	98
	2_Lexm	Het	248	198	14
	2_Vcam1	Het	72	36	36
	32_Gpm6a	Het	448	418	20
	3_Slc34a1	Het	570	498	24
	4_Abo	Het	124	98	8
21104	11_Oaf	Homo	48	4	4
	18_Eme1	Het	1014	774	52
	18_Krt16	Homo	2782	58	16
	18_Rnft1	Het	106	94	74
	1_Gimap6	Het	54	36	22
	1_Rsl24d1	Het	106	96	92
	20_Scn11a	Het	18	12	2
	2_Cd5l	Homo	774	78	74
	2_Lexm	Homo	472	14	8
	2_Vcam1	Het	66	58	14
	32_Gpm6a	Homo	846	42	8
	3_Slc34a1	Homo	1142	8	8
	4_Abo	Homo	160	2	2
21105	11_Oaf	Het	152	48	12
	18_Eme1	Het	1560	506	494
	18_Krt16	Het	1318	1314	22
	18_Rnft1	Het	316	278	22
	1_Gimap6	Het	132	50	12
	1_Rsl24d1	Het	284	136	12
	20_Scn11a	Homo	92	2	2
	2_Cd5l	Homo	1046	138	112
	2_Lexm	Het	366	322	18
	2_Vcam1	Het	166	82	54
	32_Gpm6a	Het	684	674	28
	3_Slc34a1	Het	748	706	12
	4_Abo	Homo	300	10	8

21106	11_Oaf	Homo	316	10	7
	18_Eme1	Het	936	242	237
	18_Krt16	Het	2487	1954	74
	18_Rnft1	Het	932	508	335
	1_Gimap6	Homo	213	16	12
	1_Rsl24d1	Het	668	552	489
	20_Scn11a	Homo	328	22	12
	2_Cd5l	Het	944	531	45
	2_Lexm	Het	392	315	130
	2_Vcam1	Het	56	35	33
	32_Gpm6a	Het	312	294	99
	3_Slc34a1	Homo	1470	90	76
	4_Abo	Het	22	20	2
21107	11_Oaf	Het	58	23	22
	18_Eme1	Het	500	235	128
	18_Krt16	Het	1394	975	199
	18_Rnft1	Het	219	192	170
	1_Gimap6	Homo	27	3	3
	1_Rsl24d1	Het	846	155	86
	20_Scn11a	Homo	234	12	11
	2_Cd5l	Het	221	190	136
	2_Lexm	Het	197	194	86
	2_Vcam1	Het	28	16	9
	32_Gpm6a	Het	469	401	17
	3_Slc34a1	Homo	1024	69	39
	4_Abo	Homo	27	2	2
21110	11_Oaf	Het	40	17	16
	18_Eme1	Het	537	303	113
	18_Krt16	Het	1336	940	143
	18_Rnft1	Het	509	203	196
	1_Gimap6	Het	104	12	5
	1_Rsl24d1	Het	445	398	40
	20_Scn11a	Homo	257	14	9
	2_Cd5l	Het	196	154	138
	2_Lexm	Het	206	196	90
	2_Vcam1	Het	17	13	9
	32_Gpm6a	Het	298	181	60
	3_Slc34a1	Homo	1007	55	47
	4_Abo	Het	24	12	3
21111	11_Oaf	Het	46	21	20
	18_Eme1	Het	504	219	113
	18_Krt16	Het	869	578	103
	18_Rnft1	Het	459	195	166
	1_Gimap6	Homo	84	7	6
	1_Rsl24d1	Het	526	147	76

21106	11_Oaf	Homo	220	14	8
	18_Eme1	Het	1082	388	372
	18_Krt16	Het	1778	1716	40
	18_Rnft1	Het	128	70	38
	1_Gimap6	Het	68	22	20
	1_Rsl24d1	Het	136	98	72
	20_Scn11a	Homo	22	2	2
	2_Cd5l	Het	1142	170	44
	2_Lexm	Het	270	236	154
	2_Vcam1	Het	142	90	38
	32_Gpm6a	Het	358	182	26
	3_Slc34a1	Homo	1528	10	8
	4_Abo	Het	134	114	8
21107	11_Oaf	Homo	48	8	6
	18_Eme1	Het	620	342	100
	18_Krt16	Het	672	584	68
	18_Rnft1	Het	42	32	32
	1_Gimap6	Het	28	14	2
	1_Rsl24d1	Het	136	34	8
	20_Scn11a	NA	16	2	2
	2_Cd5l	Het	204	160	28
	2_Lexm	Het	106	92	42
	2_Vcam1	Het	34	20	14
	32_Gpm6a	Het	262	166	14
	3_Slc34a1	Homo	640	6	6
	4_Abo	Homo	120	4	2
21110	11_Oaf	Het	156	52	12
	18_Eme1	Het	1368	1038	58
	18_Krt16	Het	1254	1034	14
	18_Rnft1	Het	340	130	114
	1_Gimap6	Het	68	52	36
	1_Rsl24d1	Het	180	172	6
	20_Scn11a	Homo	106	4	4
	2_Cd5l	Het	1022	406	90
	2_Lexm	Het	294	292	16
	2_Vcam1	Het	92	80	78
	32_Gpm6a	Het	576	364	20
	3_Slc34a1	Homo	1574	10	10
	4_Abo	Het	170	166	6
21111	11_Oaf	Het	126	50	12
	18_Eme1	Het	1222	448	386
	18_Krt16	Het	1480	1330	44
	18_Rnft1	Het	230	132	70
	1_Gimap6	Het	46	22	20
	1_Rsl24d1	Het	244	110	6

	20_Scn11a	Homo	171	8	6
	2_Cd5l	Het	206	194	151
	2_Lexm	Het	150	145	54
	2_Vcam1	Het	31	16	7
	32_Gpm6a	Het	169	161	56
	3_Slc34a1	Het	349	304	99
	4_Abo	NA	10	9	2
21112	11_Oaf	Homo	169	4	4
	18_Eme1	Het	505	284	139
	18_Krt16	Het	1055	1019	195
	18_Rnft1	Het	249	237	236
	1_Gimap6	Het	114	17	6
	1_Rsl24d1	Homo	1244	79	49
	20_Scn11a	Het	140	107	7
	2_Cd5l	Het	419	247	21
	2_Lexm	Homo	541	20	17
	2_Vcam1	Het	36	22	12
	32_Gpm6a	Het	262	215	78
	3_Slc34a1	Het	449	425	101
	4_Abo	Homo	27	3	2
21113	11_Oaf	Het	26	13	11
	18_Eme1	Het	241	112	98
	18_Krt16	Het	693	671	18
	18_Rnft1	Het	254	103	103
	1_Gimap6	NA	NA	NA	NA
	1_Rsl24d1	Het	181	176	37
	20_Scn11a	Homo	131	7	4
	2_Cd5l	Het	291	50	33
	2_Lexm	Het	122	100	20
	2_Vcam1	NA	12	11	5
	32_Gpm6a	Het	85	75	15
	3_Slc34a1	Homo	424	25	18
	4_Abo	NA	NA	NA	NA
21297	11_Oaf	Homo	99	4	3
	18_Eme1	Het	287	65	64
	18_Krt16	Homo	1668	38	33
	18_Rnft1	Het	379	177	174
	1_Gimap6	Homo	46	4	3
	1_Rsl24d1	Het	240	217	32
	20_Scn11a	Homo	228	11	9
	2_Cd5l	Het	155	131	130
	2_Lexm	Het	125	112	48
	2_Vcam1	NA	12	7	6
	32_Gpm6a	Homo	269	10	7
	3_Slc34a1	Homo	554	30	27

	20_Scn11a	Homo	92	4	2
	2_Cd5l	Homo	898	82	60
	2_Lexm	Het	286	240	10
	2_Vcam1	Het	126	86	82
	32_Gpm6a	Het	474	462	26
	3_Slc34a1	Het	732	696	6
	4_Abo	Het	134	124	4
21112	11_Oaf	NA	NA	NA	NA
	18_Eme1	NA	NA	NA	NA
	18_Krt16	NA	NA	NA	NA
	18_Rnft1	NA	NA	NA	NA
	1_Gimap6	NA	NA	NA	NA
	1_Rsl24d1	NA	NA	NA	NA
	20_Scn11a	NA	NA	NA	NA
	2_Cd5l	NA	NA	NA	NA
	2_Lexm	NA	NA	NA	NA
	2_Vcam1	NA	NA	NA	NA
	32_Gpm6a	NA	NA	NA	NA
	3_Slc34a1	NA	NA	NA	NA
	4_Abo	NA	NA	NA	NA
21113	11_Oaf	Homo	146	24	10
	18_Eme1	Het	1244	556	512
	18_Krt16	Het	1832	1758	46
	18_Rnft1	Het	266	138	114
	1_Gimap6	Het	26	24	22
	1_Rsl24d1	Het	212	158	10
	20_Scn11a	Homo	36	2	2
	2_Cd5l	Het	500	478	218
	2_Lexm	Het	334	258	20
	2_Vcam1	Het	92	60	54
	32_Gpm6a	Het	594	588	28
	3_Slc34a1	Homo	1448	12	12
	4_Abo	Het	216	136	12
21297	11_Oaf	NA	NA	NA	NA
	18_Eme1	NA	NA	NA	NA
	18_Krt16	NA	NA	NA	NA
	18_Rnft1	NA	NA	NA	NA
	1_Gimap6	NA	NA	NA	NA
	1_Rsl24d1	NA	NA	NA	NA
	20_Scn11a	NA	NA	NA	NA
	2_Cd5l	NA	NA	NA	NA
	2_Lexm	NA	NA	NA	NA
	2_Vcam1	NA	NA	NA	NA
	32_Gpm6a	NA	NA	NA	NA
	3_Slc34a1	NA	NA	NA	NA

	4_Abo	NA	5	1	1
21298	11_Oaf	Het	40	18	13
	18_Eme1	Het	313	178	65
	18_Krt16	Homo	1699	48	46
	18_Rnft1	Het	322	89	18
	1_Gimap6	Het	117	16	4
	1_Rsl24d1	Het	554	165	61
	20_Scn11a	Het	89	41	5
	2_Cd5l	Het	391	227	16
	2_Lexm	Het	180	116	55
	2_Vcam1	NA	6	2	2
	32_Gpm6a	Homo	202	5	5
	3_Slc34a1	Homo	639	43	29
	4_Abo	NA	5	1	1
21299	11_Oaf	Homo	91	3	2
	18_Eme1	Het	392	358	113
	18_Krt16	Homo	2712	67	54
	18_Rnft1	Het	564	522	36
	1_Gimap6	Het	8	1	1
	1_Rsl24d1	Het	309	269	33
	20_Scn11a	Homo	314	16	9
	2_Cd5l	Het	257	198	194
	2_Lexm	Homo	658	30	17
	2_Vcam1	Het	15	15	6
	32_Gpm6a	Homo	312	10	8
	3_Slc34a1	Het	309	288	78
	4_Abo	Het	7	5	1
21301	11_Oaf	Homo	10	1	1
	18_Eme1	Het	89	47	12
	18_Krt16	Het	223	136	6
	18_Rnft1	Het	47	45	29
	1_Gimap6	Het	38	32	4
	1_Rsl24d1	Het	689	296	119
	20_Scn11a	NA	7	2	1
	2_Cd5l	NA	5	4	1
	2_Lexm	NA	NA	NA	NA
	2_Vcam1	NA	NA	NA	NA
	32_Gpm6a	Homo	47	2	2
	3_Slc34a1	Homo	216	8	6
	4_Abo	NA	NA	NA	NA
21303	11_Oaf	Homo	133	5	3
	18_Eme1	Het	413	132	123
	18_Krt16	Het	709	695	104
	18_Rnft1	Het	846	258	250
	1_Gimap6	NA	NA	NA	NA

	4_Abo	NA	NA	NA	NA
21298	11_Oaf	Homo	194	54	18
	18_Eme1	Het	1258	1010	78
	18_Krt16	Homo	2846	46	22
	18_Rnft1	Het	178	82	10
	1_Gimap6	Het	54	32	6
	1_Rsl24d1	Het	190	92	10
	20_Scn11a	Het	36	10	2
	2_Cd5l	Het	748	206	22
	2_Lexm	Het	178	174	10
	2_Vcam1	Het	78	62	42
	32_Gpm6a	Homo	612	12	8
	3_Slc34a1	Homo	1612	12	10
	4_Abo	Homo	364	8	6
21299	11_Oaf	Homo	66	10	6
	18_Eme1	Het	1180	888	62
	18_Krt16	Homo	3010	44	26
	18_Rnft1	Het	262	234	14
	1_Gimap6	Het	102	48	8
	1_Rsl24d1	Het	168	146	6
	20_Scn11a	Homo	90	4	4
	2_Cd5l	Het	1028	518	164
	2_Lexm	Homo	716	28	14
	2_Vcam1	Het	144	70	60
	32_Gpm6a	Homo	1042	46	10
	3_Slc34a1	Het	600	446	8
	4_Abo	Het	186	166	6
21301	11_Oaf	Homo	184	12	12
	18_Eme1	Het	1160	800	148
	18_Krt16	Het	1228	1136	48
	18_Rnft1	Het	186	120	86
	1_Gimap6	Het	50	26	16
	1_Rsl24d1	Het	152	134	80
	20_Scn11a	Homo	56	4	2
	2_Cd5l	Het	836	340	78
	2_Lexm	Het	274	236	80
	2_Vcam1	Het	104	56	52
	32_Gpm6a	Homo	864	32	8
	3_Slc34a1	Homo	1084	10	8
	4_Abo	Homo	192	6	4
21303	11_Oaf	Homo	76	6	6
	18_Eme1	Het	1174	512	472
	18_Krt16	Het	1126	1026	24
	18_Rnft1	Het	150	92	88
	1_Gimap6	Het	66	40	6

	1_Rsl24d1	Het	268	208	203
	20_Scn11a	Homo	208	10	9
	2_Cd5l	Het	192	182	152
	2_Lexm	Het	157	149	64
	2_Vcam1	Het	26	22	15
	32_Gpm6a	Homo	290	10	8
	3_Slc34a1	Homo	579	28	27
	4_Abo	NA	9	3	2
21305	11_Oaf	Het	95	47	26
	18_Eme1	Het	680	237	202
	18_Krt16	Het	1988	1653	67
	18_Rnft1	Het	975	467	463
	1_Gimap6	Homo	141	10	8
	1_Rsl24d1	Het	909	472	168
	20_Scn11a	Homo	230	13	11
	2_Cd5l	Het	383	369	125
	2_Lexm	Het	329	240	20
	2_Vcam1	Het	44	17	8
	32_Gpm6a	Het	358	336	10
	3_Slc34a1	Homo	1140	61	46
	4_Abo	Homo	38	3	2
21306	11_Oaf	Homo	252	12	6
	18_Eme1	Het	922	531	28
	18_Krt16	Het	1268	1246	160
	18_Rnft1	Het	991	661	591
	1_Gimap6	Homo	144	7	6
	1_Rsl24d1	Het	561	560	205
	20_Scn11a	Homo	290	17	9
	2_Cd5l	Het	420	278	207
	2_Lexm	Homo	677	44	20
	2_Vcam1	Het	40	25	23
	32_Gpm6a	Het	180	128	35
	3_Slc34a1	Het	378	357	49
	4_Abo	Het	12	9	1
21307	11_Oaf	Het	76	35	33
	18_Eme1	Het	823	450	254
	18_Krt16	Het	1268	1112	252
	18_Rnft1	Het	946	361	313
	1_Gimap6	NA	5	1	1
	1_Rsl24d1	Homo	1342	79	42
	20_Scn11a	Het	176	161	9
	2_Cd5l	Het	262	225	183
	2_Lexm	Homo	675	46	16
	2_Vcam1	Het	40	21	14
	32_Gpm6a	Het	274	214	81

	1_Rsl24d1	Het	112	110	96
	20_Scn11a	Homo	84	2	2
	2_Cd5l	Het	908	468	126
	2_Lexm	Het	292	258	10
	2_Vcam1	Het	142	130	14
	32_Gpm6a	Homo	802	14	10
	3_Slc34a1	Homo	1084	10	8
	4_Abo	Het	170	154	6
21305	11_Oaf	Het	106	40	14
	18_Eme1	Het	528	246	244
	18_Krt16	Het	1584	1108	40
	18_Rnft1	Het	120	84	56
	1_Gimap6	Het	70	60	32
	1_Rsl24d1	Het	132	98	4
	20_Scn11a	Homo	36	4	2
	2_Cd5l	Het	388	386	72
	2_Lexm	Het	240	156	10
	2_Vcam1	Het	92	36	32
	32_Gpm6a	Het	244	230	16
	3_Slc34a1	Homo	1182	12	10
	4_Abo	Homo	208	6	4
21306	11_Oaf	NA	NA	NA	NA
	18_Eme1	Homo	24	2	2
	18_Krt16	Homo	24	2	2
	18_Rnft1	NA	NA	NA	NA
	1_Gimap6	NA	8	2	2
	1_Rsl24d1	NA	16	4	2
	20_Scn11a	NA	NA	NA	NA
	2_Cd5l	Homo	16	2	2
	2_Lexm	NA	NA	NA	NA
	2_Vcam1	NA	NA	NA	NA
	32_Gpm6a	NA	NA	NA	NA
	3_Slc34a1	NA	NA	NA	NA
	4_Abo	NA	NA	NA	NA
21307	11_Oaf	NA	96	18	14
	18_Eme1	Het	934	428	288
	18_Krt16	Het	1082	1062	102
	18_Rnft1	Het	252	138	118
	1_Gimap6	Het	102	76	6
	1_Rsl24d1	Homo	290	4	4
	20_Scn11a	NA	12	2	2
	2_Cd5l	Het	520	240	82
	2_Lexm	Homo	594	34	8
	2_Vcam1	Het	112	110	44
	32_Gpm6a	Het	294	248	24

	3_Slc34a1	Homo	1127	63	62
	4_Abo	Het	23	17	2
21309	11_Oaf	Homo	263	8	8
	18_Eme1	Het	978	590	225
	18_Krt16	Het	1977	1409	54
	18_Rnft1	Het	1396	1015	121
	1_Gimap6	Homo	173	8	8
	1_Rsl24d1	Homo	1442	67	41
	20_Scn11a	Homo	232	18	15
	2_Cd5l	Het	348	286	193
	2_Lexm	Homo	706	43	17
	2_Vcam1	Het	78	54	37
	32_Gpm6a	Het	240	226	70
	3_Slc34a1	Het	659	496	35
	4_Abo	Het	22	12	2
3A	11_Oaf	Homo	274	8	7
	18_Eme1	Het	567	293	168
	18_Krt16	Het	986	769	94
	18_Rnft1	Het	428	427	422
	1_Gimap6	Homo	126	5	4
	1_Rsl24d1	Het	912	278	113
	20_Scn11a	Homo	263	12	11
	2_Cd5l	Het	576	370	33
	2_Lexm	Het	169	147	55
	2_Vcam1	Het	34	22	19
	32_Gpm6a	Het	299	233	8
	3_Slc34a1	Homo	806	40	30
	4_Abo	Homo	25	2	2
3C	11_Oaf	Het	164	47	47
	18_Eme1	Het	1056	746	416
	18_Krt16	Het	1168	1091	211
	18_Rnft1	Het	940	511	440
	1_Gimap6	NA	7	2	2
	1_Rsl24d1	Het	404	313	261
	20_Scn11a	Homo	339	15	13
	2_Cd5l	Het	536	374	46
	2_Lexm	Het	194	169	80
	2_Vcam1	Het	51	40	28
	32_Gpm6a	Het	293	188	56
	3_Slc34a1	Homo	1202	59	45
	4_Abo	Het	15	11	4

	3_Slc34a1	Homo	1106	6	6
	4_Abo	Het	120	110	4
21309	11_Oaf	Homo	222	22	16
	18_Eme1	Het	1612	1134	108
	18_Krt16	Het	1830	1580	30
	18_Rnft1	Het	396	282	18
	1_Gimap6	Het	70	50	36
	1_Rsl24d1	Homo	270	4	4
	20_Scn11a	Homo	42	4	2
	2_Cd5l	Het	466	434	152
	2_Lexm	Homo	506	26	8
	2_Vcam1	Het	128	62	52
	32_Gpm6a	Het	512	460	24
	3_Slc34a1	Het	800	706	8
	4_Abo	Het	238	226	4
3A	11_Oaf	NA	NA	NA	NA
	18_Eme1	NA	14	4	4
	18_Krt16	Het	14	6	2
	18_Rnft1	NA	NA	NA	NA
	1_Gimap6	NA	NA	NA	NA
	1_Rsl24d1	NA	NA	NA	NA
	20_Scn11a	NA	NA	NA	NA
	2_Cd5l	NA	NA	NA	NA
	2_Lexm	NA	NA	NA	NA
	2_Vcam1	NA	NA	NA	NA
	32_Gpm6a	NA	NA	NA	NA
	3_Slc34a1	NA	8	2	NA
	4_Abo	NA	NA	NA	NA
3C	11_Oaf	Het	110	36	4
	18_Eme1	Het	1196	382	380
	18_Krt16	Het	842	830	120
	18_Rnft1	Het	164	124	108
	1_Gimap6	Het	32	28	26
	1_Rsl24d1	Het	98	88	84
	20_Scn11a	Homo	44	4	2
	2_Cd5l	Het	590	254	24
	2_Lexm	Het	234	150	32
	2_Vcam1	Het	38	38	32
	32_Gpm6a	Het	426	332	44
	3_Slc34a1	Homo	794	76	10
	4_Abo	Het	166	96	12

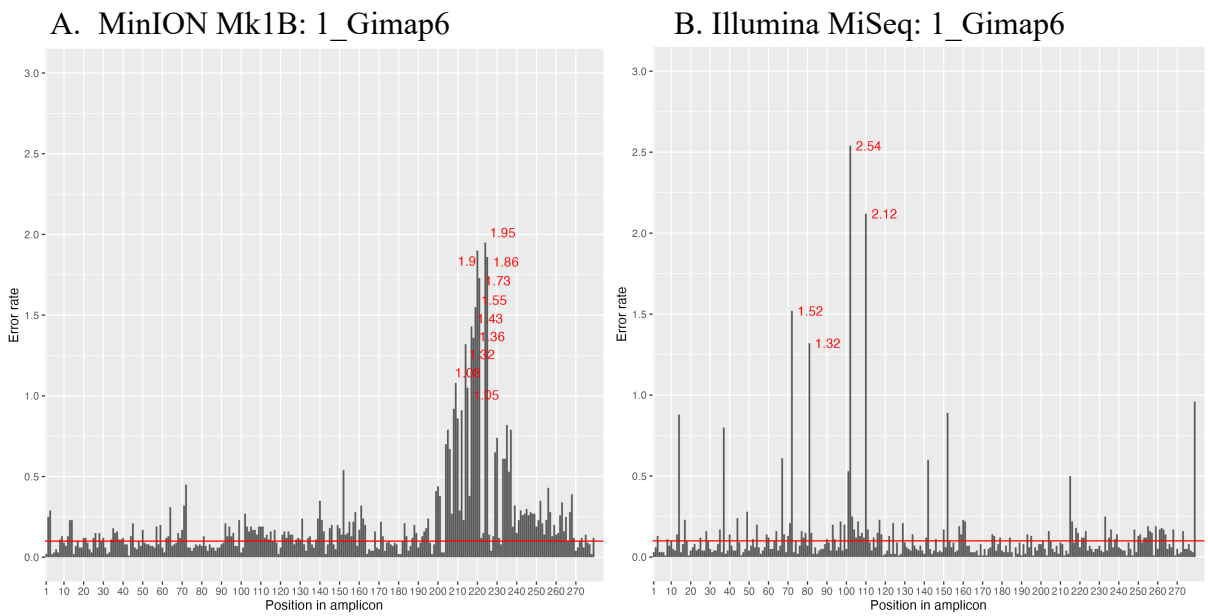


Figure B4 Average error rate (%) at each position along the 1_Gimap6 microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

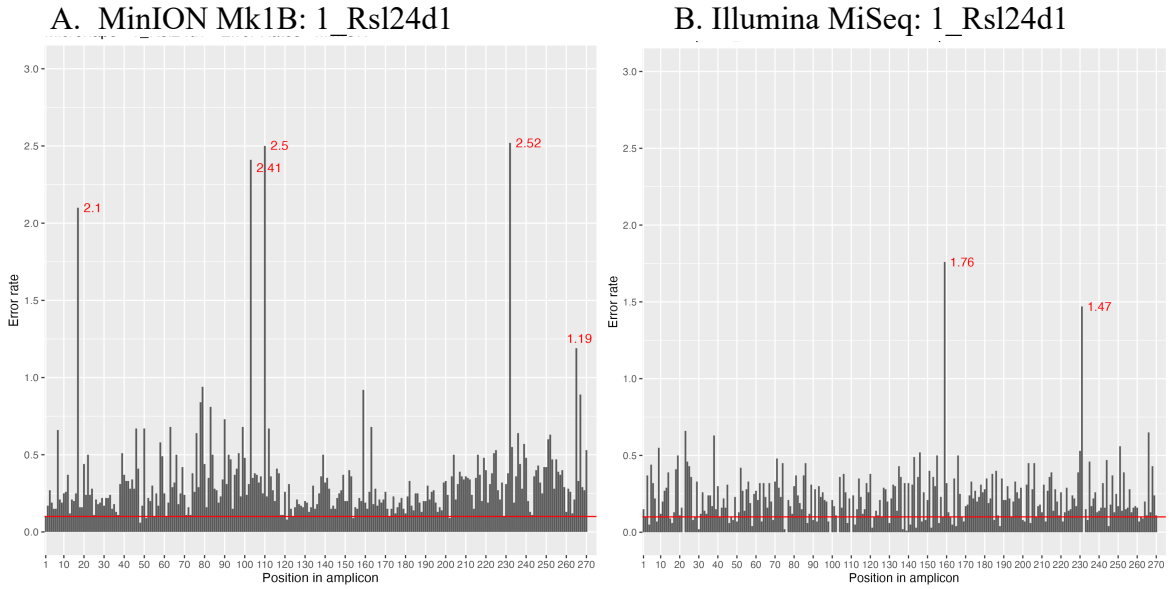


Figure B5 Average error rate (%) at each position along the 1_Rsl24d1 microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

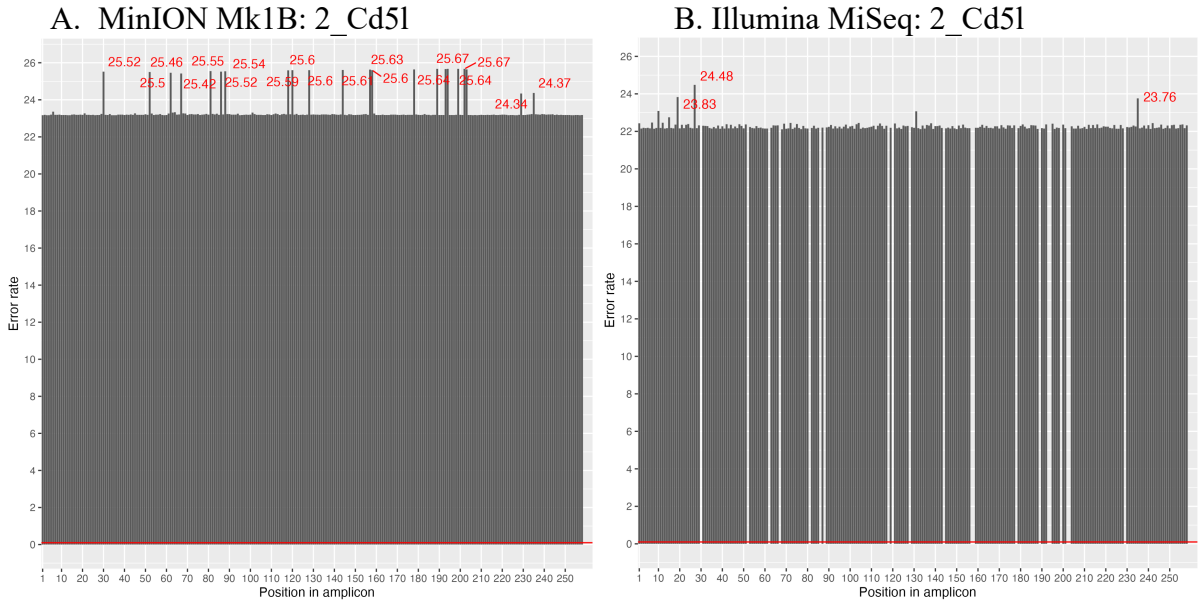


Figure B6 Average error rate (%) at each position along the 2_Cd5l microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

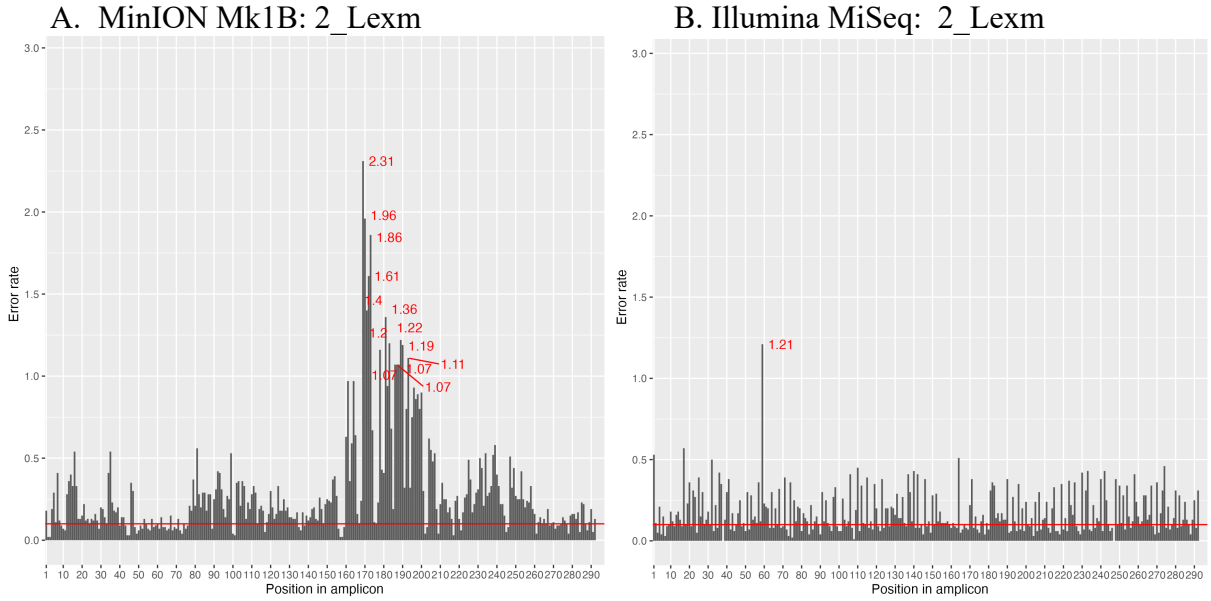


Figure B7 Average error rate (%) at each position along the 2_Lexm microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

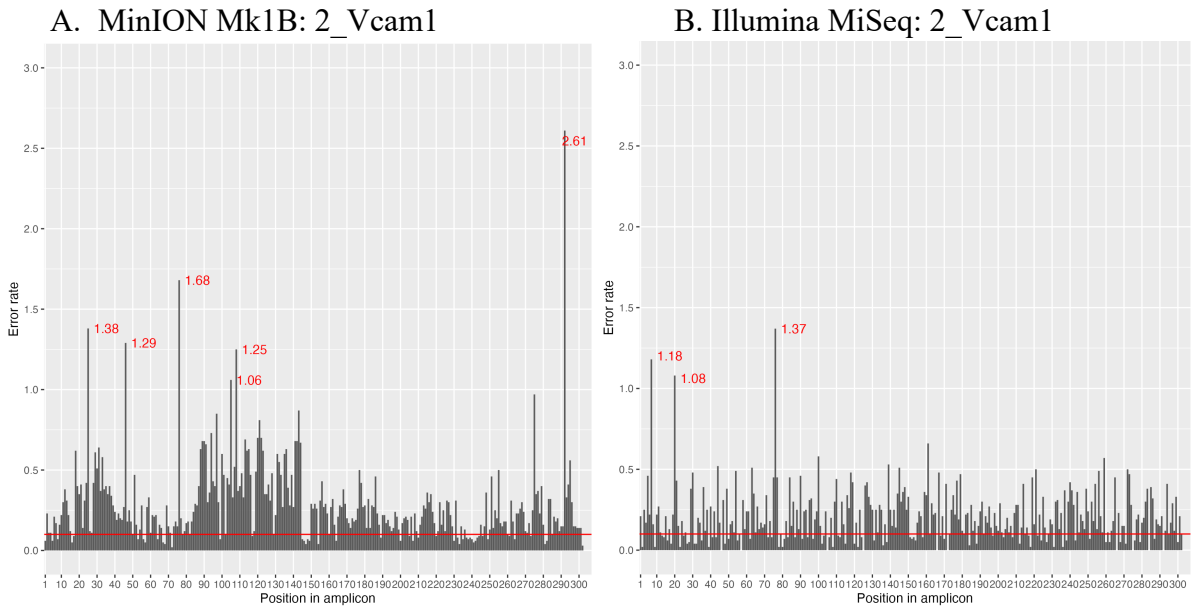


Figure B8 Average error rate (%) at each position along the 2_Vcam1 microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

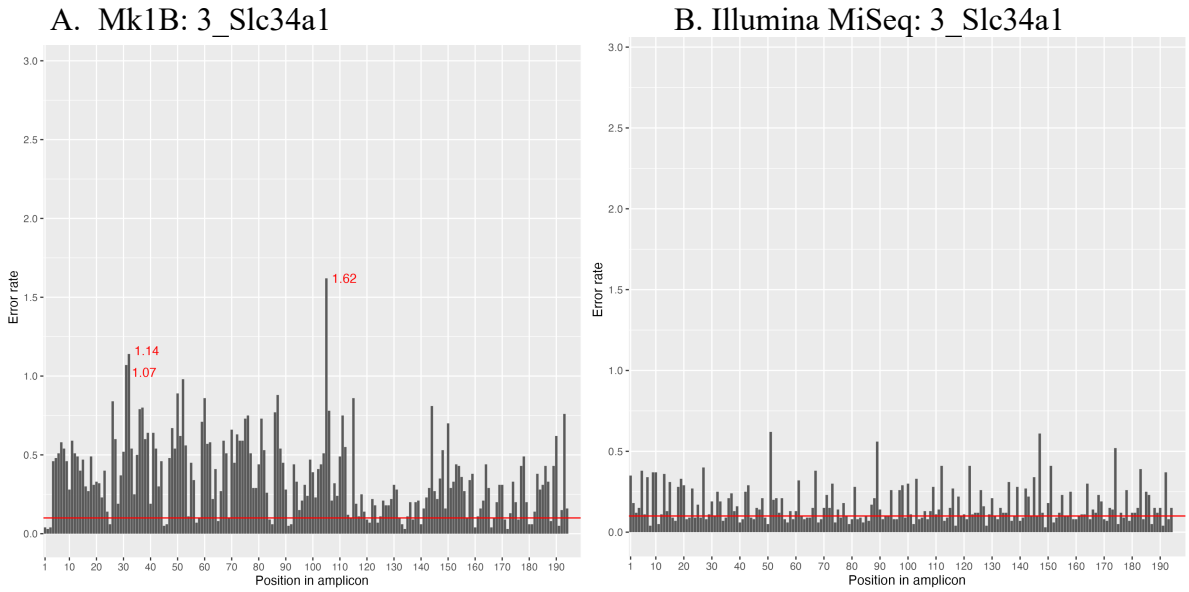


Figure B9 Average error rate (%) at each position along the 3_Slc34a1 microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

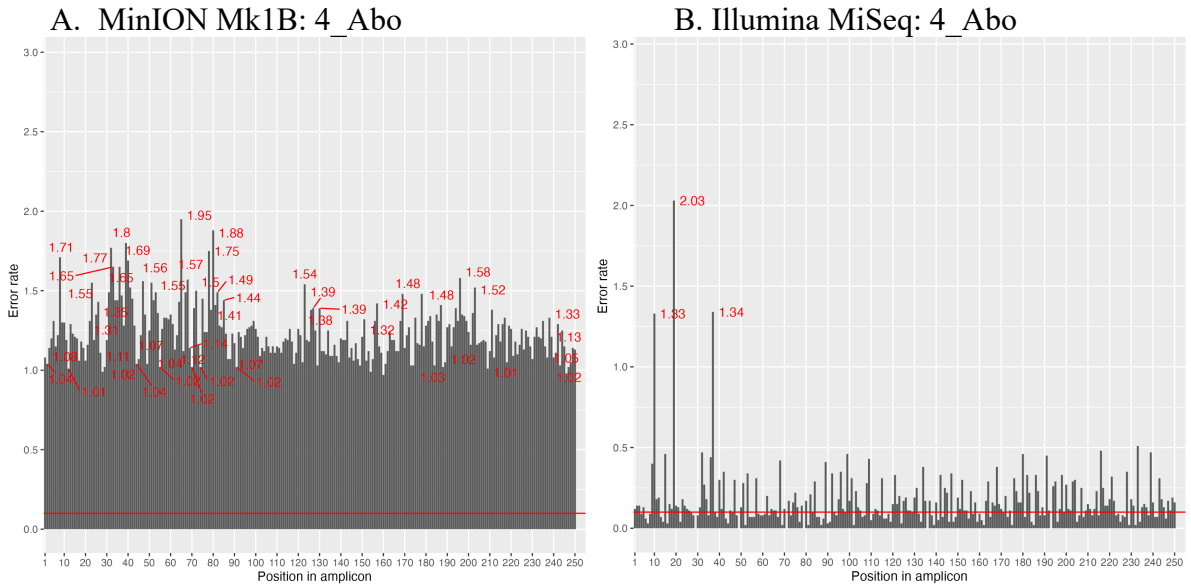


Figure B10 Average error rate (%) at each position along the 4_Abo microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

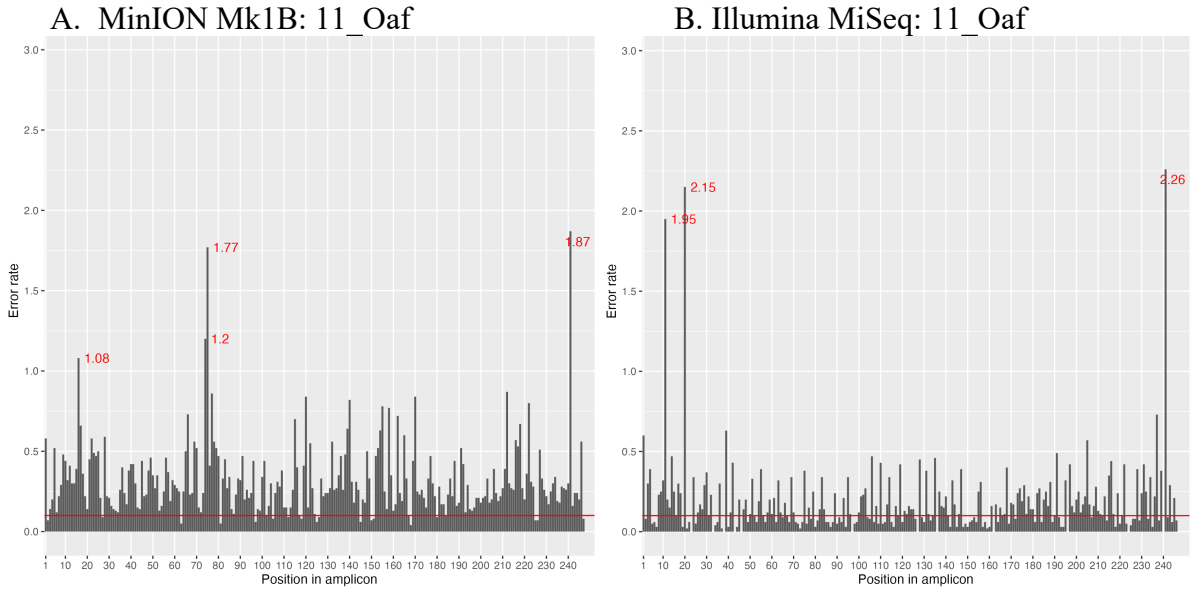


Figure B11 Average error rate (%) at each position along the 11_Oaf microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

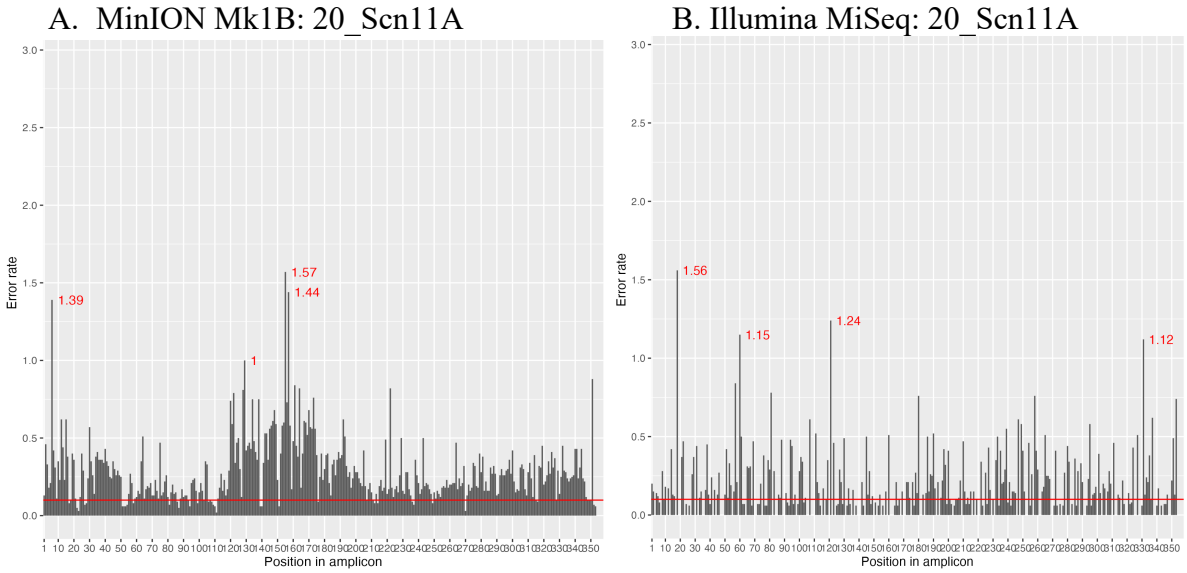


Figure B12 Average error rate (%) at each position along the 20_Scn11A microhaplotype amplicon across caribou samples sequenced with a MinION Mk1B (A) and an Illumina MiSeq (B).

Table B4 Cost estimate of sequencing 25 microhaplotype amplicons in 96 caribou samples on a MinION Mk1B sequencer using the ONT Ligation Sequencing Amplicons protocol, Native Barcoding Kit 96 V14 (SQK-NBD114-96), and an R10.4.1 flow cell. Prices are shown in $CAD.

MinION Mk1B Expenses							
Item	**Amount Item Comes With**	**Cost of Item**	**Cost Per Amount**	**Amount Needed for 96 Samples**	**Cost for 96 Samples**	**Cost per Sample**	**Cost per Locus**
2x Qiagen Type-It PCR Master Mix for All Reactions	25000 µL	$2,697.82	$0.11	3740 µL	$403.59	$4.20	$0.17
IDT Standard Primers for All Reactions	2500 µL	$22.00	$0.01	935 µL	$8.23	$0.09	Negligible
Qubit™ Reagents	500 Samples	$492.00	$0.98	96 Samples + 1 DNA Library	$95.06	$0.99	$0.04
Native Barcoding Kit 96 V14 (SQK-NBD114.96)	288 Barcodes (3 sets of 96)	$1,092.17	$3.79	96 Barcodes	$364.06	$3.79	$0.15
NEBNext Ultra II End Repair / dA-tailing Module (NEB, cat # E7546L)	672 µL Buffer, 288 µL Enzyme Mix	$1,211	$1.80	168 µL Buffer, 72 µL Enzyme Mix	$302.75	$3.15	$0.13
NEB Blunt/TA Ligase Master Mix (NEB, cat # M0367L)	1250 µL	$644	$0.52	480 µL	$247.30	$2.58	$0.10
Quick T4 DNA Ligase in NEBNext® Quick Ligation Module (NEB, cat # E6056L)	500 µL Ligase, 1000 µL Buffer	$2,116	$4.23	5 µL Ligase, 10 µL Buffer	$21.16	$0.22	$0.01
Flow Cell R10.4.1	1 Flow Cell	$956.84	$956.84	1 Flow Cell	$956.84	$9.97	$0.40
Total					**$2,399.37**	**$24.99**	**$1.00**

Table B5 Cost estimate of sequencing 25 microhaplotype amplicons in 96 caribou samples on an Illumina MiSeq sequencer using a modified 16S Illumina Amplicon protocol, the Illumina DNA Prep Kit, and a V2 flow cell. Prices are shown in $CAD.

Illumina MiSeq Expenses							
Item	**Amount Item Comes With**	**Cost of Item**	**Cost Per Amount**	**Amount Needed for 96 Samples**	**Cost for 96 Samples**	**Cost per Sample**	**Cost per Locus**
2x Qiagen Type-It PCR Master Mix for All Reactions	25000 µL	$2,697.82	$0.11	4770 µL	$514.74	$5.36	$0.21
IDT Standard Primers for All Reactions	2500 µL	$22.00	$0.01	892.5µL	$7.85	$0.08	Negligible
AMPure XP Beads for All Reactions	60000 µL	$2,031.00	$0.03	2155 µL	$72.95	$0.76	$0.03
Indexes for All Reactions	1152 Samples (3 sets of 384)	$3,012.00	$2.61	96 Samples	$251.00	$2.61	$0.10
Qubit™ Reagents	500 Samples	$492.00	$0.98	1 DNA Library	$0.98	$0.01	Negligible
Illumina V2 500 Kit	384 Samples	$2,211.00	$2,211.00	1 Kit	$2,211.00	$23.03	$0.92
Total					**$3,058.52**	**$31.86**	**$1.27**

Table B6 Cost estimate of sequencing 25 microhaplotype amplicons in 384 caribou samples on an Illumina MiSeq sequencer using a modified 16S Illumina Amplicon protocol, the Illumina DNA Prep Kit, and a V2 flow cell. Prices are shown in $CAD.

Illumina MiSeq Expenses							
Item	**Amount Item Comes With**	**Cost of Item**	**Cost Per Amount**	**Amount Needed for 384 Samples**	**Cost for 384 Samples**	**Cost per Sample**	**Cost per Locus**
2x Qiagen Type-It PCR Master Mix for All Reactions	25000 µL	$2,697.82	$0.11	19080 µL	$2,058.98	$5.36	$0.21
IDT Standard Primers for All Reactions	2500 µL	$22.00	$0.01	3570 µL	$31.42	$0.08	Negligible
AMPure XP Beads for All Reactions	60000 µL	$2,031.00	$0.03	8620 µL	$291.79	$0.76	$0.03
Indexes for All Reactions	1152 Samples (3 sets of 384)	$3,012.00	$2.61	384 Samples	$1,004.00	$2.61	$0.10
Qubit™ Reagents	500 Samples	$492.00	$0.98	4 DNA Libraries	$3.94	$0.01	Negligible
Illumina V2 500 Kit	384 Samples	$2,211.00	$2,211.00	1 Kit	$2,211.00	$5.76	$0.23
Total					**$5,601.12**	**$14.59**	**$0.58**