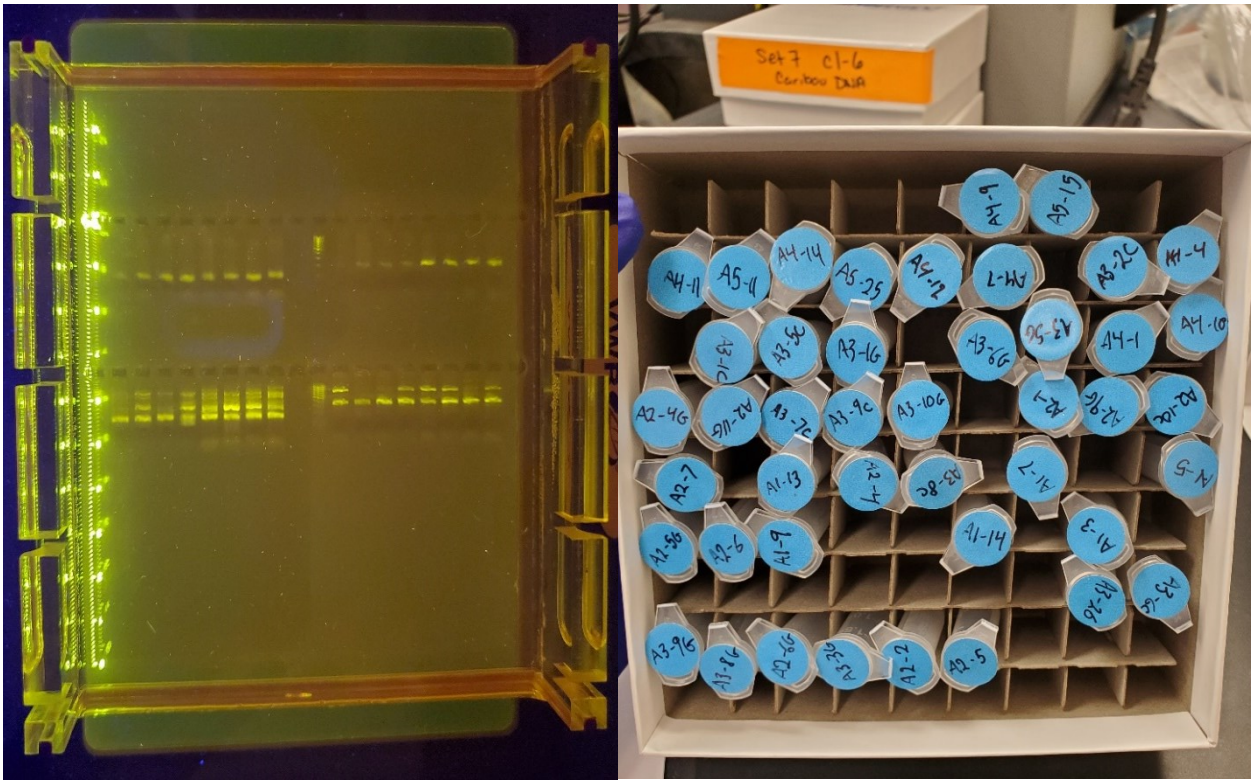
LANDSCAPE GENETIC ANALYSIS OF POPULATION STRUCTURE AND BARRIERS TO GENE FLOW IN BOREAL WOODLAND CARIBOU (*RANGIFER TARANDUS CARIBOU)*

by

Nicole N. Anderson



FACULTY OF NATURAL RESOURCES MANAGEMENT
LAKEHEAD UNIVERSITY
THUNDER BAY, ONTARIO

January 16, 2023

LANDSCAPE GENETIC ANALYSIS OF POPULATION STRUCTURE AND BARRIERS TO GENE FLOW IN BOREAL WOODLAND CARIBOU (*RANGIFER TARANDUS CARIBOU)*

by

Nicole N. Anderson

A Thesis Submitted in Fulfillment of the
Requirements for the Degree of
Masters of Science in Forestry

Faculty of Natural Resources Management
Lakehead University

January 16, 2023

ABSTRACT

Anderson, N.N. 2023. Landscape Genetic Analysis of Population Structure and Barriers to Gene Flow in Boreal Woodland Caribou (*Rangifer tarandus caribou*). 168pp +XII.

Keywords: boreal woodland caribou, *Rangifer tarandus caribou,* landscape genetics, population genetic structure, landscape resistance, fine-scale differentiation, gene flow

This study examines patterns of population genetic structure and gene flow of boreal woodland caribou (*Rangifer tarandus caribou*), which are experiencing declining population sizes across North America. Compared to previous studies, I used fine-scale landscape genetic analyses with intensive sampling to identify genetic subdivisions within a single range and anthropogenic and natural drivers of genetic discontinuity. The Brightsand Range of Ontario is among the southernmost boreal woodland caribou populations and contains actively managed and unmanaged forests. This range provided a unique opportunity to examine the drivers of population subdivision using fecal DNA samples (n = 788) previously obtained from non-invasive surveys. I used 12 microsatellite markers to investigate genetic diversity, identify patterns of genetic structure, and delineate barriers to gene flow. I found high connectivity among most sites, with low but significant population genetic substructure (F_{st}=0.009, p<0.001). The Mantel test identified a weak pattern of isolation by distance, and genetic clustering algorithms failed to identify a biologically meaningful pattern of population substructure. MEMGENE analysis and multiple regression analysis based on univariate resistances in CIRCUITSCAPE indicated that wildfires acted as a barrier to gene flow, with sites separated by burned areas having higher genetic differentiation than expected due to isolation by distance alone. The POPGRAPH analysis identified genetically isolated sites among the managed portion of the range, and CIRCUITSCAPE analysis showed that the range is highly fragmented within the managed portion and contains limited connectivity corridors, whereas the unmanaged portion had high connectivity throughout. Overall, this study suggests that boreal woodland caribou are weakly genetically differentiated across the Brightsand Range, with isolation by distance and isolation by resistance contributing to variation in allele frequencies. However, while genetic differentiation was weak, conservation efforts will be required within the managed forest area to reduce the loss of genetic diversity by improving landscape connectivity.

LIBRARY RIGHTS STATEMENT

In presenting this thesis in partial fulfillment of the requirements for the MScF degree at Lakehead University in Thunder Bay, I agree that the University will make it freely available for inspection.

This thesis is made available by my authority solely for the purpose of private study and research and may not be copied or reproduced in whole or in part (except as permitted by the Copyright Laws) without my written authority.

Signature: _____

Date: _____

ACKNOWLEDGEMENTS

First, I would like to thank Ashley Thomson, my thesis supervisor, she has been the most outstanding mentor and supporter, and I am incredibly fortunate to have had her support throughout this process. Despite unavoidable absences during the project, her strength and dedication to assisting me throughout my journey have been beyond appreciated. She is an inspiration, and I hope to be the woman she is.

I also would like to thank my committee members Brian McLaren and Stephen Hecnar and the external examiner Adam Algar for their feedback, comments and time.

I thank others who assisted with this project, including Peter Hettinga, for providing his unpublished genotyping rules for guidance. And the timely responses from authors who clarified their research or gave assistance with their programs.

I would also like to acknowledge the financial support I have received throughout this project and thank the funding agencies, including the Ontario Graduate Scholarship, Resolute Forest Products Inc. for the Graduate Environmental Forestry Award, and the Natural Resources Management faculty for the Faculty Research Scholarship.

This research was supported by funding from Environment and Climate Change Canada, the Ministry of Environment Conservation, and Parks, and partner organizations, including Resolute, Domtar, Weyerhaeuser, Eacom, and the Forest Products Association of Canada.

Lastly, I would like to thank my family and my partner for keeping me accountable and for their support in balancing the pandemic, mental health and this two-year project. Without them, I would not be where I am today.

CAUTION TO THE READER

This MScF thesis has been through a formal process of review and comment by two faculty members and an external examiner. It is made available for loan by the Faculty of Natural Resources Management for the purpose of advancing the practice of professional and scientific forestry.

The reader should be aware that the opinions and conclusions expressed in this document are those of the student and do not necessarily reflect the opinions of either the thesis advisor, external examiner, committee members, the faculty or Lakehead University.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1:
LITERATURE REVIEW

LITERATURE REVIEW

GENETIC CONSEQUENCES OF POPULATION FRAGMENTATION

Habitat fragmentation is a driving force of reduced genetic diversity within populations (Rivera-Ortíz et al. 2015). Fragmentation is generally associated with reduced gene flow, leading to isolated subpopulations that become genetically differentiated and experience a loss of genetic diversity over time (Frankham 2019). In addition, small populations isolated from larger populations for multiple generations experience increased inbreeding and loss of genetic diversity (Keller and Waller 2002). Without increased gene flow, small populations under the effects of inbreeding and reduced genetic diversity are at increased risk of extinction (Frankham et al. 2010b; Rivera-Ortíz et al. 2015).

POPULATION GENETIC STRUCTURE

Patterns of Population Structure

Individual species' natural ecology, behaviour and local geography influence the patterns of population genetic structure (Hedrick 2012). Population genetic subdivision often results from reduced gene flow due to natural geographic dispersal barriers such as rivers, lakes, mountains, or patches of unsuitable habitat (McLoughlin et al. 2004; Funk et al. 2005; Parks et al. 2015). For example, forest types were found to influence white-crowned sparrows (*Zonotrichia leucophrys* Forster, JR) both morphologically and genetically, where the size of individuals depended on the forest type and the available

food sources (Welke et al. 2021). Behavioural dynamics, such as varying male and female dispersal patterns and site fidelity, may also influence population structure. For example, high site fidelity and limited male dispersal significantly contribute to the population genetic structure of timber rattlesnakes (*Crotalus horridus* L.) (Clark et al. 2008).

While natural population fragmentation occurs, anthropogenic disturbance is a leading cause of habitat fragmentation for many species (Titus et al. 2014). Where anthropogenic land development, conversion, or resource extractions change the landscape, reduced habitat connectivity may lead to reduced gene flow among population fragments (Fischer et al. 2012). For example, anthropogenic landscape alterations, including motorways, agriculture and urban landscapes, inhibit gene flow among populations of mountain goats (*Oreamnos americanus* Blainville) in addition to natural resistances of water-based landscapes (Parks et al. 2015). Motorways were also identified as barriers to gene flow in desert bighorn sheep (*Ovis canadensis nelsoni* Merriam) as they were associated with rapid loss of genetic diversity in isolated subpopulations (Epps et al. 2005).

Restoration and maintenance of gene flow among populations are critical to reducing the adverse genetic effects of small population sizes in population fragments (Frankham et al. 2010b). Through identifying connectivity corridors, barriers and pinch-points within landscapes, efforts to restore connectivity can be implemented (Proft et al. 2018). For example, management approaches to restoring connectivity corridors with barriers impeding connectivity, such as motorways, may include wildlife crossings with overpasses and underpasses (Clevenger and Waltho 2005). As motorways are known

to fragment landscapes inhibiting gene flow, as previously mentioned (Epps et al. 2005; Parks et al. 2015), wildlife crossings can support movement across motorways (Sawaya et al. 2013; Soanes et al. 2018). Maintaining and restoring connectivity among small or isolated populations have also been shown to reduce genetic diversity loss among fragmented landscapes and the extinction risk in tigers (*Panthera tigris* L.) (Thatte et al. 2018). Restoring unsuitable and degraded habitats can increase and support connectivity between target species and fragmented landscapes (Clauzel et al. 2015).

Quantifying Population Genetic Subdivision

Examining population structure has become essential to understanding gene flow within a population (Vonholdt et al. 2010). Two key population genetic structure patterns are isolation by distance (IBD) (Wright 1943) and isolation by resistance (IBR) (McRae 2006). In natural landscapes, many species exhibit IBD, where genetic differentiation increases with geographic distance among populations (Wright 1943). A driver of IBD is the limitation of a species' dispersal across the species range (Slatkin 1993). In contrast, IBR is the effect of natural or anthropogenic disturbances or barriers inhibiting gene flow between populations (McRae 2006).

Statistical methods for identifying patterns of population genetic structure include the calculation of genetic diversity statistics, such as allele frequencies, expected and observed heterozygosity, and inbreeding coefficients. Gene flow among populations is primarily measured with F-statistics, which partitions inbreeding within and among populations (Wright 1969). Proposed by Sewall Wright (1969), the F-statistics provides a measure of inbreeding within individuals (F_{IT}), within individuals relative to subpopulations (F_{IS}), and among subpopulations relative to the total (F_{ST}). The F_{ST} value

is an indicator of the level of population subdivision; the higher the F_{ST}, the higher the population subdivision. In 1987, Nei successfully linked F-statistics with heterozygosities using simple equations, including expected and observed heterozygosities (Nei 1987). Therefore, when the observed heterozygosity is lower than expected, it indicates inbreeding.

Another approach to understanding population genetic structure is the Analysis of Molecular Variance (AMOVA). AMOVA (Excoffier et al. 1992) identifies the hierarchical partitioning of genetic variance based on allele frequencies. The user assigns individuals to hierarchical groups, for example, populations and regions, and AMOVA then identifies the amount of genetic variation within and between each group. The F-statistics and AMOVA provide a foundation for understanding population structure.

To identify genetic structure following an IBD pattern, the Mantel test (Mantel 1967) was developed based on the theory of IBD proposed by Sewall Wright (1943). This test uses distance matrices for statistically significant correlations between genetic and geographic distances. However, researchers have criticized this method for erroneous *p* values, autocorrelation, and bias within the Mantel test (Guillot and Rousset 2013). Nevertheless, this test remains an important tool for identifying the presence of IBD (Guillot and Rousset 2013), as it can be effective as a simple test.

Population Structure Models

With the growth of population genetic studies, researchers have developed a variety of computational approaches to define population genetic structure. Standard approaches to examine population structure are individual-based Bayesian clustering

models. STRUCTURE (Pritchard et al. 2000) is one such model that does not require prior population information, such as geographic information, to detect population structure. As a result, STRUCTURE has become a fundamental tool for identifying population genetic structure patterns and has been used in thousands of studies (e.g. Estes-Zumpf et al., 2010; Jensen et al., 2020; Loxterman, 2011). Other models, such as TESS (Chen et al. 2007) and BAPS (Corander et al. 2008), are also often used but require prior population information, unlike STRUCTURE.

STRUCTURE's advantage over spatial approaches is that it has an improved ability to assign clusters among individuals with high admixture (Chen et al. 2007), while TESS, a spatial model, has the advantage of accounting for spatial autocorrelation (François and Durand 2010). BAPS differs from TESS and STRUCTURE because it is spatially explicit and uses spatial priors directly (Corander et al. 2003). Its advantage is that it does not assume the number of populations (i.e. clusters) (Corander et al. 2003). However, each model is based on different assumptions and has various advantages and limitations in detecting population structure. Therefore, using more than one Bayesian model is advised to detect population structure (Latch et al. 2006). Furthermore, while Latch et al. (2006) demonstrated that BAPS and STRUCTURE were consistently able to identify the number of clusters correctly, they both begin to lose success when $F_{ST} \leq 0.02$. Therefore, studies with weak population genetic structures require multiple models to verify the accuracy of the results, as models may otherwise identify spurious clusters (e.g. BAPS; Corander et al., 2008).

Inferences of population genetic structure within wild populations are challenging to confirm. Guillot et al. (2009) highlighted that a model's assignment of individuals to a

population cannot be validated and could generate false population clusters. Therefore, individual-based clustering models are best used to understand processes within a population and in conjunction with other models for verification. Users can also check conformity with allele frequencies and the Hardy-Weinberg equilibrium to confirm that clusters are significant (Guillot et al. 2009). In addition, consideration of reasonable biological and geographical interpretations by the researcher is recommended while examining the model's results. For example, in studies where the simple Mantel test can detect IBD, STRUCTURE may overestimate clusters present within the model (Frantz et al. 2009).

Moran's eigenvector maps (MEM) and population graphs are alternative approaches to understanding population structure. MEMGENE is a newer model that utilizes MEM to detect subtle and potentially cryptic spatial patterns based on genotypes and coordinates (Galpern et al. 2014). This model does not assign individuals to populations/clusters but seeks genetic differentiation across the landscape. While this model cannot be used to define population structure explicitly, it can support and assist with understanding population structure. POPGRAPH uses a graph theory approach to examine intraspecific genetic connectivity among individuals or populations (Dyer and Nason 2004). This model is unique as it only uses individual genotypes and assigned populations to analyze genetic connectivity, where edges between nodes (populations) denote connectivity. This model can be useful for detecting isolated individuals or populations, but it is not as commonly used (Manel and Holderegger 2013). Nevertheless, such models for estimating population genetic structure have been effective in numerous studies, and researchers have made expert assumptions on

correlations between population structure and potential landscape barriers inhibiting gene flow (e.g., Ball et al. 2010b).

LANDSCAPE GENETICS

Utility of Landscape Genetics in Conservation

Landscape genetics is a relatively new field of study, named in 2003, and has effectively combined landscape ecology and population genetic disciplines (Manel et al. 2003). Landscape genetics is used to identify the effects of landscape variation on population structure and gene flow (Manel et al. 2003; Holderegger and Wagner 2008), providing fundamental knowledge on the impact of (dis)connected landscapes on gene flow (Storfer et al. 2007). It is essential to understand that landscape genetics is not a novel field of study, but a branch developed over time and recently classified. Researchers in past studies sought to explain population genetic structure within and across landscapes (Manel et al. 2003). Over the last two decades, it has become a widely-used approach for identifying landscape disturbances, including natural or anthropogenic sources, that may act as barriers to gene flow (Manel et al. 2003; Storfer et al. 2007; Holderegger and Wagner 2008).

Anthropogenic and natural landscape features can explain population structure, genetic connectivity, and spatial genetic differentiation. For example, genetic clustering algorithms STRUCTURE, GENELAND and BAPS identified motorways as a barrier to gene flow among red deer (*Cervus elaphus* L.) (Frantz et al. 2012). However, motorways did not significantly influence gene flow in wild boar (*Sus scrofa* L.) (Frantz et al. 2012). In addition, a study of an Australian marsupial identified that land clearings for farmland

inhibited gene flow between fragments of isolated vegetated forest patches (Lada et al. 2008).

By correlating genetic discontinuity and spatial heterogeneity, landscape genetic studies can identify anthropogenic or natural barriers to gene flow among populations (Manel et al. 2003; Storfer et al. 2007; Holderegger and Wagner 2008). Because landscape genetic studies are primarily concerned with patterns of population genetic structure, they rely on analytical approaches that can detect gene flow across heterogeneous landscapes and explain spatial genetic variation (Storfer et al. 2007; Holderegger and Wagner 2008).

Landscape Genetic Models

Landscape genetic analyses often combine population genetic structure analyses with geographic locations and spatial data on landscape composition, such as landcover (Spear et al. 2010). One common landscape genetic modelling approach is the least-cost path analysis used to calculate movement costs based on landscape resistance (Adriaensen et al. 2003). Resistance is a measure of the cost for an individual to move through a space (Sawyer et al. 2011). Other modelling approaches include partial Mantel tests correlating matrices of landscape or geographic distances with genetic distances (Manel and Holderegger 2013). Researchers have also used multiple regression approaches. A newer model incorporating raw genetic data is MEMGENE, which explains spatial genetic variation patterns using genetic data and landscape rasters (Galpern et al. 2014). MEMGENE (Galpern et al. 2014) was designed to detect fine-scale spatial genetic patterns present within populations by identifying spatial genetic variation across landscape resistances. The model is based on Moran's

eigenvector maps and regresses genetic distance against raw predictors instead of geographic distances (Galpern et al. 2014). The model then generates MEMGENE variables which are eigenvectors used to identify the significance of the spatial variables. MEMGENE differs from multivariate models in that it incorporates genetic and spatial data to identify autocorrelation and fine spatial patterns (Galpern et al. 2014). While it has not been used as often as the world-renowned CIRCUITSCAPE model, MEMGENE effectively tests the statistical significance of spatial resistances across landscapes using genetic data (e.g. Priadka et al. 2019).

An alternative model, CIRCUITSCAPE (Shah and McRae 2008), uses only landscape rasters to explain patterns of gene flow across landscapes based on circuit theory (Shah and McRae 2008; McRae et al. 2014). McRae (2006) proposed a framework for applying circuit theory to landscape genetics studies where resistance values between two points are calculated using random walks instead of least cost-path analyses. The proposed method uses circuit theory for connectivity analyses based on the foundation that within circuit networks, there are multiple pathways through which a current can flow (McRae et al. 2008). The circuit network scenario is analogous to gene flow in natural landscapes as the movement of individuals between populations is not limited to a single path, but landscape resistance influences available pathways (McRae and Beier 2007). The circuit theory approach is applied in the software package CIRCUITSCAPE (Shah and McRae 2008), where the network consists of nodes and edges representing connectivity between features (i.e. core areas, sites, individuals) (McRae et al. 2008). The circuit's currents identify the flow between nodes through resistors, and based on random walks, the current density between nodes is used to identify

landscape corridors (McRae et al. 2008). CIRCUITSCAPE has been used worldwide to help solve conservation questions (Dickson et al. 2019), such as predicting patterns of gene flow across landscapes using landscape rasters (McRae 2006). The success of the CIRCUITSCAPE model in explaining gene flow without genetic data has been fundamental for conservationists (Dickson et al. 2019).

The ability to identify population genetic structure and delineate the features contributing to spatial genetic variation has become a powerful tool for conservation. However, it remains under-utilized for conservation management in literature (Bowman et al. 2016). With population genetic structure and landscape genetics, researchers can identify cryptic, isolated, and fragmented populations represented by reduced genetic diversity and connectivity to implement appropriate management strategies to support the population's viability. These studies can also assist with threatened and endangered species to obtain the best available information to protect the species and reduce the likelihood of extirpation or extinction within a species (Bowman et al. 2016).

Influence of Scale

Success in identifying drivers of population genetic subdivision is often scale-dependent (Cushman and Landguth 2010; Galpern et al. 2012b; Landguth and Schwartz 2014). Variables that influence the ability to detect population genetic subdivision include spatial scale (Cushman and Landguth 2010; Galpern et al. 2012b), sampling density (Landguth and Schwartz 2014), and time lag (Landguth et al. 2010). For example, a coarse grain size among spatial scales can reduce the statistical power of landscape genetic studies using rasters to define features/barriers inhibiting gene flow among the populations of interest (Cushman and Landguth 2010). At the same

time, the sampling density across a study area can also affect the interpretation of the results. For example, sparse sampling across a range can result in an overestimation of genetic drift, making it appear that populations are significantly subdivided when the true pattern is of IBD (Landguth and Schwartz 2014). In contrast, clustered sampling can lead to overestimating the effect of landscape barriers on gene flow (Landguth and Schwartz 2014). Alternatively, time lags may limit the ability to detect new genetic patterns among populations which can take 1 – 15 generations (Landguth et al. 2010). Additionally, among species with large dispersal ranges, time may cause a species to rapidly lose genetic evidence of past landscape barriers in less than 15 years, indicating that landscape genetic studies could effectively identify current barriers to gene flow (Landguth et al. 2010).

BOREAL WOODLAND CARIBOU BEHAVIOUR AND ECOLOGY

Boreal woodland caribou (*Rangifer tarandus caribou* Gmelin) (hereafter referred to as boreal caribou), are sedentary forest-dwelling caribou found in the boreal forest of Canada (Thomas and Gray 2002). The species preferred habitat consists of large areas of undisturbed, lichen-rich, mature forest (O'Brien et al. 2006; Environment Canada 2011a) with adjacent expanses of peatlands for calving (Rettie and Messier 2000; Thomas and Gray 2002; Environment Canada 2011a). Calving occurs in late May and early June (Ferguson and Elkie 2004a) where females disperse to shorelines and lake islands which may assist with predator avoidance to reduce calf mortality (Bergerud 1985; Carr et al. 2011). In the fall, the polygamous species (L'Italien et al. 2012; McFarlane et al. 2021) forms small groups of mixed sex for the rutting period and

throughout the winter season (Fuller and Keith 1981; Rettie and Messier 1998). Boreal caribou conceive one calf per year with high pregnancy rates. However calf survival rates are low (Stuart-smith et al. 1997; Rettie and Messier 1998).

Boreal caribou are listed as a threatened species under the Canadian Species at Risk Act (SARA) and the Ontario Endangered Species Act (ESA). Evidence has shown that boreal caribou populations face declining population sizes and reduced genetic diversity associated with habitat fragmentation (Environment Canada 2011a; Thompson et al. 2019). In addition, boreal caribou are particularly sensitive to and avoid anthropogenic drivers of habitat fragmentation such as linear features (e.g. roads) and recent harvest areas (Dyer et al. 2001, 2002; DeCesare et al. 2012; Beauchesne et al. 2013). While boreal caribou favour black spruce stands, they will select jack pine stands and other forest types based on predator avoidance and forage opportunities in natural landscapes (Rettie and Messier 2000). They also exhibit higher selectivity for mature forests when anthropogenic disturbances are present nearby (Moreau et al. 2012). A primary driver of habitat selectivity among boreal caribou may be apparent competition (Wittmer et al. 2007) due to landscape disturbances, including roads, harvests, and wildfire (Wittmer et al. 2007; Courbin et al. 2009; Latham et al. 2011). In addition, lake edges and islands are associated with predator avoidance mechanisms (Bergerud 1985; Cumming and Beange 1987).

Daily movements of boreal caribou vary significantly between seasons and individuals (Ferguson and Elkie 2004a). A study of radio-collared boreal caribou in northwestern Ontario demonstrated that the most significant movement occurred during the early winter and spring seasons (Ferguson and Elkie 2004a). While average daily

movement rates were reported lowest at 0.9 and 1.1 km per day during late winter and calving seasons and up to 2.5 km per day during the early winter season in the study of Ferguson and Elkie (2004a), boreal caribou have been reported to move up to 54.6 km in a single day (Rudolph and Drapeau 2012).

Anthropogenic disturbances have been shown to influence caribou behaviour and movement (Moreau et al. 2012; Beauchesne et al. 2014). For example, boreal caribou generally avoid roads and recent harvest areas (Dyer et al. 2001; Beauchesne et al. 2013; Priadka et al. 2019). An analysis by Environment Canada (2011b) found that the minimum distance from roads where suitable habitat can be found was 500 m, while an Alberta study identified the maximum significant avoidance distances from roads was approximately 250 m (Dyer et al. 2001). However, other studies suggest that greater distances from roads are more representative of caribou land use. For example, one study showed that functional habitat loss begins within 1 km of a logging road (Schindler et al. 2007), while another found that most individuals avoided areas within 2 to 3 km of active roads (Cumming and Hyer 1998). Dyer et al. (2002) concluded that roads are not a complete barrier to the movement of individuals but are semi-permeable to individual movement based on GPS collar tracking data. In addition to roads, industrial linear features such as seismic lines are avoided by boreal caribou. It is likely that the avoidance of linear features, including roads, is due to their facilitation of wolf movement (Latham et al. 2011; Fryxell et al. 2020).

Another anthropogenic disturbance impacting caribou distribution is forest harvesting (Rettie and Messier 2000; Moreau et al. 2012). A study by Vors et al. (2007) found that caribou were extirpated from areas affected by forest harvesting within two

decades following the disturbance. However, the effects of forest clear-cuts as barriers to caribou movement are not immediate and begin as regeneration occurs (Rettie and Messier 2000; Beauchesne et al. 2013). The avoidance of regenerating clear-cuts by boreal caribou appears to be linked to apparent competition, as the use of regenerating forests by moose increases predation risk from wolves (Courbin et al. 2009). Like recent harvest areas, recently burned areas are often rich in moose browse, and caribou may avoid these areas due to increased predation risk (Wittmer et al. 2007). However, it has also been shown that caribou may demonstrate strong site fidelity to recent wildfire-disturbed areas (Dalerum et al. 2007), as lichens may continue to be available for five years post-fire (Schaefer and Pruitt 1991). Additionally, caribou can coexist with natural wildfire disturbances but may be extirpated from areas where harvest disturbance exceeds historical fire patterns (Stewart et al. 2020)

Landscape features, such as waterbodies, have also been reported to influence caribou distribution (Bergerud 1985; Ferguson and Elkie 2005). In winter, caribou have been found to select areas with medium-sized lakes (5-100 ha in size) (Ferguson and Elkie 2005). However, individuals were generally found to avoid larger lakes. During calving seasons, caribou use lake edges and islands, which are also associated with predator avoidance (Bergerud 1985; Cumming and Beange 1987). The use of lakes likely explains why a study by Vors et al. (2007) found no significant influence of lakes on caribou range occupancy. A landscape genetics study by Priadka et al. (2019) found that waterbodies were a significant factor in explaining patterns of spatial genetic variation in boreal caribou and, in some areas, lakes were as important as roads in explaining patterns of gene flow. While waterbodies can contribute to resistance of gene

flow, studies suggest that they do not represent a complete barrier to the movement of boreal caribou (Galpern et al. 2012b, 2014).

Brightsand Range of Ontario

The Brightsand Range of Northwestern Ontario contains some of the southernmost boreal caribou populations (MNRF 2014). It is approximately 22,000 km^2 in size and contains an abundance of lakes with islands and irregular shorelines ideal for calving, shallow soils and bedrock with jack pine and black spruce forests, and a short fire cycle. While the northern range provides ideal natural habitat within protected landscapes, the southern range has a history of intensive forest management and forest disturbances. Forest management occurs in the south and west within the Black Spruce Forest, Caribou Forest, English River Forest, Lac Seul Forest, and Lake Nipigon Forest. In 2011, a total of 43.5% of the range was classified as disturbed habitat (10.4% natural, 33.1% anthropogenic, and 2.6% natural and anthropogenic overlap). The 2011 Range Assessment identified a declining population trend and a minimum animal count of 224 individuals, with likely greater than 250 individuals present within the range (MNRF 2014). The highest caribou densities are found in the northern portion of the range, in and adjacent to the protected forests of Wabakimi Provincial Park. In contrast, the southern portion of the range demonstrates low occupancy by boreal caribou.

Figure 1. Location of the Brightsand Range in Ontario (MNRF 2014).

Landscape Genetic Studies of Boreal Woodland Caribou

Within the literature, numerous studies have successfully identified landscape features associated with resistance or barriers to the movement of boreal caribou (Gubili et al. 2017; Priadka et al. 2019). For example, Priadka et al. (2019) found that waterbodies and roads significantly influenced the spatial genetic variation of boreal caribou populations across Alberta, Saskatchewan and Manitoba. Contrastingly, a study by Weckworth et al. (2013) indicated that caribou populations in Alberta and Saskatchewan were not significantly affected by anthropogenic disturbances when

resource selection functions and effective population sizes were accounted for (Weckworth et al. 2013).

The population structure of boreal caribou is often studied at broad spatial scales (e.g. Yannic et al. 2016; Priadka et al. 2019; Thompson et al. 2019). Pelletier et al. (2019) examined the fine-scale population genetic structure of a caribou population in Quebec that had been isolated from the main distribution since the 1950's. The study identified significant genetic subdivisions in the second of two sampling periods and suggested that genetic structuring was linked to valleys between populations and a main road exacerbating the effect of a valley geographically separating sub-populations. However, Galpern et al. (2012b) analyzed a relatively fine spatial scale and found that genetic patterns may only be significant at some spatial grains of landscape data. Interestingly, roads were identified as posing resistance to gene flow even at larger grains, while the influence of other anthropogenic and natural disturbances was only detectable at more restricted spatial grains (Galpern et al. 2012b).

Landscape genetics studies from Manitoba and Saskatchewan have identified that waterbodies and roads may pose barriers to gene flow in boreal caribou (Galpern et al. 2012b; Priadka et al. 2019). In contrast, an Alberta study did not find a significant effect of anthropogenic disturbance on population genetic differentiation (Weckworth et al. 2013). Currently, limited information exists on patterns of gene flow in relation to landscape disturbance for boreal caribou in Ontario. One population genetic study of Ontario/Manitoba boreal caribou found higher inbreeding and reduced genetic diversity in portions of the study area with higher disturbance levels (Thompson et al. 2019). Disturbed populations located in the southern extent of the study area were found to be

highly isolated, exhibit low connectivity, and were more genetically differentiated than northern populations. While the authors concluded that high levels of anthropogenic disturbance likely contributed to lower genetic diversity and greater differentiation of caribou populations from a southern population within the Brightsand Range, they did not specifically test the influence of different natural and anthropogenic disturbance types on gene flow.

Molecular Markers

The use of molecular markers has been an effective tool for examining population genetic diversity and structure (Schwartz et al. 2007). Molecular markers can provide valuable insights into genetic diversity estimates, including allelic diversity (allelic richness), expected and observed heterozygosity, allele frequencies, fixation indices and inbreeding coefficients. Technological advancements have also allowed the use of molecular markers for understanding population structure and gene flow across landscapes. Microsatellites are a type of genetic marker that is widely used for studies of neutral population genetic diversity and structure in plants and animals (Gemmell et al. 1997; Vieira et al. 2016). Microsatellites, also known as short tandem repeats (STRs) or simple sequence repeats (SSRs) (Schlötterer 2004; Putman and Carbone 2014), are genetic markers found mainly in non-coding genomic regions (Ellegren 2004) and are highly polymorphic due to their relatively rapid mutation rate (Ellegren 2004). Their hypervariability likely occurs due to slippage during DNA replication, causing repeats to be inserted or deleted (Tautz 1989). Microsatellites were the marker of choice for most population genetic studies up until recently, but their use has decreased significantly

over the last decade due to technological advances enabling the widespread use of single nucleotide polymorphism (SNPs) (Schlötterer 2004; Grover and Sharma 2016).

Non-Invasive DNA Sampling

Fecal DNA sampling is an effective method for obtaining genetic data for species that are difficult to sample (Schwartz et al. 2007; Woodruff et al. 2015). As the collection of fecal samples does not require the costly and time-consuming physical capture of an individual, this increases the total number of samples that can be collected and minimizes risks to animal welfare (Banks and Piggott 2022). Non-invasive sampling approaches are preferred for the genetic studies of boreal caribou as they are listed as threatened federally and provincially. As part of the *Recovery Strategy for the Woodland Caribou (Rangifer tarandus caribou), Boreal Population, in Canada,* it is a priority to minimize disturbances during research or monitoring methods and use the least intrusive methods possible (Environment Canada 2012). For boreal caribou, winter surveys are preferred as they increase the visibility of caribou ‘signs’ (i.e. tracks, cratering, and individuals) (Courtois et al. 2003b) and allow for increased preservation of fecal samples due to below-freezing temperatures (Woodruff et al. 2015). However, genetic analysis of non-invasively collected samples can be challenging if poor quality DNA template is obtained (Maudet et al. 2004; Waits and Paetkau 2005; Woodruff et al. 2015). For example, the degradation of fecal DNA samples can occur due to climate, weather, season, age, and collection and storage conditions. Poor-quality DNA templates can reduce the success in amplifying DNA and the accuracy of genotyping (Waits and Paetkau 2005; Woodruff et al. 2015). While some past fecal DNA studies have suffered from issues due to sample cross-contamination and genotyping error,

several approaches can be used to reduce these problems (Waits and Paetkau 2005). For example, winter collection of fecal samples can provide a high-quality DNA template (and minimize genotyping errors due to allele dropout) as cold temperatures help preserve DNA (Maudet et al. 2004; Ball et al. 2007; Hettinga et al. 2012). Recent studies of boreal caribou have demonstrated high success in using fecal DNA samples (Priadka et al. 2019; Thompson et al. 2019) provided that they include methods, such as repeat sample analysis, to minimize genotyping error (Ball et al. 2007, 2010).

CHAPTER 2:
LANDSCAPE GENETIC ANALYSIS OF POPULATION STRUCTURE AND BARRIERS TO GENE FLOW IN BOREAL WOODLAND CARIBOU (*RANGIFER TARANDUS CARIBOU*)

INTRODUCTION

Wildlife habitat connectivity is critical for species conservation in anthropogenically altered habitats (Weckworth et al. 2013). Anthropogenic landscape changes are a leading cause of habitat loss, degradation, and fragmentation (Rivera-Ortíz et al. 2015), directly influencing the demography and genetics of populations (Epps et al. 2005). Habitat fragmentation is often associated with reduced gene flow among population fragments, leading to increased rates of loss of genetic diversity and inbreeding depression (Loxterman 2011; Rivera-Ortíz et al. 2015). Thus, identifying landscape factors influencing genetic discontinuity has become a central goal of studies of threatened species (Loxterman 2011; Jensen et al. 2020). Landscape genetic studies are frequently used to examine how landscape features influence the genetic structure of populations (Emel et al. 2019). For instance, landscape genetic studies have been used to examine the contributions of isolation by distance (IBD) (Tammeleht et al. 2010; Gariboldi et al. 2016; Priadka et al. 2019) and isolation by resistance (IBR) to the structure of populations threatened by anthropogenic and natural disturbance (Tammeleht et al. 2010; Priadka et al. 2019). Successful identification of disturbances associated with reduced gene flow via landscape genetic approaches has become fundamental to the design of strategies for species' management and recovery (Proft et al. 2018).

Habitat fragmentation is associated with declining population sizes and reduced genetic diversity in threatened boreal woodland caribou (*Rangifer tarandus caribou*)

(Environment Canada 2011a; Thompson et al. 2019). Boreal woodland caribou are sensitive to disturbances such as anthropogenic linear features and recent harvest areas (Dyer et al. 2001, 2002; DeCesare et al. 2012; Beauchesne et al. 2013). Populations occupying fragmented ranges in Manitoba and Ontario show higher inbreeding and reduced genetic diversity than those from ranges with low disturbance levels (Thompson et al. 2019). Similarly, the genetic structure of populations in Manitoba and Saskatchewan was significantly influenced by waterbodies and roads (Priadka et al. 2019). In contrast, the genetic structure of populations from Alberta was not significantly affected by anthropogenic disturbance, including roads, linear features, and forest harvests (Weckworth et al. 2013). Contrasting results among studies could be a consequence of the differing scale at which these studies were conducted, as research has shown that success in identifying drivers of population subdivision is often scale-dependent (Cushman and Landguth 2010; Galpern et al. 2012b; Landguth and Schwartz 2014).

The Brightsand Range of northwestern Ontario is among the southernmost continuous boreal woodland caribou ranges (MNRF 2014). It includes a large portion of Wabakimi Provincial Park in the north and multiple active Forest Management Units in the west and south. *The Integrated Range Assessment for Woodland Caribou and their Habitat: Brightsand Range 2011* identified that 43.5% of the range was disturbed by various anthropogenic and natural disturbances (MNRF 2014). It was also considered uncertain if the range could support a self-sustaining boreal woodland caribou population. Currently, there is little available research on population structure and landscape genetics for boreal woodland caribou in the Brightsand Range due to a lack

of range-level studies. A population genetic study of Ontario and Manitoba boreal woodland caribou populations revealed that groups of individuals located in the southern portion of the boreal caribou range in northwestern Ontario are genetically isolated with comparatively lower genetic diversity than more northern groups (Thompson et al. 2019). However, the broad interprovincial scale of the analysis combined with small sample sizes from the Brightsand range limits the utility of this study for understanding population structure and gene flow at the individual range level.

My objective for this study was to use fine-scale, intensive sampling of boreal woodland caribou within the Brightsand Range of Ontario to examine patterns of population genetic diversity, genetic substructure, and gene flow. My predictions were that (i) sites from the northern portion of the range will exhibit significantly higher genetic diversity than sites from the southern range, (ii) sites from the northern and southern portions of the range will be significantly genetically differentiated, (iii) sites from the southern range will exhibit lower levels of gene flow compared to northern sites, and (iv) anthropogenic disturbances, including roads and harvest areas, increase levels of among-site variation compared to a pattern of isolation by distance alone.

The results of this study will be useful in setting conservation objectives for boreal woodland caribou by identifying sub-populations that may be suffering from reduced genetic diversity and limited gene flow. Furthermore, understanding the spatial genetic structure and landscape connectivity of boreal woodland caribou can improve the restoration or conservation of connectivity corridors and the management of landscape barriers.

METHODS

GENETIC SAMPLING

Fecal DNA samples were obtained as part of a noninvasive mark-recapture study conducted in the winter of 2020 (Thomson et al. 2021). A fixed-wing aircraft was used to conduct reconnaissance surveys following latitudinal flight lines spaced at five-kilometre intervals (Figure 2). Two observers recorded information and GPS coordinates for all caribou sightings and signs (i.e., tracks/trails, craters, slushing pits, and observed individuals) during the reconnaissance flights. The information was relayed to a second crew operating out a helicopter which was used to land at sampling sites for fecal pellet collection. A total of 1.4 fecal pellet clusters per estimated number of individuals was collected at each site to increase the probability of sampling all animals represented by the fecal pellets at that location. Samples were collected to avoid DNA contamination by collecting fecal pellets from distinct pellet groups using single-use tools. Each sample was labelled with a unique identifier, including the Site ID and Pellet Group ID (Appendix I). In total, 788 caribou fecal pellet samples were collected during the winter 2020 sampling period.

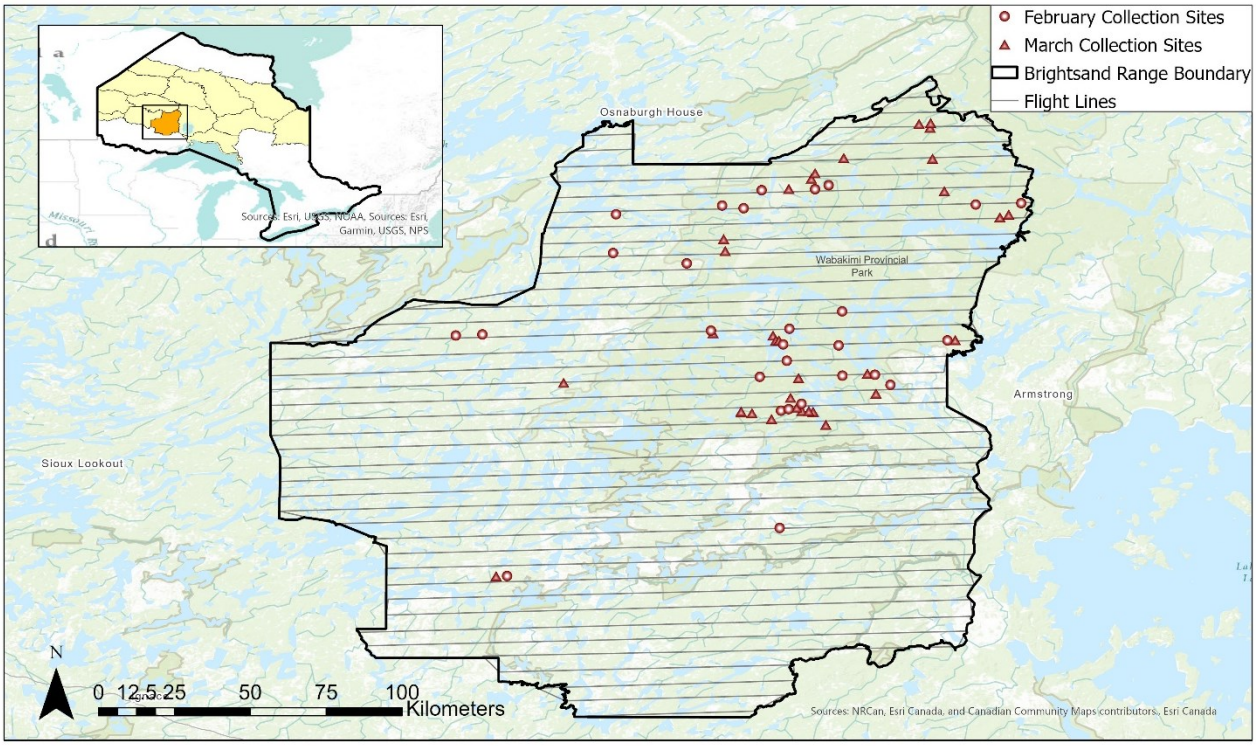


Figure 2. Location of 5-km aerial transects and fecal pellet sampling sites.

MOLECULAR ANALYSIS

Genomic DNA was extracted from the mucosal membrane of fecal samples using the DNeasy Blood and Tissue Kit (Qiagen) following a protocol (Appendix II) modified from Ball et al. (2007). Extracted DNA samples were diluted to approximately 20 ng/µl prior to amplification. The polymerase chain reaction (PCR) was used to determine the multi-locus genotype for each sample based on the analysis of 15 microsatellite markers: RT1, RT5, RT6, RT7, RT9, RT24, RT27, RT30 (Wilson et al. 1997), BM4513, BM6506, BM848, BM888, Map2C (Bishop et al. 1994), BMS1788 (Yannic et al., 2014 - modified from Cronin et al., 2005) and NVHRT30 (Røed and Midthjell 1998). Three multiplex reactions containing the following markers were used: Multiplex 1 – RT6, BM6506, BMS1788, NVHRT30, RT24; Multiplex 2 – RT27, RT9,

RT7, RT5; Multiplex 3 – Map2c, RT1, BM888, BM848, BM4513, RT30. PCR reaction mixtures contained Qiagen Multiplex PCR Master Mix 2x, forward and reverse primers, nuclease free water, and DNA at a reaction volume of 10 µl per sample (Appendix III). The thermocycling protocol of Yannic et al. (2016) was used and included an initial denaturing at 95°C for 15 minutes, followed by 35 cycles of 94°C for 45 seconds, 54°C for 90 seconds, and 72°C for 60 seconds; and a final extension of 72°C for 30 minutes. Samples were analyzed at the Toronto Centre for Applied Genomics (Toronto, Ontario) on an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, California) to determine microsatellite peak profiles.

Alleles were scored using GENEIOUS PRIME software version 2022.1.1 (Biomatters Inc, San Diego, California) following formalized scoring rules (Appendix IV). The R-package MSATALLELE v.1.05 (Alberto 2009) was used to determine the most likely bin assignments and bin names for each locus. Two people independently scored microsatellite profiles based on a consensus of genotyping rules regarding the morphology of allele peaks. Numerous samples underwent repeat amplification to reduce genotyping errors to ensure that the genotyping process and rules were consistent and accurate. A final error rate test was completed on 140 samples (~17% of samples) that were re-amplified and genotyped. Any score difference, except missing data, was identified as an error.

ALLELEMATCH v.2.5.1 (Galpern et al. 2012a) was used to identify unique and matching genetic profiles. Allele profile matches with a full-sib probability (Psib) <0.001 (Woods et al. 1999), and no more than two mismatching alleles were considered as matching genotypes calculated by ALLELEMATCH. Based on the results, duplicate

genotypes within sites were identified and removed. For example, if an individual was represented by multiple fecal pellet locations from the same location, then the duplicate genotypes from that location were be removed. Duplicate genotypes originating from different sites were retained, as this represents an individual that was marked at multiple sampling locations. The pruned dataset contained a total of 427 genotypes representing 340 individuals distributed across 61 sites. Unless otherwise stated, all analyses presented herein are based on this dataset of 427 genotypes with duplicate within-site genotypes removed. MICROCHECKER v2.2.3 (Van Oosterhout et al. 2004) was used to identify genotyping errors, null alleles, and large allele dropout among unique genotypes. To utilize the program, alleles were renamed strictly for MICRO-CHECKER using a standard naming convention (Appendix V). However, the name of the allele in some cases merely represents the distance of an allele from another (rounded) as opposed to the actual size of the allele.

I used the R package POPPR v.2.9.1 (Kamvar et al. 2015) to examine the power of the selected microsatellite loci to identify unique genotypes. This was done to ensure that the number of unique genotypes would not increase exponentially with additional microsatellite markers. The function *genotype_curve()*, randomly selects loci for each number of available loci 1000 times. The distribution of the identifiable number of unique genotypes per number of loci was then displayed within a box plot to determine the maximum number of identifiable unique genotypes in the dataset.

GENETIC DIVERSITY

I used GENEPOP v.1.1.7 (Rousset 2008) to test for Hardy Weinberg Equilibrium (HWE) and linkage disequilibrium using only unique genotypes (across-site duplicates removed, n = 340). Hardy-Weinberg equilibrium tests were conducted using the function *test_HW()*, based on 5,000 dememorizations, 2,500 batches, and 5,000 iterations. Tests of linkage disequilibrium (LD) were conducted using the function *test_LD(),* a log-likelihood ratio statistic (G-test), with 5,000 dememorizations, 2,500 batches, and 5,000 iterations. Bonferroni correction was not applied. GENALEX v.6.51b2 (Peakall and Smouse 2012) was used to determine the number of alleles (A), allelic richness (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), and the inbreeding coefficient (F) for all sites where at least five genotypes were sampled (total of 39 sites).

I conducted an Analysis of Molecular Variance (AMOVA) (Excoffier et al. 1992) using GENALEX v.6.51b2 (Peakall and Smouse 2012) to examine whether allele frequencies differed significantly among regions, and among sites within regions. Regions were used to group individuals from sites sampled within the managed or unmanaged portions of the Brightsand Range (Figure 3). The AMOVA results were used to estimate how genetic variation is distributed among regions, among sites within regions, within sites, among individuals within sites, and within individuals. In addition, F-statistics were calculated to determine the extent of genetic differentiation for each hierarchical partition. The F-statistics were also used to calculate the pairwise genetic differentiation between groups of individuals sampled at each site.

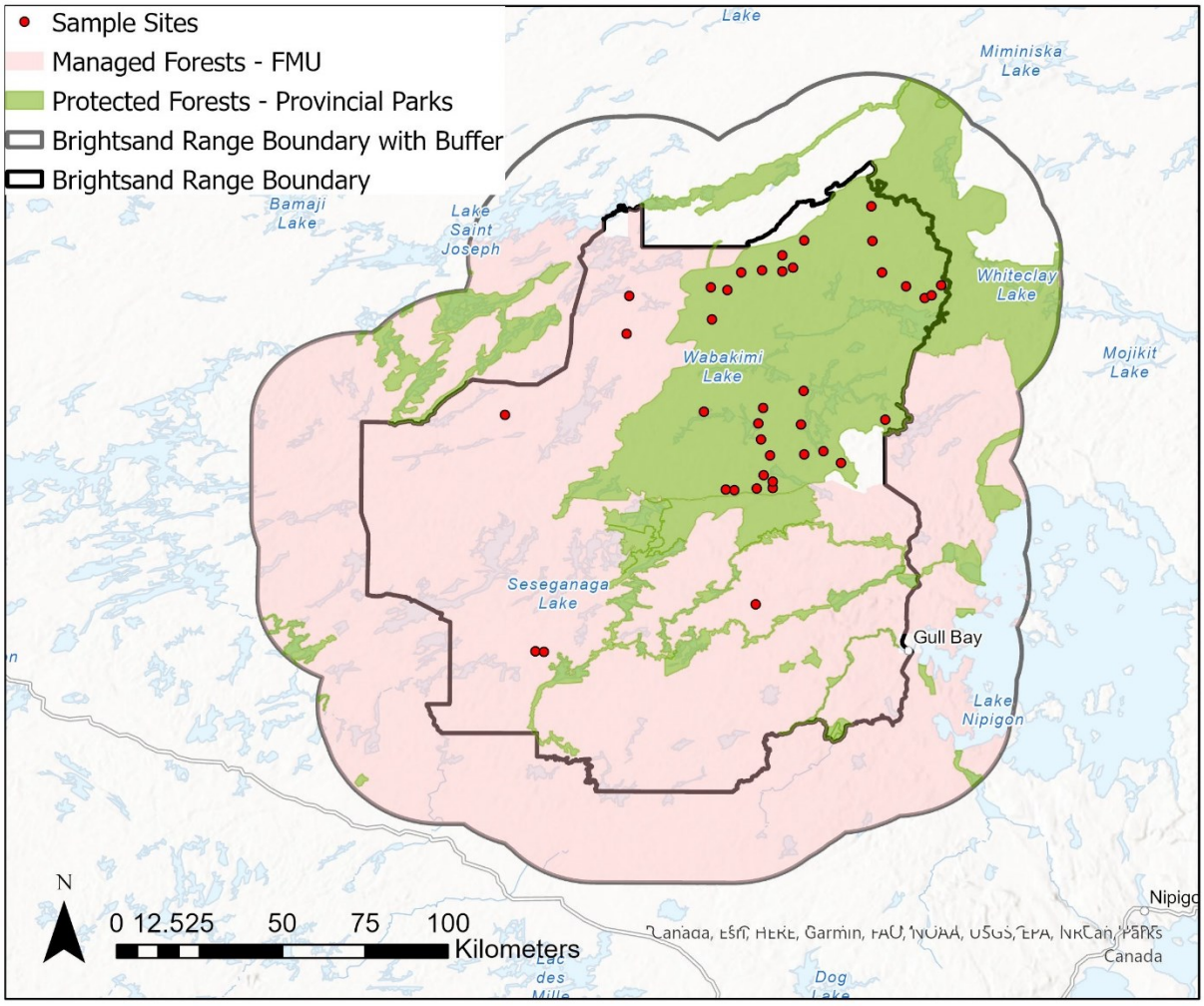


Figure 3. Map of Brightsand Range of Ontario with unmanaged forest areas (Provincial Parks) in green and managed forest areas (FMU = Forest Management Units). Uncoloured regions are not assigned to either category. Including sample sites analyzed throughout the study (n ≥ 5).

The Mantel test (Mantel 1967) was used to determine whether sites demonstrate a significant pattern of IBD among sites (Wright 1943). I regressed F_{ST} (genetic distance) against geographical distance using the R package *ADE4* (Chessel et al. 2004) with the function *mantel.rtest()* (Dray et al. 2022) with 9,999 permutations. I tested the assumptions of linearity and homoscedasticity with diagnostic plots using a linear model between the distance matrices. I also tested the assumptions of homoscedasticity in R-

package *CAR* (Fox and Weisberg 2019) with the non-constant variance score test (*ncvTest()*).

POPULATION GENETIC STRUCTURE

Multiple software programs relying on slightly different algorithms and assumptions were used to examine patterns of population genetic structure. First, I used STRUCTURE v.2.3.4 (Pritchard et al. 2000), an individual-based Bayesian clustering software, to examine the genetic population structure by identifying the number of genetic clusters (K) present within the dataset. STRUCTURE is a commonly used Bayesian clustering model that utilizes a Markov chain Monte Carlo (MCMC) algorithm based on the following assumptions (i) loci markers are unlinked, (ii) loci are in linkage equilibrium, and (iii) populations are in a state of Hardy-Weinberg equilibrium (Pritchard et al. 2000). This model uses population designations (i.e., site number) as a spatial prior but does not use spatially explicit information (Pritchard et al. 2000). I tested the population structure using the Admixture Model as there was an assumption of high admixture due to the high dispersal ability of caribou (Cumming and Beange 1987; Rudolph and Drapeau 2012). The model ran for 100,000 MCMC iterations and a burn-in period of 100,000 for each K value ranging from 1 to 10. A total of 20 replicates for each K value were completed. Next, I used STRUCTURE HARVESTER Web v0.6.94 (Earl and vonHoldt 2012) to identify the likely number of genetic clusters (K) using the Evanno method (Evanno et al. 2005). The Evanno method is an ad hoc method that is more likely to identify the number of clusters using the “modal value of ΔK”, except when K = 1 (Evanno et al. 2005). When K = 1, the modal value of K is used with other information

specified within the STRUCTURE documentation (Pritchard et al. 2010) to identify if the actual number of clusters is K = 1. The graphical and summation of cluster assignments were obtained using CLUMPAK (Kopelman et al. 2015), an automated amalgamation of CLUMPP (Jakobsson and Rosenberg 2007) and DISTRUCT (Rosenberg 2004) to remove label switching, identify multi modality and visualize the results.

Second, I used TESS V2.3.1 (Chen et al. 2007; Durand et al. 2009), a spatially explicit Bayesian clustering software. TESS differs from STRUCTURE in that it assumes spatial autocorrelation, IBD is accounted for by autocorrelation residuals, and geographic clines are accounted for by trend surfaces (François and Durand 2010). The model uses an MCMC algorithm similar to STRUCTURE but has a prior distribution utilizing a Log-Gaussian random field for its admixture model (François and Durand 2010). To estimate the likely number of clusters, I used the Admixture Model with 15 independent runs of each K_{max} value ranging from 2 to 10 with 30,000 iterations and 20,000 burn-in iterations. I also used the following parameters: a conditional auto-regressive (CAR) variance of 1.0, a linear trend as there was a slight IBD pattern, and a spatial interaction parameter of 0.6 to account for spatial autocorrelation. Finally, I used the Deviance Information Criterion (DIC) to identify the effective number of clusters, which is analogous to the STRUCTURE ΔK values (François and Durand 2010). The effective number of clusters was identified where the DIC value plateaus when plotted against K_{max} (Durand et al. 2009). CLUMPAK (Kopelman et al. 2015) was also used to visualize the results across runs of K_{max}.

The third Bayesian clustering model that I used was BAPS v.6 (Corander et al. 2008). This model uses a different algorithm approach than STRUCTURE and TESS, using

a stochastic optimization algorithm instead of MCMC when possible (François and Durand 2010). This method greatly accelerates computation time. The model uses a prior distribution inspired by Markov Random fields and requires or generates parental populations for mixture analyses (François and Durand 2010). For the admixture analysis, a non-spatial prior distribution is used (François and Durand 2010). To run the program, I used the spatial clustering of individuals population mixture analysis by generating unique coordinates where there were matching coordinates. For consistency of sampling locations to avoid missing data, I used the site coordinate for each individual. I varied K between 2-10 with ten replicates for each K value, then used the admixture analysis based on the mixture analysis results. I used the default minimum number of individuals of five, with 50,000 iterations, a default of 50 reference individuals from each cluster, and 1,000 iterations for reference individuals. Lastly, unlike STRUCTURE or TESS, BAPS explicitly states the likely number of individuals.

SPATIAL GENETIC VARIATION AND GENETIC CONNECTIVITY

I used MEMGENE v.1.0.2 (Galpern et al. 2014) as an additional method of examining population structure. Compared to the previous models, MEMGENE is not a Bayesian approach and was designed to detect weak or cryptic population structures in highly mobile species such as caribou (Galpern et al. 2014). MEMGENE uses an approach that combines Moran's eigenvector maps (MEM) with a multivariate regression approach to identify neighbourhoods among locations (Galpern et al. 2014). MEMGENE ultimately determines if the genetic variation is potentially caused by landscape variables that could impact gene flow among individuals. First, I used the R

package MEMGENE (Galpern et al. 2014) to identify the MEMGENE variables that reflect the spatial genetic structure through the function *mgQuick(),* using MEM eigenvectors (Galpern and Peres-Neto 2014). MEMGENE variables are the products of a redundancy analysis based on a principle component analysis framework using a genetic distance matrix and MEM eigenvectors (Peres-neto and Galpern 2014). Simply put, the MEMGENE variables identify spatial genetic variation based on genotypes and coordinates. Then, I estimated how much the spatial patterns explain the genetic variation within the range using the object created with *mgQuick()* to calculate the adjusted R-squared value (Galpern and Peres-Neto 2014).

Population Graphs (Dyer and Nason 2004) was used to examine the genetic connectivity among sites using the R packages *GSTUDIO* (Dyer 2014) and *POPGRAPH* (Dyer and Nason 2004). This model uses a graph-theoretic approach to explore the intraspecific genetic structure of a species. I produced two networks of genetic connectivity with and without coordinate locations, where the network represents genetic covariance between sites. Sampling sites with more than four individuals were used for analyses based on the requirements of the software packages.

RESISTANCE LAYERS

To conduct the landscape analyses, I obtained geospatial data from Ontario GeoHub (Ontario 2022) to create raster layers in ArcGIS Pro v3.0.0 (ESRI 2022). Ontario GeoHub is an online database that provides open-source data for the province of Ontario. The Ontario Ministry of Northern Development, Mines, Natural Resources and Forestry makes the data available on behalf of Land Information Ontario (Ontario

2022). From the data obtained from Ontario GeoHub, I created four types of resistance layers to boreal caribou to explain spatial genetic variation within the population. These layers are roads, waterbodies, wildfire, and young forest that is predominately created by recent forest harvest.

I used the Ontario Road Network (ORN) Road Net Element (OMNRF 2001) data to create the roads resistance layer. Based on sensitivity analyses conducted by Environment Canada (Environment Canada 2011b), a buffer of 500 m was applied to all roads. Roads within this layer are limited to managed forests and include secondary and forestry roads. No major highways are located within the range.

The Ontario Hydro Network (OHN) – Waterbody (OMNR 2010) spatial data was used to create two different water layers composed of (i) lakes > 150 ha, and (ii) lakes > 750 ha, termed as small and large waterbodies, respectively. The two layers with different-sized waterbodies were used to explore if larger lake sizes limit resistance to caribou movement. This is based on the high use of smaller lakes by boreal caribou for calving and predator avoidance (Ferguson and Elkie 2005). I removed all water features except lakes, as the literature focuses on the use of lakes by caribou.

The Ontario Fire Disturbance Area (OMNRF 1960) spatial data layer was used to generate the wildfire layer. This layer contains the location of wildfires larger than 200 ha that occurred until 1998 and fires greater than 40 ha that occurred from 1999 until the end of 2019. For the resistance layer, I used only those fires less than 40 years old as per the guidance in the federal recovery strategy for boreal caribou (Environment Canada 2012).

To create a layer that predominately represents forest harvests based on young forest landcover, I used the Forest Resources Inventory (OMNRF 2020) for forest management units overlapping the Brightsand Range and immediate vicinity, including Black Spruce, Caribou, Dog River Matawin, English River, Lac Seul, Lake Nipigon, and Ogoki. In addition, I obtained Forest Resource Inventory data for Wabakimi Provincial Park, which encompasses a large area of the northeastern portion of the Brightsand Range. The forest resource inventory data is primarily based on forest polygon updates from 2008 - 2011 (Appendix VII). I used polygons defined as forest types with overstorey canopies ≤ 35 years based on the classification of young forests (OMNR 2014). In addition, I erased young forest polygons as a result of wildfire disturbances. Within this layer, there were small, scattered areas of forest ≤ 35 years old that are not forest harvest patches that remain in the layer. Therefore, the layer represents young forest that has predominately resulted from forest harvesting.

Lastly, I used the Caribou Range Boundary (OMECP 2010) to delineate the range extent of the Brightsand Range of Ontario. In addition, I added a 30 km buffer to the range to reduce the effect of artificial boundaries caused by boundaries on a study area (Koen et al. 2010). Within this buffer, all landscape resistance layer data was also obtained except for a small tract of land within the northwest for the young forest layer. This layer was not interpolated as this section was predominately covered by a wildfire disturbance and would have been removed from the layer described above.

LANDSCAPE ANALYSIS

I used MEMGENE (Galpern et al. 2014), an R package, to detect fine-scale and potentially cryptic patterns of gene flow within the range. Based on previous analyses, I found that subtle genetic structure and genetic discontinuity occurred within the Brightsand Range. However, the spatial clustering analyses I used were not designed to identify landscape factors that can explain the spatial patterns of genetic variation. While MEMGENE was previously used to identify patterns of spatial genetic structure, it can also be used to identify landscape variables that contribute significantly to genetic variation across the range. For the analysis, least-cost paths are calculated based on distances between sites and are used to calculate MEM using the input resistance layers (Galpern et al. 2014). A Euclidean distance model is automatically included within the analysis to account for potential IBD patterns. The output of the landscape analysis partitions the spatial genetic variation among spatial predictors, model patterns, coordinates, or residuals. With MEMGENE, I evaluated the effect of roads, wildfires, harvests, and waterbodies on the spatial genetic distribution of individuals. I used the layers described above and tested the resistance cost values of 1, 10, 50, and 100 for each disturbance layer. To create the rasters, I converted the vector files to rasters using a cell size of 100 m^2, each layer with a different cost value. Based on the MEMGENE outputs, I evaluated which resistance cost value and layer best explains the observed pattern of spatial genetic variation.

Next, I used CIRCUITSCAPE (Shah and McRae 2008) to predict patterns of gene flow across the landscape. The CIRCUITSCAPE model is an IBR approach based on circuit theory where landscapes resist gene flow across the study area, acting as resistors within a circuit model (McRae et al. 2008). Random walks represent multiple

pathways through which an individual can move, thus highlighting potential corridors of likely gene flow within the landscape (McRae et al. 2008). Visual observation of the model will assist with delineating potentially critical corridors. I ran two models using CIRCUITSCAPE based on the results of MEMGENE. The first model uses the value of the resistance layers identified by MEMGENE that explained a significant proportion of the genetic variation across the range. Layers that did not significantly explain spatial variation were also included, with the lowest resistance cost value of 1. All resistance layers were included as they have been shown to affect the distribution of boreal caribou in literature (Dyer et al. 2002; Ferguson and Elkie 2005; Wittmer et al. 2007; DeCesare et al. 2012; Beauchesne et al. 2013; Priadka et al. 2019). The second model was used for exploratory purposes to further evaluate potential patterns of gene flow based on all resistance layers, regardless of whether they were considered significant based on the MEMGENE analysis. The rationale is that, based on the low number of samples from the managed forest, it is possible that MEMGENE was unable to detect a significant contribution of some layers (i.e., roads and young forest) because most sampling sites did not occur in proximity to these disturbances. Thus, in the second CIRCUITSCAPE model, I included all resistance layers and set the resistance cost values for the layers that did not explain spatial variation according to the MEMGENE analysis to a value of 10.

Before running the CIRCUITSCAPE analysis, all the resistance cost layers were merged in ArcGIS. I included an additional layer in the merged dataset, named the IBD layer, to account for the potential effects of IBD on gene flow within the Brightsand Range. For this layer, the value of each cell was set to 1, such that increasing distance

between sites will linearly increase the cost of movement (resistance) between sites. Therefore, I used the sum of all overlapping cells between raster layers to differentiate the IBD layer from landscape resistances when amalgamating multiple resistance layers to calculate each new raster cell value. Therefore, the values compound and prevent the model from identifying equivalent resistance values between areas of no natural or anthropogenic resistance and present resistance. Consequently, raster cells with resistance features are always greater than 1. To run CIRCUITSCAPE, I used a raster data type for the landscape resistance map at a 100 m resolution and a text file listing all sample sites with coordinates. Next, I calculated the current map using the pairwise mode and the cells nearest eight neighbours to generate a connectivity network.

Using the R package *CAR* (Fox and Weisberg 2019), I also statistically evaluated univariate resistance cost values 10 and 50 for each resistance type (wildfires, small waterbodies, harvest, and roads) and IBD. Each univariate layer was also converted from vector format to raster in ArcGIS with a resolution of 1 km. I used correlation analysis to identify which CIRCUITSCAPE resistance value for each layer was most strongly correlated to pairwise F_{ST} values between sites. Then, with each layer type and the data from the highest correlated resistance cost, I used a multiple regression model to identify which resistance variables significantly explained the pattern of F_{ST} within the Brightsand Range.

First, I evaluated the distribution of variables and adjusted variables data accordingly with a log transformation (+0.01) to improve the normality of the variables where required. I then examined correlations among predictor variables to identify multicollinearity among the data. Next, through the use of Variance Inflation Factors

(VIF) I sequentially removed variables with the highest VIF value, therefore removing multicollinearity until the VIF was <2 (Zuur et al. 2010). Based on the remaining variables, I examined the significance of each variable for interpretation. The diagnostic plots were examined to verify normality and homogenous residuals.

RESULTS

MOLECULAR ANALYSIS

The marker BM4513 was difficult to score due to inconsistencies among profile patterns with allele peak heights and stutter problems, so it was excluded from the analysis. In addition, BMS1788 and NVHRT30 presented challenges due to the presence of peak artifacts. While genotyping scoring rules resulted in consistent scoring among individuals and replicated PCRs for these loci, MICRO-CHECKER identified possible null alleles. Therefore, these loci were excluded from further analyses. No null alleles, allele dropout, or genotyping errors were detected for the remaining loci. In total, 12 out of 15 loci were used for analyses. The genotype accumulation curve (Figure 4) confirmed that enough microsatellite loci were used so that no new genotypes would present with additional loci. This plateau was reached at five loci, and at nine randomly selected loci, all unique multilocus genotypes were consistently identified.

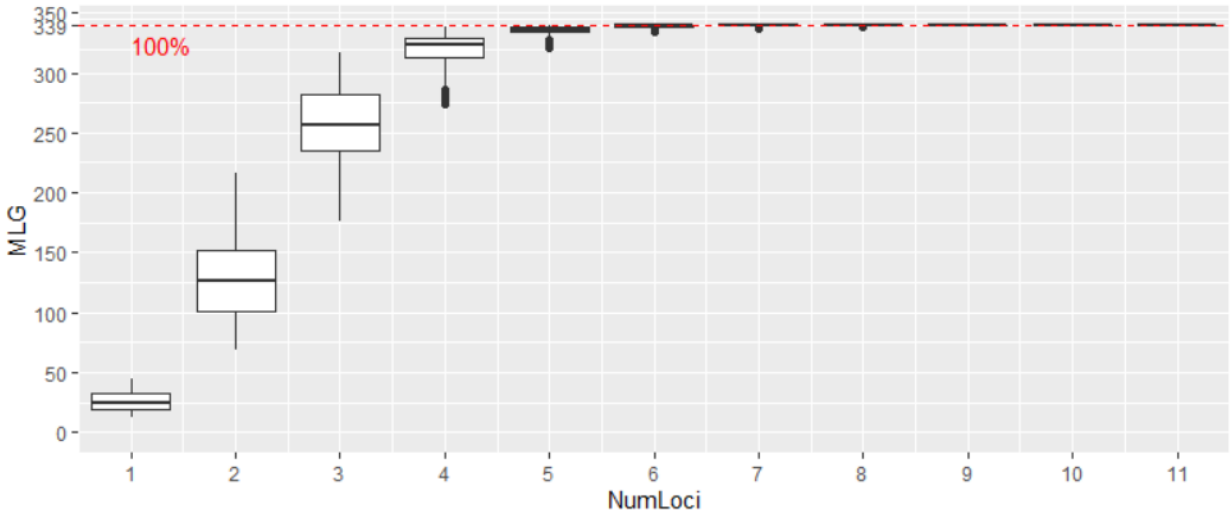


Figure 4. Genotype accumulation curve indicating the number of unique multilocus genotypes identified based on a random selection of loci over 1000 iterations.

Based on the 12-locus genotypes, ALLELEMATCH identified 340 unique genotypes among 788 samples, three of which were removed due to high missing data. Therefore, 337 unique genotypes were used for analyses. Including individuals that were found at more than one site, the dataset used for analyses included a total of 424 genotypes unless otherwise stated. Of the remaining samples, a single sample and locus contained missing data (0.025% of loci across unique samples). In addition to a low missing data rate, there was a low genotyping error rate. The genotyping error rate was calculated based on a random selection of 17.8% (n = 140) of the samples for which multiple PCR amplifications were conducted. The error rate was < 0.01 across all twelve loci (Appendix VI).

GENETIC DIVERSITY

There was no substantial deviation from HWE within loci and sites. LD was not detected across loci ($p > 0.05$), indicating that the twelve loci assort independently of one another. The number of alleles at each locus ranged from 5.00 (BM6506) to 11.00 (RT24, BM888) for an overall mean of 8.33 alleles per locus. The mean allelic diversity, N_a, ranged from 3.67 to 5.83, averaging 4.69. The effective number of alleles, N_e, ranged from 2.73 to 3.59, with an average of 3.08. The mean observed heterozygosity, H_o, ranged from 0.602 to 0.764 and was generally higher than the mean expected heterozygosity, H_e, which ranged from 0.577 to 0.695. Lastly, the mean Fixation Index (inbreeding coefficient, F) was -0.0763 and ranged between -0.241 and 0.0697. Appendix VI contains complete genetic diversity statistics by the site. The spatial distribution of the expected heterozygosity is mapped in Figure 5, which presents no spatial pattern across the range for all sample sizes and for sample sizes ≥ 10.

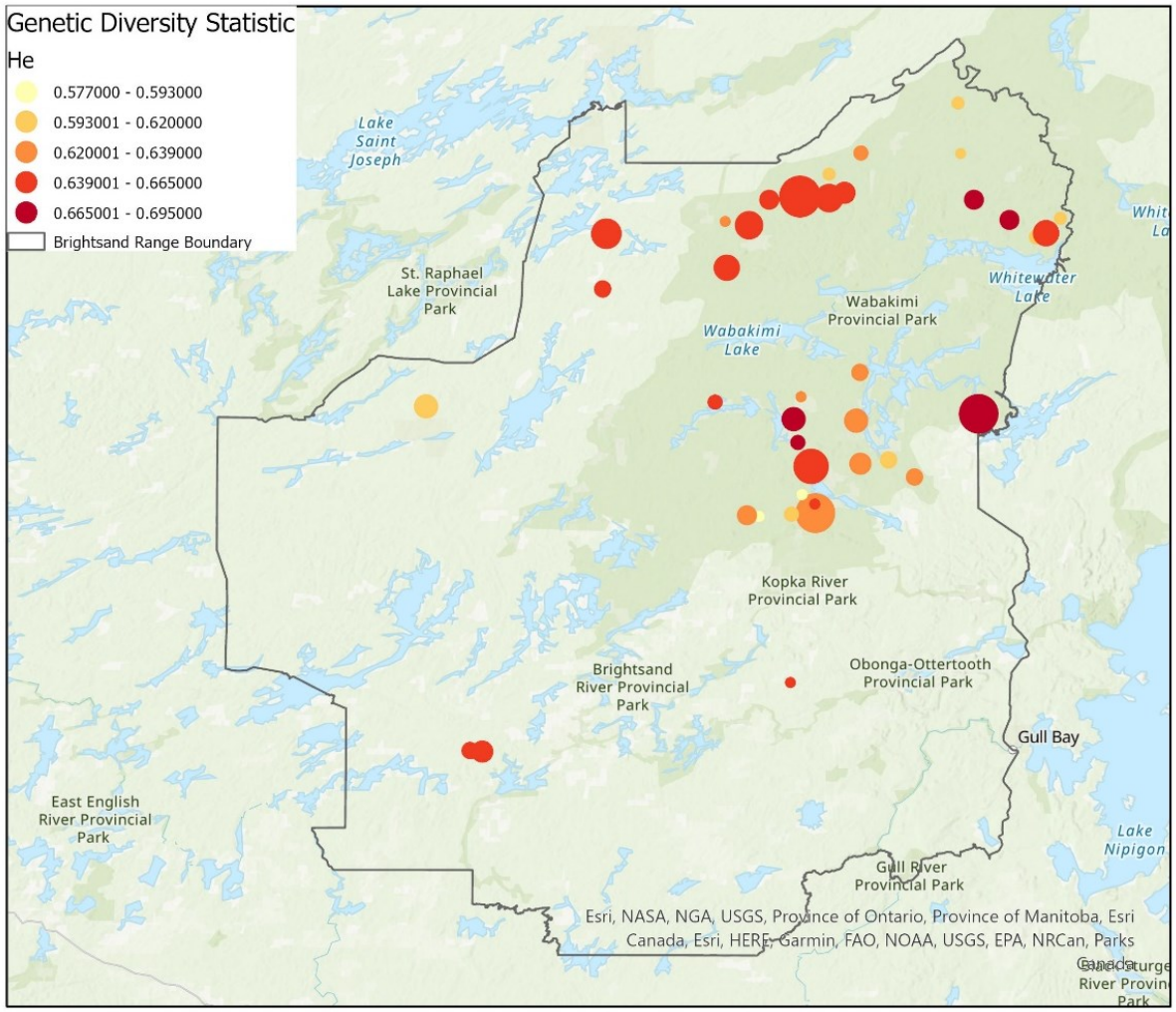


Figure 5. Spatial distribution of expected heterozygosity across the Brightsand Range. Scale is based on natural breaks, with darker values indicating higher expected heterozygosity. The size of the points indicates the sample size.

POPULATION GENETIC STRUCTURE

The AMOVA analysis did not reveal substantial variation in allele frequencies among managed and unmanaged forest sites, with a total variance of 0.30% among regions (Table 1). In addition, AMOVA did not detect substantial variation among sites, as only 0.54% of the total variance was partitioned among sites within regions. The remaining 99.16% of the variance occurred within individuals. Therefore, less than 1%

of the variation was found among regions and sites. The calculated F-statistic values were low but significant among sites and regions, with a value of 0.009 for F_{ST} ($p < 0.001$) and 0.004 ($p < 0.05$) for F_{RT} (Table 2). However, the F-statistic values were not significant within individuals (F_{IS} and F_{IT}), where $p > 0.05$. Additionally, F_{ST}-based pairwise genetic comparisons indicated that the southernmost sample sites were significantly differentiated from multiple sites located among the northernmost portion of the range (Appendix VI).

The Mantel test indicated a weak but significant pattern of IBD across the range ($r = 0.207$, $p < 0.05$) (Figure 6). Visual inspection of the diagnostic plot of Residuals vs Fitted confirmed linearity among the data. Furthermore, homoscedasticity was present within the data identified by the diagnostic plot Scale-Location and confirmed with the Non-constant Variance Score Test ($p > 0.05$).

Table 1. AMOVA results indicating the partitioning of molecular variation among regions, sites within regions, individuals within sites, and within individuals.

Source	df	Sum Square Difference	Mean Square Difference	Variance Component	Total Variance (%)
Among Regions (Managed and Unmanaged forests)	1	7.858	7.858	0.018	0.42%
Among Sites	37	165.276	4.467	0.021	0.52%
Among Individuals	323	1314.948	4.071	0	0%
Within Individuals	362	1491.500	4.120	4.120	99.06%
Total	723	2979.581		4.159	100.00%

Table 2. F-statistic values from AMOVA for managed and unmanaged forest regions, sites, and individuals.

Source	F-Statistics	Value	Probability (rand >= data)
Among regions	F_{RT}	0.004	0.013*
Among sites within regions	F_{SR}	0.005	0.018*
Within sites relative to the total population	F_{ST}	0.009	0.001*
Within individuals relative to the sites	F_{IS}	-0.006	0.717
Within individuals relative to the total population	F_{IT}	0.004	0.360

* denotes values significant at the 0.05 level

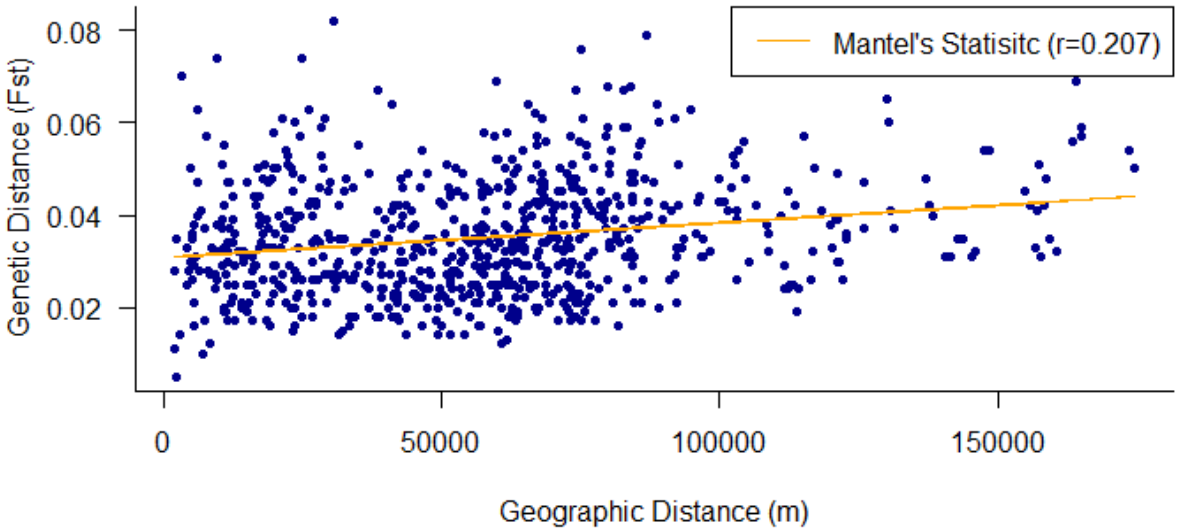


Figure 6. Pairwise comparisons of genetic distance (F_{ST}) and geographic distance (metres) matrices between all sites with n ≥ 5.

The optimal number of clusters detected by STRUCTURE was K = 2 based on the ΔK calculations and individual assignment proportions (Appendix VI presents the log-likelihood (L(K)) (Pritchard et al. 2000) and ΔK (Evanno et al. 2005) calculated by STRUCTURE HARVESTER). The ΔK chart presented a bimodal distribution at K = 2 and K = 8, while the mean LnP(K) was challenging to interpret. However, based on the STRUCTURE documentation and individual cluster assignment proportions (Figure 7), high admixture was present among all individuals when K = 8 but not at K = 2, suggesting that K = 2 was the most likely number of clusters. Cluster assignments for K = 2 indicated that individuals from the southern portion of the range were mostly assigned to Cluster 1 (orange), while individuals from the northern extent of the range were primarily assigned to cluster 2 (blue) (Figure 8). Individuals from central sites were generally admixed between Clusters 1 and 2.

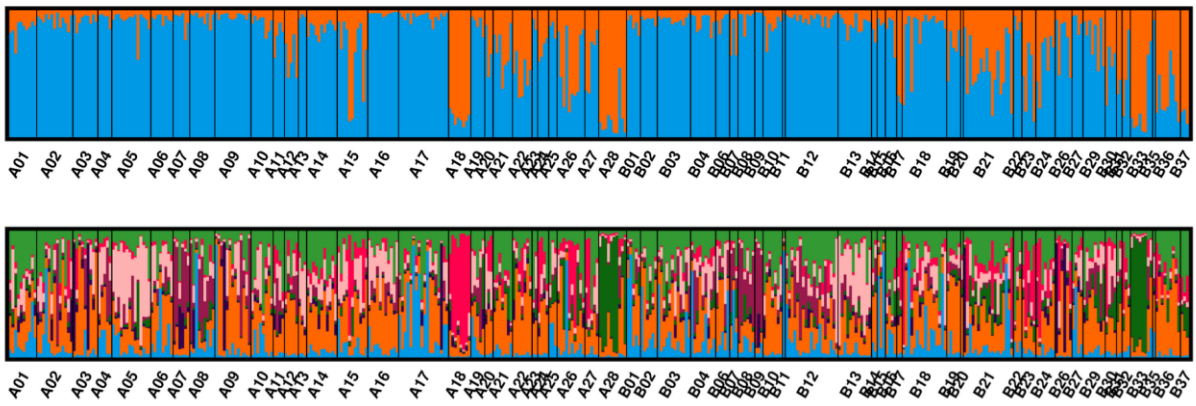


Figure 7. STRUCTURE cluster assignment for K = 2 (top) and K = 8 (bottom).

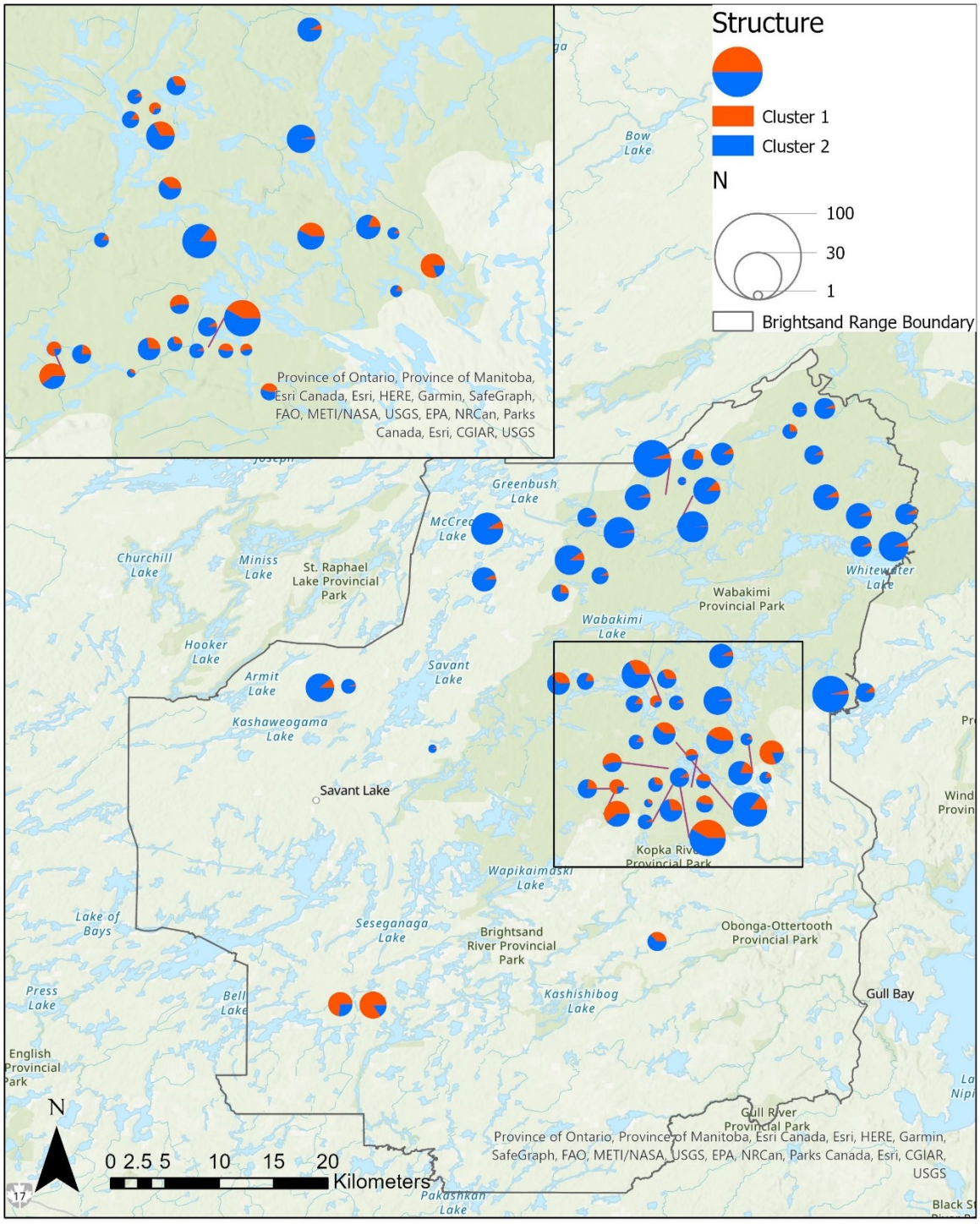


Figure 8. STRUCTURE cluster assignments across the Brightsand Range. Cluster 1 is orange, and Cluster 2 is blue, corresponding to cluster assignments in Figure 7, where K = 2. The size of each circle indicates the number of samples within the site.

The TESS algorithm identified the likely number of clusters as $K_{max} = 8$, where the DIC value began to plateau (Figure 9). The lowest DIC value occurred at $K_{max} = 10$ (Figure 9); however, there was no substantial difference in individual cluster assignments between $K_{max} = 8$ through 10. The admixture partition generally assigned most of the genetic variation to a single population for each individual. While the DIC value indicated a high number of clusters, an inspection of the assignment probabilities suggested a single population with some admixed individuals, as demonstrated in the assignment probabilities plot (Figure 10).

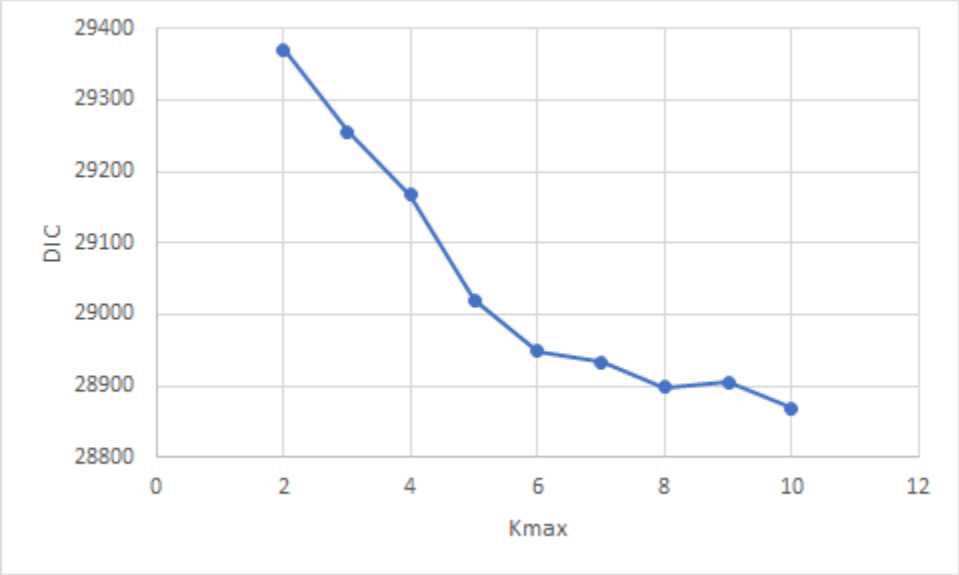


Figure 9. Average DIC across 15 runs per K_{max} plotted against K_{max} for the TESS analysis.

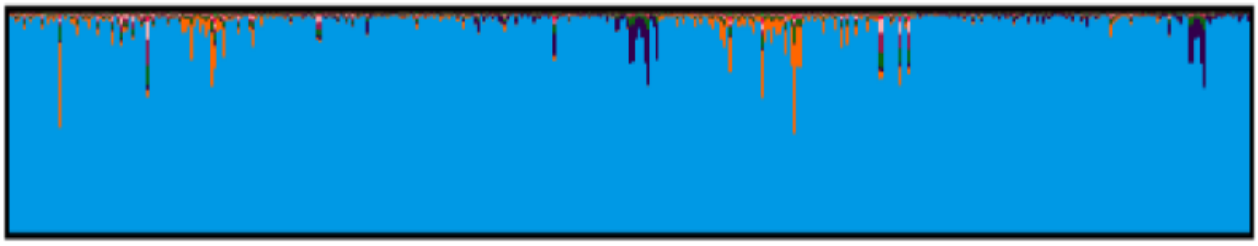


Figure 10. TESS individual cluster assignment where K = 8.

The BAPS analysis identified three genetic clusters, with most individuals assigned to Cluster 3 (blue) (Figure 11). Cluster 3 contained individuals from throughout most of the range, whereas Cluster 2 (green) contained individuals from three sites in the central portion of the range, and Cluster 1 (red) contained individuals from the two sites in the southwest portion of the range and two sites within the central portion of the range.

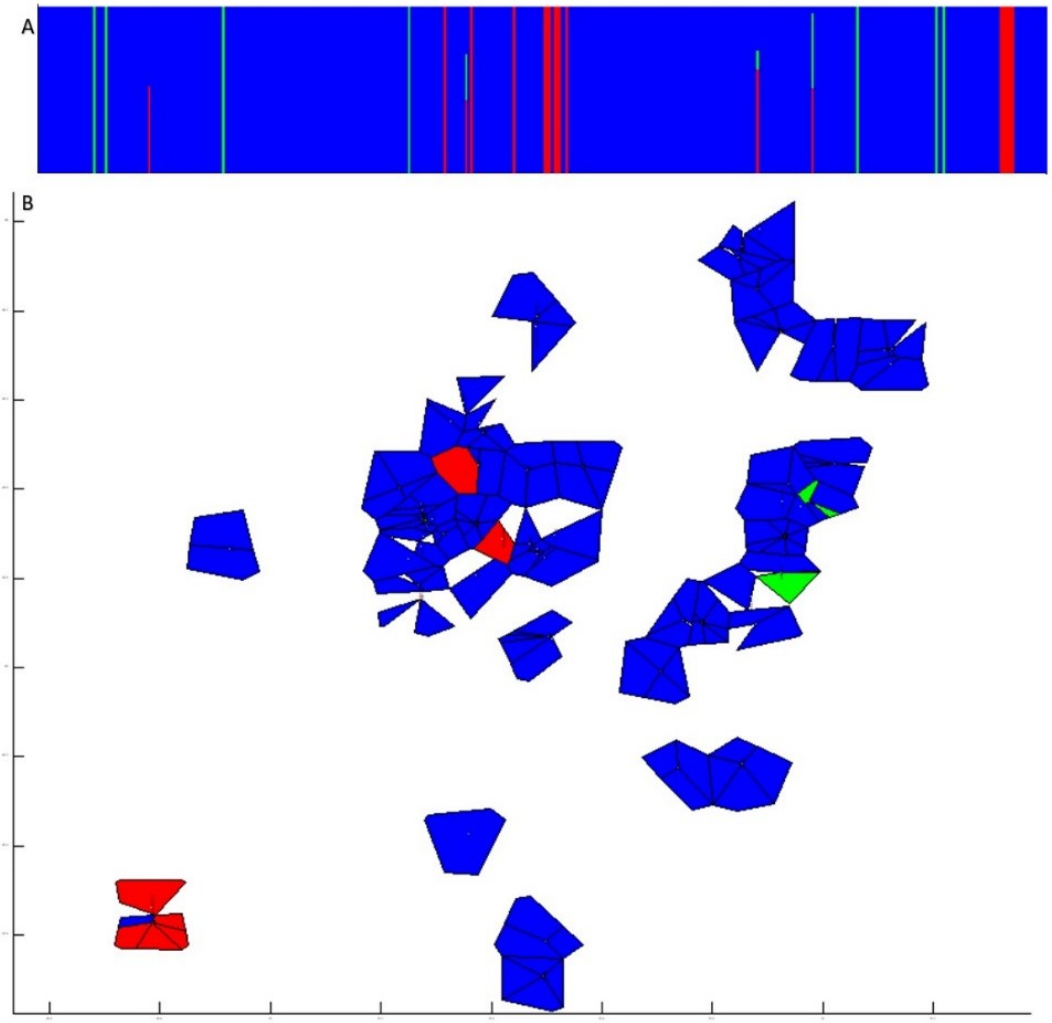


Figure 11. BAPS cluster assignments identified by varying colour (red = Cluster 1, green = cluster 2, and blue = Cluster 3). A) Admixture partition for each individual with three clusters. Admixed individuals have partial cluster assignments. Non-significantly admixed individuals are assigned to a single cluster within the admixture plot. B) Spatial Voronoi tessellation of membership of individuals to three genetic clusters.

SPATIAL GENETIC VARIATION AND GENETIC CONNECTIVITY

The results of the MEMGENE analysis indicated a weak association of genetic variation explained by spatial genetic patterns (R-Squared = 0.011). Calculated eigenvalues were 0.378, 0.241, and 0.180 for MEMGENE variables 1 through 3, respectively. Eigenvalue 1 explained the most spatial variation and depicted the clearest and largest amount of spatial genetic pattern, with sites in the northeast (black circles) more genetically differentiated from sites in the southwestern portion of the range (white circles) (Figure 12). However, this within-population spatial genetic structure association was weak. MEMGENE variables 2 and 3 had even weaker associations between genetic variation and geographic sample location. Eigenvalue 2 clustered individuals in the central portion of the range while surrounding individuals were more differentiated, and Eigenvalue 3 showed an east-to-west differentiation pattern.

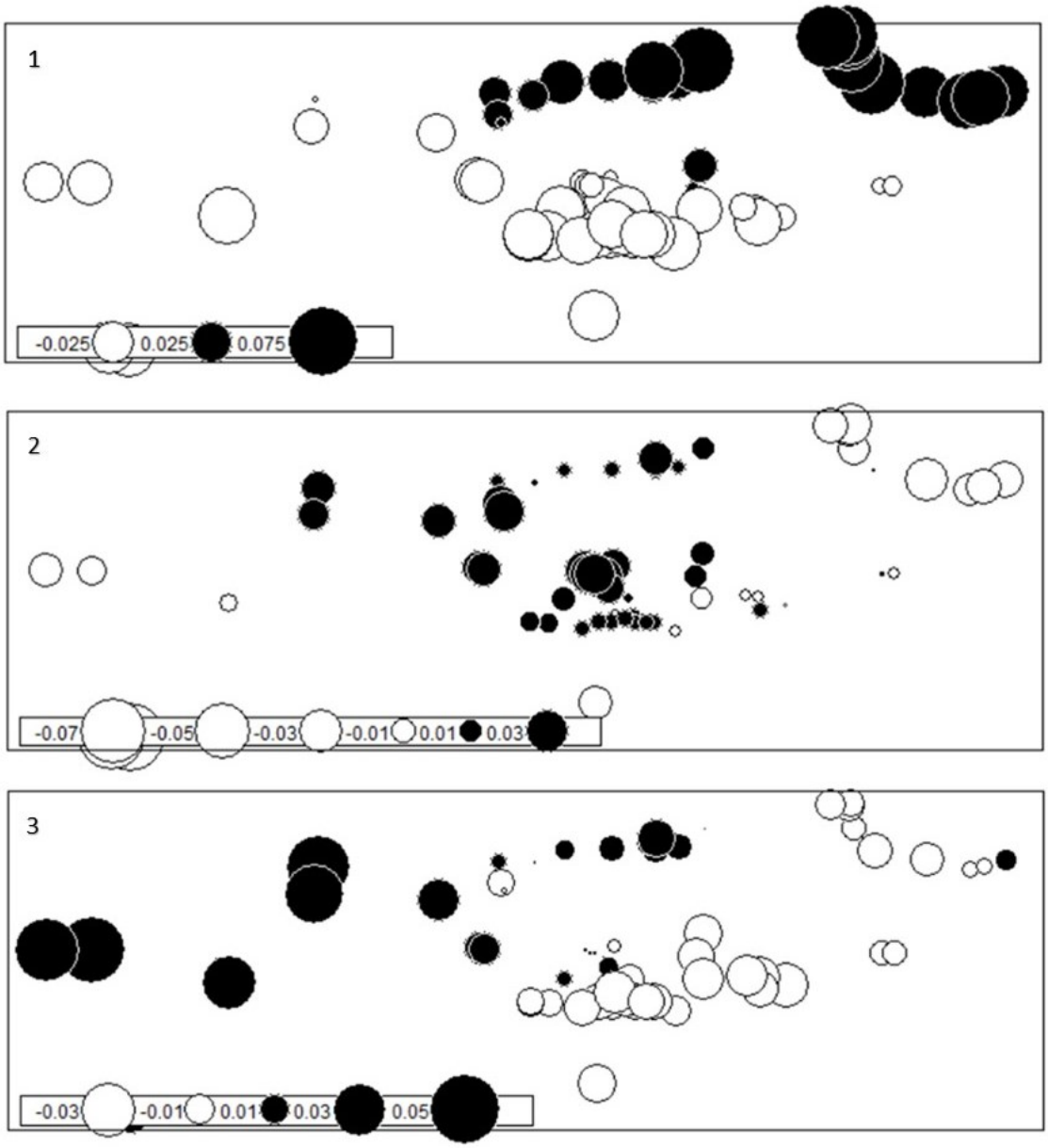


Figure 12. MEMGENE variables 1, 2, and 3 are plotted on latitude and longitude axes. Differences in colour and the size of circles represent genetic differentiation.

POPGRAPH outputs indicated that all the sampling locations within the Brightsand Range were connected to at least one other site (Figure 13). Within the central network, sites A02, A16, A21, B03, and B36 had the greatest number of connections (5 connections), while sites A05, A26, A27, and B29 were peripheral, with only one connection to the central network. Two managed forest sites (A28 and B33) were isolated from the central network but maintained a single connection between them (Figure 14). The number of connections was generally greater among sites within the central and northern portions of the range, with lesser connectivity between these portions. In contrast, the greatest dysconnectivity was present in the southernmost portion of the range. Two of the peripheral sites were located within the managed forest (A05 and A27), while two were located less than 13 km from the western border of the unmanaged forest (A26 and B29) but were located among the central cluster of sample sites (Figure 14).

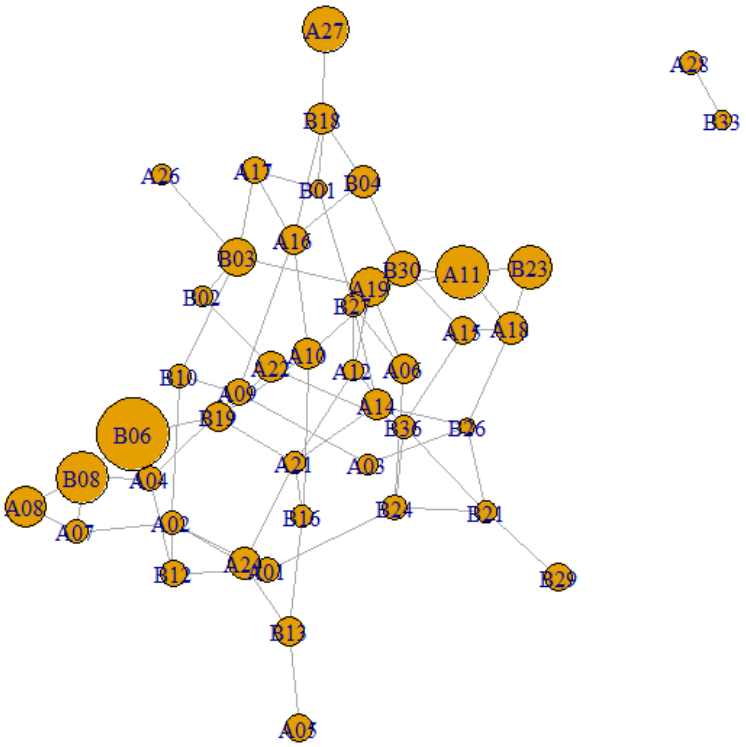


Figure 13. POPGRAPH connectivity network where node sizes represent the genetic variation within sites and edge lengths represent the conditional genetic distances between sites.

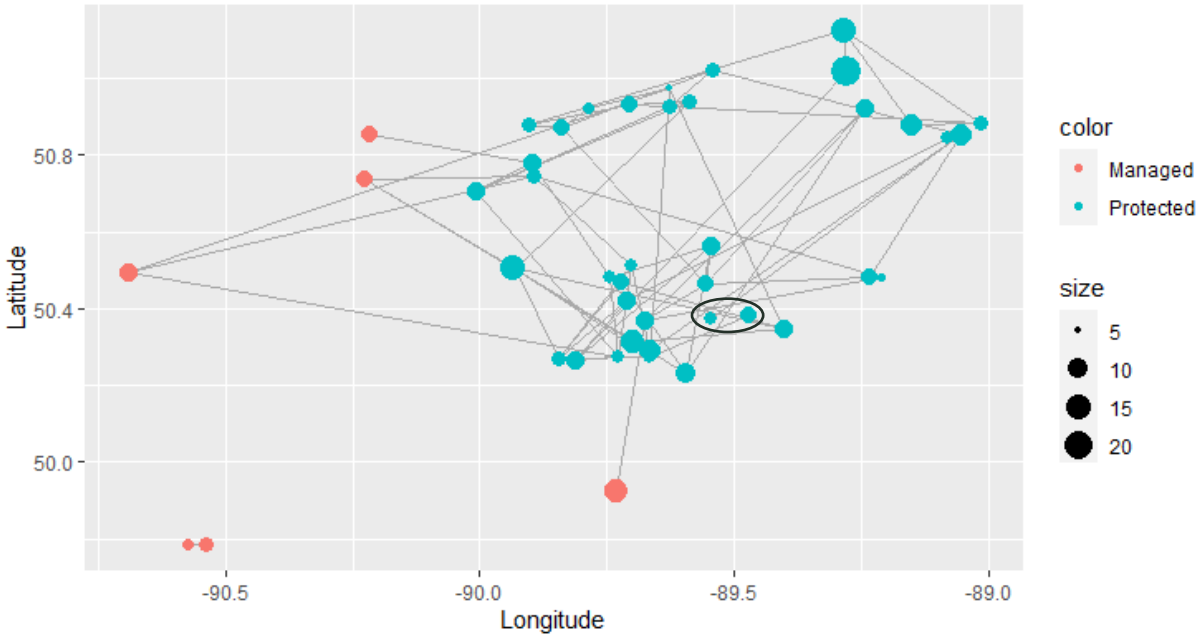


Figure 14. POPGRAPH output depicting genetic connectivity among sites with geographic locations indicated by X and Y axis values. Sites connected by lines are inferred to be connected by gene flow, whereas sites not connected by lines are inferred to be genetically isolated from one another. Blue nodes represent protected forests, orange nodes represent managed forests. Node size represents the genetic variation within sites. Edges represent geographic distances and genetic connectivity between sites. The circled sites indicate the peripheral sites within the protected forest.

LANDSCAPE ANALYSIS

MEMGENE did not detect significant Moran's eigenvectors of each resistance cost value for roads, small waterbodies, large waterbodies, and young forests (predominately harvested). Therefore, those layers did not significantly explain spatial genetic differentiation across the range. Wildfire was the only statistically significant layer that partitioned spatial genetic variation and for all resistance costs. The model Fire50, where the value represents the resistance cost value, best explained the partitioning of spatial genetic variation across wildfire disturbances (Table 3). The Fire50 model explained the highest proportion of variation using spatial predictors and nearly double the Euclidean model (IBD model). The spatial pattern within the Fire50 model

also explained more spatial variation than coordinates, indicating a stronger spatial relationship explained by the natural disturbance wildfire than a linear pattern. While the Euclidean distance model was statistically significant, it explained the lowest proportion of spatial predictors based on Euclidean distances and the model was confounded between the model pattern and coordinates (linear pattern). Therefore, the Euclidean model (IBD) does not best explain the spatial genetic variation within the range. Additionally, all wildfire models were statistically significant and indicated greater explanatory power using the wildfire resistance layer than the Euclidean model to explain spatial genetic variation across the range.

Table 3. Proportions of variation across four layers of fire disturbance with resistance values of 1,10,50,100 and a Euclidean model. Proportions explain [abc] spatial predictors (MEM eigenvectors); [a] model patterns; [c] coordinates (does not use Moran's eigenvectors to describe linear patterns); [b] confounded between [a] and [c]; and [d] residual. P denotes the *p*-value indicating the significance.

Model	[abc]	P[abc]	[a]	P[a]	[c]	P[c]	[b]	[d]
Euclidean	0.0122	0.002	0.00447	0.01	0.00229	0.044	0.0054	0.988
Fire 1	0.0171	0.002	0.00943	0.002	0.00902	0.002	-0.00133	0.983
Fire 10	0.0158	0.002	0.00809	0.002	0.00935	0.002	-0.00166	0.984
Fire 50	0.0221	0.002	0.01441	0.002	0.01025	0.002	-0.00256	0.978
Fire 100	0.0205	0.002	0.01278	0.002	0.01016	0.002	-0.00247	0.98

Based on the results of MEMGENE, I used resistance layers wildfire (50), small waterbodies (1), young forest (1), and roads (1) in CIRCUITSCAPE to visually examine the connectivity across the landscape in the first CIRCUITSCAPE model (Figure 15). To present the results, I overlayed the collection sites over the CIRCUITSCAPE output to further highlight the sample sites. Within the map, the darkest areas primarily represent

fire disturbances avoided by the connectivity network. Sample sites were often close to wildfire areas but did not occur within fire-disturbed areas. This model showed high connectivity among all sites, with some slightly weaker connectivity in the southern portion of the range.

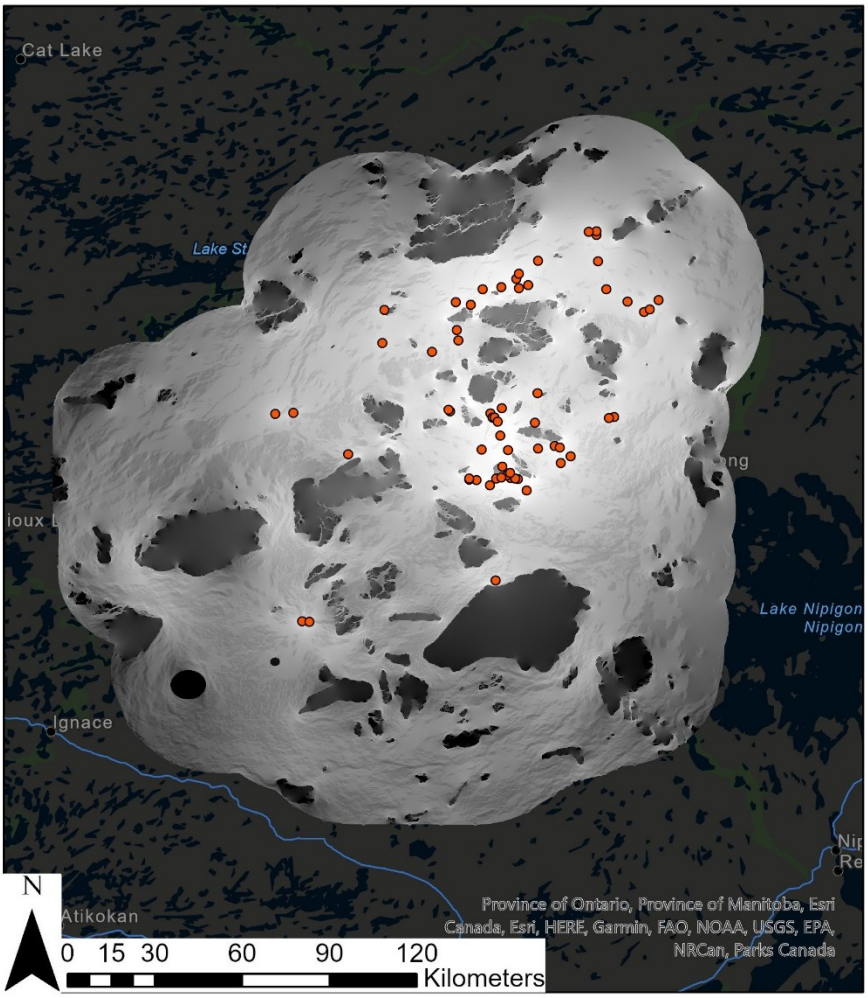


Figure 15. CIRCUITSCAPE base model output based on a map of the Brightsand range with a 30 km buffer raster with resistance values compounded. Wildfire (50), Roads (1), Waterbodies (1), and Young Forest (predominately harvest) (1).

The second model (Figure 16) was based on an exploratory analysis to explore potentially critical corridors for habitat connectivity within the managed forest. The resistance cost values for roads, small waterbodies and young forests were increased to 10 while the wildfire layer remained at 50. This model further highlighted corridors of connectivity among the managed forest. The southernmost sites appeared to have less connectivity to other sites and contained weak points within the connectivity corridors.

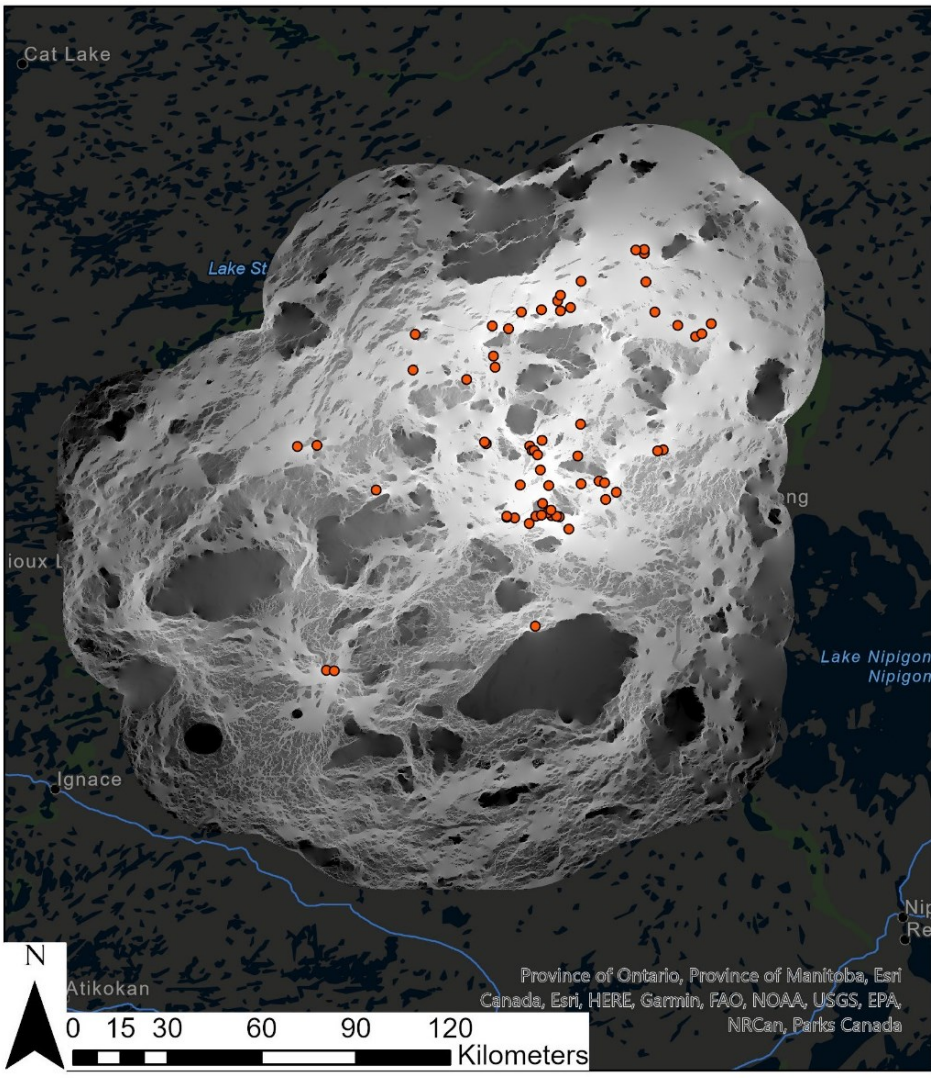


Figure 16. CIRCUITSCAPE exploratory output current map. Based on a map of the Brightsand range with a 30 km buffer raster with resistance values compounded. Wildfire (50), Roads (10), Waterbodies (10), Young Forest (predominately harvest) (10)

An evaluation of the correlation among pairwise resistances and F_{ST} values indicated that the correlation was greater with resistance costs of 10 for all resistance types (wildfire, harvest, roads, and small waterbodies). However, the data was skewed based on the distribution of the resistance data for roads and wildfires. Therefore, a log10(+0.1) transformation was applied to improve the normality of the data. An evaluation of covariation between predictor variables indicated that covariation was present. IBD was highly correlated with the small waterbodies and harvest layers, and harvest was also highly correlated with small waterbodies Appendix VI.

The final model, based on the removal of all variables with VIF > 2, included only wildfire and roads (Table 4) and was statistically significant ($F_{2,738} = 24.293$, $p < 0.0001$; Appendix VI). However, the model only explained 6.18% of the variance in genetic differentiation (F_{ST}). Furthermore, the partial regression plots indicated that genetic differentiation decreased significantly with wildfire, while roads had no significant effect on genetic differentiation (Figure 17).

Table 4. Estimates of multivariate regression model. Multiple R^2 value was 0.0618.

	Estimate	Standard Error	t value	Pr(>|t|)
(Intercept)	0.0343	0.000491	69.7	<0.001*
fire10T	0.0244	0.00417	5.85	<0.001*
roads10T	0.00410	0.00388	1.06	0.291

*significant at the 0.001 level

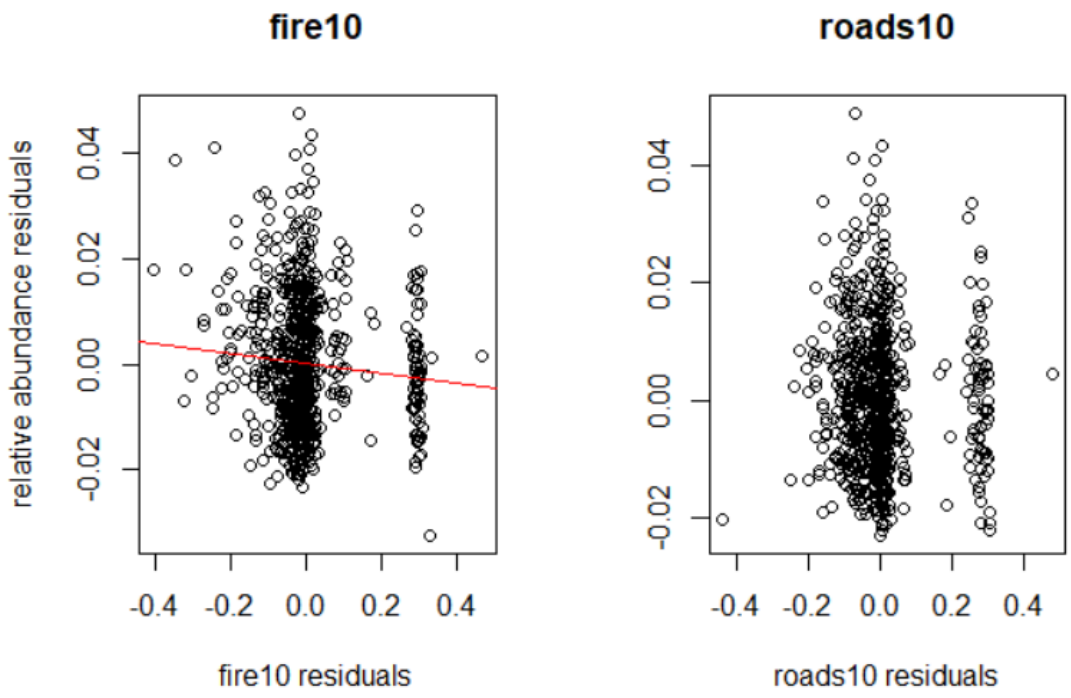


Figure 17. Partial regression plots for wildfire and roads. The red line indicates a significant regression among the data.

DISCUSSION

Genetic diversity is important among populations of small sizes as they are at greater risk of increased inbreeding and reduced genetic diversity (Keller and Waller 2002) and loss of genetic diversity is greater among isolated populations (Courtois et al. 2003a). Within this study, there was no distinct pattern of genetic diversity across the Brightsand Range and expected heterozygosity values varied between 0.577 - 0.695. The mean H_e was 0.641, which was lower than average among continuous populations of boreal caribou across Saskatchewan, Manitoba and Quebec, where $H_e > 0.7$ (Courtois et al. 2003a; Ball et al. 2010; Yannic et al. 2016). Within the same studies, isolated populations had comparative H_e values to my study, for example, the lower Interlake region in Manitoba ($H_e = 0.63$) (Ball et al. 2010) and Val d'Or in Quebec ($H_e = 0.57$) (Yannic et al. 2016). However, the study by Thompson et al. (2019) demonstrated overall lower H_e values across Ontario, where the mean H_e was 0.678 (calculated based on results), which is comparable to my findings.

Within the Brightsand Range, Thompson et al. (2019) previously demonstrated that a southern group of two populations was genetically isolated and had lower genetic diversity. The H_e value of the lower Ignace population found in the managed forest of the current study was 0.64, which was equivalent to the overall mean H_e of my study. At the same time, the maximum H_e value in my study was 0.695, which was 0.005 greater than the H_e of the northern Wabakimi area population in the study by Thompson et al. (2019).

In this study, approximately one-third of the sites contained less than five individuals, one-third had five to eight individuals, and one-third had at least ten individuals. Since small sample sizes ($n < 10$) are often associated with higher variability in genetic diversity estimates (Hale et al. 2012), estimates of genetic diversity for many of the sites in this study should be interpreted with some caution. However, H_e values in this study for sites with at least 10 individuals ranged from 0.613 to 0.675 and were within the range of previously calculated H_e values with similar sample sizes in the study by Thompson et al. (2019) for the Brightsand Range. Furthermore, there was no indication that the low numbers of individuals within sites were due to sampling methods, as 1.4 samples were collected per estimated number of individuals at each site. Instead, small samples within the Brightsand Range can be explained by the low occupancy of boreal caribou (MNRF 2014), and boreal caribou are commonly found in small group sizes of approximately six or fewer individuals during winter months (Fuller and Keith 1981; Stuart-smith et al. 1997).

While acknowledging the potential influence of sites with smaller group sizes, there was no distinct pattern of genetic diversity in the managed portion of the range due to limited sampling sites. However, the higher H_e among the three sites in the southernmost portion of the range was unexpected. Therefore, the prediction that sites from the northern portion will exhibit significantly higher genetic diversity than sites from the southern range was not supported due to a lack of spatial patterns of genetic diversity across the range. Overall, there was reduced genetic diversity among the Brightsand Range compared to surrounding northern populations based on past studies (Ball et al. 2010; Thompson et al. 2019). As low genetic diversity decreases the ability of

a population to adapt to changing environments (Frankham et al. 2010a), this finding is concerning for the population as it may indicate that the population has begun to suffer from genetic erosion.

The results of this study suggest that boreal caribou within the Brightsand Range exhibit low population genetic subdivision and are generally well-connected by gene flow. The AMOVA and F-statistics results indicated the occurrence of a weak yet significant population genetic substructure ($F_{ST} = 0.009$, $p < 0.001$). AMOVA identified that nearly all genetic variation was found within individuals (99.06%), which could suggest a panmictic population under normal circumstances (Kamvar et al. 2017). However, the F_{ST} value from AMOVA ($F_{ST} = 0.009$) was significant at $p < 0.001$. While the significant ($p < 0.05$) pairwise F_{ST} values ranged from 0.027 to 0.082 between sites, the maximum F_{ST} within the study is comparative to the significant pairwise F_{ST} found between subpopulations McGerrigle and Logan/Albert ($F_{ST} = 0.078$ and 0.75 respectively) of an isolated population in Quebec. However, in that study, the majority of genetic variation was among subpopulations opposed to within individuals of the current study (Pelletier et al. 2019). Additionally, a north-to-south gradient of genetic differentiation was identified among sites by the genetic structure analyses. The Mantel test identified a significant IBD pattern across the Brightsand Range ($r = 0.207$, $p < 0.05$). However, sites among study areas with greater distances and genetic differentiation can skew the results of IBD analyses (Rivers et al. 2005). Within the study area of the Brightsand Range, there was reduced sampling among the southern managed forest resulting in a large gap of geographic distance between the southernmost sites and the remaining sites. A Mantel test based on samples in the

unmanaged forest verified a significant IBD pattern within the range, albeit with a weaker regression coefficient (results not shown). The IBD pattern within the Brightsand Range was further supported by MEMGENE variable 1 (eigenvalue 1), where the pattern of spatial genetic variation reflects a north-to-south gradient. Additionally, pairwise genetic distances (F_{ST}) indicated that sites in the southernmost range were significantly genetically differentiated from multiple sites located at the northernmost extent of the range. Taken together, these results suggest that sites from the northern and southern portions of the range are significantly but weakly genetically differentiated from one another due to IBD.

In this study, only duplicate genotypes within sites were removed from the data, compared to other studies that removed all duplicate genotypes (McFarlane et al. 2021). Removing all duplicate samples for the fine-scale analysis of this study would have caused an incorrect interpretation of the results as it would artificially inflate the level of among-site genetic differentiation. A preliminary analysis with all duplicate samples removed resulted in much higher calculated values of pairwise F_{ST} among sites (results not shown), and consequently, downstream analyses may have had differing interpretations. Therefore, by only removing duplicate genotypes within sites, the results of this study should more accurately reflect true patterns of relatedness of individuals within and between sites in this highly mobile species (Wilson et al. 2019). Large-scale population genetic studies typically remove all matching genotypes where non-invasive mark and recapture sampling methods are used (e.g. McLoughlin et al. 2004; Ball et al. 2010; Drake et al. 2018). However, many studies do not clearly describe how duplicate samples were managed (e.g. Klütsch et al. 2016; Yannic et al. 2016). Alternatively, it is

possible that previous broad-scale population genetic studies of woodland caribou did not capture duplicate genotypes between sites due to large distances between sampling locations.

Population genetic structure and landscape genetic studies are commonly completed at broad ranges, including multiple boreal caribou ranges (Weckworth et al. 2013; Yannic et al. 2016; Thompson et al. 2019). This study is the first in Ontario to be completed at a fine-scale level of a single delineated caribou range, the Brightsand Range. The range is located among the southernmost continuous boreal caribou populations (MNRF 2014), and little was known about the population's genetic structure.

The results of the genetic clustering analyses were inconsistent between models, with the inferred number of subpopulations equal to 2, 3, and 8 for STRUCTURE, BAPS, and TESS, respectively. Variation in the inferred number of subpopulations is not uncommon among Bayesian clustering methods (Ball et al. 2010; Yannic et al. 2016) and may be due to low among-site genetic differentiation ($F_{ST} = 0.009$, $p < 0.001$). Studies have demonstrated that low F_{ST} (≤ 0.02) substantially decreases the accuracy of STRUCTURE and BAPS (Latch et al. 2006). STRUCTURE estimated the least number of clusters (K = 2) among the range and assigned the majority of individuals to a single cluster (n = 283), which, based on the model documentation, an asymmetrical proportion of individual assignments supports a true assignment to two population clusters (Pritchard et al. 2010). However, when the cluster assignments were plotted on a map, the assignment proportions appeared as a north-to-south gradient. This indicates that the assignments may have been influenced by an IBD pattern previously detected and may have produced spurious clusters (Pritchard et al. 2010). While BAPS

is normally consistent with STRUCTURE in terms of identifying the correct number of clusters (Latch et al. 2006), in this study, BAPS identified an additional population group but similarly assigned most individuals to a single cluster (n = 395). In contrast, TESS identified eight genetic clusters based on the DIC value. However, the individual assignment probabilities also showed that most individuals (n = 397) were assigned to a single cluster. This may have resulted from the inability to identify a meaningful number of clusters based on a no-admixture analysis prior to the admixture analysis and the best-perceived but arbitrary selection of parameters. Also, TESS has been shown to perform inconsistently for data under a strong IBD pattern (Guillot 2009).

While the population assignment models did not select the same number of clusters, clustering patterns can be used for interpretation to find a consensus between models (Ball et al. 2010). The consensus among Bayesian clustering methods where most individuals were assigned to a single population supports high gene flow among the population within the Brightsand Range with some genetic structuring. It may be possible that the small proportion of statistically significant genetic structuring is a result of the limited genetic isolation among the two southernmost sites, as indicated by the POPGRAPH analysis. The results showed connectivity throughout the range except for the two isolated sites, which were further reflected within the cluster membership maps from BAPS and STRUCTURE where the southernmost sites were more genetically differentiated from the population. Taken together, these results support the prediction that sites from the southern range experience lower levels of gene flow compared to northern sites. Additionally, this was consistent with a previously identified isolated population within the Brightsand Range close to the population identified by Thompson

et al. (2019). Therefore, maintaining the genetic connectivity of boreal caribou to the central network of connectivity may reduce the loss of genetic diversity and prevent the extirpation of the species from the area facing increased landscape fragmentation (Thatte et al. 2018).

Among the landscape genetics analyses, both MEMGENE and a multivariate regression identified wildfire as the only significant resistance variable to gene flow across the range. In both instances, the models explained small amounts of genetic variation (Memgene $R^2 = 0.011$; multivariate regression $R^2 = 0.0618$) and presented different results regarding the resistance cost that best explains genetic variation. The difference in results may be because all genotypes were included in the MEMGENE analysis while the multivariate regression was limited to sites where $n \geq 5$. However, the wildfire disturbance layer remained statistically significant in both models. Within the range, the distribution of wildfire appears to be associated with a natural break between core sampling areas (and sightings). Within this natural break, wildfires ranged from the mid-1990s to 2018, with the top three largest burn areas occurring in 2011, 2018, and 1996. The remaining resistance layers, including roads, small waterbodies and harvest, were not statistically significant within the Brightsand Range. Therefore, this study did not support the prediction that anthropogenic disturbances, including roads and harvest areas, increase levels of among-site variation compared to a pattern of IBD alone. The lack of significant effect may be attributed to a genetic time lag where no genetic change has occurred (Landguth et al. 2010), despite the continuous presence of disturbances. This may be because effective population sizes are large enough that it

would take numerous generations for drift to cause genetic differentiation between populations that maintain gene flow (Charlesworth 2009).

Wildfire has been previously identified to significantly contribute to the resistance of gene flow (Priadka et al. 2019). It is possible that the increased distance between core sampling areas and the significant genetic differentiation from IBD between sites inflated the correlation between wildfire and genetic distance between sampling sites. At the time of sample collections, this natural break of wildfire disturbances between core sampling areas contains extensive wildfire disturbances from 1996, 2011, and the most recent in 2018. However, as caribou are known to avoid post-fire landscapes > 5 years due to reduced lichen supply (Schaefer and Pruitt Jr 1991) and predator avoidance caused by apparent competition from regenerating stands (Wittmer et al. 2007), this may have led to the change in the distribution of boreal caribou within the study. However, wildfire's effect may persist for many years as reports on boreal caribou indicate that individuals avoid burned landscapes for several decades after a fire (Environment Canada 2012).

Roads and harvests were expected to contribute to significant IBR within the managed forest, as they have been shown to act as a barrier to gene flow and movement among caribou populations (Dyer et al. 2001, 2002; Vors et al. 2007; Beauchesne et al. 2013; Pelletier et al. 2019; Priadka et al. 2019). One potential explanation for the lack of significant effect of roads on allele frequency variation among sites in this study is that it included only a small number of sites within the managed forest area due to low occupancy. Thus, the majority of pairwise comparisons of resistance costs did not include road resistance as roads are limited to managed

forests, and most samples were within unmanaged forests. This may also explain why harvest areas did not significantly explain genetic variation, even though forest harvesting has a similar effect to wildfires by creating young forests leading to avoidance due to apparent competition (Wittmer et al. 2007).

While I expected that waterbodies might contribute to reduced gene flow among sites, lakes did not explain a significant portion of allele frequency variation among caribou in this study. Previous studies have found that lakes reduce gene flow among caribou populations however, these studies were at very broad spatial scales i.e. province-wide (Ball et al. 2010; Priadka et al. 2019). Therefore, the differences in scale may account for the contrasting results between studies i.e. the effect of lakes may only be apparent at larger spatial scales (Cushman and Landguth 2010). Alternatively, the insignificant effect of lakes on gene flow may be attributed to the fact that islands in large lakes form an important component of caribou habitat during the calving season (Carr et al. 2011). Additionally, female caribou are strong swimmers known to move between islands during the calving period to escape predators (Bergerud 1985; Carr et al. 2011). Contrastingly, while studies have found that woodland caribou generally avoid lakes during winter (Ferguson and Elkie 2004b; O'Brien et al. 2006), another study from Northern Ontario found that caribou used frozen lakes and actively selected lakes with greater area and perimeter (Ferguson and Elkie 2005). While the available studies of winter space-use and habitat selection of woodland caribou do not provide a clear answer as to whether caribou actively avoid lakes during the winter, the lack of a significant IBR effect of lakes in this study suggests that they do not act as a significant barrier to gene flow of woodland caribou in the Brightsand Range.

Based on the CIRCUITSCAPE analyses, the range was well connected within the first model with lower resistance costs for waterbodies, young forest (harvests), and roads, while the exploratory model had increased dysconnectivity and distinguished corridors among the range. The CIRCUITSCAPE model maps demonstrated that the model could underestimate resistance within the landscape without incorporating known barriers to caribou movement based on the literature. Resistance costs used in CIRCUITSCAPE analyses are sometimes arbitrarily selected based on expert opinion (Spear et al. 2010). However, within this study, I used the statistically significant resistance costs from MEMGENE and increased resistance costs for the exploratory model based on literature review. The lower connectivity in the managed portion of the range within the exploratory model better reflects the genetic data where the southernmost sites were isolated from the rest of the network. It is hypothesized that this genetic differentiation is due to increased disturbances, including roads and harvest areas, as 33.1% of the range is disturbed by anthropogenic sources and was concentrated among the managed forest areas of the range (MNRF 2014). However, sparse sampling present among the managed forest due to low occupancy where roads and harvests are located likely contributed to the inability of the statistical models to detect significant genetic variation.

This study identified wildfire disturbances and IBD as drivers of genetic variation within the Brightsand Range. Both factors contributed to a north-to-south pattern of allele frequency variation, although differentiation between sites was generally low. A previous study across a broad geographic range only identified three genetic clusters, suggesting that the broad mobility of caribou promotes long-distance gene flow such

that populations are not differentiated over narrow geographic ranges (Thompson et al. 2019). While the study did not identify IBD as a driving variable of genetic variation, another study supported IBD as a driver at the individual level (Priadka et al. 2019). Therefore, an IBD pattern may be supported at the fine-scale level. However, due to the weak associations, further sampling would be required to disentangle the effects of wildfire disturbances and IBD. It is also recommended that the study area be extended to surrounding populations to assist in understanding the fine-scale genetic structure of boreal caribou and further verify if the greater genetic differentiation of the northern sites within the Brightsand Range is a result of wildfire or IBD.

CONCLUSION

Overall, this landscape genetic analysis found weak genetic differentiation between sites of boreal woodland caribou within the Brightsand Range of Ontario. Weak but significant genetic differentiation between sites in the northern and southernmost portion of the range was explained by a combination of isolation by distance (IBD) and isolation by resistance (IBR), contributing to variation in allele frequencies. Analyses of genetic connectivity indicated that the two southernmost sampling locations were genetically isolated from the remainder of the network, and landscape connectivity analyses suggested a high degree of landscape fragmentation within the managed forest area. Taken together, these analyses suggest that southernmost sites of boreal caribou are at risk of genetic erosion due to reduced gene flow with sites in the more northern portions of the range. Thus, I suggest that conservation efforts for boreal caribou in the Brightsand Range should be focused on improving genetic connectivity among sites to improve the long-term probability of population persistence.

LITERATURE CITED

Adriaensen F, Chardon JP, De Blust G, et al (2003) The application of “least-cost” modelling as a functional landscape model. Landsc Urban Plan 64:233–247. https://doi.org/10.1016/S0169-2046(02)00242-6

Alberto F (2009) MsatAllele-1.0: An R package to visualize the binning of microsatellite alleles. J Hered 100:394–397. https://doi.org/10.1093/jhered/esn110

Ball MC, 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. Conserv Genet 11:2131–2143. https://doi.org/10.1007/s10592-010-0099-3

Ball MC, Pither R, Manseau M, et al (2007) Characterization of target nuclear DNA from faeces reduces technical issues associated with the assumptions of low-quality and quantity template. Conserv Genet 8:577–586. https://doi.org/10.1007/s10592-006-9193-y

Banks SC, Piggott MP (2022) Non-invasive genetic sampling is one of our most powerful and ethical tools for threatened species population monitoring: a reply to Lavery et al. Biodivers Conserv 31:723–728. https://doi.org/10.1007/s10531-022-02377-x

Beauchesne D, Jaeger JAG, St-Laurent M-H (2014) Thresholds in the capacity of boreal caribou to cope with cumulative disturbances: Evidence from space use patterns. Biol Conserv 172:190–199. https://doi.org/10.1016/j.biocon.2014.03.002

Beauchesne D, Jaeger JAG, St-Laurent MH (2013) Disentangling woodland caribou movements in response to clearcuts and roads across temporal scales. PLoS One 8:e77514. https://doi.org/10.1371/journal.pone.0077514

Bergerud AT (1985) Antipredator strategies of caribou: dispersion along shorelines. Can J Zool 63:1324–1329. https://doi.org/10.1139/z85-199

Bishop MD, Kappes SM, Keele JW, et al (1994) A Genetic Linkage Map for Cattle. Genetics 619–639

Bowman J, Greenhorn JE, Marrotte RR, et al (2016) On applications of landscape genetics. Conserv Genet 17:753–760. https://doi.org/10.1007/s10592-016-0834-5

Carr NL, Rodgers AR, Kingston SR, Lowman DJ (2011) Use of island and mainland shorelines by woodland caribou during the nursery period in two northern Ontario parks. Rangifer 49–61. https://doi.org/10.7557/2.31.2.1989

Charlesworth B (2009) Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10:195–205. https://doi.org/10.1038/nrg2526

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Mol Ecol Notes 7:747–756. https://doi.org/10.1111/j.1471-8286.2007.01769.x

Chessel D, Dufour AB, Thioulouse J (2004) The ade4 Package - I: One-table Methods.

R News 4:5–10
Clark RW, Brown WS, Stechert R, Zamudio KR (2008) Integrating individual behaviour and landscape genetics: The population structure of timber rattlesnake hibernacula. Mol Ecol 17:719–730. https://doi.org/10.1111/j.1365-294X.2007.03594.x
Clauzel C, Bannwarth C, Foltete JC (2015) Integrating regional-scale connectivity in habitat restoration: An application for amphibian conservation in eastern France. J Nat Conserv 23:98–107. https://doi.org/10.1016/j.jnc.2014.07.001
Clevenger AP, Waltho N (2005) Performance indices to identify attributes of highway crossing structures facilitating movement of large mammals. Biol Conserv 121:453–464. https://doi.org/10.1016/j.biocon.2004.04.025
Corander J, Sirén J, Arjas E (2008) Bayesian spatial modeling of genetic population structure. Comput Stat 23:111–129. https://doi.org/10.1007/s00180-007-0072-x
Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics 163:367–374. https://doi.org/10.1093/genetics/163.1.367
Courbin N, Fortin D, Dussault C, Courtois R (2009) Landscape management for woodland caribou: The protection of forest blocks influences wolf-caribou co-occurrence. Landsc Ecol 24:1375–1388. https://doi.org/10.1007/s10980-009-9389-x
Courtois R, Bernatchez L, Ouellet J-P, Breton L (2003a) Significance of caribou (Rangifer tarandus) ecotypes from a molecular genetics viewpoint. Conserv Genet 4:393–404. https://doi.org/10.1023/A:1024033500799
Courtois R, Gingras A, Dussault C, et al (2003b) An aerial survey technique for the forest-dwelling ecotype of Woodland Caribou, Rangifer tarandus caribou. Can Field-Naturalist 117:546–554. https://doi.org/10.22621/cfn.v117i4.823
Cronin MA, MacNeil MD, Patton JC (2005) Variation in Mitochondrial DNA and Microsatellite DNA in Caribou (Rangifer tarandus) in North America. J Mammal 86:495–505
Cumming HG., Beange D. B (1987) Dispersion and Movements of Woodland Caribou Near Lake Nipigon , Ontario. J Wildl Manage 51:69–79
Cumming HG, Hyer BT (1998) Experimental log hauling through a traditional caribou wintering area. Rangifer 241. https://doi.org/10.7557/2.18.5.1562
Cushman SA, Landguth EL (2010) Scale dependent inference in landscape genetics. Landsc Ecol 25:967–979. https://doi.org/10.1007/s10980-010-9467-0
Dalerum F, Boutin S, Dunford JS (2007) Wildfire effects on home range size and fidelity of boreal caribou in Alberta, Canada. Can J Zool 85:26–32. https://doi.org/10.1139/Z06-186
DeCesare NJ, Hebblewhite M, Schmiegelow F, et al (2012) Transcending scale dependence in identifying habitat with resource selection functions. Ecol Appl 22:1068–1083. https://doi.org/10.1890/11-1610.1
Dickson BG, Albano CM, Anantharaman R, et al (2019) Circuit-theory applications to connectivity science and conservation. Conserv Biol 33:239–249. https://doi.org/10.1111/cobi.13230
Drake CC, Manseau M, Klütsch CFC, et al (2018) Does connectivity exist for remnant boreal caribou (Rangifer tarandus caribou) along the Lake Superior Coastal Range? Options for landscape restoration. Rangifer 38:13–26.

https://doi.org/10.7557/2.38.1.4124
Dray S, Dufour A-B, Thioulouse J (2022) ade4: Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences
Durand E, Jay F, Gaggiotti OE, François O (2009) Spatial inference of admixture proportions and secondary contact zones. Mol Biol Evol 26:1963–1973. https://doi.org/10.1093/molbev/msp106
Dyer RJ (2014) gstudio: Analyses and functions related to the spatial analysis of genetic marker data
Dyer RJ, Nason JD (2004) Population Graphs: The graph theoretic shape of genetic structure. Mol Ecol 13:1713–1727. https://doi.org/10.1111/j.1365-294X.2004.02177.x
Dyer SJ., O' Neill JP., Wasel SM., Boutin S (2001) Avoidance of Industrial Development by Woodland Caribou. J Wildl Manage 65:531–542
Dyer SJ, O'Neill JP, Wasel SM, Boutin S (2002) Quantifying barrier effects of roads and seismic lines on movements of female woodland caribou in northeastern Alberta. Can J Zool 80:839–845. https://doi.org/10.1139/z02-060
Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7
Ellegren H (2004) Microsatellites: Simple sequences with complex evolution. Nat Rev Genet 5:435–445. https://doi.org/10.1038/nrg1348
Emel SL, Olson DH, Knowles LL, Storfer A (2019) Comparative landscape genetics of two endemic torrent salamander species, Rhyacotriton kezeri and R. variegatus: implications for forest management and species conservation. Conserv Genet 20:801–815. https://doi.org/10.1007/s10592-019-01172-6
Environment Canada (2011a) Recovery Strategy for the Woodland Caribou (Rangifer tarandus caribou), Boreal Population, in Canada [Proposed]. Environment Canada, Ottawa
Environment Canada (2012) Recovery Strategy for the Woodland Caribou (Rangifer tarandus caribou), Boreal population, in Canada. Environment Canada, Ottawa
Environment Canada (2011b) Scientific Assessment to Inform the Identification of Critical Habitat for Woodland Caribou (Rangifer tarandus), Boreal Population, in Canada: 2011 Update. Ottawa, Ontario, Canada
Epps CW, Palsbøll PJ, Wehausen JD, et al (2005) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. Ecol Lett 8:1029–1038. https://doi.org/10.1111/j.1461-0248.2005.00804.x
ESRI (2022) ArcGIS Pro v3.0
Estes-Zumpf WA, Rachlow JL, Waits LP, Warheit KI (2010) Dispersal, gene flow, and population genetic structure in the pygmy rabbit (brachylagus idahoensis). J Mammal 91:208–219. https://doi.org/10.1644/09-MAMM-A-032R.1
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x
Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479–491. https://doi.org/10.1093/genetics/131.2.479

Ferguson SH, Elkie PC (2004a) Seasonal movement patterns of woodland caribou (Rangifer tarandus caribou). J Zool 262:125–134. https://doi.org/10.1017/S0952836903004552

Ferguson SH, Elkie PC (2004b) Habitat requirements of boreal forest caribou during the travel seasons. Basic Appl Ecol 5:465–474. https://doi.org/10.1016/j.baae.2004.04.011

Ferguson SH, Elkie PC (2005) Use of Lake Areas in Winter by Woodland Caribou. Northeast Nat 12:45–66

Fischer J, Lindenmayer DB, Hobbs RJ (2012) Landscape pattern and biodiversity. In S. A. Levin, S. R. Carpenter, & H. C. J. Godfray (Eds.), The Princeton Guide to ecology. Princeton University Press

Fox J, Weisberg S (2019) An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA

François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. Mol Ecol Resour 10:773–784. https://doi.org/10.1111/j.1755-0998.2010.02868.x

Frankham R (2019) Conservation Genetics. In: Fath B (ed) Encyclopedia of Ecology. Elsevier, pp 182–390

Frankham R, Ballou J, Briscoe DA (2010a) Genetic Diversity. In: Introduction to Conservation Genetics, Second. Cambridge University Press, New York, pp 41–65

Frankham R, Ballou JD, Briscoe DA (2010b) Population fragmentation. In: Introduction to Conservation Genetics, Second. Cambridge University Press, New York, pp 309–336

Frantz AC, Bertouille S, Eloy MC, et al (2012) Comparative landscape genetic analyses show a Belgian motorway to be a gene flow barrier for red deer (Cervus elaphus), but not wild boars (Sus scrofa). Mol Ecol 21:3445–3457. https://doi.org/10.1111/j.1365-294X.2012.05623.x

Frantz AC, Cellina S, Krier A, et al (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? J Appl Ecol 46:493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Fryxell JM, Avgar T, Liu B, et al (2020) Anthropogenic Disturbance and Population Viability of Woodland Caribou in Ontario. J Wildl Manage 84:636–650. https://doi.org/10.1002/jwmg.21829

Fuller TK, Keith LB (1981) Woodland Caribou Population Dynamics in Northeastern Alberta. J Wildl Manage 45:197–213

Funk WC, Blouin MS, Corn PS, et al (2005) Population structure of Columbia spotted frogs (Rana luteiventris) is strongly affected by the landscape. Mol Ecol 14:483–496. https://doi.org/10.1111/j.1365-294X.2005.02426.x

Galpern P, Manseau M, Hettinga P, et al (2012a) Allelematch: An R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. Mol Ecol Resour 12:771–778. https://doi.org/10.1111/j.1755-0998.2012.03137.x

Galpern P, Manseau M, Wilson P (2012b) Grains of connectivity: Analysis at multiple spatial scales in landscape genetics. Mol Ecol 21:3996–4009. https://doi.org/10.1111/j.1365-294X.2012.05677.x

Galpern P, Peres-Neto P (2014) MEMGENE package for R: Tutorials. R Help 1–14
Galpern P, Peres-Neto PR, Polfus J, Manseau M (2014) MEMGENE: Spatial pattern detection in genetic distance data. Methods Ecol Evol 5:1116–1120. https://doi.org/10.1111/2041-210X.12240
Gariboldi MC, Túnez JI, Failla M, et al (2016) Patterns of population structure at microsatellite and mitochondrial DNA markers in the franciscana dolphin (Pontoporia blainvillei). Ecol Evol 6:8764–8776. https://doi.org/10.1002/ece3.2596
Gemmell NJ, Allen PJ, Goodman SJ, Reed JZ (1997) Interspecific microsatellite markers for the study of pinniped populations. Mol Ecol 6:661–666. https://doi.org/10.1046/j.1365-294X.1997.00235.x
Grover A, Sharma PC (2016) Development and use of molecular markers: Past and present. Crit Rev Biotechnol 36:290–302. https://doi.org/10.3109/07388551.2014.959891
Gubili C, Mariani S, Weckworth B V., et al (2017) Environmental and anthropogenic drivers of connectivity patterns: A basis for prioritizing conservation efforts for threatened populations. Evol Appl 10:199–211. https://doi.org/10.1111/eva.12443
Guillot G (2009) On the inference of spatial structure from population genetics data. Bioinformatics 25:1796–1801. https://doi.org/10.1093/bioinformatics/btp267
Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. Mol Ecol 18:4734–4756. https://doi.org/10.1111/j.1365-294X.2009.04410.x
Guillot G, Rousset F (2013) Dismantling the Mantel tests. Methods Ecol Evol 4:336–344. https://doi.org/10.1111/2041-210x.12018
Hale ML, Burg TM, Steeves TE (2012) Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. PLoS One 7:e45170. https://doi.org/10.1371/journal.pone.0045170
Hedrick P (2012) Population genetics and ecology. In S. A. Levin, S. R. Carpenter, & H. C. J. Godfray (Eds.), The Princeton Guide to ecology. Princet Univ Press
Hettinga PN, Arnason AN, Manseau M, et al (2012) Estimating size and trend of the north interlake woodland caribou population using fecal-DNA and capturerecapture models. J Wildl Manage 76:1153–1164. https://doi.org/10.1002/jwmg.380
Holderegger R, Wagner HH (2008) Landscape Genetics. Bioscience 58:199–207
Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806. https://doi.org/10.1093/bioinformatics/btm233
Jensen EL, Tschritter C, de Groot PVC, et al (2020) Canadian polar bear population structure using genome-wide markers. Ecol Evol 10:3706–3714. https://doi.org/10.1002/ece3.6159
Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6:1–10. https://doi.org/10.3389/fgene.2015.00208
Kamvar ZN, Everhart SE, Grünwald NJ (2017) Population genetics and genomics in R: AMOVA. https://grunwaldlab.github.io/Population_Genetics_in_R/AMOVA.html
Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends Ecol Evol 17:230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Klütsch CFC, Manseau M, Trim V, et al (2016) The eastern migratory caribou: The role of genetic introgression in ecotype evolution. R Soc Open Sci 3:. https://doi.org/10.1098/rsos.150469

Koen EL, Garroway CJ, Wilson PJ, Bowman J (2010) The effect of map boundary on estimates of landscape resistance to animal movement. PLoS One 5:e11785. https://doi.org/10.1371/journal.pone.0011785

Kopelman NM, Mayzel J, Jakobsson M, et al (2015) Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191. https://doi.org/10.1111/1755-0998.12387

L'Italien L, Weladji RB, Holand Ø, et al (2012) Mating group size and stability in reindeer rangifer tarandus: The effects of male characteristics, sex ratio and male age structure. Ethology 118:783–792. https://doi.org/10.1111/j.1439-0310.2012.02073.x

Lada H, Thomson JR, Mac Nally R, Taylor AC (2008) Impacts of massive landscape change on a carnivorous marsupial in south-eastern Australia: Inferences from landscape genetics analysis. J Appl Ecol 45:1732–1741. https://doi.org/10.1111/j.1365-2664.2008.01563.x

Landguth EL, Cushman SA, Schwartz MK, et al (2010) Quantifying the lag time to detect barriers in landscape genetics. Mol Ecol 19:4179–4191. https://doi.org/10.1111/j.1365-294X.2010.04808.x

Landguth EL, Schwartz MK (2014) Evaluating sample allocation and effort in detecting population differentiation for discrete and continuously distributed individuals. Conserv Genet 15:981–992. https://doi.org/10.1007/s10592-014-0593-0

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes Jr. OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conserv Genet 7:295–302. https://doi.org/10.1007/s10592-005-9098-1

Latham ADM, Latham MC, Boyce MS, Boutin S (2011) Movement responses by wolves to industrial linear features and their effect on woodland caribou in northeastern alberta. Ecol Appl 21:2854–2865. https://doi.org/10.1890/11-0666.1

Loxterman JL (2011) Fine scale population genetic structure of pumas in the Intermountain West. Conserv Genet 12:1049–1059. https://doi.org/10.1007/s10592-011-0208-y

Manel S, Holderegger R (2013) Ten years of landscape genetics. Trends Ecol Evol 28:614–621. https://doi.org/10.1016/j.tree.2013.05.012

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. Trends Ecol Evol 18:189–197. https://doi.org/10.1016/S0169-5347(03)00008-9

Mantel N (1967) The Detection of Disease Clustering and a Generalized Regression Approach. Cancer Res 27:209–220

Maudet C, Luikart G, Dubray D, et al (2004) Low genotyping error rates in wild ungulate faeces sampled in winter. Mol Ecol Notes 4:772–775. https://doi.org/10.1111/j.1471-8286.2004.00787.x

McFarlane S, Manseau M, Wilson PJ (2021) Spatial familial networks to infer demographic structure of wild populations. Ecol Evol 11:4507–4519. https://doi.org/10.1002/ece3.7345

McLoughlin PD, Paetkau D, Duda M, Boutin S (2004) Genetic diversity and relatedness

of boreal caribou populations in western Canada. Biol Conserv 118:593–598. https://doi.org/10.1016/j.biocon.2003.10.008
McRae BH (2006) Isolation By Resistance. Evolution (N Y) 60:1551–1561. https://doi.org/10.1554/05-321.1
McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. Proc Natl Acad Sci U S A 104:19885–19890. https://doi.org/10.1073/pnas.0706568104
McRae BH, Dickson BG, Keitt TH, Shah VB (2008) USING CIRCUIT THEORY TO MODEL CONNECTIVITY IN ECOLOGY, EVOLUTION, AND CONSERVATION. Ecol Soc Am 89:2712–2724
McRae BH, Shah VB, Mohapatra T (2014) CIRCUITSCAPE User Guide. Nat Conserv 30
MNRF (2014) Integrated Range Assessment for Woodland Caribou and their Habitat: Brightsand Range 2011. Species at Risk Branch, Thunder Bay, Ontario
Moreau G, Fortin D, Couturier S, Duchesne T (2012) Multi-level functional responses for wildlife conservation: The case of threatened caribou in managed boreal forests. J Appl Ecol 49:611–620. https://doi.org/10.1111/j.1365-2664.2012.02134.x
Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York
O'Brien D, Manseau M, Fall A, Fortin MJ (2006) Testing the importance of spatial configuration of winter habitat for woodland caribou: An application of graph theory. Biol Conserv 130:70–83. https://doi.org/10.1016/j.biocon.2005.12.014
OMECP (2010) Caribou Range Boundary. Area classification data [computer file]. Ontario Ministry of the Environment, Conservation and Parks. Last updated September 8, 2022
OMNR (2010) Ontario Hydro Network - Waterbodies. Land classification data [computer file]. Ontario Ministry of Natural Resources and Forestry. Last updated May 17, 2022
OMNR (2014) Forest Management Guide for Boreal Landscapes. Queen's Printer for Ontario, Toronto
OMNRF (2001) Ontario Road Network (ORN) Road Net Element. Land classification data [computer file]. Ontario Ministry of Natural Resources and Forestry - Provincial Mapping Unit. Last updated May 24, 2022
OMNRF (1960) Fire Disturbance Area. Land classification data [computer file]. Ontario Ministry of Natural Resources and Forestry. Last updated March 7, 2022
OMNRF (2020) Forest Resource Inventory. Land classification data [multiple computer files]. Ontario Ministry of Natural Resources and Forestry. Last updated January 19, 2022
Ontario (2022) Ontario GeoHub. Queen's Printer for Ontario. https://geohub.lio.gov.on.ca/. Accessed 11 Jul 2022
Parks LC, Wallin DO, Cushman SA, McRae BH (2015) Landscape-level analysis of mountain goat population connectivity in Washington and southern British Columbia. Conserv Genet 16:1195–1207. https://doi.org/10.1007/s10592-015-0732-2
Peakall R, Smouse PE (2012) GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pelletier F, Turgeon G, Bourret A, et al (2019) Genetic structure and effective size of an endangered population of woodland caribou. Conserv Genet 20:203–213. https://doi.org/10.1007/s10592-018-1124-1

Peres-neto AP, Galpern P (2014) Package ‘ memgene .’ 1–13

Priadka P, Manseau M, Trottier T, et al (2019) Partitioning drivers of spatial genetic variation for a continuously distributed population of boreal caribou: Implications for management unit delineation. Ecol Evol 9:141–153. https://doi.org/10.1002/ece3.4682

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Pritchard JK, Wen X, Falush D (2010) Documentation for structure software: Version 2.3. 39

Proft KM, Jones ME, Johnson CN, Burridge CP (2018) Making the connection: expanding the role of restoration genetics in restoring and evaluating connectivity. Restor Ecol 26:411–418. https://doi.org/10.1111/rec.12692

Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecol Evol 4:4399–4428. https://doi.org/10.1002/ece3.1305

Rettie WJ, Messier F (2000) Hierarchical habitat selection by woodland caribou: Its relationship to limiting factors. Ecography (Cop) 23:466–478. https://doi.org/10.1111/j.1600-0587.2000.tb00303.x

Rettie WJ, Messier F (1998) Dynamics of woodland caribou populations at the southern limit of their range in Saskatchewan. Can J Zool 76:251–259. https://doi.org/10.1139/cjz-76-2-251

Rivera-Ortíz FA, Aguilar R, Arizmendi MDC, et al (2015) Habitat fragmentation and genetic variability of tetrapod populations. Anim Conserv 18:249–258. https://doi.org/10.1111/acv.12165

Rivers NM, Butlin RK, Altringham JD (2005) Genetic population structure of Natterer’s bats explained by mating at swarming sites and philopatry. Mol Ecol 14:4299–4312. https://doi.org/10.1111/j.1365-294X.2005.02748.x

Røed KH, Midthjell L (1998) Microsatellites in reindeer, Rangifer tarandus, and their use in other cervids. Mol Ecol 7:1773–1776. https://doi.org/10.1046/j.1365-294x.1998.00514.x

Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. Mol Ecol Notes 4:137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Rousset F (2008) GENEPOP’007: A complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Rudolph TD, Drapeau P (2012) Using movement behaviour to define biological seasons for woodland caribou. Rangifer 295–307. https://doi.org/10.7557/2.32.2.2277

Sawaya MA, Clevenger AP, Kalinowski ST (2013) Demographic Connectivity for Ursid Populations at Wildlife Crossing Structures in Banff National Park. Conserv Biol 27:721–730. https://doi.org/10.1111/cobi.12075

Sawyer SC, Epps CW, Brashares JS (2011) Placing linkages among fragmented

habitats: Do least-cost models reflect how animals use landscapes? J Appl Ecol 48:668–678. https://doi.org/10.1111/j.1365-2664.2011.01970.x
Schaefer JA, Pruitt Jr WO (1991) Fire and Woodland Caribou in Southeastern Manitoba. Wildl Monogr 3–39
Schindler DW, Walker D, Davis T, Westwood R (2007) Determining effects of an all weather logging road on winter woodland caribou habitat use in south-eastern Manitoba. Rangifer 209–217
Schlötterer C (2004) The evolution of molecular markers - Just a matter of fashion? Nat Rev Genet 5:63–69. https://doi.org/10.1038/nrg1249
Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol 22:25–33. https://doi.org/10.1016/j.tree.2006.08.009
Shah VB, McRae BH (2008) Circuitscape : A Tool for Landscape Ecology. Proc 7th Python Sci Conf 62–65. https://doi.org/10.1111/j.1523-1739.2008.00942.x
Slatkin M (1993) Isolation by Distance in Equilibrium and Non-Equilibrium Populations. Evolution (N Y) 47:264–279
Soanes K, Taylor AC, Sunnucks P, et al (2018) Evaluating the success of wildlife crossing structures using genetic approaches and an experimental design: Lessons from a gliding mammal. J Appl Ecol 55:129–138. https://doi.org/10.1111/1365-2664.12966
Spear SF, Balkenhol N, Fortin MJ, et al (2010) Use of resistance surfaces for landscape genetic studies: Considerations for parameterization and analysis. Mol Ecol 19:3576–3591. https://doi.org/10.1111/j.1365-294X.2010.04657.x
Stewart FEC, Nowak JJ, Micheletti T, et al (2020) Boreal Caribou Can Coexist with Natural but Not Industrial Disturbances. J Wildl Manage 84:1435–1444. https://doi.org/10.1002/jwmg.21937
Storfer A, Murphy MA, Evans JS, et al (2007) Putting the “landscape” in landscape genetics. Heredity (Edinb) 98:128–142. https://doi.org/10.1038/sj.hdy.6800917
Stuart-smith K, Bradshaw CJA, Boutin S, et al (1997) Woodland Caribou Relative to Landscape Patterns in Northeastern Alberta. J Wildl Manage 61:622–633
Tammeleht E, Remm J, Korsten M, et al (2010) Genetic structure in large, continuous mammal populations: The example of brown bears in northwestern Eurasia. Mol Ecol 19:5359–5370. https://doi.org/10.1111/j.1365-294X.2010.04885.x
Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471. https://doi.org/10.1093/nar/gkp1211
Thatte P, Joshi A, Vaidyanathan S, et al (2018) Maintaining tiger connectivity and minimizing extinction into the next century: Insights from landscape genetics and spatially-explicit simulations. Biol Conserv 218:181–191. https://doi.org/10.1016/j.biocon.2017.12.022
Thomas D., Gray D. (2002) update COSEWIC status report on the woodland caribou Rangifer tarandus caribou in Canada, in COSEWIC assessment and update status report on the Woodland Caribou Rangifer tarandus caribou in Canada. Comm Status Endanger Wildl Canada Ottawa 1–98
Thompson LM, Klütsch CFC, Manseau M, Wilson PJ (2019) Spatial differences in genetic diversity and northward migration suggest genetic erosion along the boreal

caribou southern range limit and continued range retraction. Ecol Evol 9:7030–7046. https://doi.org/10.1002/ece3.5269

Thomson A, Hettinga P, Anderson N (2021) Estimation of Census Population Size for Woodland Caribou in the Brightsand Range, Ontario, Using Fecal DNA Mark-Recapture Analysis. Prepared for Environment and Climate Change Canada and the Canadian Wildlife Service

Titus VR, Bell RC, Becker CG, Zamudio KR (2014) Connectivity and gene flow among Eastern Tiger Salamander (Ambystoma tigrinum) populations in highly modified anthropogenic landscapes. Conserv Genet 15:1447–1462. https://doi.org/10.1007/s10592-014-0629-5

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Vieira MLC, Santini L, Diniz AL, Munhoz C de F (2016) Microsatellite markers: What they mean and why they are so useful. Genet Mol Biol 39:312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Vonholdt BM, Stahler DR, Bangs EE, et al (2010) A novel assessment of population structure and gene flow in grey wolf populations of the Northern Rocky Mountains of the United States. Mol Ecol 19:4412–4427. https://doi.org/10.1111/j.1365-294X.2010.04769.x

Vors LS, Schaefer JA, Pond BA, et al (2007) Woodland Caribou Extirpation and Anthropogenic Landscape Disturbance in Ontario. J Wildl Manage 71:1249–1256. https://doi.org/10.2193/2006-263

Waits LP, Paetkau D (2005) Noninvasive Genetic Sampling Tools for Wildlife Biologists: a Review of Applications and Recommendations for Accurate Data Collection. J Wildl Manage 69:1419–1433. https://doi.org/10.2193/0022-541x(2005)69[1419:ngstfw]2.0.co;2

Weckworth B V, Musiani M, Decesare NJ, et al (2013) Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. Proc R Soc B Biol Sci 280:1–9

Welke CA, Graham B, Conover RR, et al (2021) Habitat-linked genetic structure for white-crowned sparrow (Zonotrichia leucophrys): Local factors shape population genetic structure. Ecol Evol 11:11700–11717. https://doi.org/10.1002/ece3.7887

Wilson GA, Strobeck C, Wu L, Coffin JW (1997) Characterization of microsatellite loci in caribou Rangifer tarandus, and their use in other artiodactyls. Mol Ecol 6:697–699. https://doi.org/10.1046/j.1365-294X.1997.00237.x

Wilson KS, Pond BA, Brown GS, Schaefer JA (2019) The biogeography of home range size of woodland caribou Rangifer tarandus caribou. Divers Distrib 25:205–216. https://doi.org/10.1111/ddi.12849

Wittmer HU, McLellan BN, Serrouya R, Apps CD (2007) Changes in landscape composition influence the decline of a threatened woodland caribou population. J Anim Ecol 76:568–579. https://doi.org/10.1111/j.1365-2656.2007.01220.x

Woodruff SP, Johnson TR, Waits LP (2015) Evaluating the interaction of faecal pellet deposition rates and DNA degradation rates to optimize sampling design for DNA-based mark-recapture analysis of Sonoran pronghorn. Mol Ecol Resour 15:843–854. https://doi.org/10.1111/1755-0998.12362

Woods JG, Paetkau D, Lewis D, et al (1999) Genetic tagging of free-ranging black and brown bears. Wildl Soc Bull 27:616–627

Wright S (1943) Isolation by Distance. Genetics 28:114–138

Wright S (1969) Evolution and the Genetics of Populations. Vol 2. The Theory of Gene Frequencies. The University of Chicago Press, Chicago

Yannic G, Pellissier L, Ortego J, et al (2014) Genetic diversity in caribou linked to past and future climate change. Nat Clim Chang 4:132–137. https://doi.org/10.1038/nclimate2074

Yannic G, St-Laurent MH, Ortego J, et al (2016) Integrating ecological and genetic structure to define management units for caribou in Eastern Canada. Conserv Genet 17:437–453. https://doi.org/10.1007/s10592-015-0795-0

Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol 1:3–14. https://doi.org/10.1111/j.2041-210x.2009.00001.x

APPENDICES

APPENDIX I

SUMMARY OF COLLECTION PERIODS

February Sample Collection Period

Site ID	Number of Samples Collected	Unique Genotypes within Site	Region	Latitude	Longitude
A01	14	10	Protected	[redacted]	
A02	15	13	Protected		
A03	19	9	Protected		
A04	14	5	Protected		
A05	30	14	Managed		
A06	11	8	Managed		
A07	8	6	Protected		
A08	15	9	Protected		
A09	18	13	Protected		
A10	9	8	Protected		
A11	9	4	Protected		
A12	9	5	Protected		
A13	7	3	Managed		
A14	22	11	Managed		
A15	17	11	Protected		
A16	16	11	Protected		
A17	31	18	Protected		
A18	11	8	Protected		
A19	5	5	Protected		
A20	3	3	Protected		
A21	11	7	Protected		
A22	9	7	Protected		
A23	4	2	Protected		
A24	7	4	Managed		
A25	5	3	Protected		
A26	15	10	Protected		
A27	6	5	Managed		
A28	16	10	Managed		

March Sample Collection Period

Site ID	Number of Samples Collected	Unique Genotypes within Site	Region	Latitude	Longitude
B01	5	4	Protected		
B02	11	7	Protected		
B03	36	12	Protected		
B04	19	9	Protected		
B06	13	5	Protected		
B07	3	3	Protected		
B08	10	6	Protected		
B09	3	3	Protected		
B10	9	7	Protected		
B11	1	1	Protected		
B12	47	19	Protected		
B13	46	12	Protected		
B14	2	2	Protected		
B15	3	3	Protected		
B16	7	4	Protected		
B17	2	2	Protected		
B18	46	16	Protected		
B19	11	5	Protected		
B20	2	1	Protected		
B21	40	18	Protected		
B22	6	3	Protected		
B23	5	5	Protected		
B24	14	7	Protected		
B26	8	6	Protected		
B27	5	4	Protected		
B29	18	8	Protected		
B30	7	4	Protected		
B31	3	2	Protected		
B32	7	3	Protected		
B33	20	8	Managed		
B35	2	1	Managed		
B36	17	9	Protected		
B37	5	3	Protected		

APPENDIX II

EXTRACTION PROTOCOL DOCUMENT

CARIBOU FECAL DNA EXTRACTION PROTOCOL

This protocol was used for the extraction of DNA from frozen caribou fecal pellets. A QiagenDNeasy Kit was used to complete the extraction following a modification of the procedure of Ball et al. 2007.

BEFORE STARTING:

- Buffer AW1 and AW2 in the DNEasy kit come as concentrates. The amount of 95-100% ethanolthat must be added to each bottle is listed on the front and should be added before the extraction is begun.
- A water bath should be preheated before step 3 to 65 degrees.
- Buffer AL and ethanol used in step 5 can be pre-mixed as one stock solution of equal partsethanol and AL buffer.

PROCEDURE:

1. Collect the mucosal membrane from the outside of four pellets taken from the same sample, using a different toothpick for each pellet (use toothpicks that have been cut inhalf so that they fit into the 2 ml microcentrifuge tube used in step 2). Place the four toothpicks into a paper envelope for storage until use or use immediately in step two.

2. Using tweezers that are cleaned with ethanol after each sample, place the four toothpicks from each sample into a clean 2 ml microcentrifuge tube. Add 400 µL of buffer ATL and 40 µL of proteinase K to each tube. Vortex each tube for 5-10 seconds to mix.

   If there is any white precipitate in the ATL bottle, allow the bottle to sit in a 70°C waterbath for 1 minute.

3. Place tubes in a water bath at 65°C for two hours, making sure to mix the samples for thefirst hour by removing them from the bath and shaking or vortexing every ten minutes. After two hours, use tweezers to remove the

toothpicks. Add an additional 40 µL of proteinase K to each tube. Leave the samples in the water bath overnight (around 12-14 hours) at 56°C.discard the toothpicks and centrifuge samples at 11000 x g for 5 minutes to pellet leftover matter at the bottom of the tube. Carefully pipette off the liquid into a new 1.5 ml microcentrifuge tube and discard the used 2 ml tube.

4. Pipette 400 µL of buffer AL and 400 µL of 95-100% ethanol into each tube (if using apre-made stock of 1:1 AL and ethanol, pipette 800 µL of the stock mixture into each tube). Mix thoroughly with the pipette.

5. Transfer the mixture from the step 5 into a DNeasy mini spin column placed in a 2mL collection tube (from the Qiagen kit). Centrifuge the samples at 6000 x g for 1 minute. Remove the central spin column from the 2 ml collection tube and place it into a new 2 mlcollection tube (from the Qiagen kit). Discard the flow-through and the old collection tube.

6. Add 500 µL of buffer AW1 to the spin column and centrifuge at 6000 x g for 1 minute.Remove the central spin column from the 2ml flow tube and place it into a new 2 ml collection tube (from the Qiagen kit). Discard the flow-through and the old collection tube.

7. Add 500 µL of buffer AW2 to the spin column and centrifuge at 20000 x g for 3 minutes.Place the mini spin column in a clean 1.5 ml microcentrifuge tube (cut the cap off of the tube at the bend joint so that the column can fit better). *When removing the mini spin column from the collection tube, do not allow the membrane to come into contact with the flow-through, as this will cause ethanol to be carried over. If ethanol carryover occurs, empty the collection tube and reuse it with thespin column for another centrifugation at 20 000 x g for 1 minute.*

8. Heat 100ml of H_2O to 70°C in a glass beaker and allow the bottle of buffer AE to sit in the water bath for 2 minutes. Pipette 65 µL of buffer AE directly onto the membrane in the DNeasy mini spin column and allow it to incubate at room temperature for 3 minutes.Centrifuge at 6000 x g for 1 minute.

9. Pipette the product collected in the 1.5 ml microcentrifuge tube into a new 1.5 ml microcentrifuge tube with a lid and label. Discard the DNeasy mini spin column. Storeat -20°C.

References

Ball MC, Pither R, Manseau M, et al (2007) Characterization of target nuclear DNA from faeces reduces technical issues associated with the assumptions of low-quality and quantity template. Conserv Genet 8:577–586. doi: 10.1007/s10592-006-9193-y

APPENDIX III

MULTIPLEX RECIPES

Multiplex 1	1x (µl)
Qiagen Multiplex PCR Master Mix 2x	5
Nuclease Free Water	0.4
BM6506-Forward	0.2
BM6506-Reverse	0.2
BMS1788- Forward	0.1
BMS1788- Reverse	0.1
NVHRT30- Forward	0.15
NVHRT30- Reverse	0.15
RT24- Forward	0.2
RT24- Reverse	0.2
RT6- Forward	0.4
RT6- Reverse	0.4
DNA	2.5
Total	10

Multiplex 2	1x (µl)
Qiagen Multiplex PCR Master Mix 2x	5
Nuclease Free Water	1.2
RT27 – Forward	0.1
RT27 – Reverse	0.1
RT5 – Forward	0.1
RT5 – Reverse	0.1
RT7 – Forward	0.15
RT7 – Reverse	0.15
RT9 – Forward	0.3
RT9 – Reverse	0.3
DNA	2.5
Total	10

Multiplex 3	1x (µl)
Qiagen Multiplex PCR Master Mix 2x	5
Nuclease Free Water	0.1
RT30 – Forward	0.2
RT30 – Reverse	0.2
BM848 – Forward	0.3
BM848 – Reverse	0.3
BM888 – Forward	0.1
BM888 – Reverse	0.1
Map2C – Forward	0.2
Map2C – Reverse	0.2
RT1 – Forward	0.2
RT1 – Reverse	0.2
BM4513 – Forward	0.2
BM4513 – Reverse	0.2
DNA	2.5
TOTAL	10

MICROSATELLITE MARKER INFORMATION

Marker	Reference	Expected size*	Dye	Forward (5'-3')	Reverse (5'-3')
RT6	Wilson et al. 1997	88-120	6FAM	TTCCTCTTACTCATTCTTGG	CGGATTTTGAGACTGTTAC
BM6506	Bishop et al. 1994	188-222	6FAM	GCACGTGGTAAAGAGATGGC	AGCAACTTGAGCATGGCAC
BMS1788*	Yannic et al 2014 - modified from Cronin et al. 2005	103-143	VIC	ATTCATATCTACGTCCAGATTCAGATTTCTTG	GGAGAGGAATCTTGCAAAGG
NVHRT30	Røed & Midthjell 1998	159–177	VIC	GTGGAGCATTGTGTATGTGT	GCCCCCACTGTGTTTT
RT24	Wilson et al. 1997	215-227	NED	TGTATCCATCTGGAAGATTTCAG	CAGTTTAACCAGTCCTCTGTG
RT27	Wilson et al. 1997	135-155	6FAM	CCAAAGACCCAACAGATG	TTGTAACACAGCAAAAGCATT
RT9	Wilson et al. 1997	116-128	VIC	TGAAGTTTAATTTCCACTCT	CAGTCACTTTCATCCCACAT
RT7	Wilson et al. 1997	216-234	VIC	CCTGTTCTACTCTTCTTCTC	ACTTTTCACGGGCACTGGTT
RT5	Wilson et al. 1997	143-171	NED	TGGTTGGAAGGAAAACTTGG	CCTCTGCTCCTCAAGACAC
Map2C	Bishop et al. 1994	81-119	6FAM	TTTACCAGACAGTTTAGTTTTGAGC	AAGGATTCTGTCTGATACCACTTAG
RT1	Wilson et al. 1997	208-244	6FAM	TGCCTTCTTTCATCCAACAA	CATCTTCCCATCCTCTTTAC
BM888	Bishop et al. 1994	158-210	VIC	AGGCCATATAGGAGGCAAGCTT	CTCGGTCAGCTCAAAACGAG
BM848	Bishop et al. 1994	352-384	NED	TGGTTGGAAGGAAAACTTGG	CCTCTGCTCCTCAAGACAC
BM4513	Bishop et al. 1994	118-150	PET	TCAGCAATTCAGTACATCACCC	GCGCAAGTTTCCTCATGC
RT30	Wilson et al. 1997	181-207	PET	CACTTGGCTTTTGGACTTA	CTGGTGTATGTATGCACACT

*Expected Size – in Rangifer tarandus, references for expected size in Rangifer tarandus for markers from Bishop et al 1994 are from https://datadryad.org/stash/dataset/doi:10.5061/dryad.9qh56

APPENDIX IV

ALLELE DISTRIBUTIONS

Bins are assigned based on MsatAllele and have been adjusted to allow bins to capture alleles that vary up to ±0.4 from the samples analysed.

Table Headings:

Bins = MSATALLELE assigned bin name
Min and Max = Minimum and Maximum extent of the allele peak sizes
New Min and New Max = Adjusted bin sizes to include an extra 0.4 bp at both extents
New Range = Bin size in bp

Multiplex 1

RT6

Bins	Min	Max	New Min	New Max	New Range
93	92.989	93.111	92.5	93.6	1.1
97	96.552	96.77	96.1	97.2	1.1
99	98.449	98.648	98.0	99.1	1.1
100	100.261	100.422	99.8	100.9	1.1
102	102.017	102.144	101.6	102.6	1.0
106	105.561	105.758	105.1	106.2	1.1
107	107.375	107.597	106.9	108.0	1.1
109	109.193	109.424	108.7	109.9	1.2

BM6506

Bins	Min	Max	New Min	New Max	New Range
196	195.716	195.833	195.3	196.3	1.0
198	197.515	197.789	197.1	198.2	1.1
201	201.248	201.535	200.8	202	1.2
207	206.986	207.264	206.5	207.7	1.2
209	208.881	209.218	208.4	209.7	1.3

BMS1788

Bin 129 was removed as it held a single allele peak. This locus was removed from analyses.

Bins	Min	Max	New Min	New Max	New Range
113	113.203	113.311	112.8	113.7	0.9
117	117.318	117.49	116.9	117.9	1.0
124	123.674	124.114	123.3	124.5	1.2
126	125.874	125.972	125.5	126.4	0.9
128	127.783	128.884	127.4	128.8	1.4
129	128.898	128.898			
130	128.926	130.337	128.9	130.7	1.8
132	131.878	132.342	131.5	132.7	1.3
134	133.367	134.223	133.0	134.6	1.7
136	135.701	136.013	135.3	136.4	1.1
137	137.256	138.23	136.9	138.6	1.8
143	142.464	143.407	142.1	143.8	1.7

NVHRT30

This locus was removed from analyses

Bins	Min	Max	New Min	New Max	New Range
147	147.147	147.205	146.7	147.7	1
154	154.291	154.445	153.8	154.9	1.1
157	156.528	156.716	156.1	157.2	1.1
159	158.758	159.181	158.3	159.6	1.3
161	160.89	161.145	160.4	161.6	1.2
163	163.046	163.251	162.6	163.7	1.1
165	165.193	165.415	164.7	165.9	1.2
167	167.289	167.601	166.8	168.1	1.3
169	169.371	169.644	168.9	170.1	1.2
172	171.465	172.035	171	172.5	1.5
174	173.602	173.837	173.2	174.3	1.1

RT24

Bins	Min	Max	New Min	New Max	New Range
207	206.665	206.887	206.2	207.3	1.1
209	208.848	209.033	208.4	209.5	1.1
211	210.881	210.98	210.4	211.4	1
213	213	213.038	212.6	213.5	0.9
215	215.159	215.209	214.7	215.7	1
217	217.161	217.428	216.7	217.9	1.2
219	219.196	219.611	218.7	220.1	1.4
221	221.246	221.589	220.8	222	1.2
224	223.498	223.685	223	224.1	1.1
226	225.613	225.732	225.2	226.2	1
234	234.054	234.205	233.6	234.7	1.1

Multiplex 2

RT27

Bins	Min	Max	New Min	New Max	New Range
131	131.204	131.368	130.8	131.8	1
134	133.198	133.624	132.7	134.1	1.4
145	145.332	145.47	144.9	145.9	1
148	147.499	148.038	147	148.5	1.5
150	150.399	150.548	149.9	151	1.1
153	152.775	152.902	152.3	153.4	1.1
155	155.077	155.172	154.6	155.6	1
160	159.591	159.7	159.1	160.1	1

RT9

Bins	Min	Max	New Min	New Max	New Range
103	103.307	103.486	102.9	103.9	1
116	116.056	116.242	115.6	116.7	1.1
120	119.88	120.03	119.4	120.5	1.1
122	121.745	121.971	121.3	122.4	1.1
124	123.712	123.914	123.3	124.4	1.1
128	127.649	127.846	127.2	128.3	1.1
130	129.707	129.794	129.3	130.2	0.9

RT7

Bins	Min	Max	New Min	New Max	New Range
217	216.776	217.253	216.3	217.7	1.4
219	218.708	219.079	218.3	219.5	1.2
221	220.612	220.763	220.2	221.2	1
225	224.471	224.944	224	225.4	1.4
227	226.433	226.741	226	227.2	1.2
232	232.256	232.496	231.8	232.9	1.1
234	234.236	234.408	233.8	234.9	1.1

RT5

Bins	Min	Max	New Min	New Max	New Range
142	141.93	142.173	141.5	142.5	1
150	149.565	149.772	149.1	150.1	1
154	154.281	154.482	153.8	154.8	1
157	156.581	156.671	156.1	157	0.9
161	161.051	161.144	160.6	161.5	0.9
163	163.133	163.356	162.7	163.7	1
165	165.306	165.425	164.9	165.8	0.9
167	167.41	167.567	167	167.9	0.9
170	169.54	169.723	169.1	170.1	1

Multiplex 3

Map2c

Bins	Min	Max	New Min	New Max	Range
90	90.41	90.516	90	91	1
95	94.458	94.63	94	95.1	1.1
97	96.52	96.622	96.1	97.1	1
101	100.548	100.652	100.1	101.1	1
103	102.482	102.651	102	103.1	1.1
105	104.444	104.605	104	105.1	1.1
107	106.447	106.596	106	107	1
109	108.459	108.622	108	109.1	1.1
111	110.492	110.638	110	111.1	1.1
113	112.582	112.702	112.1	113.2	1.1

RT1

Bins	Min	Max	New Min	New Max	Range
222	221.58	221.736	221.1	222.2	1.1
224	223.519	223.713	223.1	224.2	1.1
226	225.544	225.587	225.1	226	0.9
227	227.375	227.582	226.9	228	1.1
229	229.232	229.276	228.8	229.7	0.9
231	231.219	231.314	230.8	231.8	1
233	233.26	233.357	232.8	233.8	1
235	235.085	235.209	234.6	235.7	1.1
237	236.981	237.149	236.5	237.6	1.1

BM888

Bins	Min	Max	New Min	New Max	Range
172	171.611	171.868	171.2	172.3	1.1
176	175.844	175.918	175.4	176.4	1
178	177.96	178.278	177.5	178.7	1.2
180	180.06	180.332	179.6	180.8	1.2
182	182.096	182.392	181.6	182.8	1.2
184	184.133	184.416	183.7	184.9	1.2
186	186.208	186.497	185.8	186.9	1.1
188	188.329	188.582	187.9	189	1.1
190	190.372	190.558	189.9	191	1.1
192	192.373	192.631	191.9	193.1	1.2
195	194.451	194.682	194	195.1	1.1

BM848

Bins	Min	Max	New Min	New Max	Range
361	361.157	361.249	360.7	361.7	1
363	363.046	363.208	362.6	363.7	1.1
367	366.939	366.973	366.5	367.4	0.9
371	370.788	370.86	370.3	371.3	1
373	372.76	372.882	372.3	373.3	1
378	378.404	378.514	378	379	1
382	382.346	382.346	381.9	382.8	0.9
384	384.307	384.368	383.9	384.8	0.9

BM4513

Not scored due to inconsistent stutter and peak patterns.

RT30

Bins	Min	Max	New Min	New Max	Range
193	193.258	193.276	192.8	193.7	0.9
195	195.239	195.304	194.8	195.8	1
197	197.221	197.35	196.8	197.8	1
199	198.86	199.347	198.4	199.8	1.4
201	201.193	201.424	200.7	201.9	1.2
203	203.36	203.472	202.9	203.9	1
205	205.439	205.531	205	206	1
210	209.624	209.672	209.2	210.1	0.9

GENOTYPING RULES DOCUMENT

NOTE TO READER

Genotyping rules were guided based on the unpublished genotyping rules developed by Peter Hettinga and others. Due to significant variations in patterns and typical fragment sizes, adjustments were required to accommodate the dataset. Therefore, naming of alleles and bin assignments were defined by the R-Package MsatAllele as opposed to the bin assignments used by Peter Hettinga. Genotyping rules were guided but are inconsistent between studies but Peter Hettinga provided the foundation and understanding of genotyping. The General Rules stated below are adapted from the unpublished document. The individual locus based genotyping rules are based on the patterns presented among this study's dataset as guided by similar patterns within the unpublished document.

Additionally, scoring rules are included here for BMS1788 and NVHRT30; however, they were removed from analyses due to potential null alleles and the challenges and inconsistencies of scoring BMS1788.

GENERAL RULES:

- Only score peaks with a stutter. As patterns are relatively consistent among samples, peaks without stutters are non-specific peaks and may have similar peak heights to alleles.
- When alleles fall off the ladder, they are ALWAYS scored in the bin to the **RIGHT** of the peak.
- When split peaks occur, always score the second peak. Except when a split peak occurs due to artifact interference (e.g. BM6506).
- Peaks MUST be higher than 200 RFUs. **Sometimes lower peak heights are acceptable if there is a clear and distinct stutter pattern**.
- Background variation must be <30% of smallest peak (in RFUs) scored and the stutter pattern should still be distinguishable from noise.
- When alleles are only two base pairs apart the smaller allele (to the left) should be taller than the larger peak (on the right). If alleles are more than two base pairs apart the smaller allele may have a shorter peak height. Sometimes the smaller allele may be slightly shorter but must display the correct stutter pattern and be a verified pattern for the locus.

LOCI RANGES AND MULTIPLEXES

Loci Ranges and corresponding multiplexes. Ranges include buffer up to 2bp.

Multiplex 1		Multiplex 2		Multiplex 3	
Locus	Range	Locus	Range	Locus	Range
RT6 - FAM	91 – 111	RT27 – FAM	129 - 161	MAP2C - FAM	88 - 114
BM6506 – FAM	194 – 210	RT9 – VIC	102 - 130	RT1 - FAM	220 - 238
BMS1788 – VIC	111 – 145*	RT7 – VIC	216 - 236	BM888 - VIC	170-196
NVHRT30 – VIC	145* - 175	RT5 – NED	140 - 170	BM848 - NED	360-386
RT24 – NED	205 - 235			~~BM4513 - PET~~	~~110-170~~
				RT30 - PET	192-212

*Alleles for BMS1788 includes 145 while alleles for NVHRT30 start at 147 (145 is the buffer)

- Locus BM4513 profiles were inconsistent and unsuccessful in defining genotyping rules

CARIBOU PANEL FILES BY LOCUS

MULTIPLEX 1

RT6

RT6				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
93	92.5	93.6	93.1	1.1
97	96.1	97.2	96.7	1.1
99	98.0	99.1	98.6	1.1
100	99.8	100.9	100.4	1.1
102	101.6	102.6	102.1	1.0
106	105.1	106.2	105.7	1.1
107	106.9	108.0	107.5	1.1
109	108.7	109.9	109.3	1.2

BM6506

BM6506				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
196	195.3	196.3	195.8	1.0
198	197.1	198.2	197.7	1.1
201	200.8	202.0	201.4	1.2
207	206.5	207.7	207.1	1.2
209	208.4	209.7	209.1	1.3

BMS1788

BMS1788				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
113	112.8	113.7	113.3	0.9
117	116.9	117.9	117.4	1.0
124	123.3	124.5	123.9	1.2
126	125.5	126.4	126.0	0.9
128	127.4	128.7	128.1	1.3
130	128.9	130.7	129.8	1.8
132	131.5	132.7	132.1	1.2
134	133.0	134.6	133.8	1.6
136	135.3	136.4	135.9	1.1
137	136.9	138.6	137.8	1.7
143	142.1	143.8	143.0	1.7

NVHRT30

NVHRT30				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
147	146.7	147.7	147.2	1.0
154	153.8	154.9	154.4	1.1
157	156.1	157.2	156.7	1.1
159	158.3	159.6	159.0	1.3
161	160.4	161.6	161.0	1.2
163	162.6	163.7	163.2	1.1
165	164.7	165.9	165.3	1.2
167	166.8	168.1	167.5	1.3
169	168.9	170.1	169.5	1.2
172	171.0	172.5	171.8	1.5
174	173.2	174.3	173.8	1.1

RT24

RT24				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
207	206.2	207.3	206.8	1.1
209	208.4	209.5	209.0	1.1
211	210.4	211.4	210.9	1.0
213	212.6	213.5	213.1	0.9
215	214.7	215.7	215.2	1.0
217	216.7	217.9	217.3	1.2
219	218.7	220.1	219.4	1.4
221	220.8	222.0	221.4	1.2
224	223.0	224.1	223.6	1.1
226	225.2	226.2	225.7	1.0
234	233.6	234.7	234.2	1.1

MULTIPLEX 2

RT27

RT27				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
131	130.8	131.8	131.3	1.0
134	132.7	134.1	133.4	1.4
145	144.9	145.9	145.4	1.0
148	147.0	148.5	147.8	1.5
150	149.9	151.0	150.5	1.1
153	152.3	153.4	152.9	1.1
155	154.6	155.6	155.1	1.0
160	159.1	160.1	159.6	1.0

RT9

RT9				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
103	102.9	103.9	103.4	1.0
116	115.6	116.7	116.2	1.1
120	119.4	120.5	120.0	1.1
122	121.3	122.4	121.9	1.1
124	123.3	124.4	123.9	1.1
128	127.2	128.3	127.8	1.1
130	129.3	130.2	129.8	0.9

RT7

RT7				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
217	216.3	217.7	217.0	1.4
219	218.3	219.5	218.9	1.2
221	220.2	221.2	220.7	1.0
225	224.0	225.4	224.7	1.4
227	226.0	227.2	226.6	1.2
232	231.8	232.9	232.4	1.1
234	233.8	234.9	234.4	1.1

RT5

RT5				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
142	141.5	142.5	142.0	1.0
150	149.1	150.1	149.6	1.0
154	153.8	154.8	154.3	1.0
157	156.1	157.0	156.6	0.9
161	160.6	161.5	161.1	0.9
163	162.7	163.7	163.2	1.0
165	164.9	165.8	165.4	0.9
167	167.0	167.9	167.5	0.9
170	169.1	170.1	169.6	1.0

MULTIPLEX 3

MAP2C

Map2C				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
90	90	91	90.5	1.0
95	94	95.1	94.6	1.1
97	96.1	97.1	96.6	1.0
101	100.1	101.1	100.6	1.0
103	102	103.1	102.6	1.1
105	104	105.1	104.6	1.1
107	106	107	106.5	1.0
109	108	109.1	108.6	1.1
111	110	111.1	110.6	1.1
113	112.1	113.2	112.7	1.1

RT1

RT1				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
222	221.1	222.2	221.7	1.1
224	223.1	224.2	223.7	1.1
226	225.1	226	225.6	0.9
227	226.9	228	227.5	1.1
229	228.8	229.7	229.3	0.9
231	230.8	231.8	231.3	1.0
233	232.8	233.8	233.3	1.0
235	234.6	235.7	235.2	1.1
237	236.5	237.6	237.1	1.1

BM888

BM888				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
172	171.2	172.3	171.8	1.1
176	175.4	176.4	175.9	1.0
178	177.5	178.7	178.1	1.2
180	179.6	180.8	180.2	1.2
182	181.6	182.8	182.2	1.2
184	183.7	184.9	184.3	1.2
186	185.8	186.9	186.4	1.1
188	187.9	189	188.5	1.1
190	189.9	191	190.5	1.1
192	191.9	193.1	192.5	1.2
195	194	195.1	194.6	1.1

BM848

BM848				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
361	360.7	361.7	361.2	1.0
363	362.6	363.7	363.2	1.1
367	366.5	367.4	367.0	0.9
371	370.3	371.3	370.8	1.0
373	372.3	373.3	372.8	1.0
378	378	379	378.5	1.0
382	381.9	382.8	382.4	0.9
384	383.9	384.8	384.4	0.9

RT30

RT30				
Allele Name	Left Boundary	Right Boundary	Bin Center	Size Range
193	192.8	193.7	193.3	0.9
195	194.8	195.8	195.3	1.0
197	196.8	197.8	197.3	1.0
199	198.4	199.8	199.1	1.4
201	200.7	201.9	201.3	1.2
203	202.9	203.9	203.4	1.0
205	205	206	205.5	1.0
210	209.2	210.1	209.7	0.9

MICROSATELLITE PRIMER INFORMATION

Marker	Reference	Expected size in *Rangifer tarandus*	Dye	Multiplex - Yannic et al. 2016	Forward (5'-3')	Reverse (5'-3')
Multiplex 1						
RT6	Wilson et al. 1997	105-117	6FAM	b	TTCCTCTTACTCATTCTTGG	CGGATTTTGAGACTGTTAC
BM6506	Bishop et al. 1994	188-222	6FAM	b	GCACGTGGTAAAGAGATGGC	AGCAACTTGAGCATGGCAC
BMS1788*	Yannic et al 2014 - modified from Cronin et al. 2005	103-143	VIC	b	ATTCATATCTACGTCCAGATTCAGATTTCTTG	GGAGAGGAATCTTGCAAAGG
NVHRT30	Røed & Midthjell 1998	159–177	VIC	b	GTGGAGCATTGTGTATGTGT	GCCCCCACTGTGTTTT
RT24	Wilson et al. 1997	215-227	NED	b	TGTATCCATCTGGAAGATTTCAG	CAGTTTAACCAGTCCTCTGTG
Multiplex 2						
RT27	Wilson et al. 1997	135-155	6FAM	c	CCAAAGACCCAACAGATG	TTGTAACACAGCAAAAGCATT
RT9	Wilson et al. 1997	116-128	VIC	c	TGAAGTTTAATTTCCACTCT	CAGTCACTTTCATCCCACAT
RT7	Wilson et al. 1997	216-234	VIC	c	CCTGTTCTACTCTTCTTCTC	ACTTTTCACGGGCACTGGTT

RT5	Wilson et al. 1997	143-171	NED	c	TGGTTGGAAGGAAAACTTGG	CCTCTGCTCCTCAAGACAC
Multiplex 3						
Map2c	Bishop et al. 1994	81-119	6FAM		TTTACCAGACAGTTTAGTTTTGAGC	AAGGATTCTGTCTGATACCACTTAG
RT1	Wilson et al. 1997	208-244	6FAM	c	TGCCTTCTTTCATCCAACAA	CATCTTCCCATCCTCTTTAC
BM888	Bishop et al. 1994	158-210	VIC		AGGCCATATAGGAGGCAAGCTT	CTCGGTCAGCTCAAAACGAG
BM848	Bishop et al. 1994	352-384	NED		TGGTTGGAAGGAAAACTTGG	CCTCTGCTCCTCAAGACAC
BM4513	Bishop et al. 1994	118-150	PET	c	TCAGCAATTCAGTACATCACCC	GCGCAAGTTTCCTCATGC
RT30	Wilson et al. 1997	181-207	PET		CACTTGGCTTTTGGACTTA	CTGGTGTATGTATGCACACT

references for expected size in Rangifer tarandus for markers from Bishop et al 1994 are from https://datadryad.org/stash/dataset/doi:10.5061/dryad.9qh56

MORPHOLOGY

MULTIPLEX 1

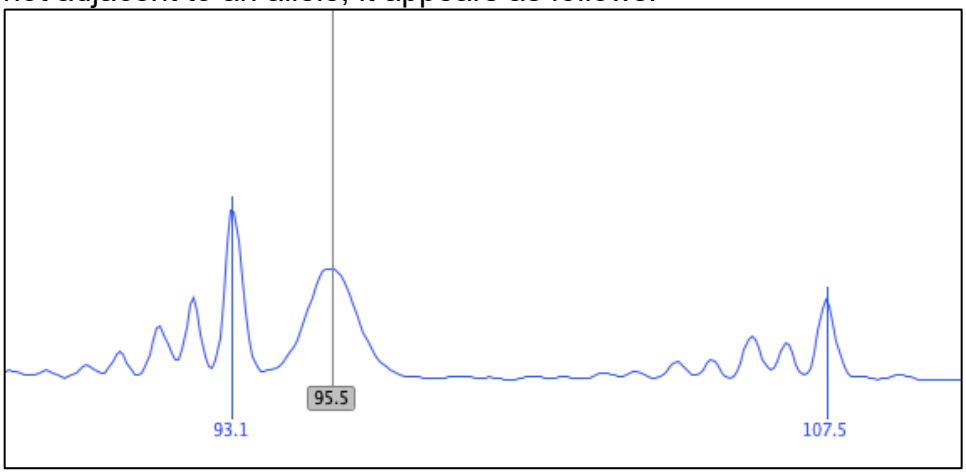
Multiplex 1 contains the most complex loci to score. Therefore, careful attention is required to score alleles correctly. When unsure, the locus is identified as missing data -99.

RT6 – FAM

RT6 contains various factors that make scoring difficult if some of the following features are not recognized. Features include differing allelic patterns, a primer dimer, and artifacts, with some achieving equivalent RFU heights as alleles.

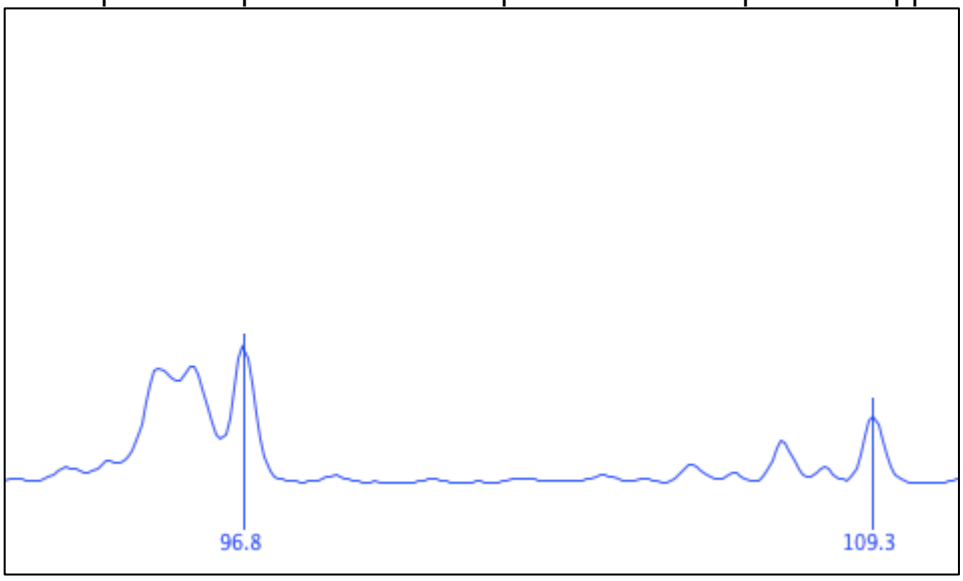
Primer Dimer ~95.5

There is a primer dimer at ~95.5 bp. Do not score this; it is not a peak. When the peak is not adjacent to an allele, it appears as follows:

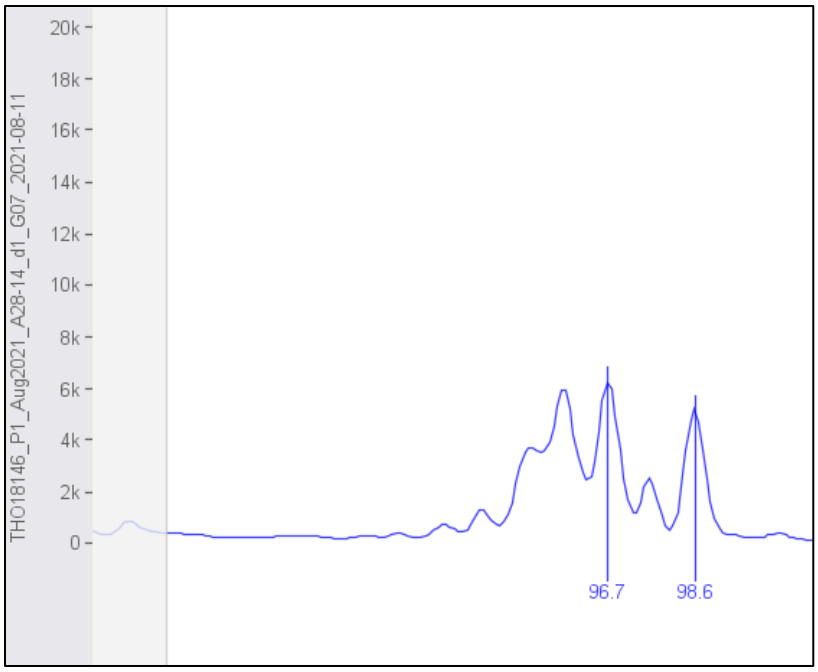


Primer Dimer Interference ~95.5

When there is an allele following the primer dimer but adjacent to the primer dimer, the stutter pattern is present within the primer dimer the pattern it appears as follows:

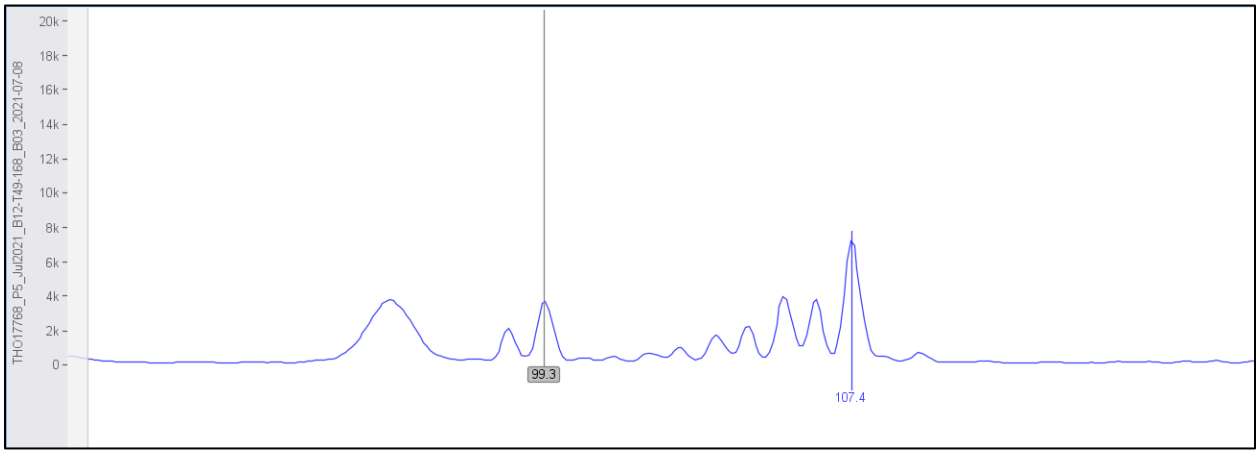


When there is a pair of alleles present adjacent and within the artifact, it appears as follows:



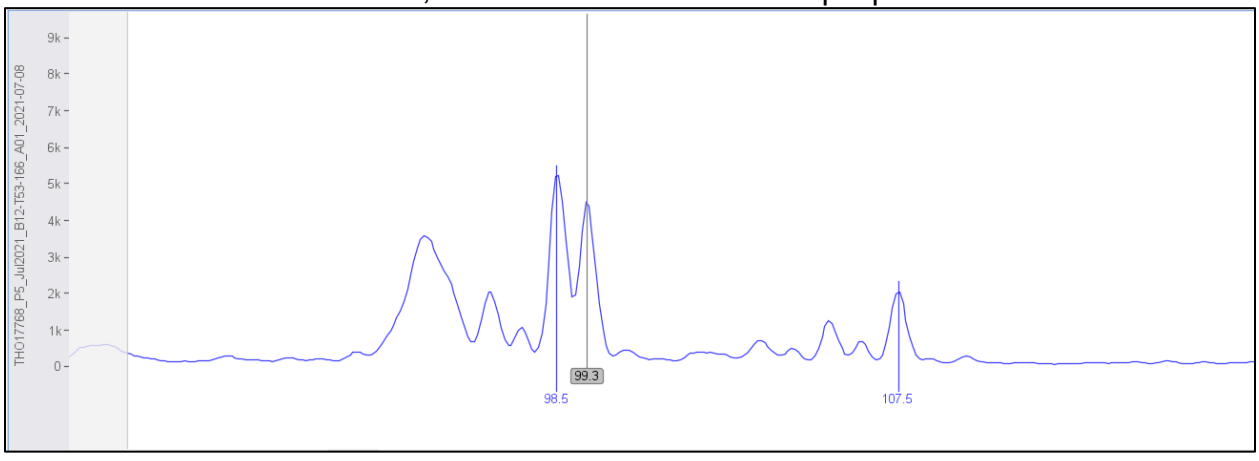
Artifact ~99.3

At 99.3, there is a common artifact that occurs with various RFU heights. It is a two-peak pattern that is NOT consistent with the allelic pattern. Therefore, this peak is not to be scored.



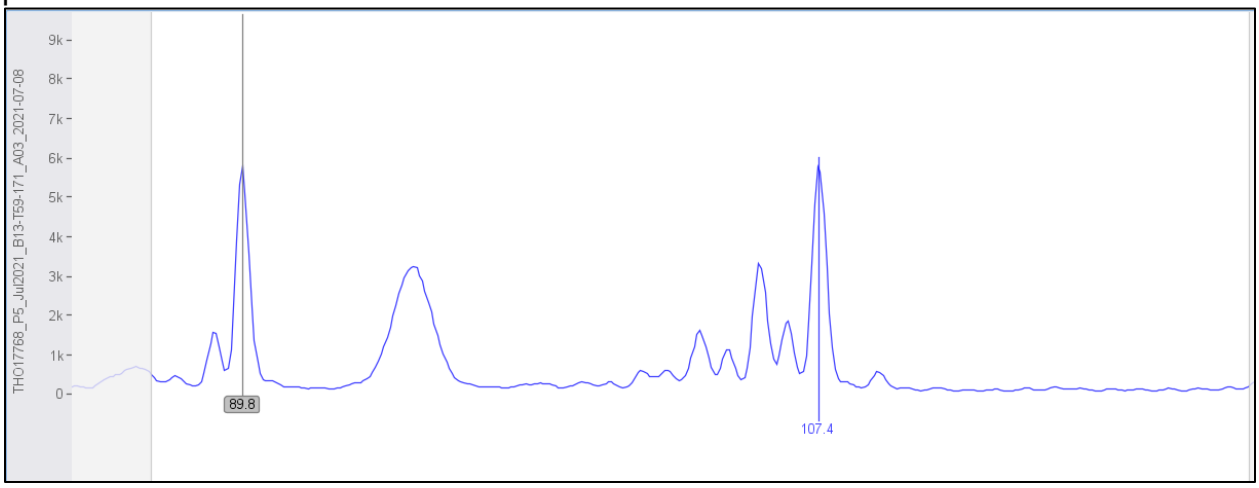
Artifact Interference ~99.3

The artifact then interferes with a common allele at 98.5. The peak that is scored in this situation is the one at 98.5 as the artifact is an abnormality to the common allelic pattern. This is because this artifact occurs 1bp from the preceding peak, and the stutter peak before the allele follows the 2 bp pattern.
It is important to note that sometimes rules will indicate that you always score to the right of a split peak. However, in this case the split peak contains an allele and an artifact, and if one were to score the artifact instead, it would result in a different genotype.
When there is an interference, we score to the left of the split peak



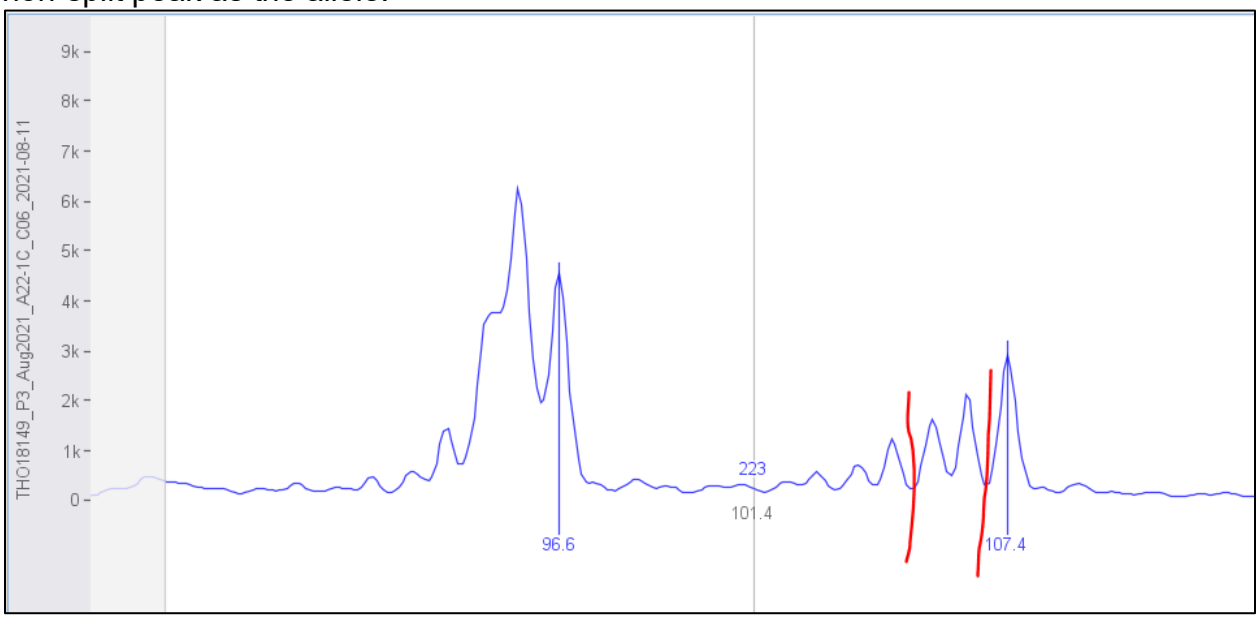
Artifact ~89.8

There is an occasional artifact at 89.8 with the same pattern as the artifact present at 99.3. To score alleles within this range it is important to identify the proper allelic pattern.

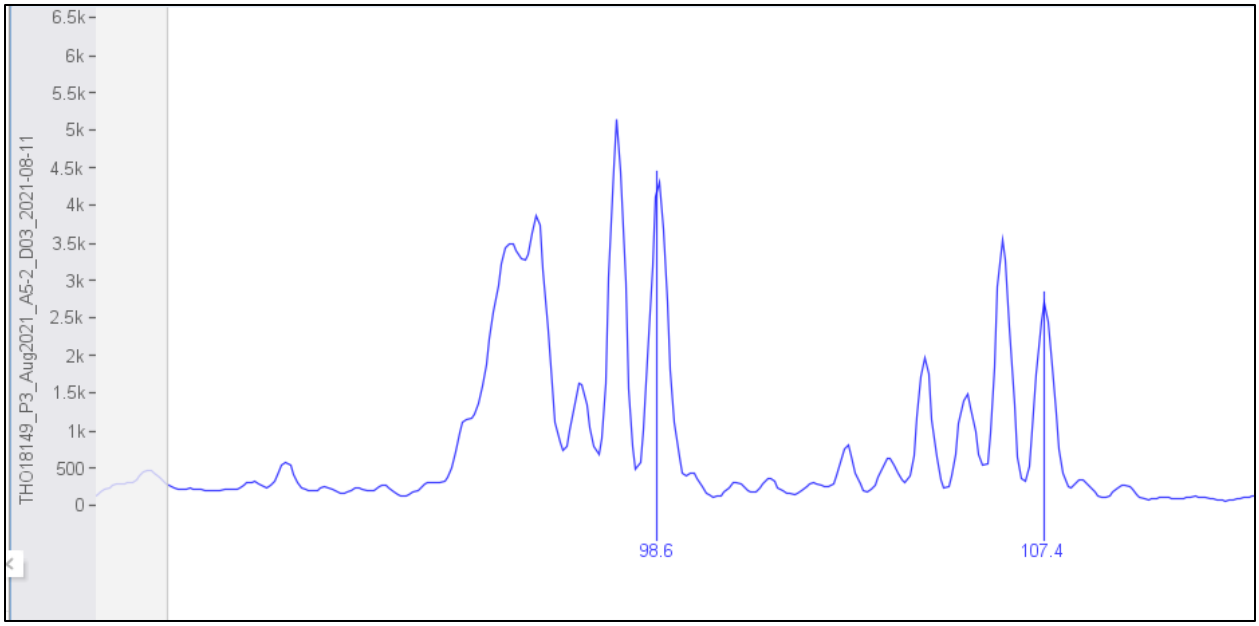


Special Case – THO18149_P3_Aug appears in other files too

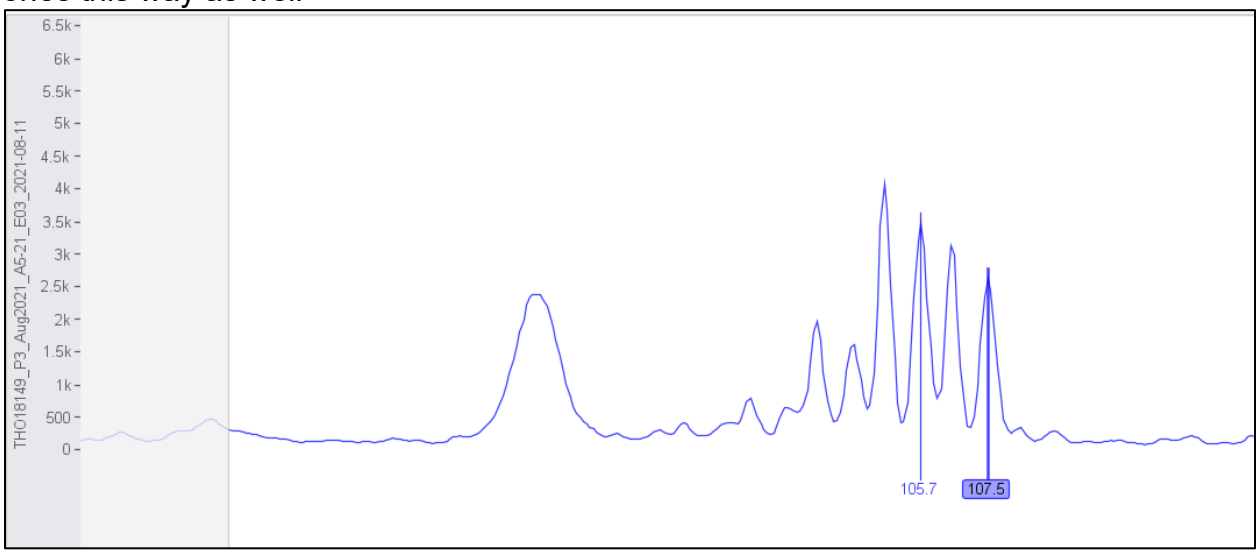
This was an extreme case of a pattern that occurred within some plates. This specific plate was re-amplified to ensure scoring was accurate. Regardless, advice based on this extreme case was included in this document. Overall this morphology was abnormal.
The Allele in these odd cases are the peaks following the preceding split peak. Sometimes the preceding split peak is higher than the allele. However, we score the non-split peak as the allele.



Additionally, the peak is often shorter than the preceding split peaks

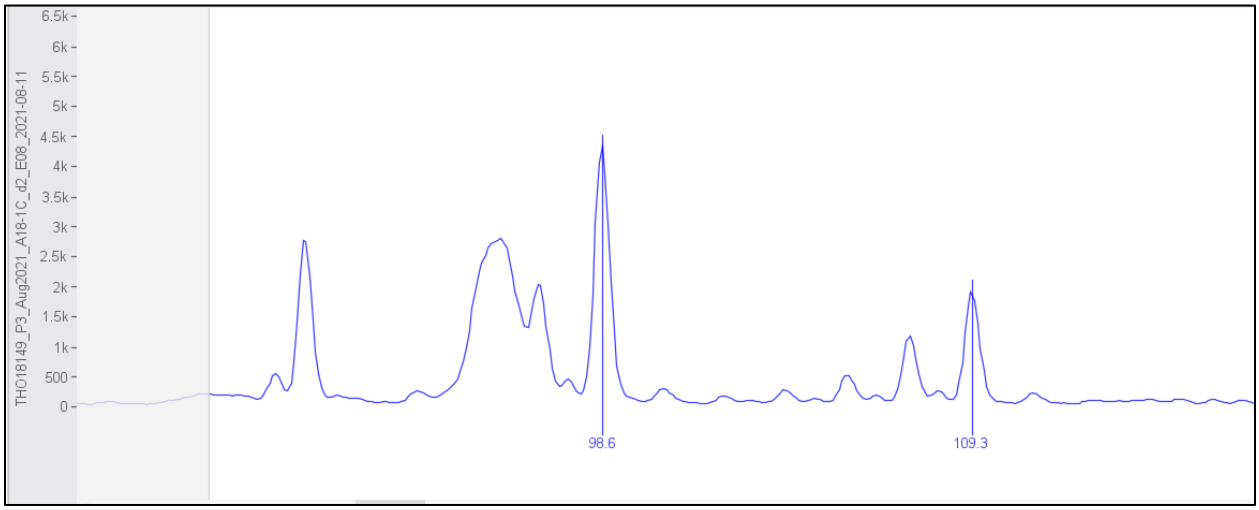


For adjacent alleles with this pattern: In this case, the preceding split peak of the second allele has merged with the preceding allele that is 2bp away. This pattern occurs in traces where the preceding peaks are lower than the allele which is why we score these ones this way as well



Special Case – Different Stutter Pattern

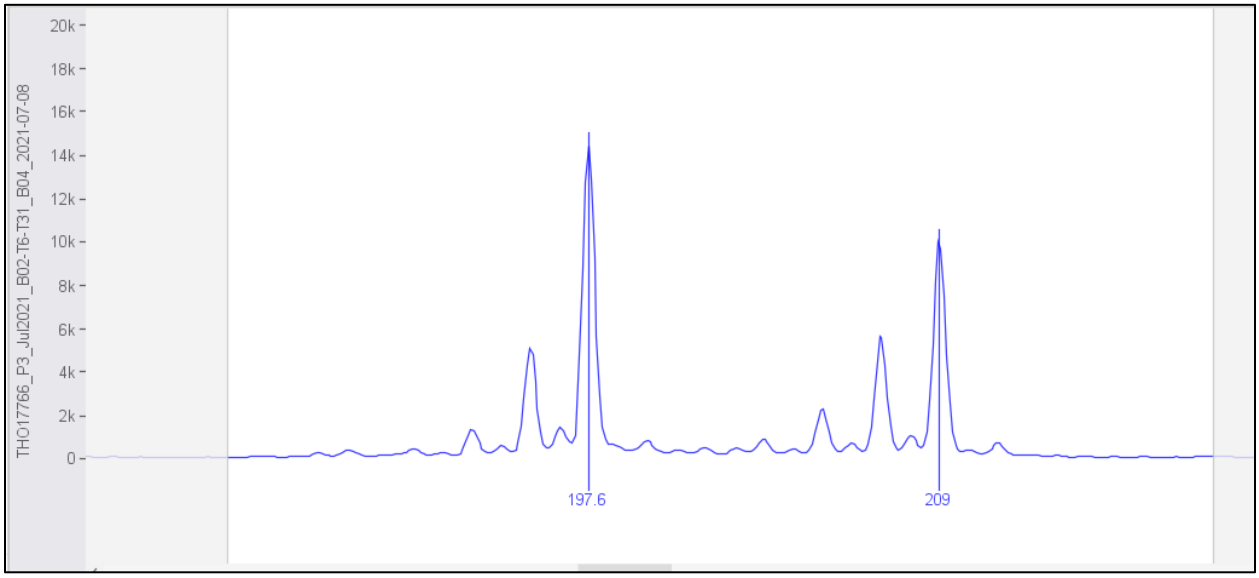
Note that there is an added common morphology pattern in this file. Sometimes the artifact masks the preceding stutter pattern for alleles in the 198.6 position and appear to be a standalone artifact that should not be scored.



BM6506

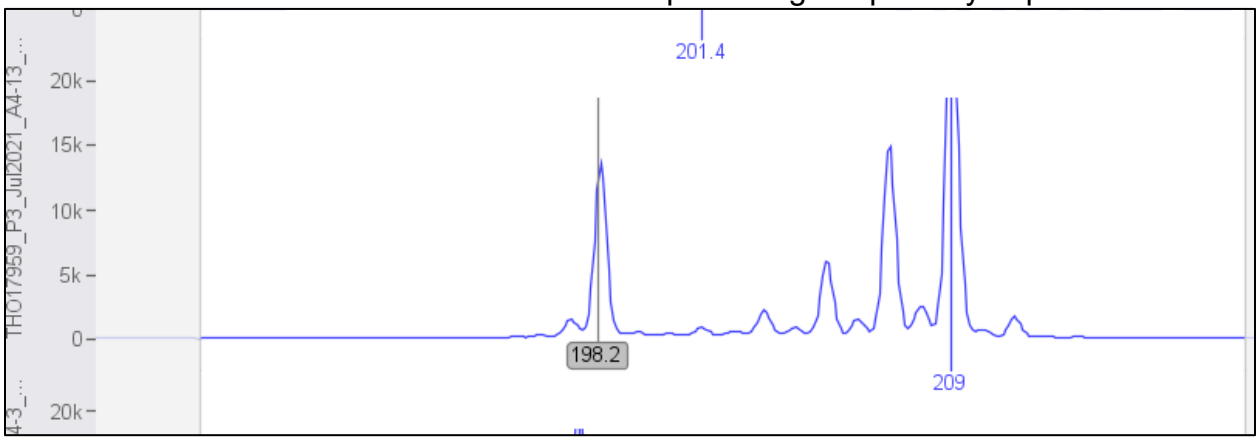
Common Morphology

The common morphology for BM6506 typically has two clear stutter peaks 2bp apart and from the allele.



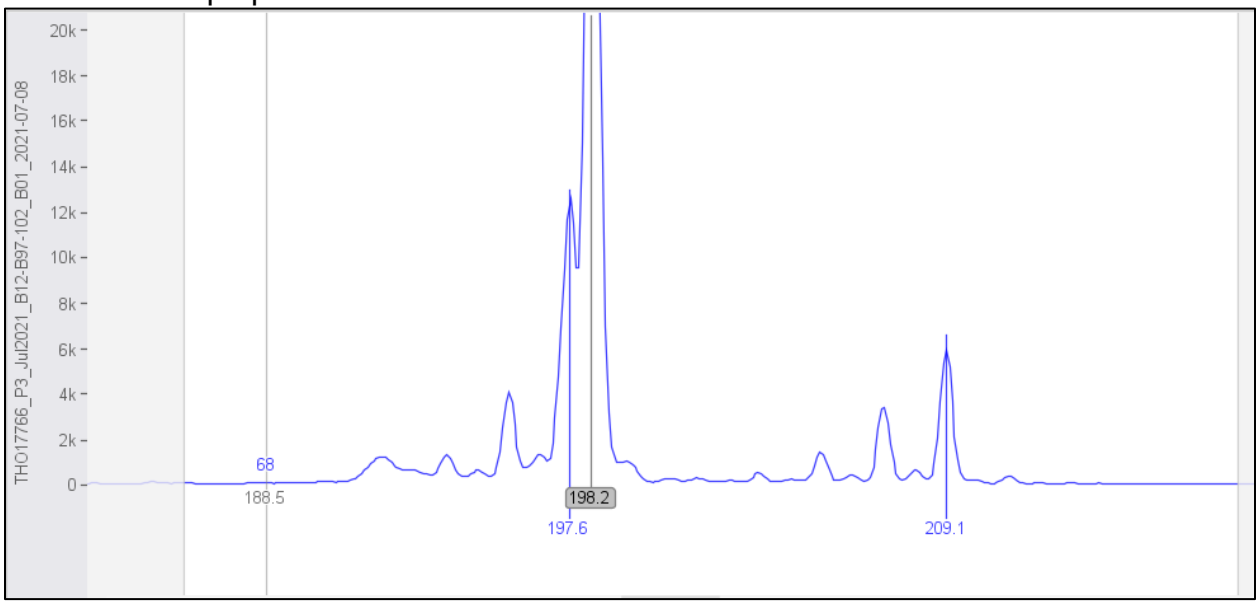
Artifact ~198.2

A common artifact at 198.2 has a minor stutter preceding the peak by 1bp.

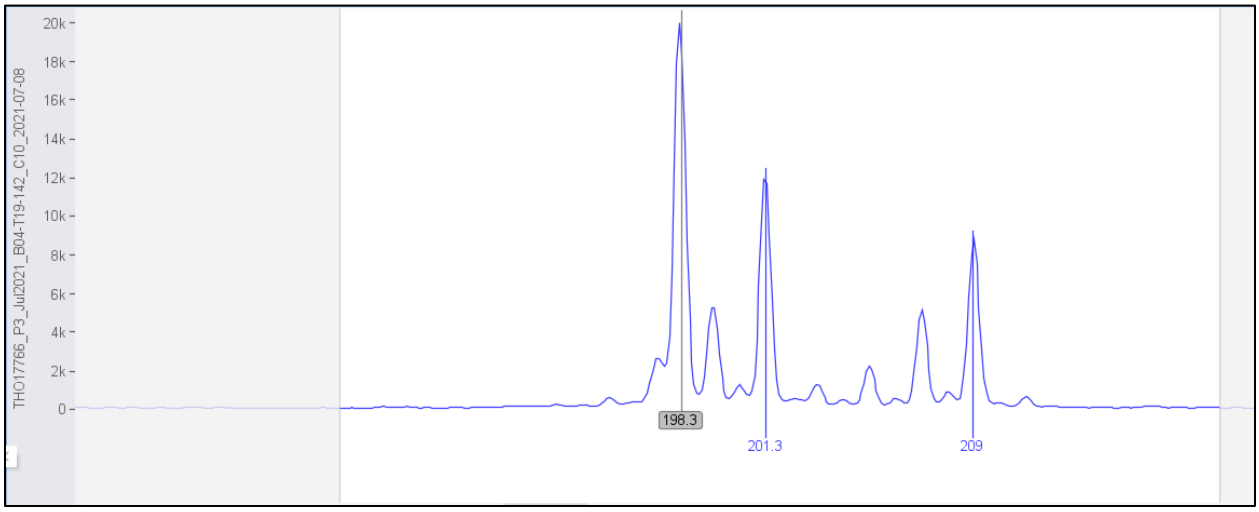


Artifact Interference ~198.2

Artifact interference is very common, but alleles are still easily identifiable. If there is an allele at ~197.6 we score to the left as the artifact is to the right of the split peak (1bp apart). The allele will always have a distinguishable preceding stutter pattern where stutters are 2bp apart.

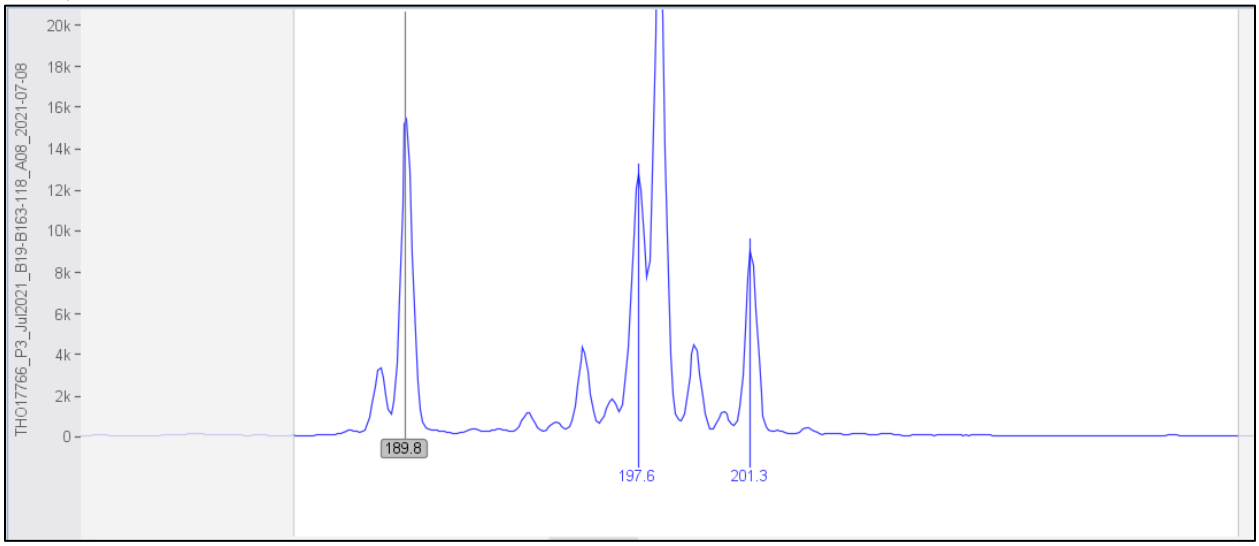


An artifact may also appear within the stutter pattern of an allele. These are usually easier to differentiate.



Artifact ~189.8

There is an artifact at 189.8; however, there are no alleles of that size. This is just a note, so it is not scored as an allele.

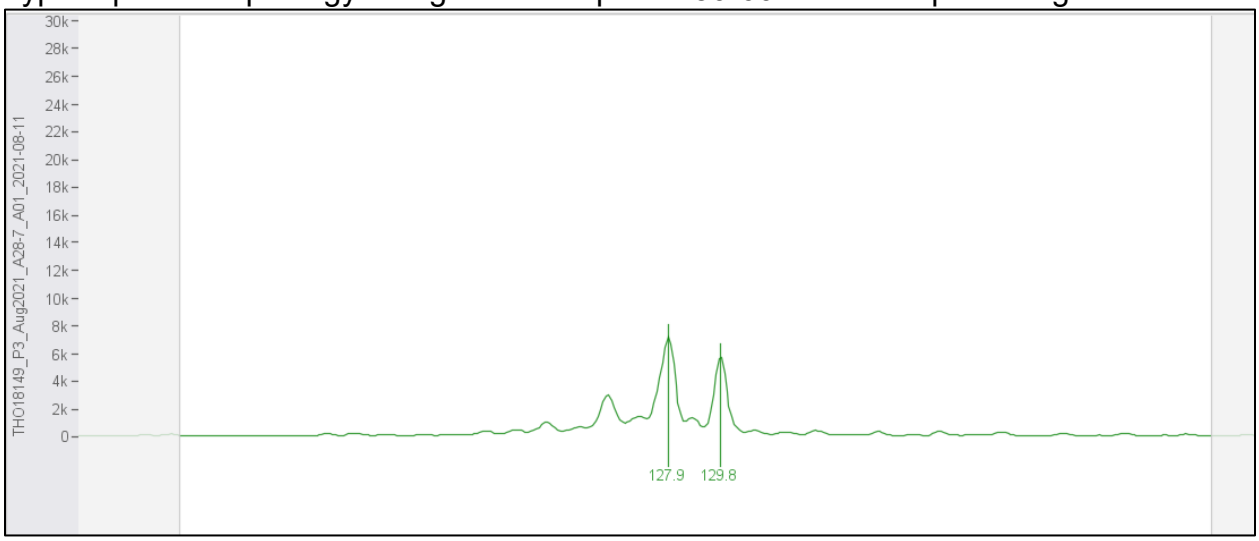


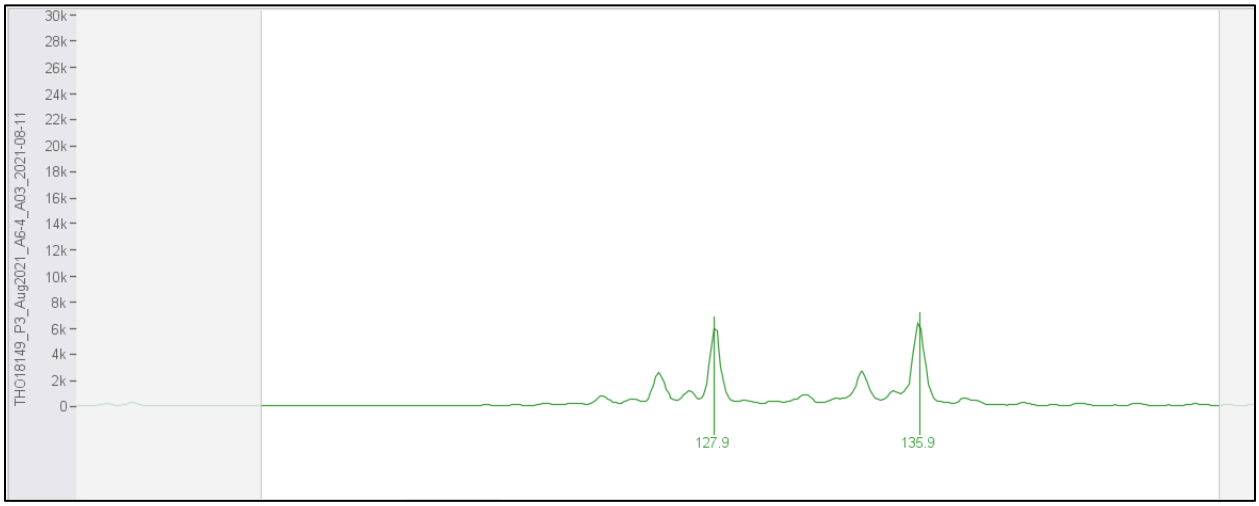
BMS1788 – VIC

BMS1788 is a challenging locus to score. There is a large variety of allelic patterns within this locus however, redoing PCR's verified what patterns constitute alleles within the locus. Therefore, careful scoring is required.

Common Morphology

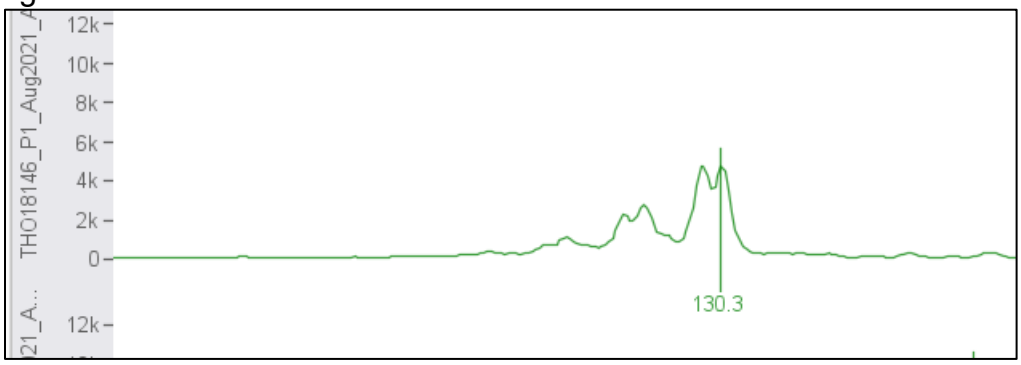
Typical peak morphology – largest stutter peak ~50-60% of allele peak height

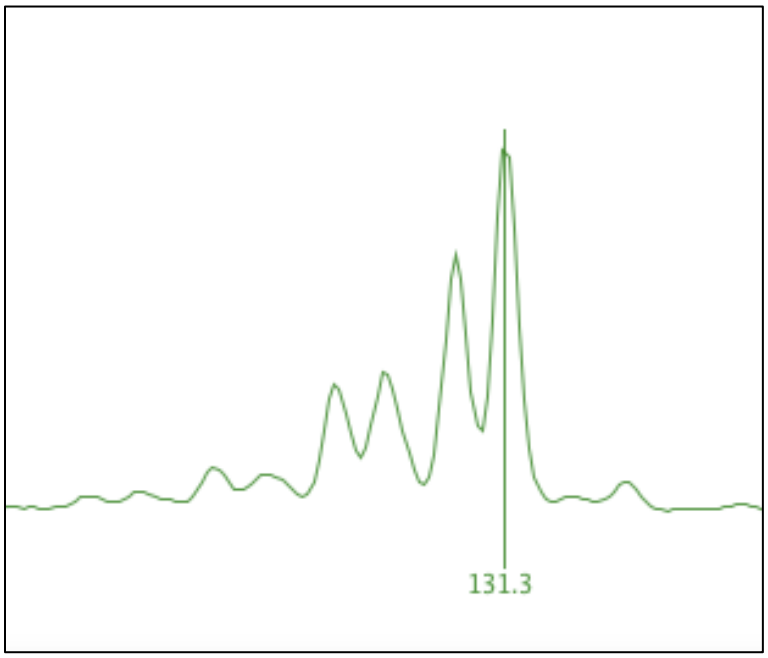




Split Peaks

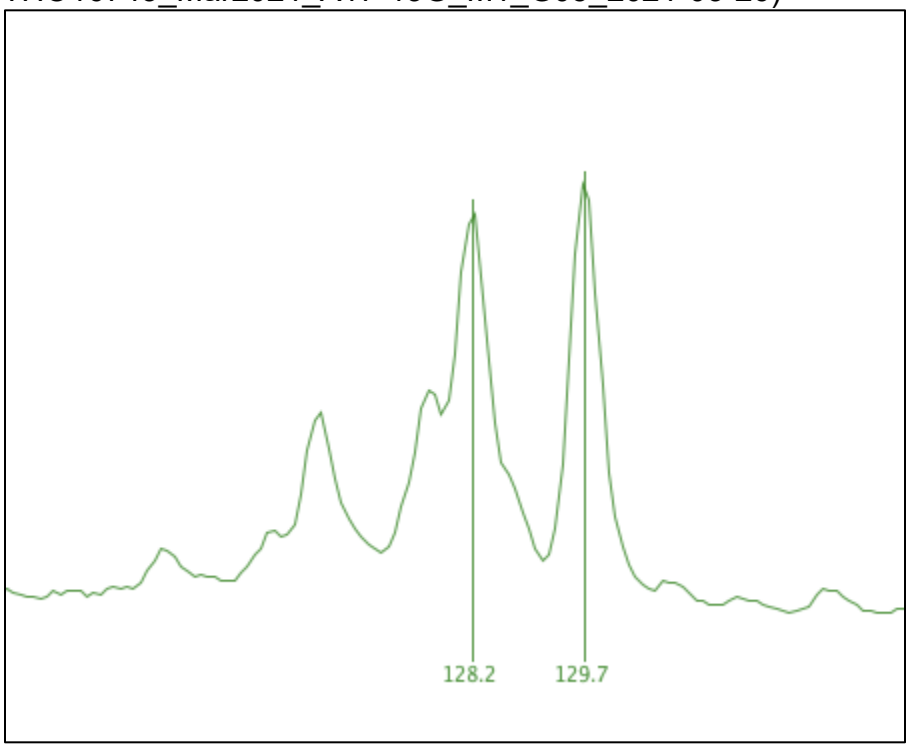
In some cases, the alleles have a split peak morphology as shown and the peak to the right is scored.



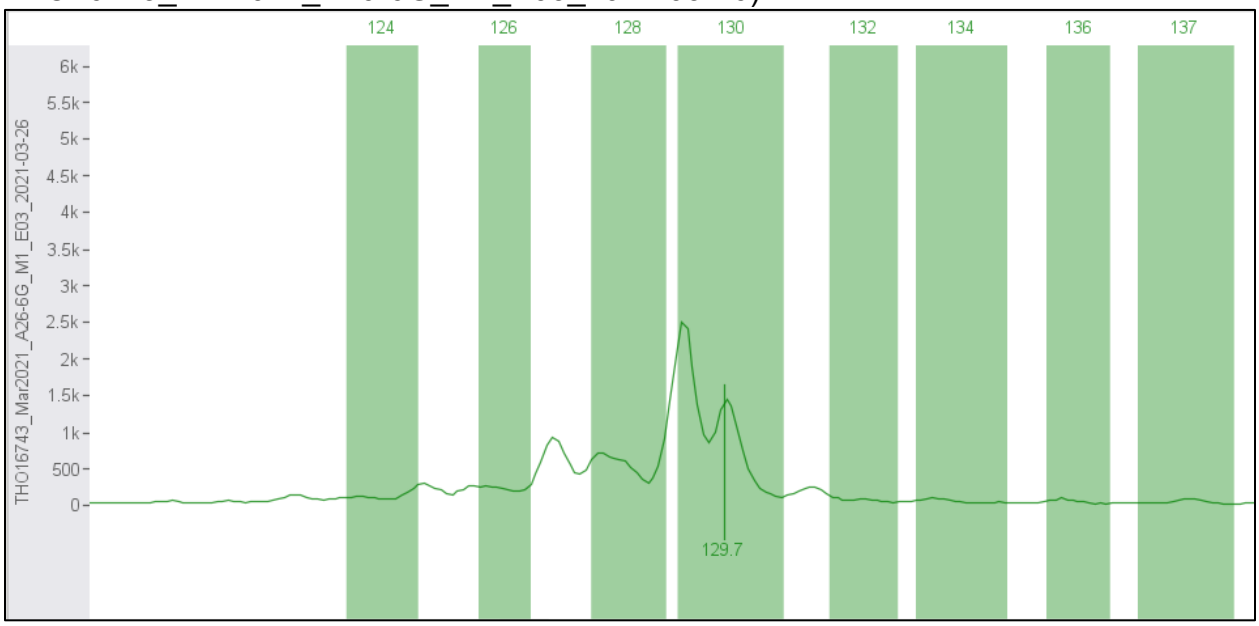


The Hook

In some cases, a large peak 2bp before the main, ‘obvious’ allele shows a hooked morphology. This peak is taller than the ‘normal’ stutter height as shown in previous figure and is considered to be an allele in this case (e.g. Sample THO16743_Mar2021_A17-13G_M1_G05_2021-03-26)

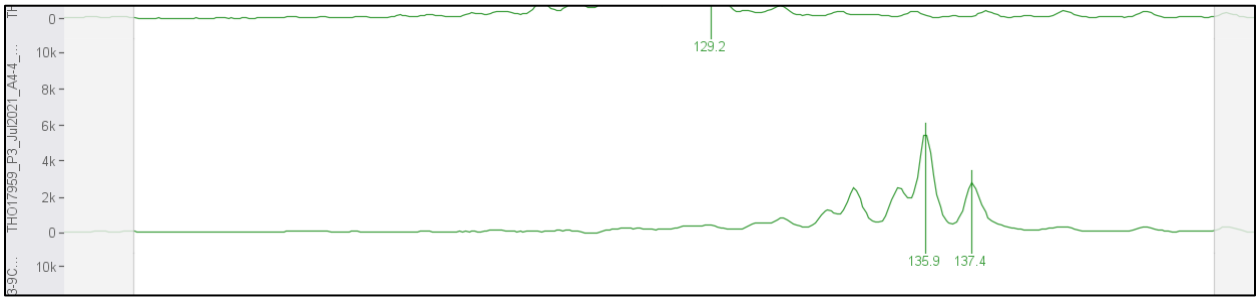


Hooked peak at ~130 bp. Score the second peak since it overlaps well with the bin (e.g. THO16743_Mar2021_A26-6G_M1_E03_2021-03-26)



Heterozygote

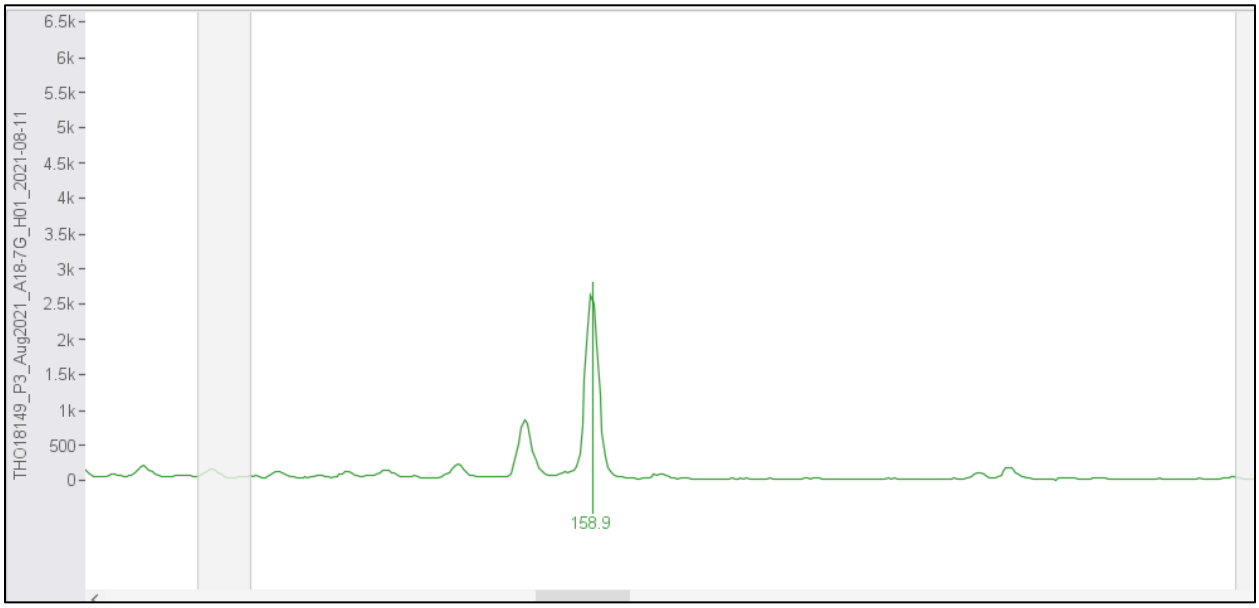
Sometimes the second allele is significantly lower than the preceding split peak allele. The second allele is also scored.

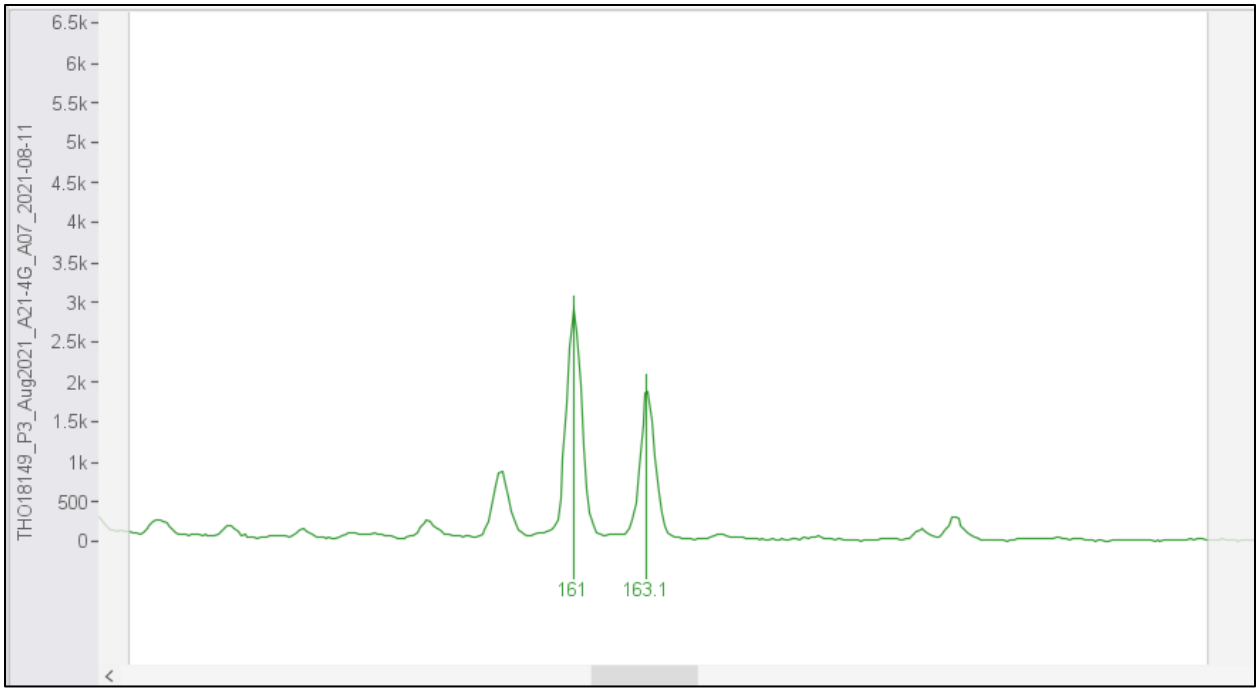


NVHRT30 – VIC

NVHRT30 contains a significant amount of noise in comparison to all other loci. It is important to be confident when scoring alleles within this locus and maintain the rule where noise “Background variation must be <30% of smallest peak (in RFUs) scored”.

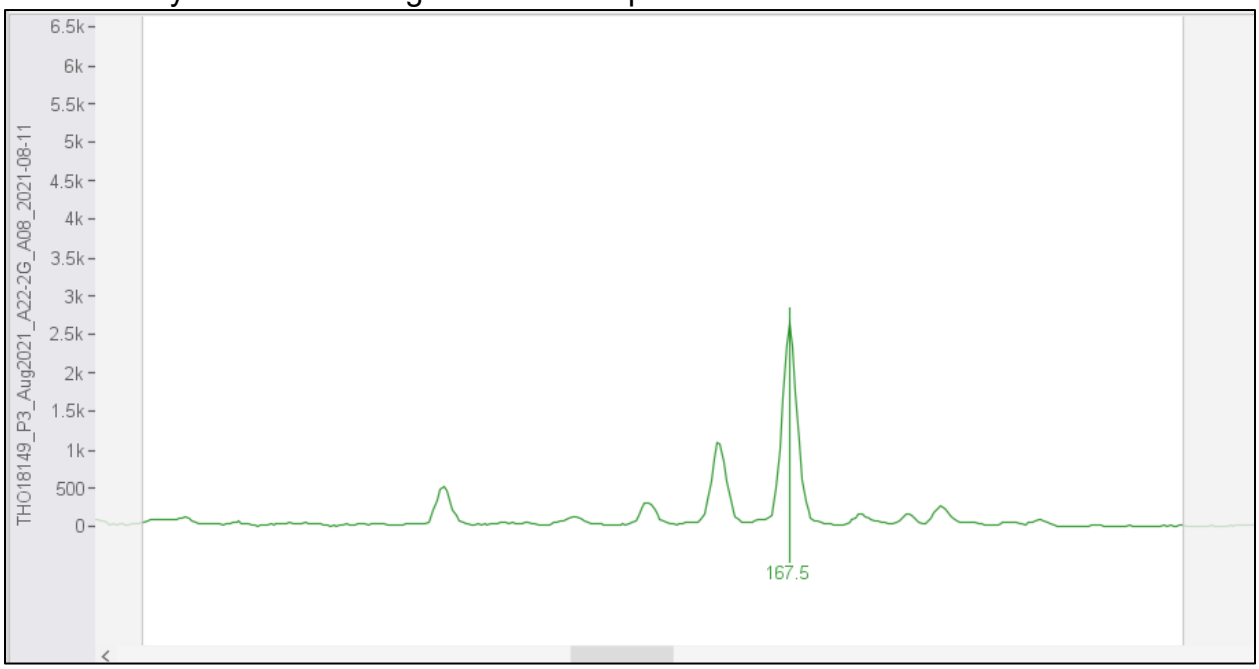
Common Morphology





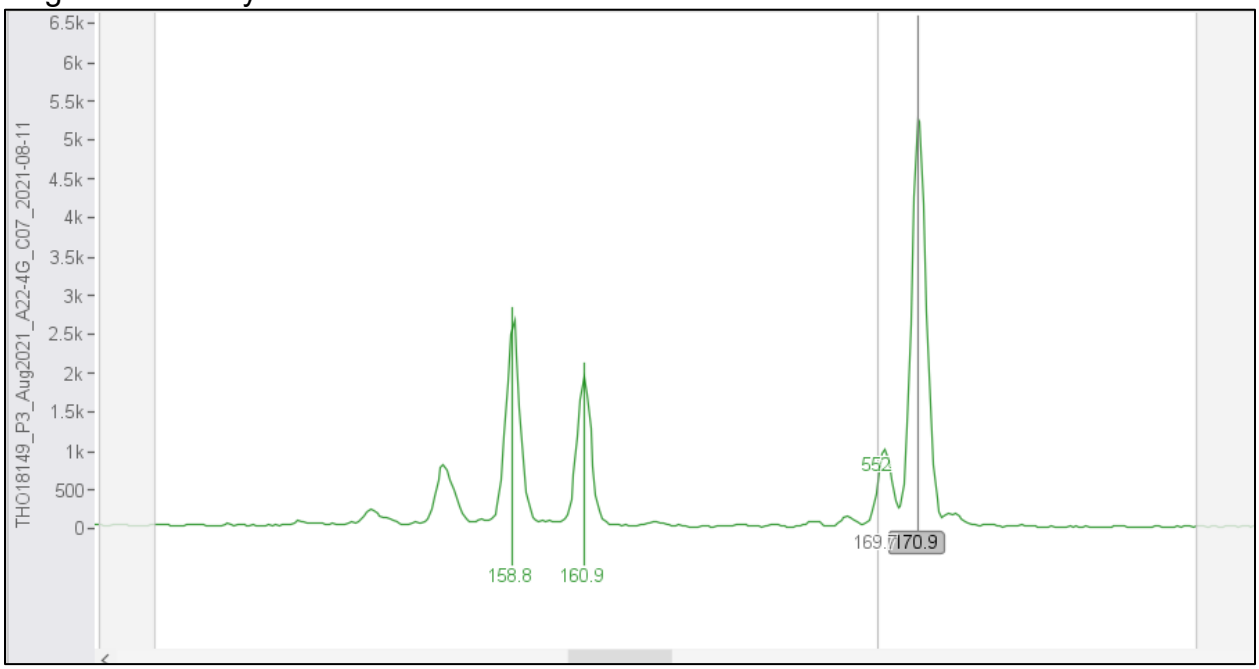
Noise

Noise is very common among this locus but peaks remain distinct

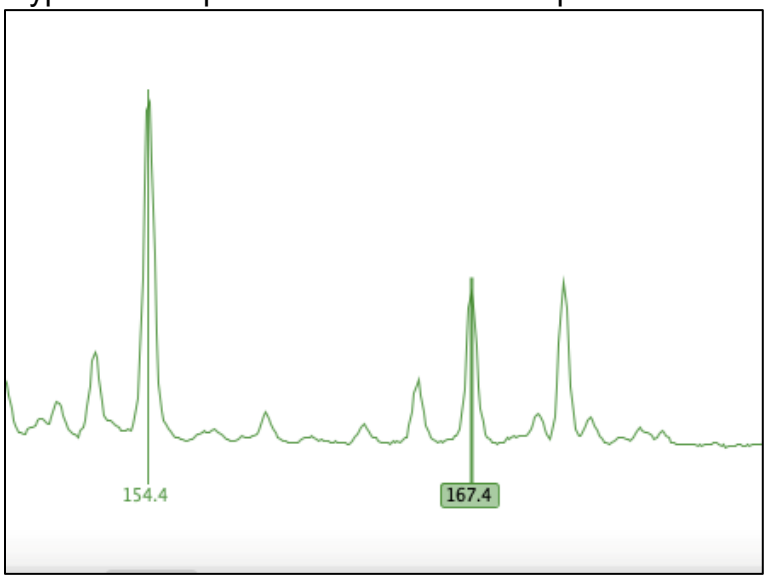


Artifact ~ 170.9/171

Standalone artifacts are not peaks. They do not maintain the pattern of a peak and the height is also very variable.

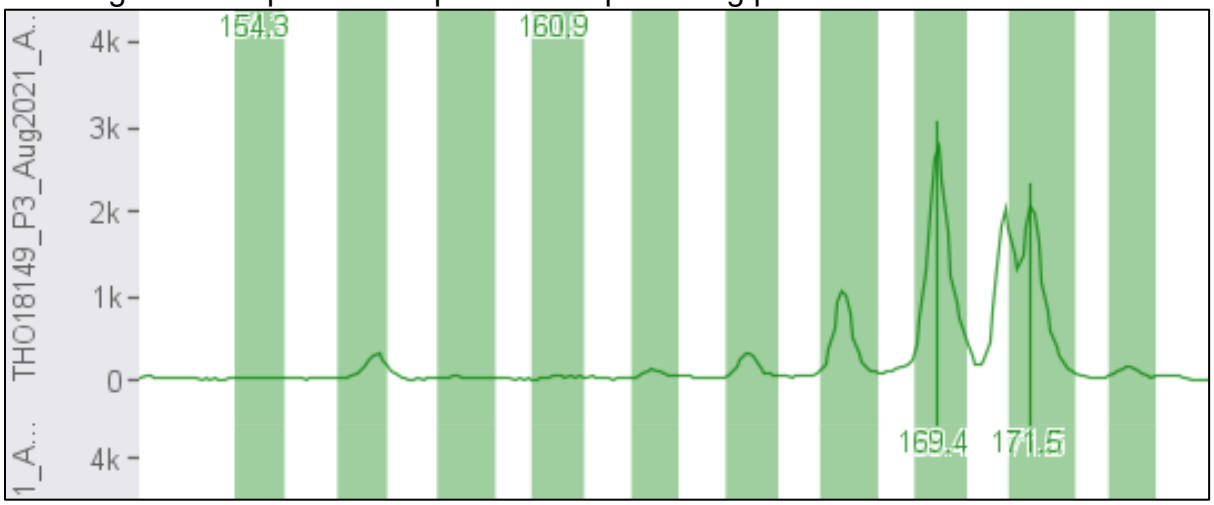


This profile shows an artifact at 171 bp, as indicated by the lack of stutter peaks and atypical allele profile. Do not score this peak.

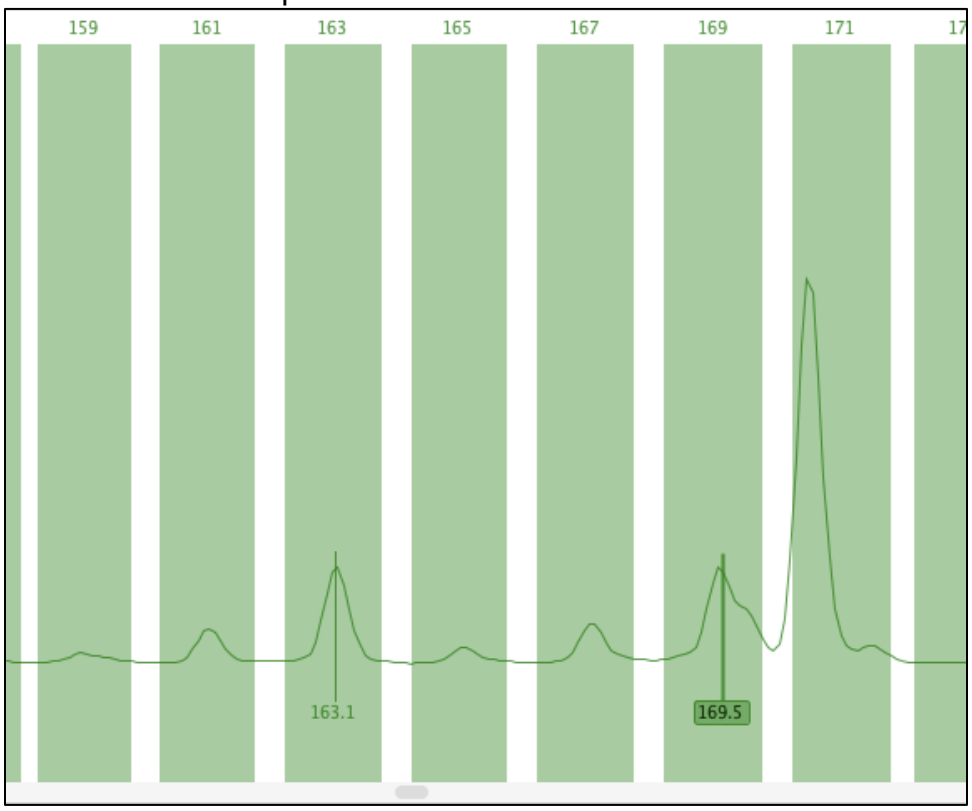


Artifact Interference ~170.9/171

Peaks do not occur at 170.9 among this set of data and are an artifact. Select the peak to the right of the split as it is 2pb from the preceding peak.



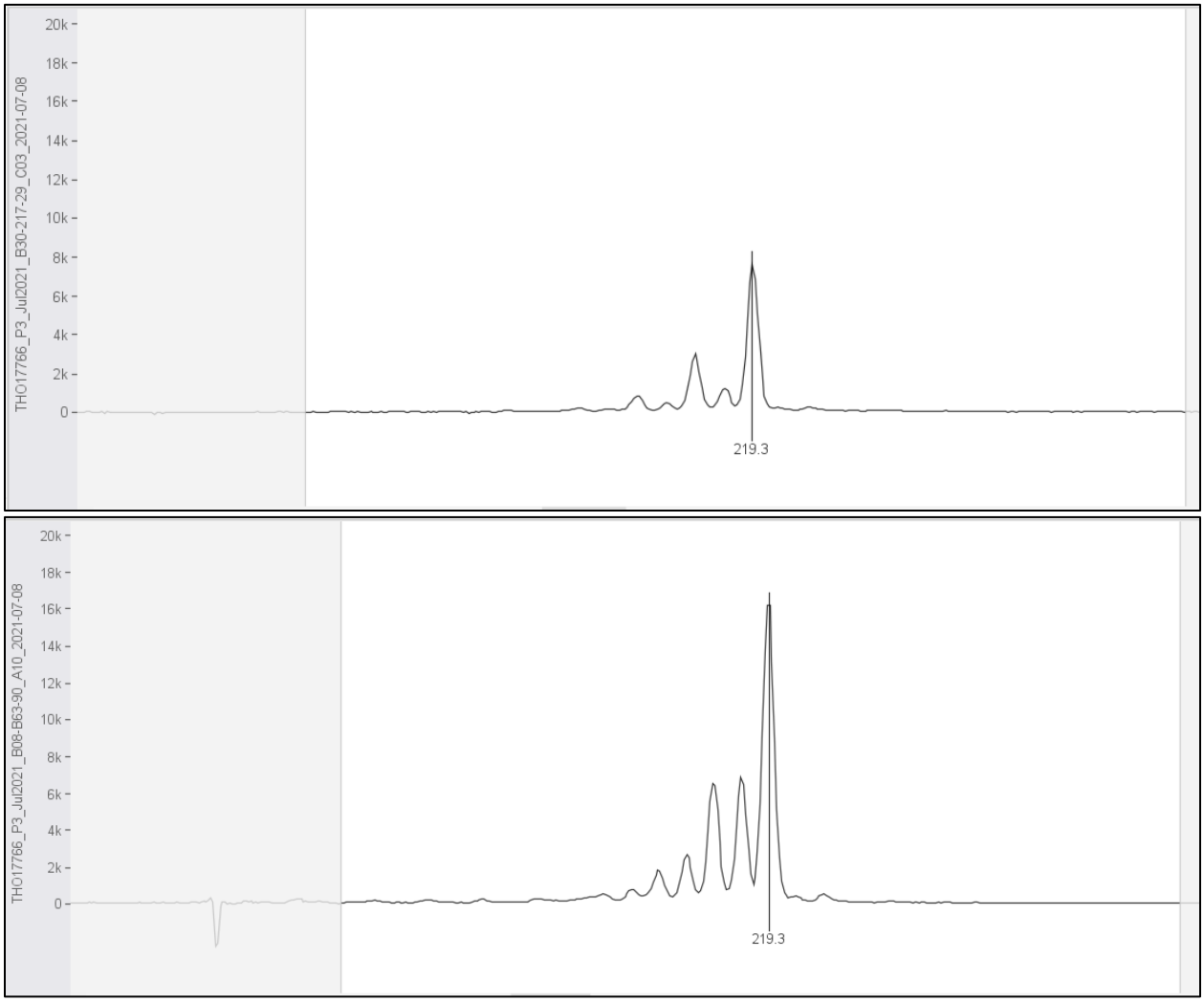
Artifact interference with allele at ~169 bp. Here, the artifact at 170.9 bp occurs adjacent to an allele at 169 bp.



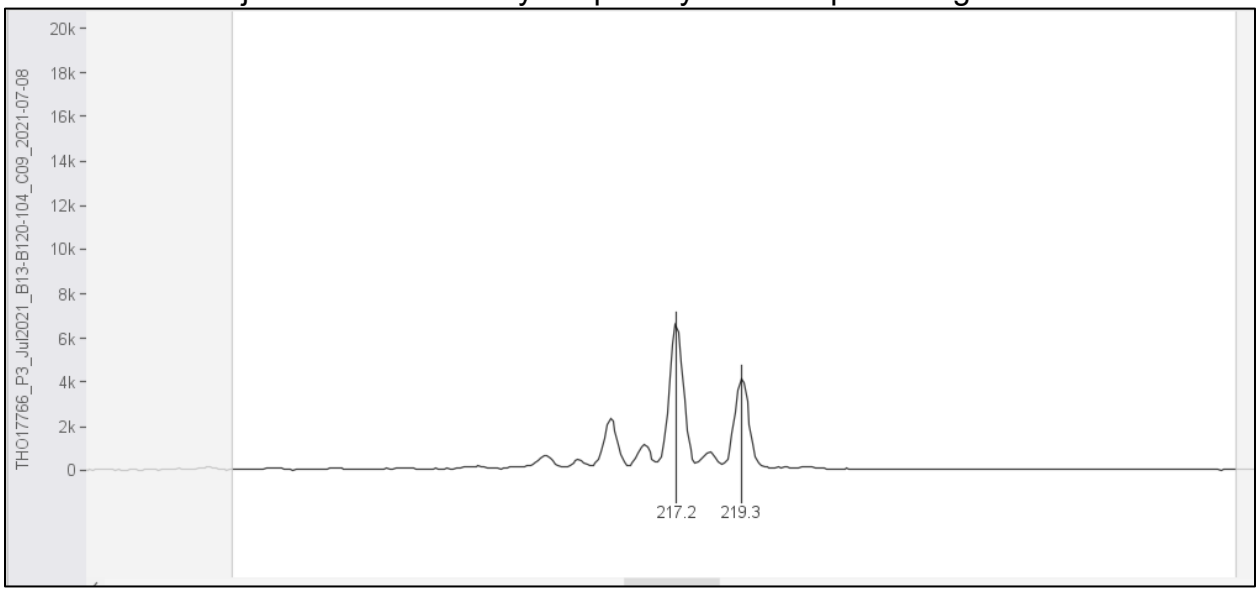
RT24

Common Morphology

There are two different types of allelic patterns within this locus. The difference being the odd bp's between the main stutters (separated by 2bp) and allele peaks. The difference is shown below and both are scored the same way. The allele is always taller than the stutters.



An immediate adjacent allele is always 2bp away from the preceding allele.

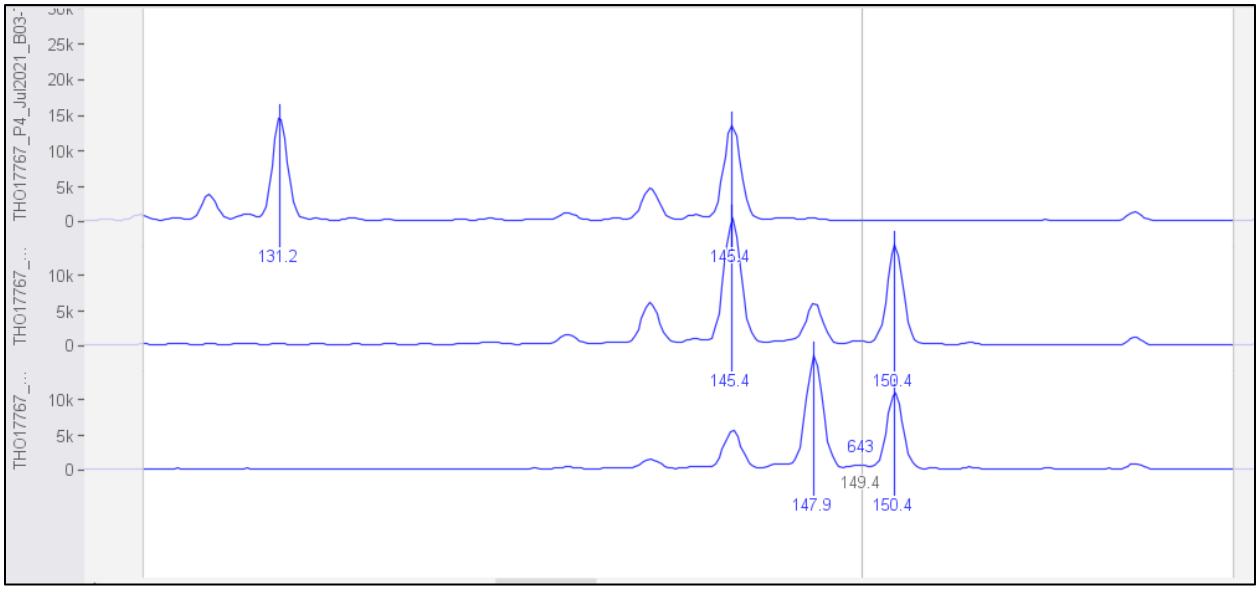


MULTIPLEX 2

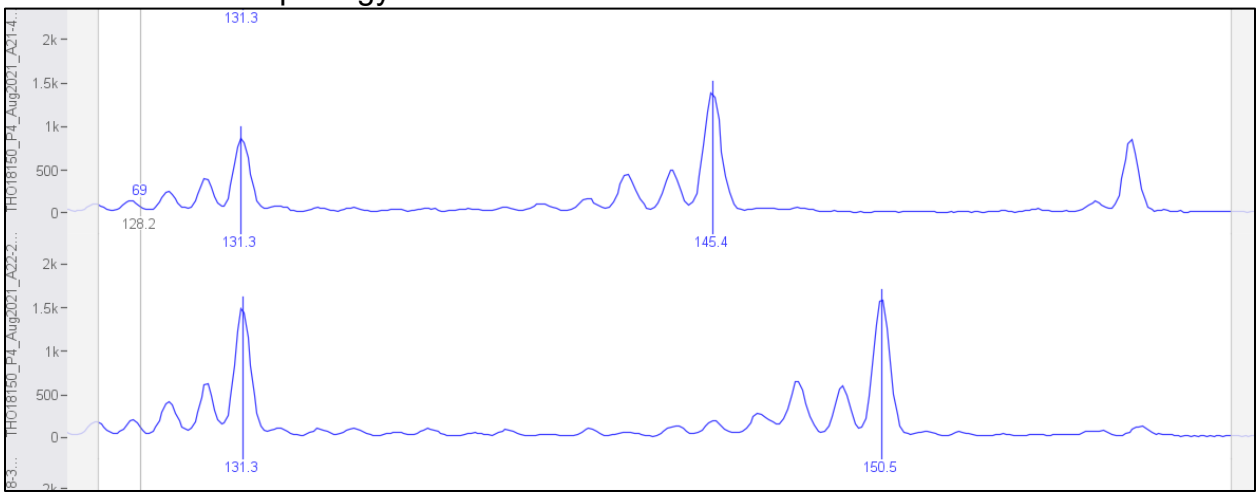
RT27

Common Morphology

The number of stutter peaks preceding the allele increases with allele size. There is also an artifact but does not interfere with the alleles.



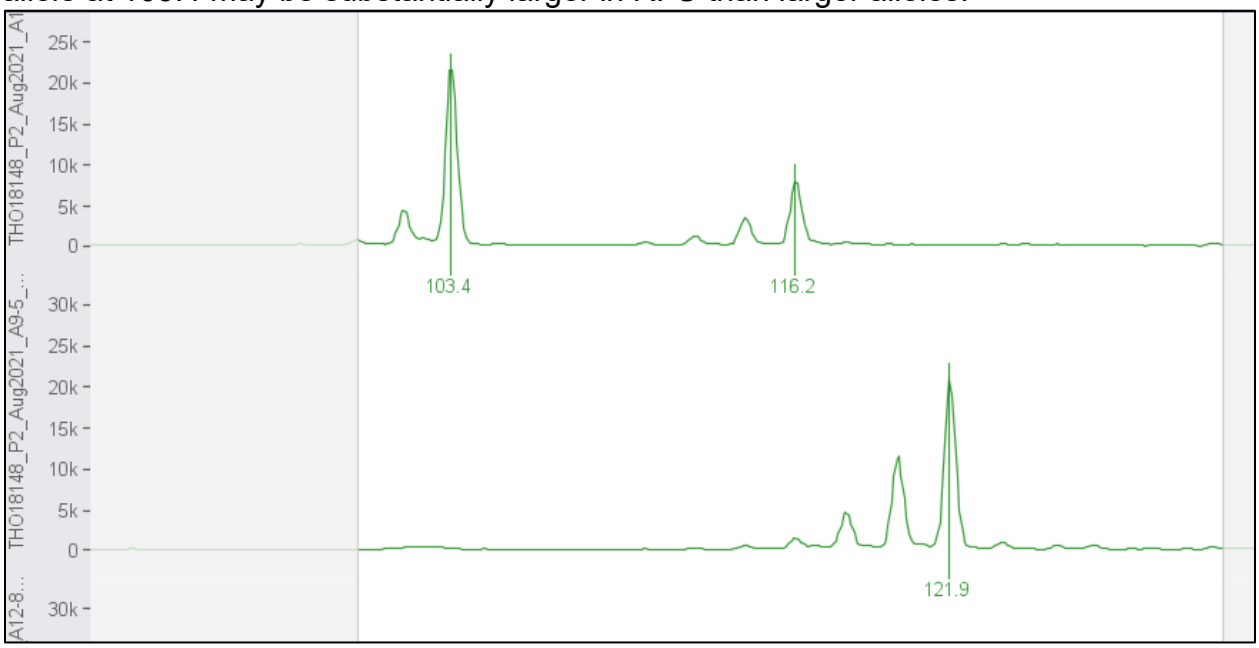
Sometimes the morphology is not as clean:



RT9-VIC

Common Morphology

The number of stutter peaks increases with allele size within this locus. The smallest allele at 103.4 may be substantially larger in RFU than larger alleles.



RT7-VIC

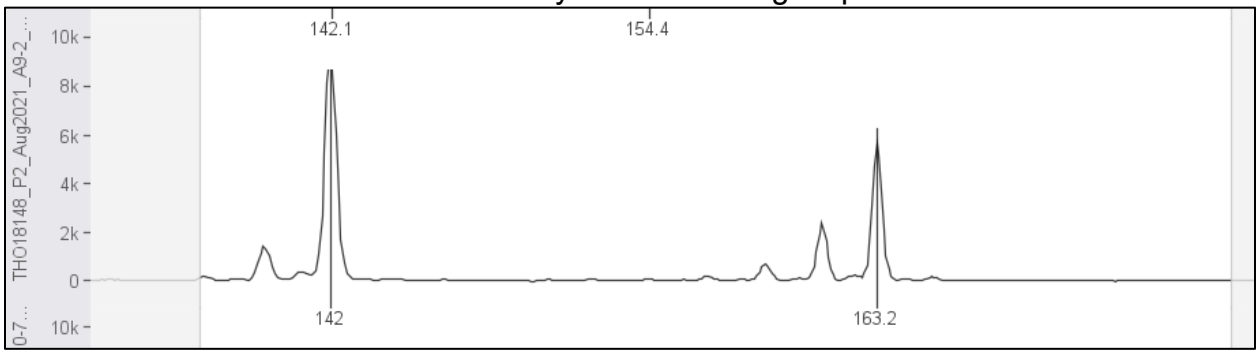
Common Morphology

Morphology contains some noise between the 2bp peak pattern.
Allele size 217 may be substantially larger and has a different stutter pattern than larger alleles. However, it does have its own distinct pattern.

RT5-NED

Common Morphology

The common morphology for this locus has a distinct pattern where the number of stutters increases with allele size. Always score the largest peak

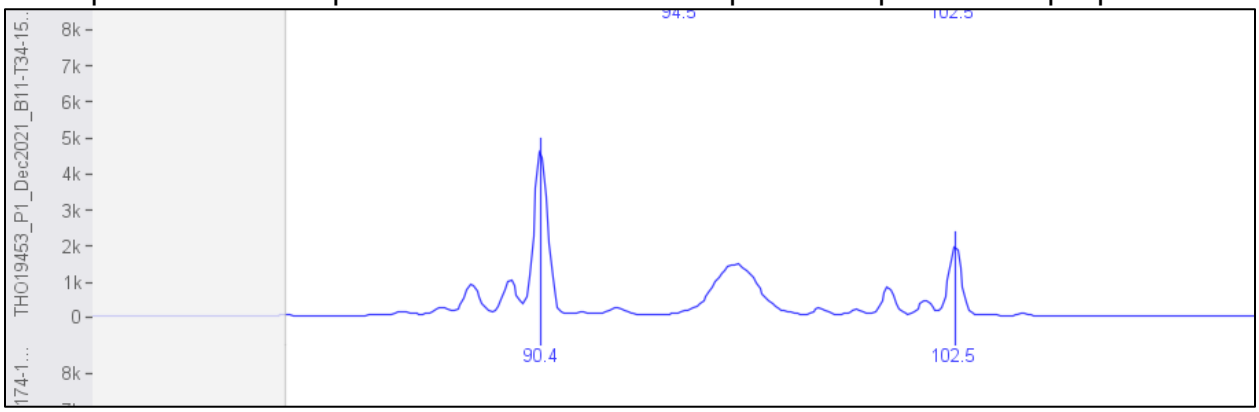


MULTIPLEX 3

MAP2C – FAM

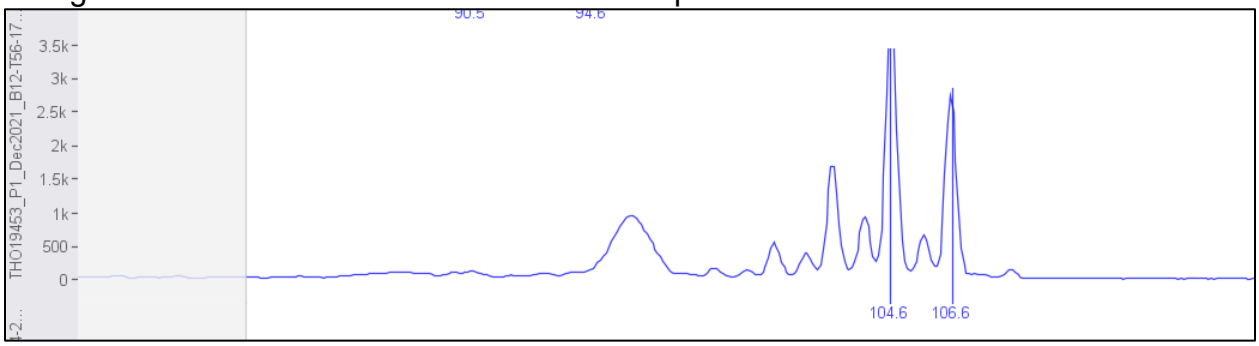
Common Morphology

The common morphology is a distinct split peak pattern which is more noticeable on larger RFU traces. The pattern contains one stutter peak 2bp preceding the allele. The minor peaks located 1bp from the allele and stutter peaks are part of the split peak.



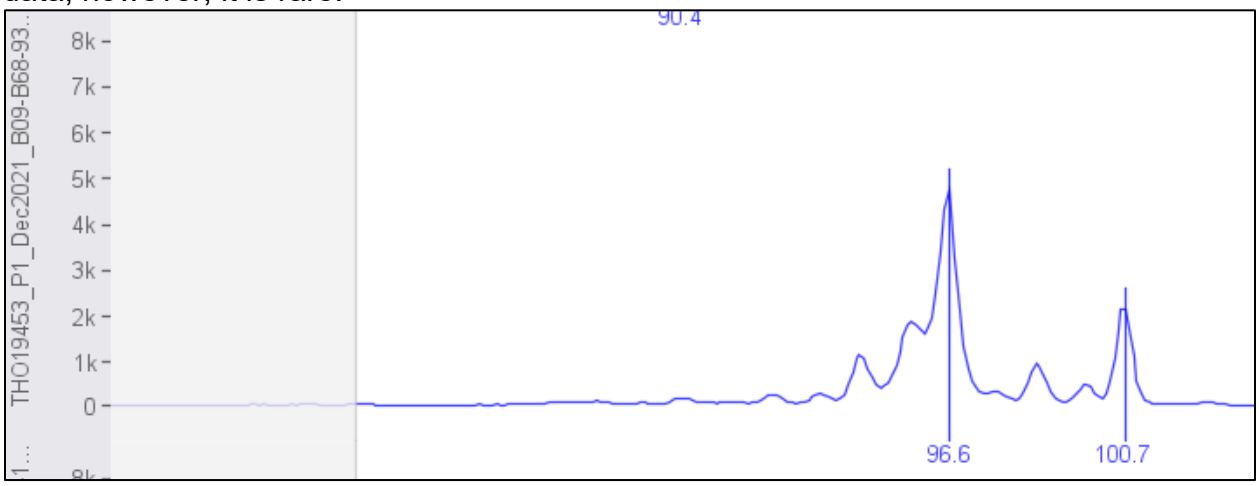
Primer Dimer ~95-96

There is a consistent presence of a primer dimer within this locus. However, there was no significant interference with the selected recipe.



Primer Dimer interference ~95-96

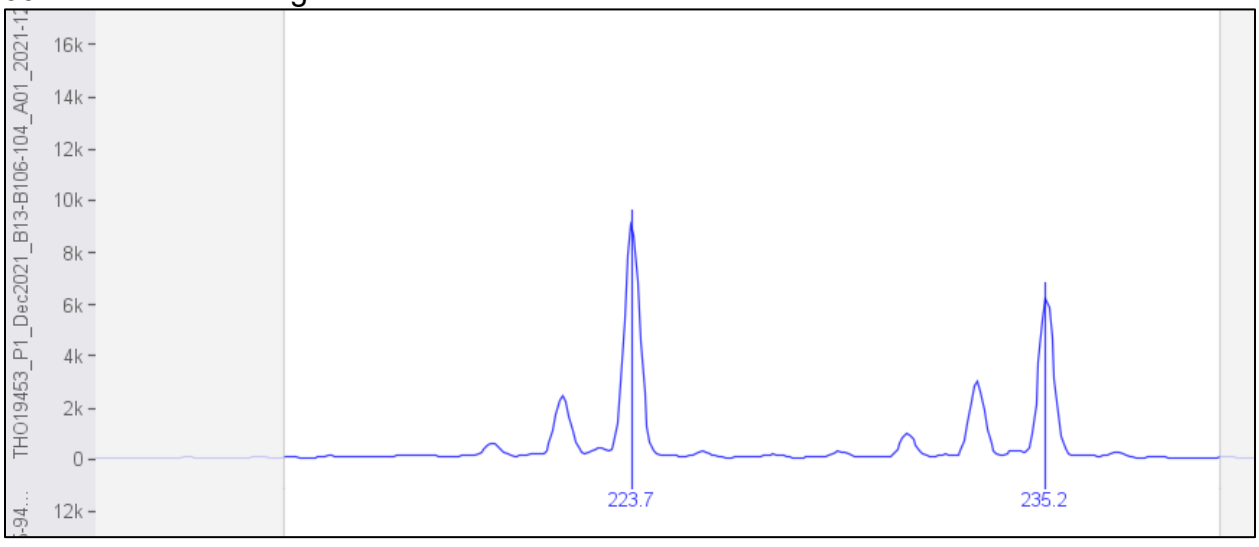
An example of a primer dimer interference looks as such. Sometimes the peaks are located on the edge of the primer dimer and sometimes in the center. The key identifier of the allele is the large peak. If un-identifiable, the locus may be scored as missing data; however, it is rare.



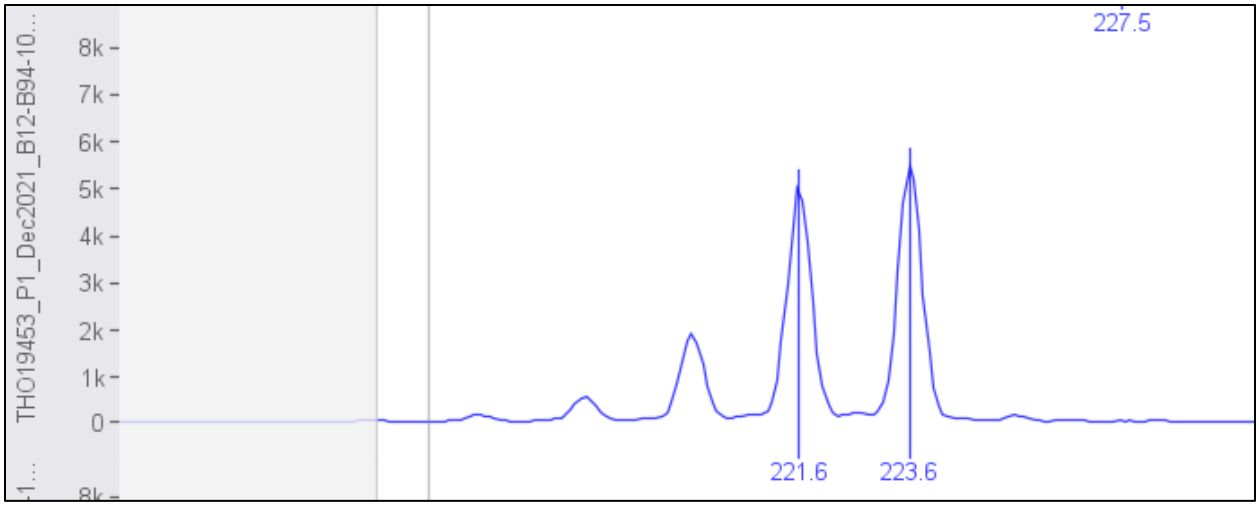
RT1 – FAM

Common Morphology

The stutter pattern preceding a peak is often 2 to 3 identifiable stutters, with more occurring with larger allele sizes. However, these stutters are often no more than 50-60% of the allele height.



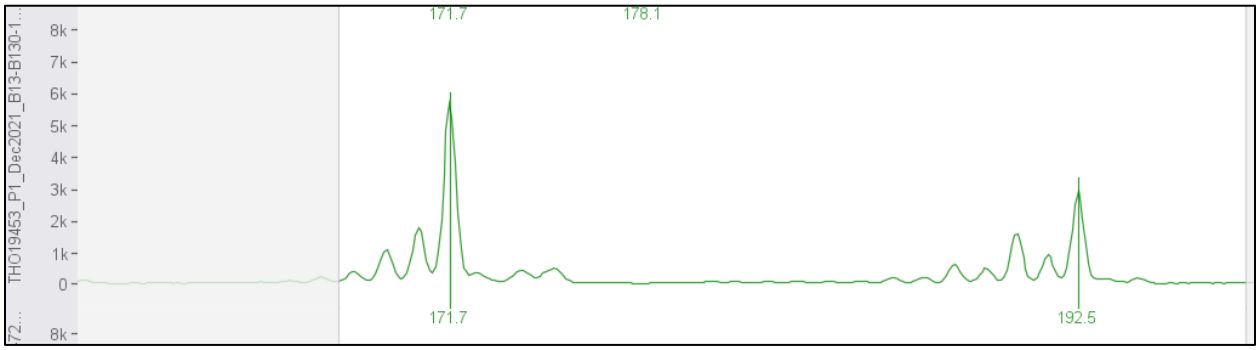
If a peak preceding an allele is not taller but within 90% of the peak height, it is scored as an allele when the remaining stutters follow the pattern. E.g.:



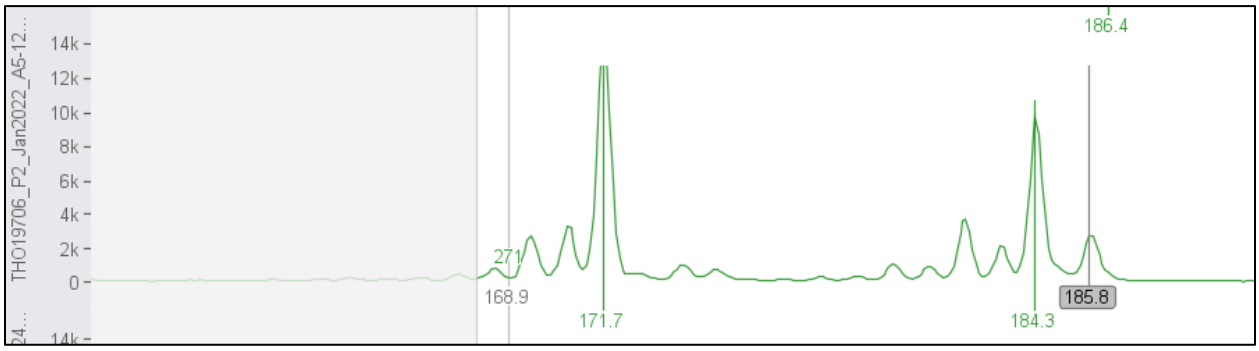
BM888 – VIC

Common Morphology

The morphology is consistent throughout this locus, where specific alleles have different patterns. The below example shows the typical difference between smaller and larger allele sizes. However, the allele peak is always significantly larger than the preceding stutters.

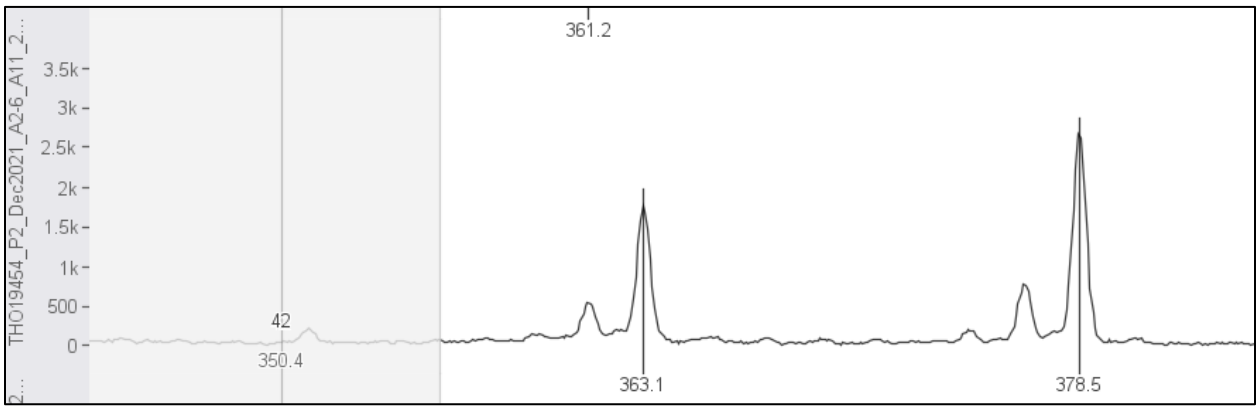


Some contain a stutter after the allele but do not present a stutter pattern and therefore are noise.



BM848 – NED

Common Morphology



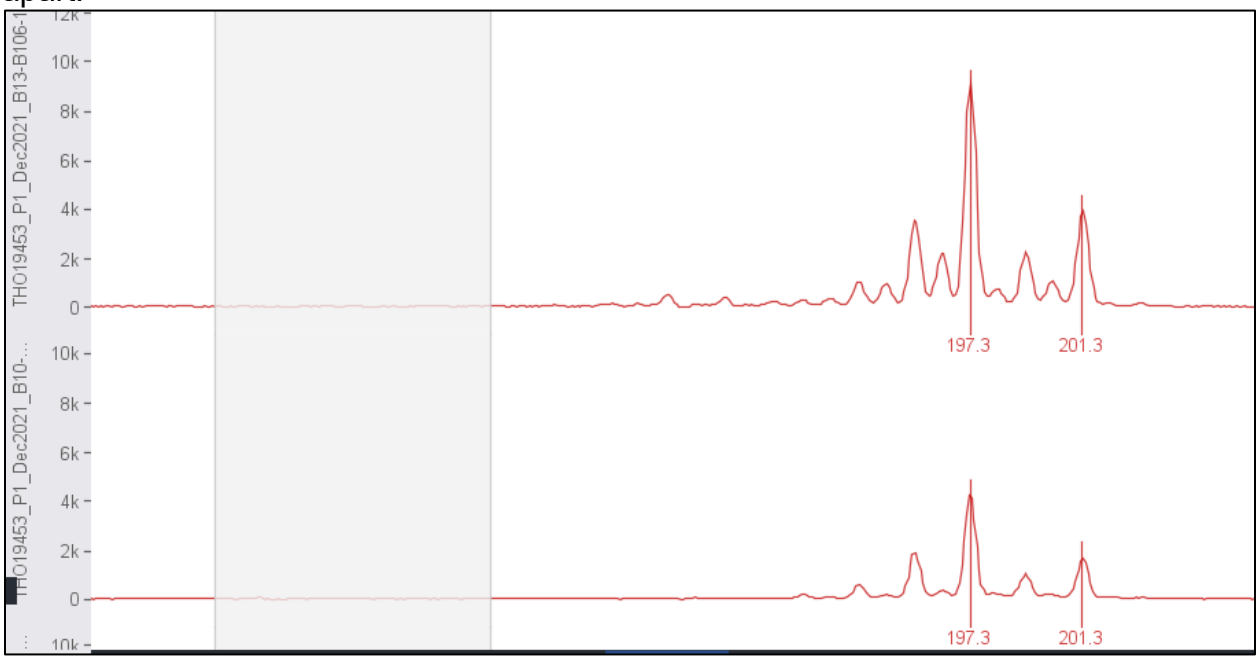
BM4513 – PET

This locus was un-scorable due to inconsistencies and a large number of stutters in the largest allele sizes. Making it difficult to distinguish heterozygotes among large allele sizes.

RT30 – PET

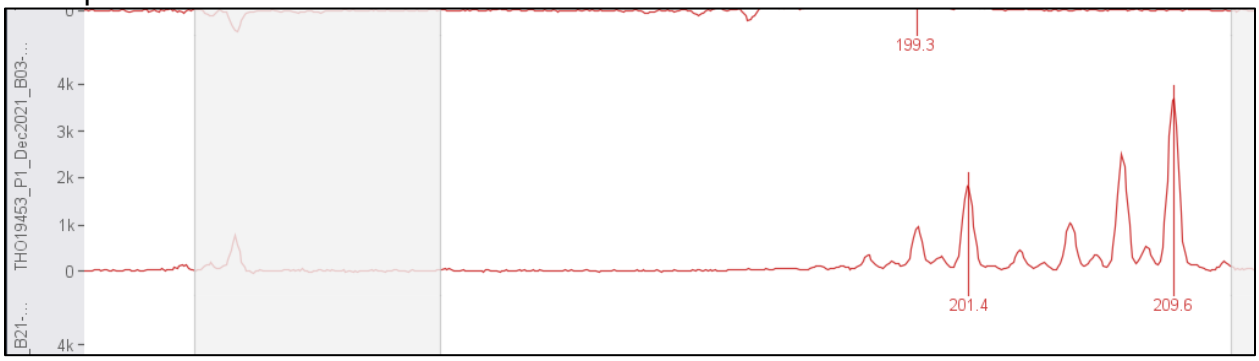
Common Morphology

This locus contains two different allelic patterns. The only difference is whether there is a presence of a minor stutter located between the stutter peaks and alleles that are 2bp apart.

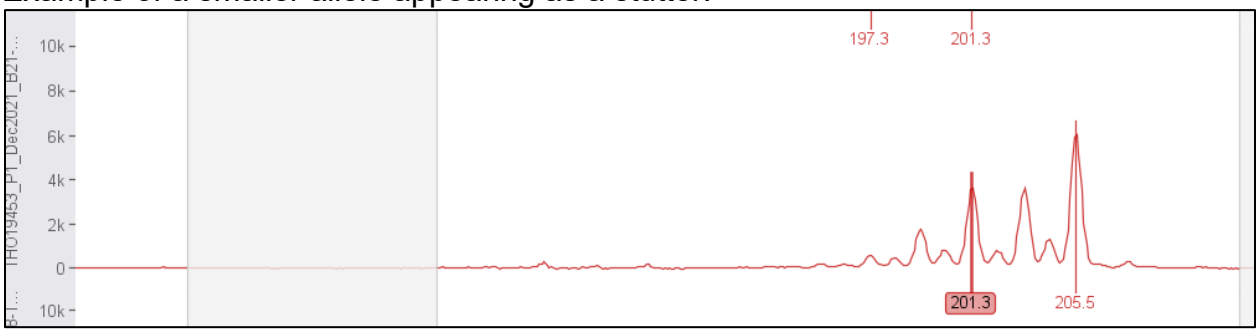


Small Alleles

Smaller alleles may have a lower RFU than a larger allele and may present as stutters at first glance but should be alleles. Alleles have a distinct two stutter pattern.
Example of a smaller allele:

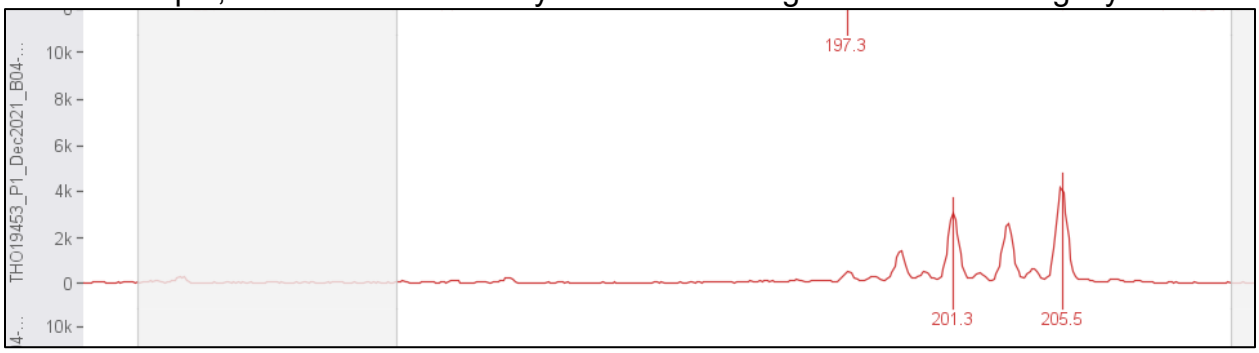


Example of a smaller allele appearing as a stutter:



Note that the height of the smaller allele matches the height of the stutter peak of the larger allele. Because the standard pattern of the larger allele has 2 stutter peaks, and in this example, it has 4 stutter peaks and the second peak matches the height of the first preceding peak, it is much more likely to be an allele. Along with this, only alleles that are only 2bp apart need to be of an equivalent height to be scored. These alleles are 4 bp apart; thus, the smaller allele can have a lower RFU. (This rule was confirmed by comparing results of the same sample)

In this example, that first allele actually exceeds the height of the stutter slightly:



REFERENCES

Thompson L, K. Smith, P. Hettinga, and P. Galpern. 2010. The Caribou Scoring Bible (unpublished). 31pp.

APPENDIX V

MICROCHECKER NAMING CONVENTION

Locus:	RT6
MSATALLELE Name	MICROCHECKER Name
93	93
97	97
99	99
100	101
102	103
106	107
107	109
109	111

Locus:	BM6506
MSATALLELE Name	MICROCHECKER Name
196	196
198	198
201	202
207	208
209	210

Locus:	BMS1788
MSATALLELE Name	MICROCHECKER Name
113	113
117	117
124	123
126	125
128	127
130	129
132	131
134	133
136	135
137	137
143	143

Locus:	NVHRT30
MSATALLELE Name	MICROCHECKER Name
147	147
154	155
157	157
159	159
161	161
163	163
165	165
167	167
169	169
172	171
174	173

Locus:	RT24
MSATALLELE Name	MICROCHECKER Name
207	207
209	209
211	211
213	213
215	215
217	217
219	219
221	221
224	223
226	225
234	233

Locus:	RT27
MSATALLELE Name	MICROCHECKER Name
131	131
134	133
145	145
148	147
150	149
153	151
155	153
160	159

Locus:	RT5
MSATALLELE Name	MICROCHECKER Name
142	143
150	151
154	155
157	157
161	161
163	163
165	165
167	167
170	169

Locus:	RT9
MSATALLELE Name	MICROCHECKER Name
103	104
116	116
120	120
122	122
124	124
128	128
130	130

Locus:	Map2C
MSATALLELE Name	MICROCHECKER Name
90	91
95	95
97	97
101	101
103	103
105	105
107	107
109	109
111	111
113	113

Locus:	RT7
MSATALLELE Name	MICROCHECKER Name
217	216
219	218
221	220
225	224
227	226
232	232
234	234

Locus:	RT1
MSATALLELE Name	MICROCHECKER Name
222	222
224	224
226	226
227	228
229	230
231	232
233	234
235	236
237	238

Locus:	RT30
MSATALLELE Name	MICROCHECKER Name
193	192
195	194
197	196
199	198
201	200
203	202
205	204
210	208

Locus:	BM888
MSATALLELE Name	MICROCHECKER Name
172	172
176	176
178	178
180	180
182	182
184	184
186	186
188	188
190	190
192	192
195	194

Locus:	BM848
MSATALLELE Name	MICROCHECKER Name
361	360
363	362
367	366
371	370
373	372
378	378
382	382
384	384

APPENDIX VI

HWE TABLE

Hardy-Weinberg Equilibrium pairwise table between sites and loci. Grey cells are sites with unique genotypes <10 and red cells are deviations from HWE, dashes are uncalculatable.

Locus	RT6...1	BM6506...	RT24...1	RT27...1	RT9...1	RT7...1	RT5...1	Map2C...1	RT1...1	BM888...1	BM848...1	RT30...1	
Site													N
A01	1	1	0.4304	0.5308	0.4293	0.4136	0.2133	0.0997	0.9982	0.2484	1	0.3872	10
A02	0.6184	0.0474	0.2709	0.7515	0.8865	0.6954	0.095	0.4887	0.1509	0.9664	0.2968	0.3915	11
A03	0.211	0.701	1	0.5083	1	1	0.8212	0.8069	0.9069	0.5632	1	1	8
A04	1	0.187	1	1	0.619	-	0.8472	1	1	0.873	1	1	5
A05	0.6102	0.2318	0.753	0.8754	0.3707	0.2159	0.9593	0.6235	0.44	0.6386	0.4683	0.3586	14
A06	0.3451	0.3277	1	0.4979	1	1	0.006	0.6702	0.9787	0.2919	1	1	7
A07	0.5323	0.3938	0.2036	1	0.5157	0.6373	1	1	0.7511	0.3267	0.0906	1	6
A08	0.9575	0.3598	0.2331	0.7661	0.1039	0.4781	0.3603	0.7207	1	0.1003	0.4412	0.0832	7
A09	0.2907	1	0.4044	0.0674	1	0.4414	0.7759	0.2729	1	0.5952	1	1	7
A10	0.6272	1	0.1985	0.2396	0.7087	0.1104	0.1492	1	0.1176	0.3286	1	0.0466	8
A11	0.3134	0.0571	-	1	0.3144	1	1	1	0.6569	0.3148	0.4287	1	4
A12	1	0.3148	1	1	1	1	1	0.4304	1	1	1	0.6585	4
A13	1	1	0.4677	1	0.4661	1	1	0.1996	1	1	1	-	3
A14	0.9171	0.024	0.6687	0.6038	0.6574	1	0.9311	0.4812	0.6649	0.4328	0.2531	0.6177	11
A15	0.9616	0.9014	0.9142	0.6017	0.2954	1	0.9906	0.4733	0.2198	0.5836	0.259	0.8965	11
A16	0.1576	0.157	0.8942	0.494	0.0556	1	0.7727	0.6665	0.5351	0.7095	0.1764	0.3197	9
A17	1	0.7944	0.4756	0.1978	0.7283	0.243	0.399	0.4184	0.4505	0.3162	0.331	0.3956	19
A18	0.3031	1	0.7818	0.4401	1	0.3282	1	0.7089	0.5817	1	0.5907	0.8512	8
A19	1	-	1	1	1	1	1	0.4673	1	1	1	0.2004	3
A20	1	-	-	-	-	1	-	-	0.334	0.3335	1	-	2
A21	0.7171	0.734	1	1	0.2537	0.1619	0.3167	1	0.3597	1	0.72	0.1055	7
A22	1	1	0.1911	0.4466	1	0.2199	1	0.0985	0.1922	0.0779	0.2121	0.2147	6
A23	1	0.3336	0.3338	1	1	-	-	1	0.3338	1	-	-	2
A24	1	1	1	1	1	1	1	1	0.0286	0.3157	1	0.3146	4
A25	1	1	0.4659	1	-	-	1	1	1	0.0664	-	1	3
A26	0.3447	0.7099	1	0.6974	0.8947	1	1	0.7362	0.3035	0.9865	0.7363	0.9204	9
A27	1	0.6953	1	0.6173	0.4935	0.1113	0.1426	0.7452	1	0.3508	0.3653	1	5
A28	0.2873	0.7332	0.0771	0.7669	0.9169	0.4095	1	1	0.6485	0.1364	0.1264	1	10
B01	-	-	-	-	-	-	-	-	-	-	-	-	1
B02	1	0.0394	0.1787	0.8503	0.44	0.4407	1	1	0.4026	0.8508	1	0.889	7
B03	1	1	0.7922	1	0.1	1	0.4808	0.7639	0.3157	0.9547	1	0.7908	6
B04	0.7405	0.0029	0.2387	0.8208	0.049	0.627	0.2174	0.7657	0.2377	0.4829	0.9007	1	8
B06	1	0.6192	0.6192	0.7475	1	-	0.6949	0.4289	0.6196	0.3529	0.3329	0.7695	5
B07	-	-	-	-	-	-	-	-	-	-	-	-	1
B08	-		0.6005	0.4664	1	1	1	1	0.6008	1	1	1	3
B10	1	0.1867	0.6535	0.1515	0.3429	1	1	0.0778	0.0524	1	1	0.5146	6
B12	0.089	0.7858	0.4109	0.8608	1	0.8249	0.5361	0.1224	0.5417	0.1307	0.7033	0.0807	13
B13	0.8382	1	0.6684	0.772	0.5632	0.849	0.9444	0.3265	0.9906	0.7935	0.038	0.042	11
B14	1	1	0.3333	-	1	-	1	0.3341	0.3331	1	1	1	2
B15	1	1	1	-	0.6008	1	1	0.6	1	1	-	0.2006	3
B16	-	-	-	-	-	-	-	-	-	-	-	-	1
B17	-	-	-	-	-	-	-	-	-	-	-	-	1
B18	0.895	0.7236	1	0.2152	0.9181	0.1053	0.4466	0.8066	0.0512	0.906	0.2199	0.3936	13
B19	1	1	-	0.3145	1	0.7714	1	1	0.4285	0.314	1	0.3149	4
B20	-	-	-	-	-	-	-	-	-	-	-	-	1
B21	0.8788	0.3632	0.0151	0.685	0.1506	0.8158	0.4976	0.9431	0.5963	0.1535	0.2583	0.7878	15
B22	-	-	-	-	-	-	-	-	-	-	-	-	1
B23	1	1	1	1	1	-	1	0.3136	1	0.3147	1	1	4
B24	1	1	1	0.6192	0.0478	1	0.6178	1	0.2895	0.8989	1	1	5
B26	-	-	-	-	-	-	-	-	-	-	-	-	1
B27	1	-	-	-	-	-	-	-	1	-	1	1	2
B29	0.5154	0.6364	0.0301	1	0.2733	1	0.6541	0.4305	0.434	0.0827	1	0.5845	6
B30	0.7725	0.3145	1	1	1	0.1427	1	1	1	0.3146	1	1	4
B31	-	-	-	-	-	-	-	-	-	-	-	-	1
B32	-	-	-	-	-	-	-	-	-	-	-	-	1
B33	-	-	-	-	-	-	-	-	-	-	-	-	1
B35	-	-	-	-	-	-	-	-	-	-	-	-	1
B36	1	0.325	0.5144	1	0.516	1	0.7917	0.5559	0.1586	0.8753	1	0.6541	6

LD

Locus pair			Chi2	df	P-Value
RT6...1	&	BM6506...	33.99637	64	0.999254
RT6...1	&	RT24...1	22.23201	62	0.999999
BM6506...	&	RT24...1	33.06451	74	0.99999
RT6...1	&	RT27...1	20.87105	58	0.999998
BM6506...	&	RT27...1	31.39319	66	0.999907
RT24...1	&	RT27...1	34.01234	64	0.999249
RT6...1	&	RT9...1	32.17753	62	0.999381
BM6506...	&	RT9...1	54.04958	74	0.960776
RT24...1	&	RT9...1	34.08399	72	0.999958
RT27...1	&	RT9...1	27.72392	62	0.99995
RT6...1	&	RT7...1	29.25567	54	0.997621
BM6506...	&	RT7...1	47.25479	68	0.973937
RT24...1	&	RT7...1	39.87847	66	0.995473
RT27...1	&	RT7...1	29.0875	60	0.999745
RT9...1	&	RT7...1	46.53359	64	0.950688
RT6...1	&	RT5...1	15.79154	56	1
BM6506...	&	RT5...1	27.04983	66	0.999995
RT24...1	&	RT5...1	34.31561	64	0.999135
RT27...1	&	RT5...1	26.73413	60	0.999938
RT9...1	&	RT5...1	24.51484	62	0.999995
RT7...1	&	RT5...1	33.0571	62	0.999056
RT6...1	&	Map2C...1	18.28748	48	0.999969
BM6506...	&	Map2C...1	27.3959	54	0.999036
RT24...1	&	Map2C...1	44.02432	52	0.776304
RT27...1	&	Map2C...1	27.19075	50	0.996489
RT9...1	&	Map2C...1	15.51128	54	1
RT7...1	&	Map2C...1	20.21975	52	0.999979
RT5...1	&	Map2C...1	16.34908	52	0.999999
RT6...1	&	RT1...1	30.61638	54	0.995697
BM6506...	&	RT1...1	26.90727	64	0.999988
RT24...1	&	RT1...1	30.50171	60	0.999457
RT27...1	&	RT1...1	22.65174	58	0.999992
RT9...1	&	RT1...1	16.33276	60	1
RT7...1	&	RT1...1	27.1679	56	0.999596
RT5...1	&	RT1...1	7.581603	56	1
Map2C...1	&	RT1...1	15.3288	48	0.999998
RT6...1	&	BM888...1	12.41942	52	1
BM6506...	&	BM888...1	22.69005	56	0.999978
RT24...1	&	BM888...1	28.8965	56	0.998992
RT27...1	&	BM888...1	9.994657	52	1
RT9...1	&	BM888...1	29.51934	56	0.998633
RT7...1	&	BM888...1	21.95034	52	0.999921
RT5...1	&	BM888...1	10.30647	54	1
Map2C...1	&	BM888...1	8.154006	50	1
RT1...1	&	BM888...1	22.9706	48	0.999169
RT6...1	&	BM848...1	30.2592	62	0.999772
BM6506...	&	BM848...1	47.32007	74	0.993332
RT24...1	&	BM848...1	29.76501	74	0.999999
RT27...1	&	BM848...1	15.26681	64	1
RT9...1	&	BM848...1	29.78843	68	0.999985
RT7...1	&	BM848...1	33.51553	66	0.999707
RT5...1	&	BM848...1	20.89848	66	1
Map2C...1	&	BM848...1	29.62929	52	0.994683
RT1...1	&	BM848...1	38.10884	60	0.987726
BM888...1	&	BM848...1	37.80882	56	0.970352
RT6...1	&	RT30...1	30.61211	54	0.995704
BM6506...	&	RT30...1	37.24156	62	0.994672
RT24...1	&	RT30...1	22.48595	62	0.999999
RT27...1	&	RT30...1	11.30102	56	1
RT9...1	&	RT30...1	35.58555	62	0.997175
RT7...1	&	RT30...1	34.37306	60	0.996823
RT5...1	&	RT30...1	34.24272	58	0.99453
Map2C...1	&	RT30...1	24.04212	54	0.999863
RT1...1	&	RT30...1	19.00506	58	1
BM888...1	&	RT30...1	8.358507	52	1
BM848...1	&	RT30...1	35.13831	62	0.997645

GENETIC DIVERSITY STATISTICS

Site characteristics including sample size and mean genetic diversity statistics. Number of individuals (N) and the mean allelic diversity (Na), mean observed heterozygosity (Ho), mean expected heterozygosity (He), and mean fixation index (F). Raw output from GENALEX.

Pop	N	Na	Ne	I	Ho	He	uHe	F
A01	10.00	4.67	3.15	1.26	0.650	0.658	0.693	0.00471
A02	13.00	5.25	3.01	1.28	0.699	0.652	0.678	-0.0731
A03	9.00	5.00	3.13	1.29	0.731	0.663	0.702	-0.100
A04	5.00	4.25	3.21	1.19	0.700	0.623	0.693	-0.127
A05	14.00	5.42	3.33	1.31	0.708	0.655	0.679	-0.0688
A06	8.00	4.92	3.26	1.29	0.708	0.654	0.698	-0.0830
A07	6.00	3.83	2.90	1.13	0.764	0.620	0.677	-0.241
A08	9.00	5.42	3.59	1.40	0.704	0.695	0.736	-0.0187
A09	13.00	5.75	3.32	1.35	0.660	0.661	0.687	-0.00275
A10	8.00	4.83	2.99	1.25	0.656	0.632	0.674	-0.0457
A12	5.00	4.17	3.03	1.19	0.733	0.627	0.696	-0.175
A14	11.00	4.50	2.75	1.16	0.674	0.613	0.642	-0.104
A15	11.00	5.17	3.26	1.34	0.682	0.675	0.707	-0.00839
A16	11.00	5.00	3.12	1.27	0.667	0.639	0.670	-0.0542
A17	18.00	5.58	3.47	1.36	0.681	0.674	0.693	-0.0111
A18	8.00	4.08	2.89	1.17	0.646	0.635	0.678	-0.00986
A19	5.00	4.25	3.17	1.23	0.750	0.647	0.719	-0.167
A21	7.00	4.08	2.73	1.14	0.655	0.616	0.664	-0.0692
A22	7.00	4.50	3.27	1.29	0.631	0.681	0.734	0.0697
A26	10.00	4.83	3.02	1.24	0.742	0.632	0.665	-0.177
A27	5.00	4.25	3.04	1.23	0.700	0.648	0.720	-0.0607
A28	10.00	4.33	3.14	1.23	0.700	0.663	0.697	-0.0509
B02	7.00	4.50	2.82	1.19	0.690	0.619	0.667	-0.107
B03	12.00	5.50	3.18	1.34	0.722	0.665	0.694	-0.0830
B04	9.00	4.75	3.22	1.30	0.685	0.675	0.715	-0.0133
B06	5.00	4.08	2.94	1.14	0.717	0.602	0.669	-0.170
B08	6.00	4.33	2.86	1.18	0.722	0.612	0.668	-0.187
B10	7.00	4.17	2.90	1.16	0.655	0.622	0.669	-0.0666
B12	19.00	5.83	3.12	1.34	0.706	0.665	0.683	-0.0627
B13	12.00	5.42	3.25	1.32	0.708	0.654	0.682	-0.0601
B18	15.92	5.58	3.12	1.31	0.643	0.654	0.675	0.00804
B19	5.00	3.83	2.80	1.10	0.650	0.593	0.659	-0.103
B21	18.00	5.50	2.98	1.25	0.681	0.629	0.647	-0.0659
B23	5.00	3.67	2.86	1.06	0.667	0.577	0.641	-0.162
B24	7.00	4.58	3.09	1.25	0.714	0.650	0.700	-0.0932
B26	6.00	4.08	2.77	1.14	0.708	0.606	0.662	-0.179
B29	8.00	4.58	3.23	1.21	0.635	0.617	0.658	-0.0362
B33	8.00	4.33	3.18	1.25	0.708	0.665	0.709	-0.0634

B36	9.00	4.25	2.97	1.17	0.602	0.625	0.662	0.0386
Mean	9.280	4.69	3.08	1.24	0.689	0.641	0.684	-0.0763

ERROR RATES

Table 5. Error rate by locus based on 140 samples (~17%).

Locus	Error Rate
RT6	0.0%
BM6506	0.0%
BMS1788	0.0%
NVHRT30	0.7%
RT24	0.0%
RT27	0.0%
RT9	0.0%
RT7	0.0%
RT5	0.0%
Map2C	0.0%
RT1	0.0%
BM888	0.0%
BM848	0.0%
RT30	0.0%
Mean	0.1%

NUMBER OF ALLELES PER LOCUS

Table 6. Number of alleles per locus.

Locus	Number of Alleles
RT6	8.00
BM6506	5.00
RT24	11.00
RT27	8.00
RT9	7.00
RT7	7.00
RT5	9.00
Map2C	10.00
RT1	9.00
BM888	11.00
BM848	7.00
RT30	8.00
Mean	8.33

PAIRWISE F_{ST}

Table of Pairwise F_{ST} where significant comparisons are highlighted in yellow. F_{ST} values are below the diagonal and significance thresholds are denoted be * -0.05, **0.01,***0.001.

	A01	A02	A03	A04	A05	A06	A07	A08
A01	0.000							
A02	0.014	0.000						
A03	0.021	0.020	0.000					
A04	0.037	0.036	0.038	0.000				
A05	0.023	0.022	0.029	0.041	0.000			
A06	0.023	0.024	0.030	0.035	0.025	0.000		
A07	0.035	0.026	0.036	0.045	0.039	0.048	0.000	
A08	0.026	0.026	0.041	0.045	0.027	0.032	0.024	0.000
A09	0.016	0.017	0.010	0.028	0.023	0.024	0.029	0.032
A10	0.023	0.020	0.030	0.048	0.023	0.034	0.044	0.031
A12	0.028	0.025	0.033	0.040	0.039	0.037	0.033	0.037
A14	0.021	0.020	0.026	0.037	0.028	0.027	0.037	0.032
A15	0.027	0.021	0.027	0.035	0.023	0.022	0.045	0.030
A16	0.028	0.026	0.024	0.024	0.030	0.029	0.034	0.035
A17	0.016	0.016	0.019	0.029	0.021	0.019	0.034	0.023
A18	0.043	0.048	0.035	0.056	0.051	0.049	0.064	0.058
A19	0.033	0.033	0.042	0.055	0.030	0.022	0.054	0.029
A21	0.021	0.024	0.033	0.046	0.030	0.033	0.049	0.039
A22	0.029	0.031	0.032	0.045	0.029	0.036	0.050	0.035
A26	0.025	0.026	0.034	0.038	0.017	0.032	0.045	0.033
A27	0.025	0.029	0.040	0.042	0.032	0.035	0.041	0.024
A28	0.031	0.035	0.042	0.065	0.030	0.038	0.056	0.041
B02	0.028	0.033	0.042	0.046	0.026	0.035	0.047	0.028
B03	0.018	0.018	0.027	0.036	0.016	0.024	0.033	0.017
B04	0.022	0.026	0.032	0.032	0.027	0.021	0.042	0.028
B06	0.045	0.043	0.049	0.054	0.044	0.046	0.060	0.048
B08	0.036	0.027	0.042	0.046	0.042	0.042	0.035	0.031
B10	0.026	0.021	0.029	0.045	0.030	0.045	0.039	0.035
B12	0.012	0.016	0.021	0.025	0.020	0.017	0.032	0.022
B13	0.034	0.026	0.034	0.051	0.016	0.028	0.049	0.032
B18	0.017	0.013	0.026	0.045	0.024	0.025	0.046	0.029
B19	0.042	0.041	0.048	0.061	0.045	0.052	0.064	0.059
B21	0.018	0.020	0.028	0.043	0.017	0.022	0.042	0.025
B23	0.041	0.047	0.049	0.054	0.036	0.052	0.059	0.046
B24	0.014	0.014	0.024	0.042	0.020	0.016	0.041	0.028
B26	0.035	0.038	0.022	0.054	0.042	0.043	0.049	0.055
B29	0.025	0.027	0.036	0.052	0.023	0.040	0.036	0.025
B33	0.032	0.035	0.040	0.060	0.030	0.036	0.057	0.042
B36	0.022	0.023	0.028	0.047	0.023	0.034	0.042	0.032

	A09	A10	A12	A14	A15	A16	A17	A18
A01								
A02								**
A03								
A04								
A05				*		*		**
A06								
A07								*
A08								*
A09	0.000							*
A10	0.024	0.000						**
A12	0.027	0.044	0.000					
A14	0.029	0.036	0.036	0.000		*		**
A15	0.022	0.029	0.036	0.025	0.000			
A16	0.014	0.027	0.032	0.033	0.030	0.000		*
A17	0.012	0.027	0.026	0.026	0.018	0.020	0.000	***
A18	0.042	0.063	0.046	0.051	0.042	0.050	0.047	0.000
A19	0.036	0.030	0.048	0.043	0.026	0.040	0.034	0.058
A21	0.031	0.033	0.030	0.025	0.035	0.040	0.025	0.057
A22	0.030	0.029	0.037	0.039	0.029	0.037	0.032	0.050
A26	0.025	0.033	0.037	0.034	0.025	0.032	0.018	0.051
A27	0.030	0.034	0.029	0.033	0.025	0.032	0.024	0.050
A28	0.041	0.045	0.043	0.043	0.036	0.054	0.039	0.051
B02	0.030	0.033	0.044	0.033	0.029	0.031	0.021	0.069
B03	0.018	0.020	0.033	0.026	0.018	0.021	0.014	0.050
B04	0.023	0.038	0.035	0.025	0.020	0.025	0.020	0.039
B06	0.032	0.049	0.050	0.057	0.039	0.033	0.035	0.076
B08	0.027	0.029	0.048	0.049	0.042	0.036	0.030	0.079
B10	0.021	0.022	0.035	0.043	0.034	0.030	0.024	0.061
B12	0.017	0.018	0.028	0.021	0.021	0.019	0.015	0.043
B13	0.029	0.027	0.048	0.041	0.024	0.038	0.027	0.057
B18	0.024	0.026	0.031	0.022	0.019	0.032	0.018	0.048
B19	0.043	0.036	0.059	0.057	0.049	0.053	0.049	0.082
B21	0.020	0.018	0.040	0.025	0.021	0.029	0.018	0.050
B23	0.046	0.047	0.051	0.047	0.044	0.050	0.036	0.061
B24	0.022	0.026	0.028	0.022	0.019	0.034	0.017	0.041
B26	0.033	0.046	0.045	0.031	0.047	0.036	0.042	0.045
B29	0.022	0.025	0.038	0.043	0.033	0.033	0.023	0.063
B33	0.037	0.042	0.046	0.041	0.031	0.048	0.036	0.056
B36	0.020	0.024	0.027	0.033	0.020	0.035	0.022	0.045

	A19	A21	A22	A26	A27	A28	B02	B03
A01								
A02						*		
A03						*		
A04						*		
A05								
A06								
A07				*		*		
A08						*		
A09						**		
A10						*		
A12								
A14				*		**		
A15						*		
A16						**		
A17						**		
A18		*		**		*	**	**
A19	0.000							
A21	0.050	0.000						
A22	0.030	0.044	0.000					
A26	0.039	0.033	0.034	0.000				
A27	0.037	0.041	0.034	0.031	0.000			
A28	0.044	0.037	0.033	0.031	0.046	0.000		
B02	0.040	0.045	0.037	0.025	0.024	0.051	0.000	
B03	0.021	0.033	0.025	0.019	0.019	0.035	0.011	0.000
B04	0.031	0.034	0.032	0.026	0.026	0.033	0.031	0.021
B06	0.053	0.060	0.057	0.038	0.047	0.069	0.040	0.033
B08	0.044	0.048	0.047	0.037	0.048	0.054	0.046	0.026
B10	0.044	0.043	0.034	0.032	0.035	0.045	0.032	0.018
B12	0.024	0.022	0.021	0.020	0.024	0.031	0.026	0.016
B13	0.031	0.041	0.030	0.025	0.043	0.038	0.037	0.023
B18	0.033	0.025	0.031	0.027	0.024	0.039	0.032	0.021
B19	0.055	0.040	0.039	0.053	0.067	0.056	0.067	0.047
B21	0.028	0.025	0.029	0.020	0.029	0.031	0.027	0.017
B23	0.070	0.050	0.047	0.029	0.046	0.049	0.042	0.045
B24	0.027	0.021	0.026	0.026	0.025	0.026	0.035	0.019
B26	0.055	0.040	0.047	0.051	0.050	0.054	0.064	0.042
B29	0.040	0.037	0.038	0.026	0.037	0.041	0.031	0.022
B33	0.039	0.039	0.035	0.032	0.047	0.005	0.048	0.032
B36	0.036	0.027	0.021	0.023	0.027	0.033	0.031	0.021

	B04	B06	B08	B10	B12	B13	B18	B19
A01						*		
A02								
A03								
A04						*		
A05								
A06								
A07						*	*	*
A08								*
A09								
A10								
A12								
A14		*	*	*		**		*
A15								
A16						*	*	*
A17						*		*
A18		*	**	*	**	**	*	*
A19								
A21						*		
A22								
A26								*
A27								
A28		*	*		*	*	**	
B02								*
B03								
B04	0.000							*
B06	0.042	0.000					*	
B08	0.041	0.047	0.000				*	
B10	0.039	0.051	0.027	0.000				
B12	0.015	0.042	0.025	0.025	0.000			
B13	0.035	0.051	0.038	0.034	0.024	0.000	**	*
B18	0.023	0.050	0.047	0.035	0.018	0.032	0.000	
B19	0.059	0.061	0.053	0.056	0.038	0.058	0.047	0.000
B21	0.024	0.040	0.032	0.035	0.017	0.021	0.019	0.039
B23	0.045	0.068	0.063	0.057	0.034	0.044	0.041	0.074
B24	0.019	0.051	0.037	0.030	0.016	0.027	0.015	0.042
B26	0.042	0.074	0.061	0.057	0.037	0.050	0.042	0.068
B29	0.037	0.038	0.031	0.029	0.025	0.032	0.032	0.043
B33	0.031	0.059	0.050	0.042	0.031	0.039	0.041	0.053
B36	0.029	0.042	0.039	0.026	0.021	0.028	0.023	0.035

	B21	B23	B24	B26	B29	B33	B36
A01							
A02							
A03							
A04							
A05							
A06							
A07	*					*	
A08				*			
A09							
A10							
A12							
A14					*	*	
A15				*			
A16	*					*	
A17				*		*	
A18	***				**	*	
A19							
A21							
A22							
A26				*			
A27							
A28	*			*			
B02				*			
B03							
B04							
B06				*			
B08		*		*			
B10							
B12							
B13				*		*	
B18						*	
B19							
B21	0.000			*			
B23	0.030	0.000					
B24	0.014	0.040	0.000				
B26	0.040	0.067	0.034	0.000	*		
B29	0.017	0.036	0.029	0.062	0.000		
B33	0.031	0.055	0.027	0.054	0.041	0.000	
B36	0.017	0.034	0.018	0.049	0.021	0.035	0.000

PLOT OF DELTA K AND LNP(K) VALUES

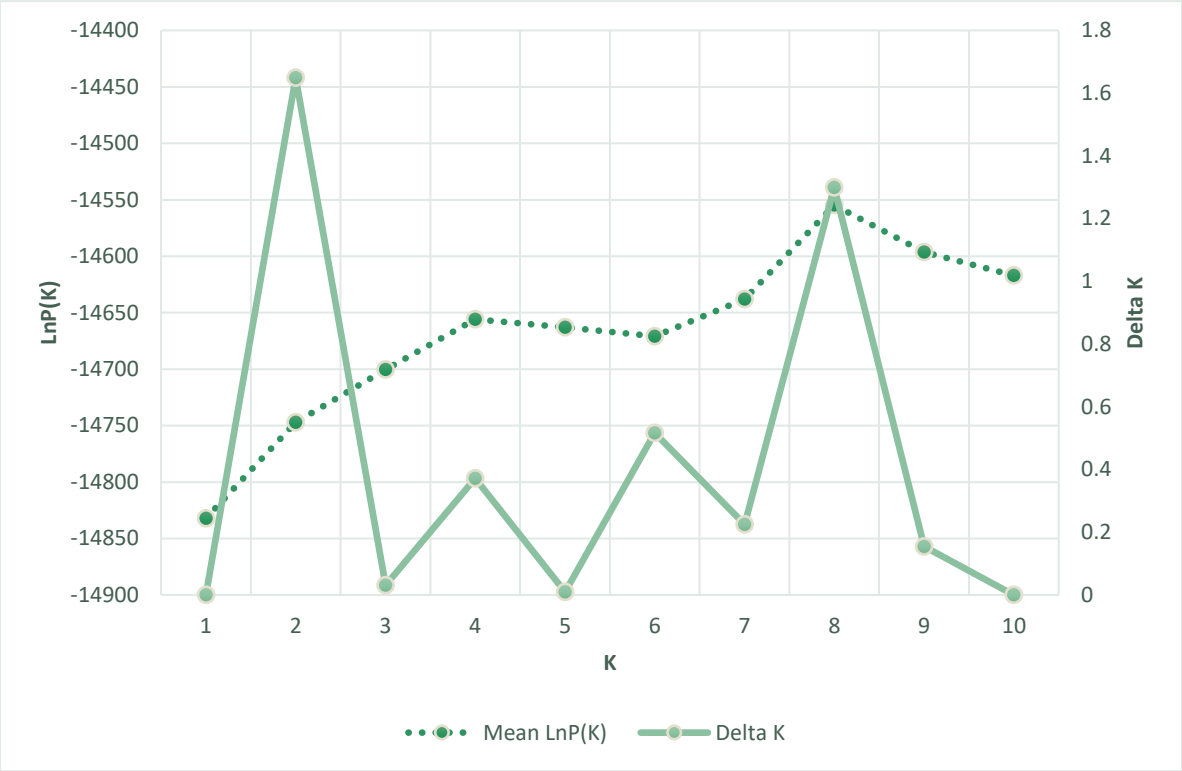


Figure 18. Plot of Delta K and LnP(K) values for varying number of population clusters (k) analyzed in STRUCTURE.

MULTIPLE REGRESSION COVARIATION PLOT

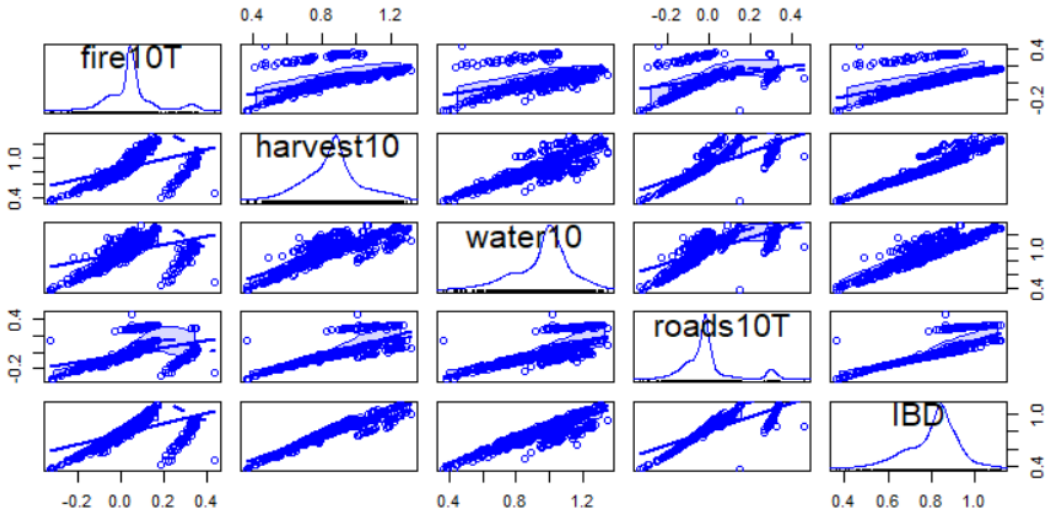


Figure 19. Pairwise correlation plots among predictor variables at a resistance value of 10.

MULTIPLE REGRESSION ANOVA TABLE

Statistically significant multivariate regression model

Source of Variation	SS	df	MS	F
Regression	0.007142624	2	0.003571312	24.29341
Residual	0.108491476	738	0.0001470074	
Total	0.1156341	740		

APPENDIX VII

DATA SOURCES

Data is all obtained from Ontario GeoHub

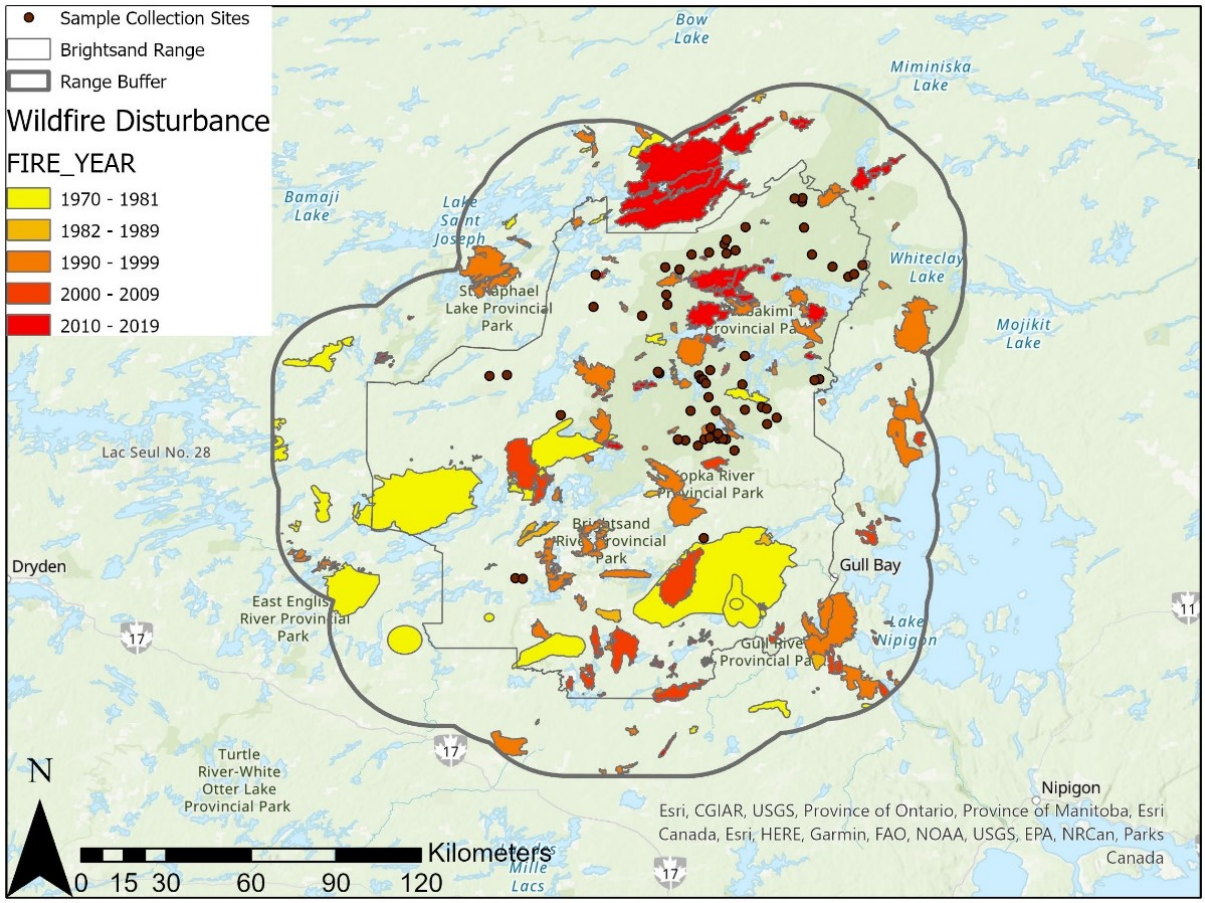
Data Type	Source	Last updated	Download Date
Roads	https://geohub.lio.gov.on.ca/datasets/mnrf::ontario-road-network-orn-road-net-element/about	May 24, 2022	May 28, 2022
Lakes	https://geohub.lio.gov.on.ca/datasets/mnrf::ontario-hydro-network-ohn-waterbody/explore	May 17, 2022	May 22, 2022
Fire	https://geohub.lio.gov.on.ca/datasets/lio::fire-disturbance-area/explore?location=51.075992%2C-91.248619%2C4.73	March 7, 2022	May 24, 2022
FRI	https://geohub.lio.gov.on.ca/maps/lio::forest-resources-inventory-packaged-products-version-2/about	January 19, 2022	May 27, 2022
Caribou Range	https://geohub.lio.gov.on.ca/datasets/lio::caribou-range-boundary/about	September 8, 2014	May 22, 2022

FRI YEAR OF SOURCE

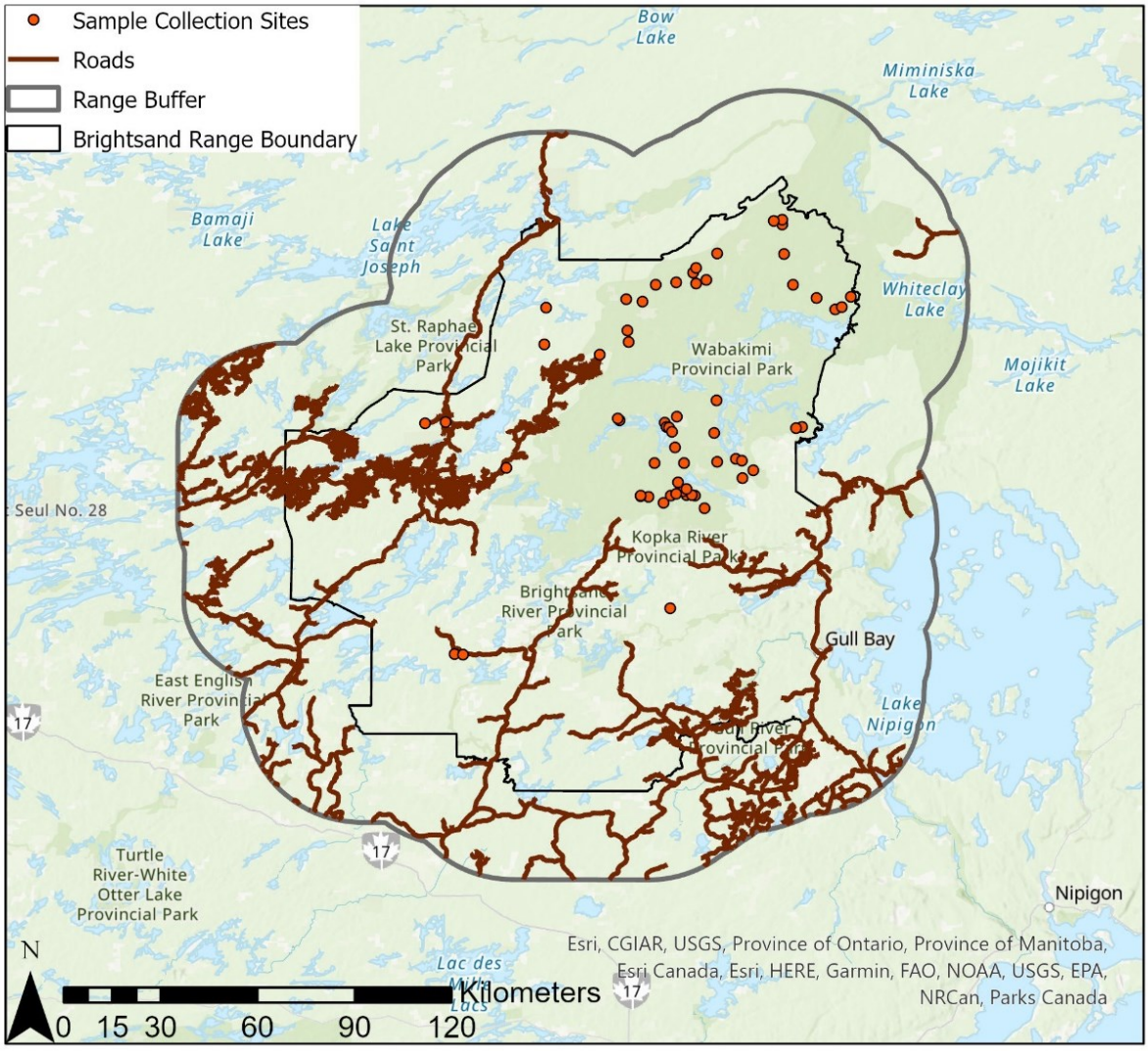
Unit	Year of Oldest Updates	Year of Most Recent Updates	Year of Most Updates (~90%)
Black Spruce Forest	2008	2015	2008
Caribou Forest	2008	2013	2010
Dog River Matawin Forest	2008	2013	2008
English River Forest	2003	2015	2010
Lac Seul Forest	2007	2015	2010-2011
Lake Nipigon Forest	2008	2015	2010
Ogoki Forest	2008	2010	2008
Wabakimi Forest	2008	2013	2008

RESISTANCE LAYERS WITH COLLECTION SITES

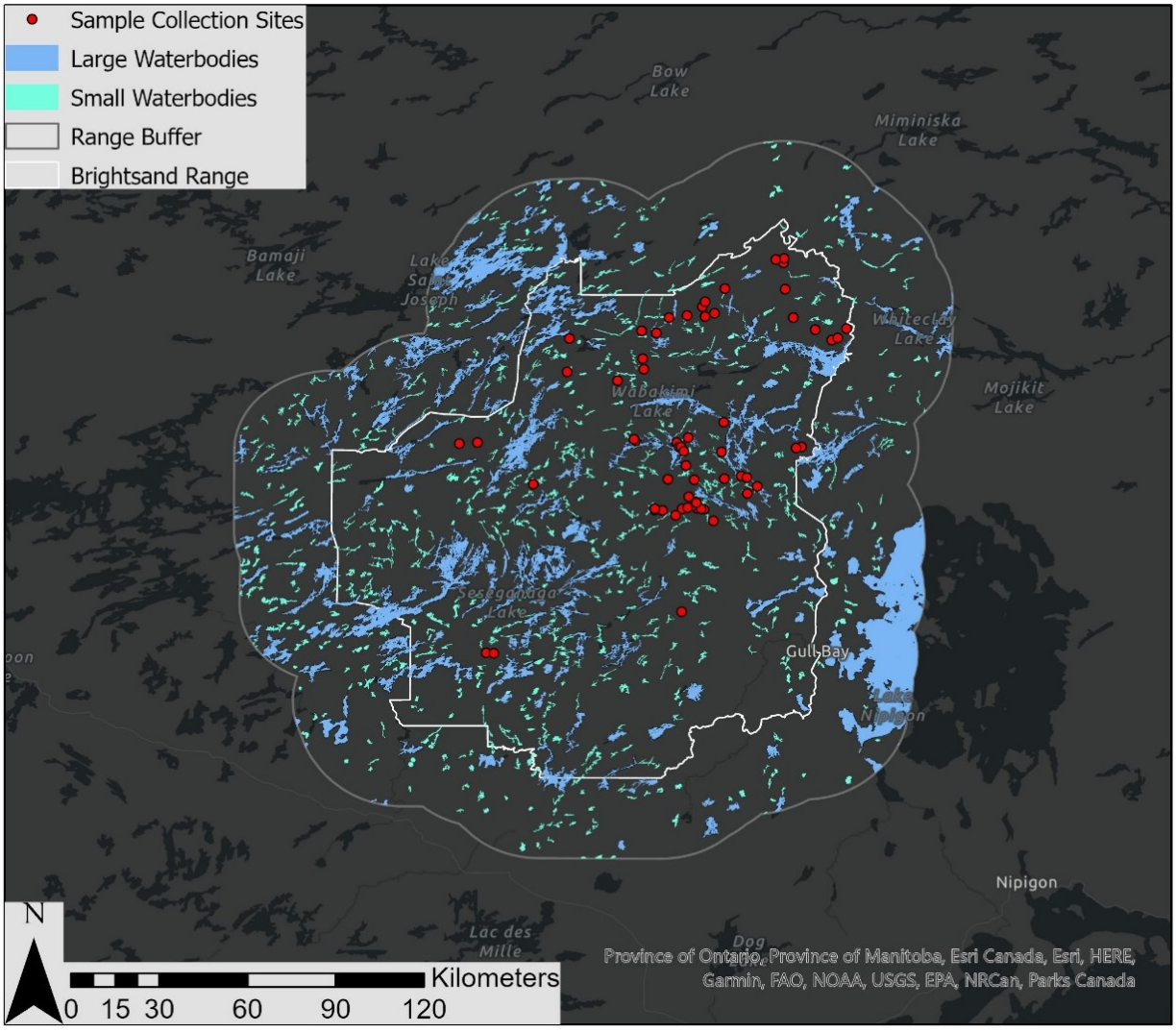
Wildfire



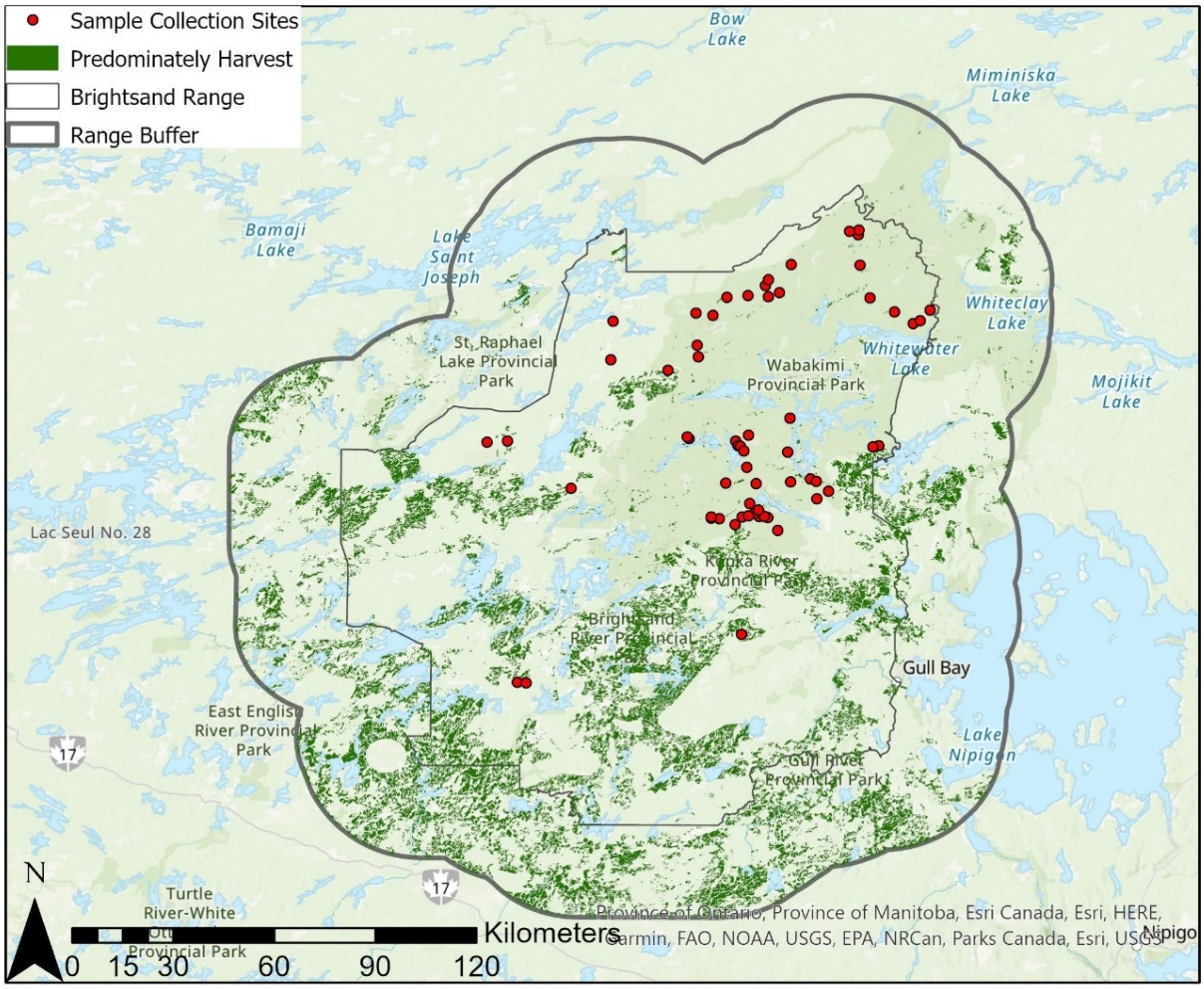
Roads


Sample Collection Sites
Roads
Range Buffer
Brightsand Range Boundary
Bow Lake
Miminiska Lake
Bamaji Lake
Lake Saint Joseph
St. Raphael Lake Provincial Park
Wabakimi Provincial Park
Whiteclay Lake
Mojikit Lake
Seul No. 28
Kopka River Provincial Park
Brightsand River Provincial Park
Gull Bay
East English River Provincial Park
Lake Nipigon
17
Turtle River-White Otter Lake Provincial Park
Nipigon
N
Kilometers
0 15 30 60 90 120
Esri, CGIAR, USGS, Province of Ontario, Province of Manitoba, Esri Canada, Esri, HERE, Garmin, FAO, NOAA, USGS, EPA, NRCan, Parks Canada

Waterbodies



Young Forest (Predominantly Harvest)



APPENDIX VIII

As a firm believer in data sharing I have included additional information and some scripts used for analyses within R Studio. These scripts are for knowledge and may not be the most efficient methods however if these were provided to me when I started it would have made the process easier.

SOFTWARE SOURCES

Software	Type	Source
GENEIOUS PRIME	Program	https://www.geneious.com/
MSATALLELE	R Package	http://alberto-lab.blogspot.com/p/code.html
ALLELEMATCH	R Package	https://cran.r-project.org/web/packages/allelematch/index.html
MICROCHECKER	Program	
POPPR	R Package	https://grunwaldlab.github.io/poppr/
GENEPOP	R Package	https://cran.r-project.org/web/packages/genepop/index.html A web version is also available https://genepop.curtin.edu.au/
GENALEX	Excel Add On	https://biology-assets.anu.edu.au/GenAlEx/Welcome.html
STRUCTURE	Program	https://web.stanford.edu/group/pritchardlab/structure.html
STRUCTURE HARVESTER	Web Program	http://taylor0.biology.ucla.edu/structureHarvester/#
CLUMPAK	Web Program	http://clumpak.tau.ac.il/
TESS	Program	http://membres-timc.imag.fr/Olivier.Francois/tess.html
BAPS	Program	Email Contact due to site maintenance – Jukka Corander

MEMGENE	R Package	https://cran.r-project.org/web/packages/memgene/index.html
POPGRAPH	R Package	https://github.com/dyerlab/popgraph
GSTUDIO	R Package	https://github.com/dyerlab/gstudio
CIRCUITSCAPE	Program	https://circuitscape.org/

ADDITIONAL INFORMATION

MSATALLELE

The data used to determine the bins are based on sample sets 1 through 8 and half set. There are no duplicate sets used.
This would include data on duplicate samples and unknown samples, but it is not used as a comprehensive list of samples for their exact genotypes. It is strictly for the purpose of identify the bin distributions and naming. The data files used to determine these bins cannot be used for any analyses related to the study.
I investigated the apparent outliers within the allele peak fragment sizes and there was no indication that those points were scored incorrectly and were just unique alleles among the remainder of samples.

MSATALLELE Script M3 Example

```
library(MsatAllele)

#importing each file

Set1=read.frag.sizes("THO19708_Set1_M3.txt","Dec", "Set1")
Set2=read.frag.sizes("THO19710_Set2_M3.txt", "Dec", "Set2")
Set3=read.frag.sizes("THO19709_Set3_M3.txt", "Jan", "Set3")
Set4=read.frag.sizes("THO19711_Set4_M3.txt", "Jan", "Set4")
Set5=read.frag.sizes("THO19453_Set5_M3.txt", "Jan", "Set5")
Set6=read.frag.sizes("THO19454_Set6_M3.txt", "Jan", "Set6")
Set7=read.frag.sizes("THO19705_Set7_M3.txt", "Jan", "Set7")
Set8=read.frag.sizes("THO19706_Set8_M3.txt", "Jan", "Set8")
SetHalf=read.frag.sizes("THO19707_SetHalf_M3.txt", "Jan", "SetHalf")



#combining each file to use as a single database

DataBase=rbind(Set1, Set2, Set3, Set4, Set5, Set6, Set7, Set8, SetHalf)
```

```
DataBase

#Allele Distribution
AlleleCum(DataBase, "Map2C", ymin=NULL, ymax = NULL, c1="darkblue",
c2="lightblue", ytsize=1, psize = 1, pch = 1)
AlleleCum(DataBase, "RT1", ymin=NULL, ymax = NULL, c1="darkblue",
c2="lightblue", ytsize=1, psize = 1, pch = 1)
AlleleCum(DataBase, "BM888", ymin=NULL, ymax = NULL, c1="darkgreen",
c2="lightgreen", ytsize=1, psize = 1, pch = 1)
AlleleCum(DataBase, "BM848", ymin=NULL, ymax = NULL, c1="Black", c2="grey",
ytsize=1, psize = 1, pch = 1)
AlleleCum(DataBase, "RT30", ymin=NULL, ymax = NULL, c1="darkred", c2="red",
ytsize=1, psize = 1, pch = 1)

##Bin Statistics
BinStat(DataBase,"Map2C")
BinStat(DataBase,"RT1")
BinStat(DataBase,"BM888")
BinStat(DataBase,"BM848")
BinStat(DataBase,"RT30")


#Histograms with code for bin assignment too
AlleleHist(DataBase, "Map2C")

RT6bins<-bin.limits(DataBase, "Map2C")
mark.bins(RT6bins, text.size = 1, yscale = 50, offtext = 0.5)

AlleleHist(DataBase, "RT1")

BM6506bins<-bin.limits(DataBase, "RT1")
mark.bins(BM6506bins, text.size = 1, yscale = 70, offtext = 0.5)

AlleleHist(DataBase, "BM888")

BMS1788bins<-bin.limits(DataBase, "BM888")
mark.bins(BMS1788bins, text.size = 1, yscale = 45, offtext = 0.5)

AlleleHist(DataBase, "BM848")

NVHRT30bins<-bin.limits(DataBase, "BM848")
mark.bins(NVHRT30bins, text.size = 1, yscale = 125, offtext = 0.5)

AlleleHist(DataBase, "RT30")

RT24bins<-bin.limits(DataBase, "RT30")
mark.bins(RT24bins, text.size = 1, yscale = 125, offtext = 0.5)
```

ALLELEMATCH

ALLELEMATCH Script

```
library(allelematch)
setwd("C:/Users/athomson/OneDrive - lakeheadu.ca/Research/MECP caribou/Caribou_R_data")
geno <- read.csv("Allelematch_input_June22.csv")
dataset <- amDataset(geno, indexColumn=1, metaDataColumn=2, missingCode="-99")
amUniqueProfile(dataset, doPlot=TRUE)
uniquegenotypes2 <- amUnique(dataset, alleleMismatch=2, doPsib="all")
summary(uniquegenotypes2, html="Allelematch_june22_12loci.html")
summary(uniquegenotypes2, csv=" Allelematch_june22_12loci.csv")
```

MICROCHECKER

MICROCHECKER is an older program that is challenging to find a reputable source to download the program. In addition, the program does not work on newer windows programs. Therefore, modifications must be made to the computer to get the program to work.

I have included here the source to the program which I used and the source of additional information to use the program. However, caution should be used prior to modifying your computer. I personally used an older computer in case I messed up the operating system.

This is the website used to download the program. I do not by any means suggest this is a safe website to download the program. However, with the age of the program a more reputable source is challenging to locate.

https://micro-checker.software.informer.com/2.2/

I used the following guide to fully install the program. However, as the guide indicates making changes to the operating system are at the risk of the user.

http://biologicallyrelevant.com/misc/microchecker.html

Newer methods are recommended due to the limitations of access. However, I am unsure of other options at this time.

POPPR

The creators of POPPR provide a thorough step by step tutorial for R-scripts with explanations. I thoroughly recommend this guide:

https://grunwaldlab.github.io/Population_Genetics_in_R/index.html

GENEPOP

To obtain the input file for GENEPOP I used GENALEX to export the required data to a GENEPOP format.

A	B	C	D	E	F	G	H	I	J	K	L
12	337	58	10	11	8	5	14	7	6	7	7
			A01	A02	A03	A04	A05	A06	A07	A08	A09
Sample	Site	RT6...1	RT6...2	BM6506...	BM6506...	RT24...1	RT24...2	RT27...1	RT27...2	RT9...1	RT9...2
A01-1	A01	99	107	198	201	217	221	131	150	122	122

To use this file:

```
library(genepop)
inputFile <- "unique_genotypes"
basic_info(inputFile, outputFile = "output_allelefrequencies.txt", verbose = interactive())
```

Hardy-Weinberg Equilibrium and Linkage Disequilibrium function scripts are available in R package documentation.

GENALEX

STRUCTURE

Sample data format. In the second row columns are, sample name, population number (only numbers), Allele A, Allele B for the respective order of the above loci

A	B	C	D	E	F	G	H	I	J	K	L	M	N
RT6	BM6506	RT24	RT27	RT9	RT7	RT5	Map2C	RT1	BM888	BM848	RT30		
A01-1	1	99	107	198	201	217	221	131	150	122	122	225	225

CLUMPAK

STRUCTURE

For structure the output results folder is used and an additional label file defining the names of the populations:

```
1    A01
2    A02
3    A03
4    A04
5    A05
6    A06
7    A07
```

TESS

The R-package POPHELPER was used to assist with formatting the TESS output files for visualization using CLUMPAK. pophelperShiny is also available for use through launching in R. Manual editing of files is also possible.
Script:

```
library(pophelper)
file.exists("c:/")
readQ ()
clumppExport(readQ("pophelper_newdir"), exportpath = choose.dir()) #manually select the output directory
```

MEMGENE

Sample data format:

x	y	RT6a	RT6b	BM6506a	BM6506b	RT24a	RT24b	RT27a	RT27b	RT9a	RT9b	RT7a	RT7b	RT5a	RT5b	MAP2Ca
-6301489	7738521	99	107	198	201	217	221	131	150	122	122	225	225	154	154	95

```
rm(list=ls())

library(memgene)
library(raster)
library(rgdal)
library(adegenet)


## Load the caribou genetic data
caribouData <- read.csv("july7_memgene.csv")

## Create objects for positional information and genotypes
caribouXY <- caribouData[ ,1:2]
caribouGen <- caribouData[, 3:ncol(caribouData)]

## Produce a proportion of shared alleles genetic distance matrix
## using the convenience wrapper function provided with the package
caribouDM <- codomToPropShared(caribouGen)

## Run the MEMGENE analysis
## May take several minutes
if (!exists("caribouAnalysis"))
  caribouAnalysis <- mgQuick(caribouDM, caribouXY)

plot(caribouXY, type="n", xlab="", ylab="", axes=FALSE)
mgMap(caribouXY, caribouAnalysis$memgene[, 3], add.plot=TRUE,
```

```
    legend=TRUE)
box()


## Finding the adjusted R-squared (i.e. the genetic variation explained by spatial
pattern)
caribouAnalysis$RsqAdj

## Find the proportional variation explained by each MEMGENE variable
caribouMEMGENEProp <- caribouAnalysis$sdev/sum(caribouAnalysis$sdev)


## Neatly print proportions for the first three MEMGENE variables
format(signif(caribouMEMGENEProp, 3)[1:3], scientific=FALSE)



##########################################################################
########

##Exploring Landscape Variables

##########################################################################
########


## use the above script to prepare the data then use the below script for
## further landscape analyses


## Prepare landscape resistance models
roads_resistanceMaps <- stack(
  (raster("roads1_cost.asc", package="memgene")),
  (raster("roads10_cost.asc", package="memgene")),
  (raster("roads50_cost.asc", package="memgene")),
  (raster("roads100_cost.asc", package="memgene"))
  )

roads_compareRasters <- mgLandscape(roads_resistanceMaps,
                        caribouDM, caribouXY, euclid=TRUE,
                        forwardPerm=200, finalPerm=500)

print(roads_compareRasters)


fire_resistanceMaps <- stack(
```

```
  (raster("fire1_cost.asc", package="memgene")),
  (raster("fire10_cost.asc", package="memgene")),
  (raster("fire50_cost.asc", package="memgene")),
  (raster("fire100_cost.asc", package="memgene"))
  )

fire_compareRasters <- mgLandscape(fire_resistanceMaps,
                             caribouDM, caribouXY, euclid=TRUE,
                             forwardPerm=250, finalPerm=500)

print(fire_compareRasters)

smallwater_resistanceMaps <- stack(
  (raster("small_waterbodies1_cost.asc", package="memgene")),
  (raster("small_waterbodies10_cost.asc", package="memgene")),
  (raster("small_waterbodies50_cost.asc", package="memgene")),
  (raster("small_waterbodies100_cost.asc", package="memgene"))
  )

smallwater_compareRasters <- mgLandscape(smallwater_resistanceMaps,
                             caribouDM, caribouXY, euclid=TRUE,
                             forwardPerm=250, finalPerm=500)

print(smallwater_compareRasters)

largewater_resistanceMaps <- stack(
  (raster("large_waterbodies1_cost.asc", package="memgene")),
  (raster("large_waterbodies10_cost.asc", package="memgene")),
  (raster("large_waterbodies50_cost.asc", package="memgene")),
  (raster("large_waterbodies100_cost.asc", package="memgene"))
  )

largewater_compareRasters <- mgLandscape(largewater_resistanceMaps,
                                caribouDM, caribouXY, euclid=TRUE,
                                forwardPerm=250, finalPerm=500)

print(smallwater_compareRasters)

harvest_resistanceMaps <- stack(
  (raster("predominately_harvest1_cost.asc", package="memgene")),
  (raster("predominately_harvest10_cost.asc", package="memgene")),
  (raster("predominately_harvest50_cost.asc", package="memgene")),
  (raster("predominately_harvest100_cost.asc", package="memgene"))
  )

harvest_compareRasters <- mgLandscape(harvest_resistanceMaps,
```

```
                caribouDM, caribouXY, euclid=TRUE,
                forwardPerm=250, finalPerm=500)

print(harvest_compareRasters)
```

POPGRAPH/GSTUDIO

Extensive tutorials are available for POPGRAPH and GSTUDIO
http://dyerlab.github.io/popgraph/
http://dyerlab.github.io/gstudio/
Do note that I often use the names that are in tutorials in my own scripts.

```
################################################################################
#####
## PopGraph R-code
################################################################################
#####

## loading the package
require(gstudio)
require(popgraph)
require(maps)
require(ggplot2)
require(igraph)
require(sp)
require(sf, Matrix, sampling, methods, magrittr, dplyr)

## process to import and save data as an object

## importing the data
caribou <- read_population( "all_genotypes_PopGraph_non-isolated.csv", type =
"column",
                    locus.columns = 4:27)
caribou

## Saving data file as a raw R object
save(caribou, file = "caribou.rda")

## load data object into R for use from appropriate folder
load("caribou.rda")
caribou

## translating genotypes to multivariate data
mv_caribou<-population_graph(caribou, stratum="Population", alpha = 0.05,
            tol = 1e-04)
mv_caribou
```

```
layout<-layout.fruchterman.reingold(mv_caribou)
plot(mv_caribou, layout=layout)

## basic popgraph output
plot(mv_caribou)


## turning an igraph object into a popgraph object
pop.g_caribou<-as.popgraph(mv_caribou)
class(pop.g_caribou)

## combining igraph with dataframe for geographic data
# lopho<-decorate_graph(mv_caribou, caribou, stratum = "Population")
baja<-read.csv("baja.csv")
summary(baja)
lopho<-decorate_graph(mv_caribou,baja, stratum = "Population")


## plotting igraph into coordinate map
require(ggplot2)
p <- ggplot()
p <- p + geom_edgeset( aes(x=Longitude,y=Latitude), lopho )
p

p <- p +  geom_nodeset( aes(x=Longitude, y=Latitude), lopho, size=4)
p


p <- ggplot() + geom_edgeset( aes(x=Longitude,y=Latitude), lopho, color="darkgrey" )
p <- p + geom_nodeset( aes(x=Longitude, y=Latitude, color=Region, size=size), lopho)
p <- p + xlab("Longitude") + ylab("Latitude")
p


c <- layout.fruchterman.reingold( lopho )
V(lopho)$x <- c[,1]
V(lopho)$y <- c[,2]
p <- ggplot() + geom_edgeset( aes(x,y), lopho, color="darkgrey" )
p <- p + geom_nodeset( aes(x, y, color=Region, size=size), lopho)
p + theme_void()
```