GENETIC NETWORKS TO INVESTIGATE STRUCTURE AND CONNECTIVITY OF CARIBOU AT MULTIPLE SPATIAL AND TEMPORAL SCALES

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

TRENT UNIVERSITY

Peterborough, Ontario, Canada

© Copyright by Cory Fournier 2022

Environmental and Life Sciences M.Sc. Graduate Program

September 2022

ABSTRACT

Genetic Networks to Investigate Structure and Connectivity of Caribou at Multiple Spatial and Temporal Scales

Cory Fournier

Understanding genetic structure, connectivity, and movement of a species is critical to management and conservation. Genetic network approaches allow the analysis of genetic information with flexibility and few prior assumptions. In chapter one, I tested the ability of individual-based genetic networks to detect fine-scale structure and connectivity in relation to sampling efforts. My findings revealed individual-based genetic networks can detect fine-scale genetic structure of caribou when using 15 highly variable microsatellite loci. Sampling levels less than 50% of the estimated population size resulted in highly disconnected networks which did not allow for accurate structure analysis; however community detection algorithms were robust in grouping closely related individuals despite low sampling. In chapter two, I used individual-based and population-based genetic networks to investigate structure, connectivity, and movement of caribou across a large study area in Western Canada. A community detection algorithm partitioned the population-based genetic network at multiple spatial scales which uncovered patterns of hierarchical genetic structure and highlighted patterns of gene flow. The hierarchical population structure results aligned with the known distribution of different caribou Designatable Units (DUs) and additional structure was found within each DU. Furthermore, individual-based networks that were constructed with a subset of samples from the Mackenzie Mountains region of the Northwest Territories revealed patterns of long-distance movement and high connectivity across the region.

Keywords: Genetic Networks, Caribou, Community Detection, Structure, Connectivity

Acknowledgements

I would like to extend gratitude to Micheline Manseau for her incredible guidance and unwavering generosity in sharing her knowledge and expertise throughout the past two years of graduate studies. I would also like to thank Paul Wilson and Deborah Simmons who continuously extended help with the conservation genomics and cross cultural collaboration aspects of this research, respectively. Leon Andrew and the Sahtu Renewable Resources Board (ʔehdzo Got'ı̨nę Gots'é Nákedı: SRRB) proved to be instrumental in facilitating this research and coordinating an educational experience that reached far beyond the norm for graduate studies, one which incorporated multiple forms of knowledge and ways of knowing, which allowed me to form questions of importance and ultimately create a thesis that will have implications far beyond the scientific community alone. This thesis incorporates samples that were funded, collected, and analyzed by many organizations including: the SRRB, Trent University, Environment and Climate Change Canada, Parks Canada, the Government of the Northwest Territories, the Government of Alberta, and the Yukon Government. Lastly, and most importantly, I would like to thank my partner, Devyn, and my family (Mom, Tiff, and Amelia) for always supporting me through all my endeavors, without hesitation.

Table of Contents

List of Figures

List of Tables

General Introduction

Increasing landscape disturbance, fragmentation and isolation caused by anthropogenic activities and climate change have resulted in growing concerns over the functional connectivity and genetic erosion of animal populations and thus their long-term viability (Fischer & Lindenmayer, 2007; Leroy et al., 2018; Mimura et al., 2017). It is essential to understand the functional connectivity of populations and dispersal of individuals to effectively manage and ensure their long-term viability (Baguette et al., 2013; Leroy et al., 2018; Van Dyck & Baguette, 2005).

Caribou in the Sahtu region of the Northwest Territories (NWT) span three Designatable Units defined by COSEWIC based on each being discrete and evolutionarily significant (COSEWIC, 2011): Barren-ground, Boreal, and Northern Mountain. Dene and Métis people have a long, interconnected history with the land and biodiversity of the area, with which comes an immense and in-depth knowledge based within their own worldviews (Andrews, MacKay, & Andrew, 2012; Andrews, MacKay, Andrew, et al., 2012; Polfus et al., 2017; Polfus et al., 2016) Dene knowledge holders refer to boreal woodland caribou in central NWT as tǫdzı and distinguish them from mountain caribou (shúhta ʔepę́) and barren-ground caribou (ʔekwę́) based on behavior, habitat selection, and morphology despite significant range overlap (Polfus et al., 2016). Additionally, Caribou in the northern extent of the boreal forest have a different evolutionary history than boreal caribou from regions found further south (Polfus et al., 2017). Given the complexity of the caribou population arrangements in the Northwest Territories and the high cost of sample collection due to the remoteness and relatively large population sizes of the area, novel methods to investigate structure, connectivity,

and movement of caribou that can be based on non-invasive sampling, and require relatively small proportions of the population sampled would be of great benefit.

There are many methods in population genetics to determine population genetic variation within and amongst populations (Bradburd & Ralph, 2019; Scribner et al., 2005) however many of these methods rely on a priori assumptions of population arrangements and do not always allow for the representation of complex genetic variation amongst all individuals/populations simultaneously (Dyer & Nason, 2004). Genetic networks have been developed to specifically address these concerns. Genetic networks analyse genetic variation across the landscape and can be used to answer questions of movement and connectivity with great flexibility and relatively few prior assumptions (Dyer & Nason, 2004; Jones & Manseau, 2022). Genetic networks can be constructed at the individual or population level and are a series of nodes (populations or individuals) connected by edges which represent genetic relationships or similarity between the nodes (Jones & Manseau, 2022). There have been recent studies utilizing genetic network approaches to analyse genetic relationships across landscapes at both a population level (Galpern et al., 2011; Garroway et al., 2008; Thompson et al., 2019) and an individual-based level (Draheim et al., 2016; Escoda et al., 2019; Escoda et al., 2017; Rollins et al., 2012). Despite the increasing use and growing interest of individual-based genetic network methods in population genetics and conservation biology, there has yet to be published studies examining the limitations and robustness of individual-based genetic networks, specifically in relation to sampling effort.

This thesis is written as two chapters. In the first chapter, I explored the limitations and robustness of individual-based genetic networks to detect fine-scale structure and connectivity in response to reduced sampling efforts. In the second chapter,

I used a combination of population-based genetic networks, individual-based genetic relatedness networks, and pedigree networks to uncover spatial patterns of population genetic structure and movement of caribou spanning the Northern Mountain, Boreal, and Barren-ground Designatable Units (DUs) in western Canada, with specific focus on the Mackenzie Mountains Region (Northern Mountain DU) of the Northwest Territories.

References

Andrews, T. D., MacKay, G., & Andrew, L. (2012). Archaeological Investigations of Alpine Ice Patches in the Selwyn Mountains, Northwest Territories, Canada. *Arctic*, *65*, 1-21.

Andrews, T. D., MacKay, G., Andrew, L., Stephenson, W., Barker, A., Alix, C., & Shuhtagot'ine Elders, T. (2012). Alpine Ice Patches and Shuhtagot'ine Land Use in the Mackenzie and Selwyn Mountains, Northwest Territories, Canada. *Arctic*, *65*, 22-42.

Baguette, M., Blanchet, S., Legrand, D., Stevens, V. M., & Turlure, C. (2013). Individual dispersal, landscape connectivity and ecological networks. *Biological Reviews*, *88* (2), 310-326. https://doi.org/10.1111/brv.12000

Bradburd, G. S., & Ralph, P. L. (2019). Spatial Population Genetics: It's About Time. *Annual Review of Ecology, Evolution, and Systematics*, *50* (1), 427-449. https://doi.org/10.1146/annurev-ecolsys-110316-022659

COSEWIC. (2011). Designatable Units for caribou (*Rangifer tarandus*) in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. 88pp.

Draheim, H. M., Moore, J. A., Etter, D., Winterstein, S. R., & Scribner, K. T. (2016). Detecting black bear source-sink dynamics using individual-based genetic graphs. *Proceedings of the Royal Society B-Biological Sciences*, *283*(1835), 9, Article 20161002. https://doi.org/10.1098/rspb.2016.1002

Dyer, R. J., & Nason, J. D. (2004). Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, *13*(7), 1713-1727. https://doi.org/10.1111/j.1365-294X.2004.02177.x

Escoda, L., Fernandez-Gonzalez, A., & Castresana, J. (2019). Quantitative analysis of connectivity in populations of a semi-aquatic mammal using kinship categories and network assortativity. *Molecular Ecology Resources*, *19*(2), 310-326. https://doi.org/10.1111/1755-0998.12967

Escoda, L., Gonzalez-Esteban, J., Gomez, A., & Castresana, J. (2017). Using relatedness networks to infer contemporary dispersal: Application to the endangered mammal Galemys pyrenaicus. *Molecular Ecology*, *26*(13), 3343-3357. https://doi.org/10.1111/mec.14133

Fischer, J., & Lindenmayer, D. B. (2007). Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, *16*(3), 265-280. https://doi.org/10.1111/j.1466-8238.2007.00287.x

Galpern, P., Manseau, M., & Fall, A. (2011). Patch-based graphs of landscape connectivity: A guide to construction, analysis and application for conservation. *Biological Conservation*, *144*(1), 44-55. https://doi.org/10.1016/j.biocon.2010.09.002

Garroway, C. J., Bowman, J., Carr, D., & Wilson, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, *1*(4), 620-630. https://doi.org/10.1111/j.1752-4571.2008.00047.x

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

Leroy, G., Carroll, E. L., Bruford, M. W., Dewoody, J. A., Strand, A., Waits, L., & Wang, J. (2018). Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evolutionary Applications*, *11*(7), 1066-1083. https://doi.org/10.1111/eva.12564

Mimura, M., Yahara, T., Faith, D. P., Vázquez-Domínguez, E., Colautti, R. I., Araki, H., Hendry, A. P. (2017). Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications*, *10*(2), 121-139. https://doi.org/10.1111/eva.12436

Polfus, J. L., Manseau, M., Klütsch, C. F. C., Simmons, D., & Wilson, P. J. (2017). Ancient diversification in glacial refugia leads to intraspecific diversity in a Holarctic mammal. *Journal of Biogeography*, *44*(2), 386-396. https://doi.org/10.1111/jbi.12918

Polfus, J. L., Manseau, M., Simmons, D., Neyelle, M., Bayha, W., Andrew, F., Wilson, P. (2016). Leghagotsenete (learning together): the importance of indigenous perspectives in the identification of biological variation. *Ecology and Society*, *21*(2), 35, Article 18. https://doi.org/10.5751/es-08284-210218

Rollins, L. A., Browning, L. E., Holleley, C. E., Savage, J. L., Russell, A. F., & Griffith, S. C. (2012). Building genetic networks using relatedness information: a novel approach for the estimation of dispersal and characterization of group structure in

social animals. *Molecular Ecology*, *21*(7), 1727-1740. https://doi.org/10.1111/j.1365-294X.2012.05492.x

Scribner, K. T., Blanchong, J. A., Bruggeman, D. J., Epperson, B. K., Lee, C.-Y., Pan, Y.-W., Luukkonen, D. R. (2005). Geographical Genetics: Conceptual Foundations And Empirical Applications Of Spatial Genetic Data In Wildlife Management. *Journal of Wildlife Management*, *69*(4), 1434-1453, 1420.

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

Van Dyck, H., & Baguette, M. (2005). Dispersal behaviour in fragmented landscapes: Routine or special movements?. *Basic and Applied Ecology*, *6*(6), 535-545. https://doi.org/10.1016/j.baae.2005.03.005

Chapter 1

Limitations and robustness of individual-based genetic relatedness networks in relation to sampling efforts

Cory Fournier

ABSTRACT

With the increasing use of genetic network analysis and graph theoretic approaches to analyse genetic relationships across landscapes, it is critical to examine the robustness of these networks, specifically in relation to reduced sampling efforts. In this chapter I investigated the effects of reduced sampling on the robustness of individual-based genetic relatedness networks. Genetic resolution and the proportion of the population that was sampled were two key factors that impacted the ability to construct meaningful and robust individual-based networks. When sampling size was reduced, individual-based networks became increasingly disconnected, impeding the ability to analyse the network using node-based or network-wide metrics. Although the number and size of communities detected in individual-based networks decreased with sample size, community detection algorithms appeared to be robust in consistently clustering closely related nodes, alluding to the ability of individual-based networks to detect groups of genetically similar individuals despite limited sample sizes.

Introduction

There are growing concerns over the functional connectivity and genetic erosion of animal populations and thus their long-term viability due to Increasing landscape disturbance, fragmentation and isolation caused by anthropogenic activities and climate change (Fischer & Lindenmayer, 2007; Leroy et al., 2018; Mimura et al., 2017). Gaining a better understanding of functional connectivity, movement, and dispersal of populations to effectively manage and ensure their long-term viability is essential (Baguette et al., 2013; Leroy et al., 2018; Van Dyck & Baguette, 2005).

Genetic analyses have allowed researchers to quantify spatial patterns of genetic variation, dispersal, gene flow, and therefore functional connectivity. Such approaches to determine genetic variation and movement across space and time are referred to as spatial population genetics (Bradburd & Ralph, 2019).

Many traditional population genetic analyses view populations as discrete, non-overlapping groups of individuals however, more continuously distributed species have been more accurately represented as gradients of differentially related individuals (Anderson et al., 2010; Evans & Cushman, 2009; Storfer et al., 2010). Using individual-based, gradient representations of populations improves conservation of animal species in complex landscapes (Segelbacher et al., 2010). Furthermore, individual-based landscape genetic methods can be better suited than population-based methods to measure contemporary genetic structure and connectivity (Landguth et al., 2010) and they require less prior knowledge or assumptions of (sub)populations.

Genetic Relatedness Networks

The increased accessibility of genetic data from wild populations has resulted in more studies using network analyses and graph theoretic approaches to analyse genetic relationships across landscapes (Jones & Manseau, 2022) . These techniques benefit from increased flexibility and require fewer a priori assumptions (Dyer & Nason, 2004; Jones & Manseau, 2022). Networks have been used in population genetic analysis at both a population level (Dyer & Nason, 2004; Galpern et al., 2011; Garroway et al., 2008; Thompson et al., 2019) and an individual level (Draheim et al., 2016; Escoda et al., 2019; Escoda et al., 2017; Rollins et al., 2012).

Individual-based genetic relatedness networks are networks built with individuals as nodes and pairwise genetic relatedness values as weighted edges between the nodes (Jones & Manseau, 2022); these networks will be the focus of this chapter. Since genetic relatedness values can be calculated for each pair of sampled individuals, the network begins as saturated, meaning that there is an edge connecting every pair of nodes in the network. To create useful networks, a biologically significant threshold value is chosen, and the network is pruned back to only include relationships above a specific order (such as parent-offspring and full siblings). After the network is pruned, all edges in the network are weighted equally depending on the relationship order that the values fall into. By treating relatedness edges categorically rather than continuously, uncertainty due to the variability in the relatedness values is mitigated. Individual-based genetic relatedness networks have been used to study dispersal (Escoda et al., 2017), barriers to dispersal (Escoda et al., 2019), and population structure (Greenbaum et al., 2019; Greenbaum et al., 2016). Additionally, similar networks built with pedigrees have been used to study

varying fitness levels (individual and familial) and demographic parameters across the landscape (McFarlane et al., 2021).

With increasing use in conservation ecology, the suite of network metrics that can be used to answer genetic-based research questions is becoming better defined (Jones & Manseau, 2022). In network theory, a community is defined as a set of nodes that are more strongly connected to each other than to nodes outside of the community (Newman, 2006). Furthermore, genetic population structure is usually defined when a population of individuals is more genetically similar to one another than to individuals outside of that population. Therefore, using network theory, one can suggest patterns of genetic population structure by partitioning communities in a network that is built with individuals as nodes, and genetic relatedness values as edges between those nodes (Greenbaum et al., 2019; Greenbaum et al., 2016). Other node-based metrics such as centrality measures (e.g. degree, eigenvector, and betweenness) define the connectedness of nodes either directly or indirectly, and depending on the type of graph, can be used to infer fitness and highlight key nodes (individuals, populations, or areas) to gene flow (Garroway et al., 2008; McFarlane et al., 2021; Rozenfeld et al., 2008). Definitions of network terminology can be found in Table 1.1. Genetic relatedness networks can also be used to visualize and analyse contemporary dispersal events both between and within structured populations. By creating networks with edges representing close relationships (i.e. parent-offspring or full siblings), one can infer that a dispersal event occurred in the recent past when two nodes in different localities are connected by an edge (Escoda et al., 2017). The assortativity coefficient (ranging from -1 to 1) can also be used to analyse how a discrete node attribute affects the tendency for nodes to share connections (Newman, 2003). If the assortativity coefficient is above 0, there are more edges between nodes that

share a specific characteristic than nodes that do not. Likewise, if the assortativity coefficient is below zero, there are fewer edges between nodes that share a specific characteristic than nodes that do not share that characteristic (Newman, 2003).

Despite the increasing use and growing interest of individual-based genetic network methods in population genetics and conservation biology, there has yet to be published studies examining the robustness of these networks, specifically in relation to reduced sampling efforts. In this study, I aimed to examine the effects of reduced sampling on the robustness of individual-based genetic relatedness networks. Specifically, I looked at the distribution of three node-based centrality metrics, the distribution of the number of communities detected, and the overall community partitioning and alignment for individual-based genetic relatedness networks built with full and subsampled datasets of two populations of caribou.

Methods

Study Areas and Data Sets

Two highly sampled populations were selected to be used in this study. Both populations had a high proportion of the estimated population sampled over multiple years. All samples were collected and genotyped at 15 variable microsatellite loci (BM848, BM888, FCB193, MAP2C, NVHRT16, OHEQ, RT1, RT5, RT6, RT7, RT9, RT13, RT24, RT27, RT30) according to methods and protocols outlined in McFarlane et al. (2021) and McFarlane et al. (2018) and references within. Summary statistics (Na, Ne, Ho, He, Fis) were calculated using the program GenAlEx 6.5 (Peakall & Smouse, 2012). All summary statistics (Table A1-A5) and Pairwise Fst values (Table A6) can be found in Appendix A.

The first population was a central mountain caribou population from Jasper, Alberta (hereafter referred to as Jasper) that was sampled annually from 2006 until 2016. Fecal samples were collected during the winter months (between October and January) for ten years consecutively. Surveys were flown in the alpine habitat, following similar contour lines each year, and were timed to align with the first snow falls to sample animals before they moved below the tree line. Prior to sampling, and continuing throughout the duration of the sampling years, the population experienced drastic declines in population size and increased fragmentation into three separate sub-populations (Tonquin, Brazeau, and Maligne) (COSEWIC, 2014; McFarlane et al., 2018). Pairwise Fst was also calculated between the three subpopulations in each of the Jasper datasets.

The second population was a Woodland Caribou population from the Boreal Plain ecozone, specifically from central Saskatchewan (hereafter referred to as SK2 Central). This data set consists of caribou fecal samples collected annually during the winters from 2017 to 2019. The sampling was conducted by systematically flying aerial transects to locate caribou cratering sites and sign (McFarlane et al., 2021). The caribou in this study area tend to be genetically structured based on IBD and IBR, and have more clinal patterns of genetic variation (Priadka et al., 2019).

Subsampling of Data Sets

A complete analysis was done on both full data sets. To mimic a lower sampling effort, I repeated the analysis while reducing the data to a single sampling year. For SK2 Central, I selected the 2017 sampling year since it had the most samples collected and it represented the highest proportion of the estimated population size out of all three years. For the Jasper data set I selected two years, 2006 and 2013. I selected a year near the

beginning of the data collection period and a year near the end to allow for representation of the population during the continued decline in population abundance and increase in fragmentation during the ten years of sampling. For all sub-sampled data sets, I calculated the proportion of the estimated population size that was represented for that respective year (table 1.2). The population abundance estimates (N) were retrieved from McFarlane et al. (2018), and Manseau et al. (2019, Unpublished) for Jasper and SK2 Central, respectively. For each individual sample year I randomly subsampled the data to 75% (when possible), 50%, and 25% of the estimated population abundance for the respective year. The random subsamples of each year were repeated 10 times each. For each dataset I estimated effective population size (Ne) using the bias-corrected method based on linkage disequilibrium (LD) (Hill, 1981; Waples, 2006; Waples & Do, 2010), as implemented in NeEstimator V2.1 (Do et al., 2014).

Table 1.1 Network terminology categorized and with respective definitions.

Term	Category	Definition
Degree Centrality	Node-based Metric	The degree centrality of each node is a basic node attribute that is directly determined by the number of adjacent edges which connect that node with other nodes in the network (Jones & Manseau, 2022).
Betweenness Centrality	Node-based Metric	Betweenness centrality is a measure of the number of shortest paths between all nodes that travel through the given node, meaning nodes with high betweenness are important to the connectivity of the network and act as bridges between otherwise disconnected components of the network (Jones & Manseau, 2022).
Eigenvector Centrality	Node-based Metric	Eigenvector centrality is a measure of influence of a node in a network, with its value representing the sum of its neighbors' centralities (Jones & Manseau, 2022).
Assortativity	Network-wide metric	The tendency for nodes to connect to other nodes that share a specific attribute (Newman, 2003).
Network Density	Network-wide metric	The ratio of the number of edges in the network compared to the number of possible edges, indicating how connected the network is as a whole (Farine & Whitehead, 2015).
Clustering Coefficient	Network-wide metric	The tendency of nodes to cluster together (Bastille-Rousseau et al., 2018).
Community Detection Algorithm	General network term	Community detection algorithms aim to partition the network into communities that are composed of nodes that are more connected to nodes in the same community than to nodes in other communities (De Meo et al., 2011; Newman, 2006).
Network Topology	General network term	Network topology is the arrangement of all elements making up a network, including the nodes and edges (Newman, 2010).

Table 1.2 Full and subsampled datasets analysed. Table displays the number of individuals in each data set, the estimated abundance (N) for the respective year for each data set (calculated from past Capture-Mark-Recapture studies), and the proportion (%) of the estimated abundance that is represented in the dataset.

Data set	Number of Individuals in dataset	Estimated Abundance (N)	Percentage of Estimated Abundance (N)
Jasper			
Complete	250	N/A	
2006	98	98	100
75% of 2006	74	98	75
50% of 2006	49	98	50
25% of 2006	25	98	25
2013	45	69	65
50% of 2013	34	69	50
25% of 2013	17	69	25
SK2 Central			
Complete	209	N/A	
2017	133	181	73
50% of 2017	91	181	50
25% of 2017	45	181	25

Pairwise genetic relatedness resolution – Network Edge Values

For each data set, pairwise genetic relatedness values (Wang, 2002) were calculated for each dyad using the R package, Related (Pew et al., 2015). Using the Related R Package, I then simulated 100 pairs of individuals of four different known relationship types (Parent-Offspring, Full Sibling, Half Sibling, and Unrelated) based on the allele frequencies of each data set. The simulated dyads with known relationships for each data set were then used to compare the degree of resolution that was to be expected for each data set, and how relatedness resolution is expected to change as the proportion of the population that is sampled decreases. The distributions of relatedness values for the four relationship types were analyzed by creating density plots. The proportion of 1st

order, 2nd order, and unrelated pairs of individuals that had relatedness values above and below a specified threshold were calculated for each data set, and for the data sets that were random subsamples, means for the ten random subsamples were calculated.

Individual-based Genetic Relatedness Network Construction

I created individual-based genetic relatedness networks for each data set using individual caribou as nodes and the pairwise genetic relatedness values (Wang, 2002) as the edges between the nodes. The networks were created with the R package igraph (Csardi & Nepusz, 2006). The networks were pruned at 0.4 in order to capture a high proportion of 1st order relationships (parent-offspring and full-siblings) while capturing as few unrelated individuals as possible. The details of this threshold value are discussed in the *Pairwise Genetic Relatedness Resolution Results* section of this chapter.

Node-based Metrics

The degree of connectedness and topology of the networks were analyzed by calculating centrality (degree, eigenvector and betweenness) metrics of each node. Explanations and definitions of node-based metrics and other network terms can be found in Table 1.1. The distributions of the centrality metrics of all nodes of each network were compared by creating histograms for each data set. I also computed the proportion of disconnected nodes in each network, which is the proportion of nodes that do not have any edges connecting them to other nodes in the network (nodes with a degree of 0). Again, for any data set that was randomly subsampled, the mean of ten random subsamples was calculated. The proportion of disconnected nodes were then plotted against the proportion of the estimated population abundance that was represented in the respective data set.

Assortativity Coefficient

The degree of assortative mixing due to the subpopulations was determined for the Jasper networks by calculating the assortativity coefficient with the subpopulation assignment of each node used as the discrete attribute. The assortativity coefficients were calculated using the igraph R package (Csardi & Nepusz, 2006).

Community Detection

For each network, I used the Louvain community detection algorithm (De Meo et al., 2011) that uses network modularity optimization to detect communities of nodes that have more edges amongst the community than with nodes in the rest of the network. I then created histograms to compare and visualize the distribution of community sizes as sampling is reduced. I also calculated the overall number of communities detected for each of the data sets, again, calculating a mean of the ten repetitions for any data set that was randomly subsampled. The number of communities were then plotted against the proportion of the estimated population abundance that was represented in the respective data set.

Community Alignment

In order to test how robust the community detection methods were as sampling is reduced, I compared the community alignment of each subsampled data set against the corresponding full year data set. I compared the community alignment by creating bipartite networks that compared the communities to which all nodes were assigned in the subsampled data set, to the communities that the same nodes were assigned to in the full year data set. An edge in the network connected each node's community in the

subsampled data set to the community of that same node in the full year's data set. The bipartite networks were created using the R package bipartite (Dormann et al., 2009).

Results

Pairwise genetic relatedness resolution – Network Edge Values

The 100 simulated pairs of individuals for each level of relatedness (parent/offspring, full sibling, half sibling, and unrelated) were plotted on a density graph to visualize the distribution of the respective relatedness values for each relationship category (Figure 1.1). Little overlap was shown between first-order (parent/offspring and full siblings) and unrelated relationships; however second order relationships (half siblings) demonstrated significant overlap of both first-order and unrelated relatedness values.

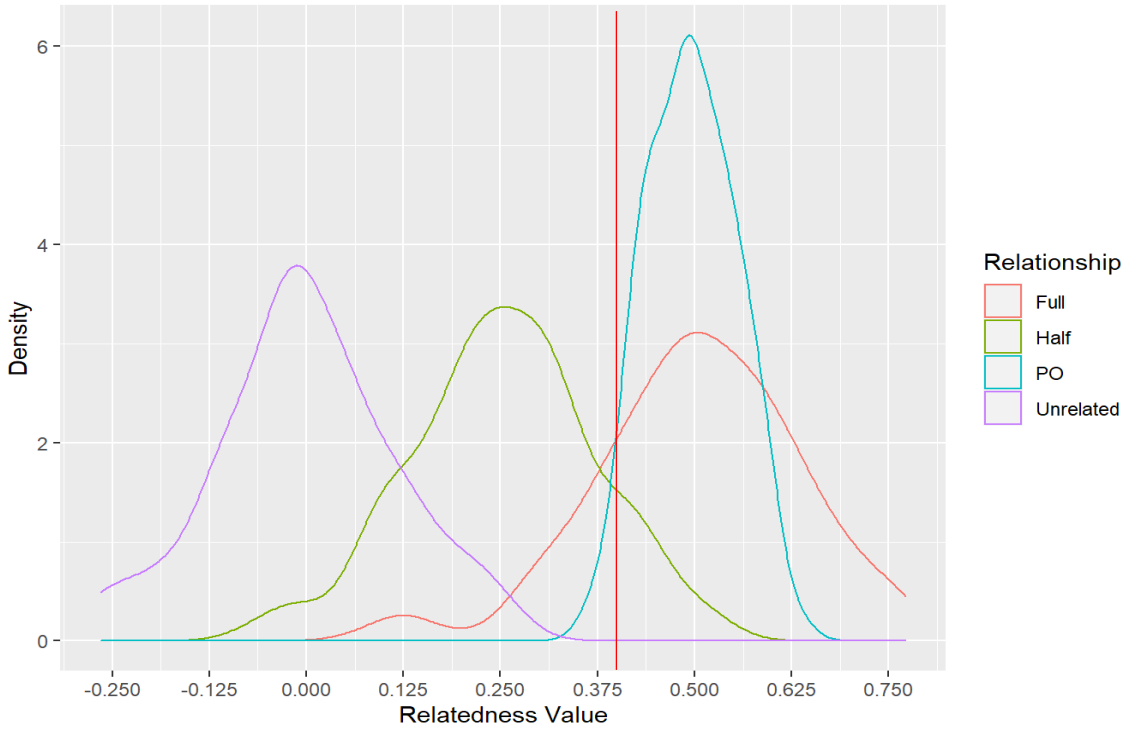


Figure 1.1 Density plot of the relatedness values (Wang, 2002) for the four simulated relationship categories (PO:Parent Offspring, Half: Half-sibling, Full: Full sibling, and Unrelated). A total of 100 pairs of each relationship category were simulated. Simulations were conducted and density plot was created using the R package Related. Red vertical line represents network threshold value of 0.4.

Table 1.3 Proportion of pairs of individuals from each relationship type that fall above and below the 0.4 ıreshold for each full and subsampled data set.

Jasper		Full	2006	2006 - 75% AVG	2005 - 50% AVG	2006 - 25% AVG	2013	2013 - 50% AVG	2013 - 25% AVG
Relationship	>= 0.4 (T or F)	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)
Full	FALSE	23	19	21.4	19.7	20.2	26	19.4	21.5
Half	FALSE	86	94	90.8	91.5	91.6	89	90.3	89.9
Parent-offspring	FALSE	5	6	3	3.3	3.2	2	4.8	3.7
Unrelated	FALSE	100	100	99.9	100	100	100	100	100
Full	TRUE	77	81	78.6	80.3	79.8	74	80.6	78.5
Half	TRUE	14	6	9.2	8.5	8.4	11	9.7	10.1
Parent-offspring	TRUE	95	94	97	96.7	96.8	98	95.2	96.3
Unrelated	TRUE	0	0	0.1	0	0	0	0	0

SK2 Central		Full	2017	2017 - 50% AVG	2017 - 25% AVG
Relationship	>= 0.4 (T or F)	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)
Full	FALSE	19	19	21.6	23.1
Half	FALSE	84	90	87.4	86.1
Parent-offspring	FALSE	7	3	8.1	8.4
Unrelated	FALSE	100	100	99.7	99.7
Full	TRUE	81	81	78.4	76.9
Half	TRUE	16	10	12.6	13.9
Parent-offspring	TRUE	93	97	91.9	91.6
Unrelated	TRUE	0	0	0.3	0.3

The exact proportion of pairs of individuals from each relationship type that were above and below the 0.4 threshold for each full and subsampled data set was calculated (Table 1.3). The proportions of relatedness values above and below 0.4 of each relationship level were very consistent even as the proportion of the population sampled was reduced to 25%. Full sibling relationships were captured at a rate of approximately 80%, parent offspring relationships were captured at rates between 91%-97%, and nearly zero unrelated pairs of individuals fell above the 0.4 threshold. The only slight variation was in the SK2 Central data set, where the proportion of relatedness values above 0.4 increased to a mean of 0.3% for unrelated individuals over the 10 repetitions for the 50% and 25% subsampled datasets (SD 0.6).

Individual-based Genetic Relatedness Network Metrics

Node-based Metrics

Degree Centrality

The distribution of node degrees for all nodes in each of the networks and their respective means were visualized by creating density plots (Figure 1.2.1). For all networks, the distribution of node degrees shifted towards 0 and the means decreased as the proportion of the population sampled was reduced.

The proportion of disconnected nodes (nodes with a degree of 0) in each network was calculated and plotted in a scatter plot against the proportion of the estimated population abundance that was represented in each respective data set (Figure 1.2.2). There was a strong increase in the proportion of disconnected nodes in the network as the proportion of the population sampled decreased. This pattern was seen across both Jasper years (2006 and 2013) as well as the 2017 SK2 Central data set.

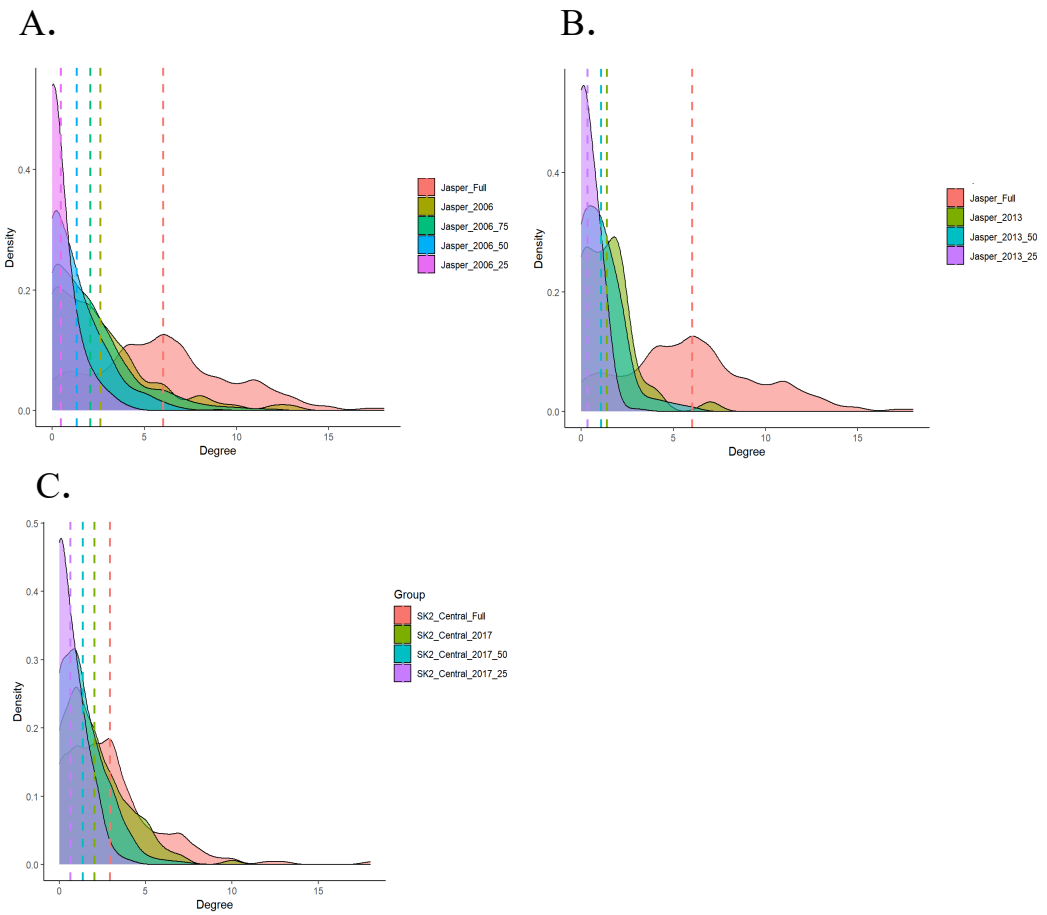


Figure 1.2.1 Density plots showing the distribution of node degrees for all nodes in each network created from the full and subsampled data sets for the (A) 2006 Jasper population, (B) the 2013 Jasper Population, and (C) the 2017 SK2 Central population. Means for each distribution are represented by hashed vertical lines.

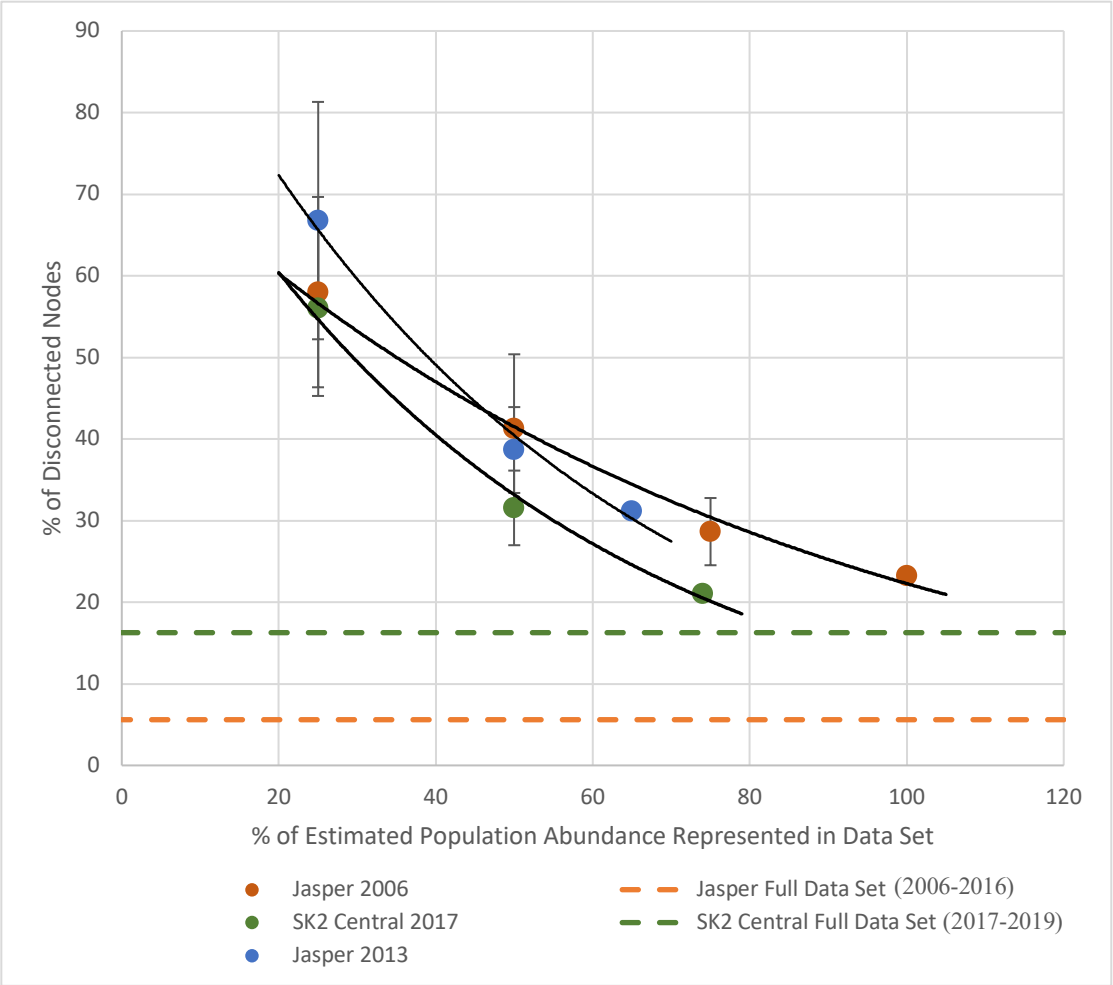


Figure 1.2.2. A plot depicting the proportion (%) of disconnected nodes (nodes with a degree of 0) in relation to the proportion of the estimated population abundance represented in the respective dataset. For randomly subsampled datasets, the mean was calculated for the 10 repetitions of each subsample and the standard deviation is shown as error bars for the respective data points. The proportion of disconnected nodes for the full data sets of Jasper (orange) and SK2 Central (green) are represented by horizontal lines.

Eigenvector Centrality

The distribution of eigenvector centrality for all nodes in each of the networks and their respective means were visualized by creating density plots (Figure 1.3). Similarly to degree centrality, for all networks, the distribution of eigenvector centralities for nodes shifted toward 0 and the means decreased as the proportion of the population sampled was reduced. An exception to this occurred when the proportion of the population that was sampled dropped to 25% for the two Jasper datasets, which resulted in a more random distribution of eigenvector centrality measures, most likely caused by the high degree of partitioning of the network.

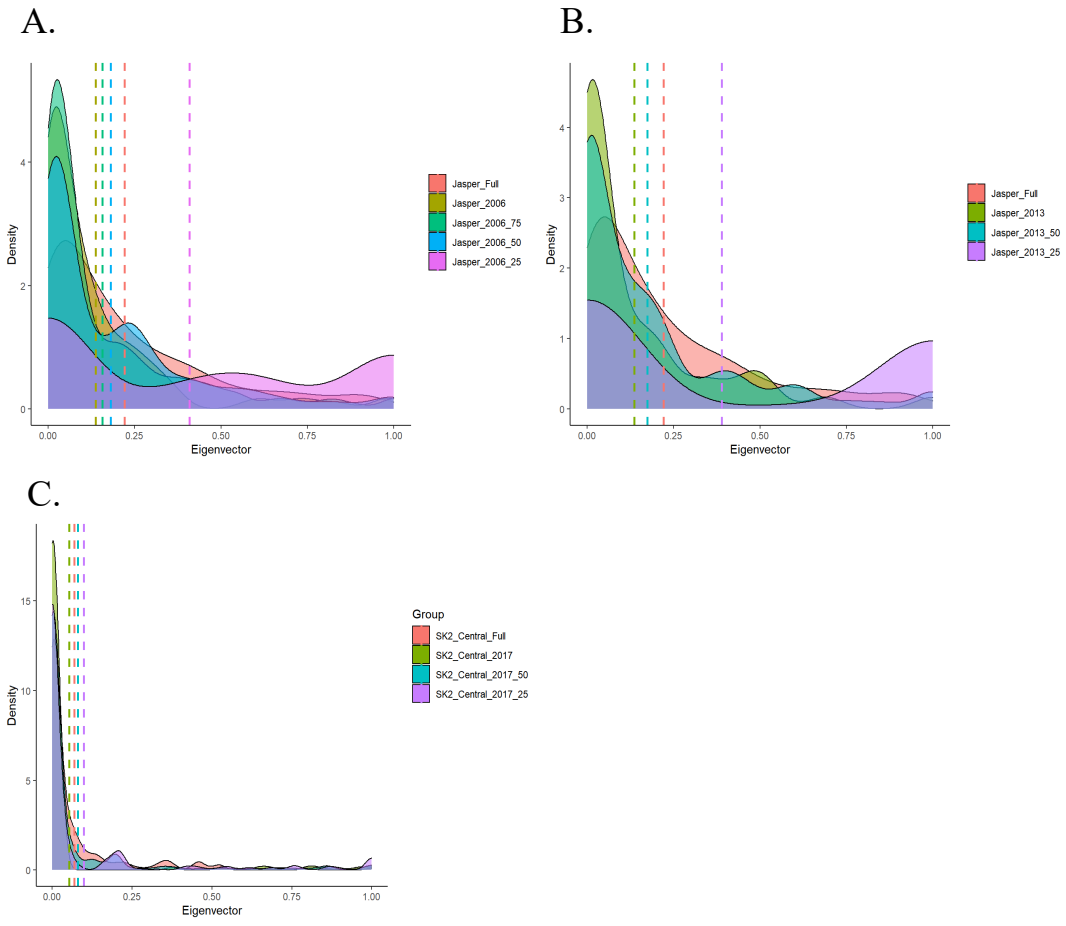


Figure 1.3. Density plots showing the distribution of eigenvector centrality for all nodes in each network created from the full and subsampled data sets for (A) the 2006 Jasper population, (B) the 2013 Jasper Population, (C) and the 2017 SK2 Central population. Means for each distribution are represented by hashed vertical lines.

Betweenness Centrality

The distribution of betweenness centralities for nodes of all networks shifted towards 0 and the means decreased as the proportion of the population sampled was reduced. Notably, in each dataset there appeared to be a threshold of the proportion of individuals sampled where betweenness centrality dropped to nearly 0 for the majority of nodes in the network. This pattern appeared at the 50% range in all three datasets.

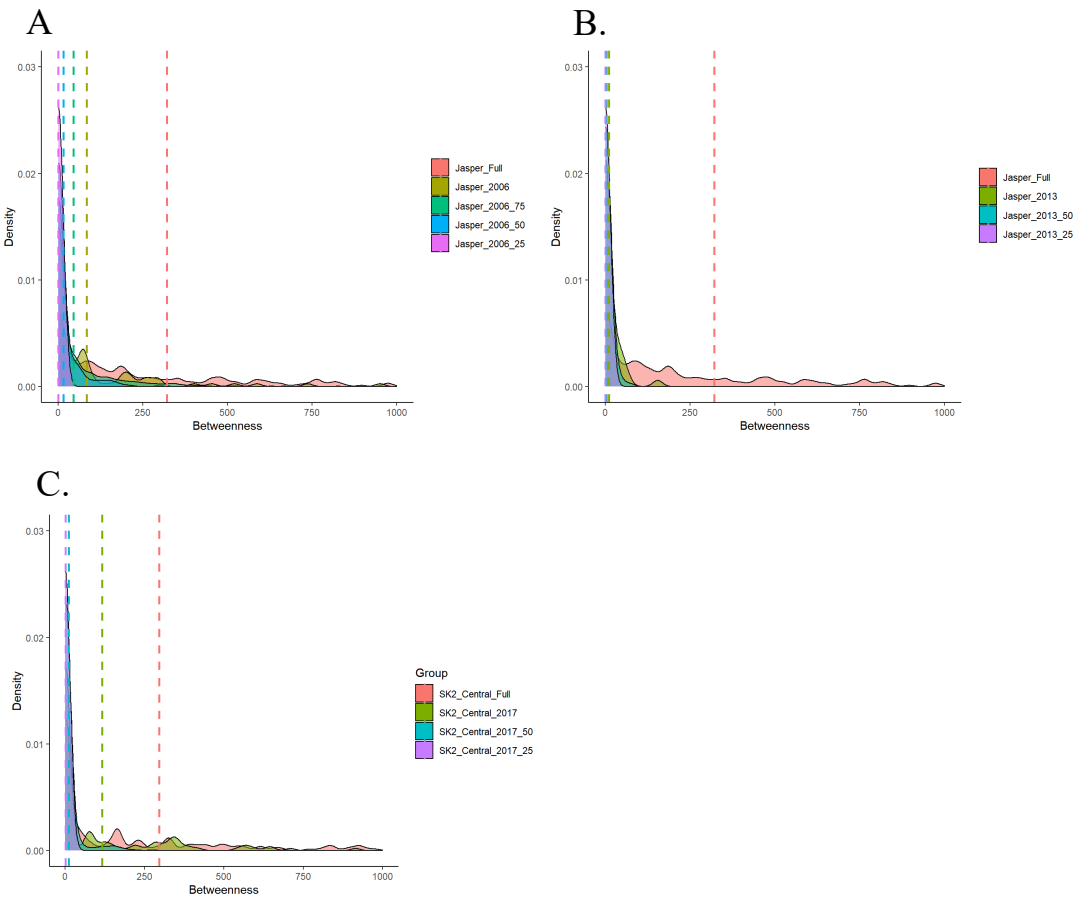


Figure 1.4. Density plots showing the distribution of betweenness centrality for all nodes in each network created from the full and subsampled data sets for (A) the 2006 Jasper population, (B) the 2013 Jasper Population, and (C) the 2017 SK2 Central population. Means for each distribution are represented by hashed vertical lines.

Network-wide metrics

Community Detection

I created histograms to visualize the distribution of number of animals in each community in the various networks created from the full and subsampled data sets (Figure 1.5.1). The distribution of community sizes shifted toward one and the general size of communities decreased as the proportion of the population sampled decreased. This pattern was shown across both the 2006 and 2013 years of the Jasper data as well as the 2017 year of the SK2 Central dataset.

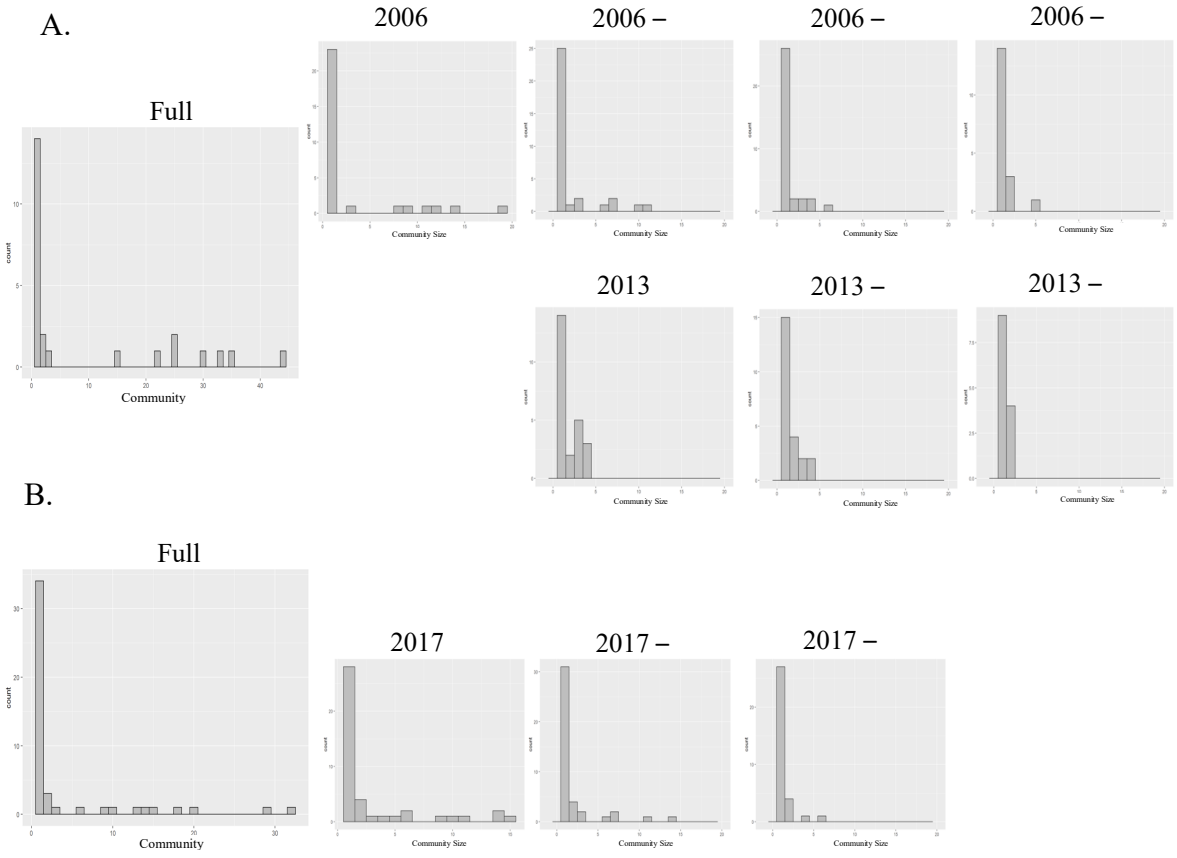


Figure 1.5.1 Histograms depicting the distribution of the number of animals in each community (community size) in the various networks created from the full and subsampled data sets of (A) the Jasper population and (B) the SK2 Central population.

In addition to looking at the distribution of community sizes for each network, I also calculated the overall number of communities detected for each of the data sets, again, calculating a mean of the ten repetitions for any data set that was randomly subsampled. I plotted the number of communities against the proportion of the estimated population abundance that was represented in the respective data set (Figure 1.5.2). There was a general plateau of the number of communities being detected until approximately 50% of the estimated population was sampled, after which as the proportion of the estimated population abundance represented in the data set decreased, the number of communities began to decrease. This pattern was the same across all data sets.

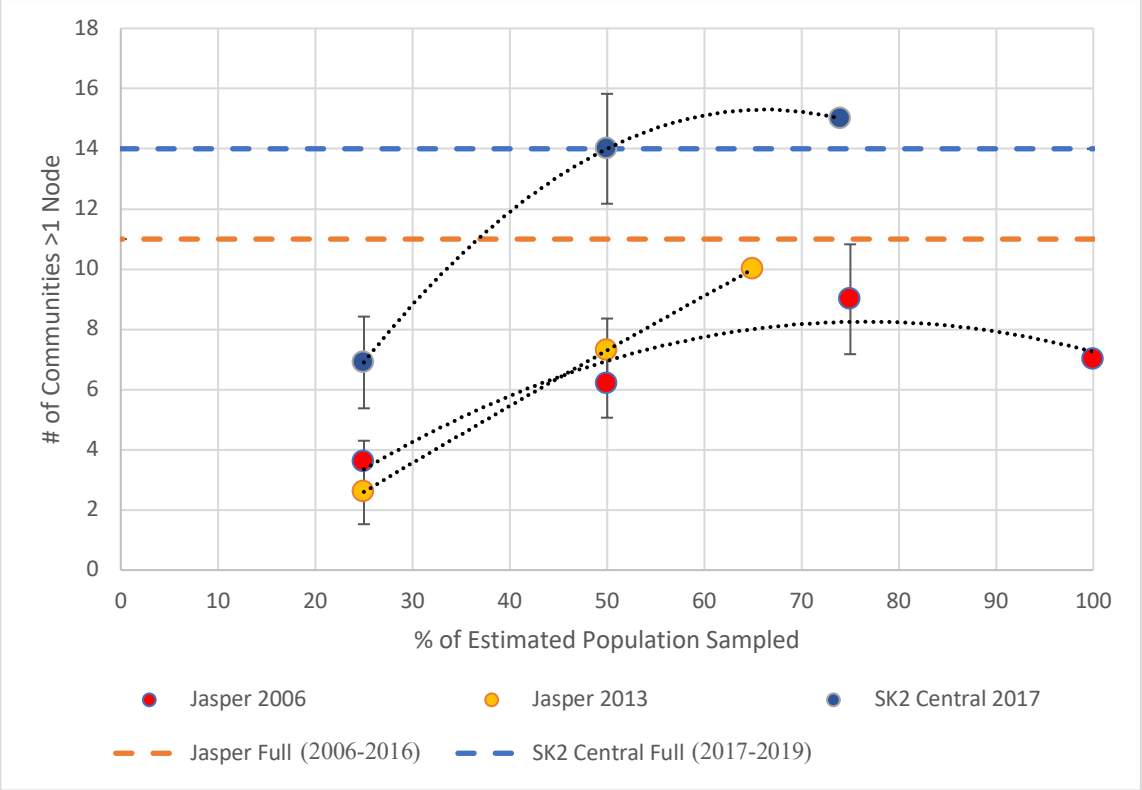


Figure 1.5.2 A plot depicting the number of communities detected by the Louvain community detection algorithm in relation to the proportion of the estimated population abundance represented in the respective dataset. For randomly subsampled datasets, the mean was calculated for the 10 repetitions of each subsample and the standard deviation is shown as error bars for the respective data points. The number of communities detected for the full data sets of Jasper (orange) and SK2 Central (green) are represented by horizontal lines.

Community Alignment

I compared the community alignment by creating bipartite networks that compared the communities to which all nodes were assigned in the subsampled data set, to the communities that the same nodes were assigned to in the full year data set (Figures 1.6.1 – 1.6.3). The majority of nodes were clustered in the same communities even as the sampling was reduced down to 25%. Communities broke up into smaller communities but consisted of nodes that were consistently grouped together. General community alignment across full and subsampled datasets was seen across all years for the Jasper and SK2 Central populations.

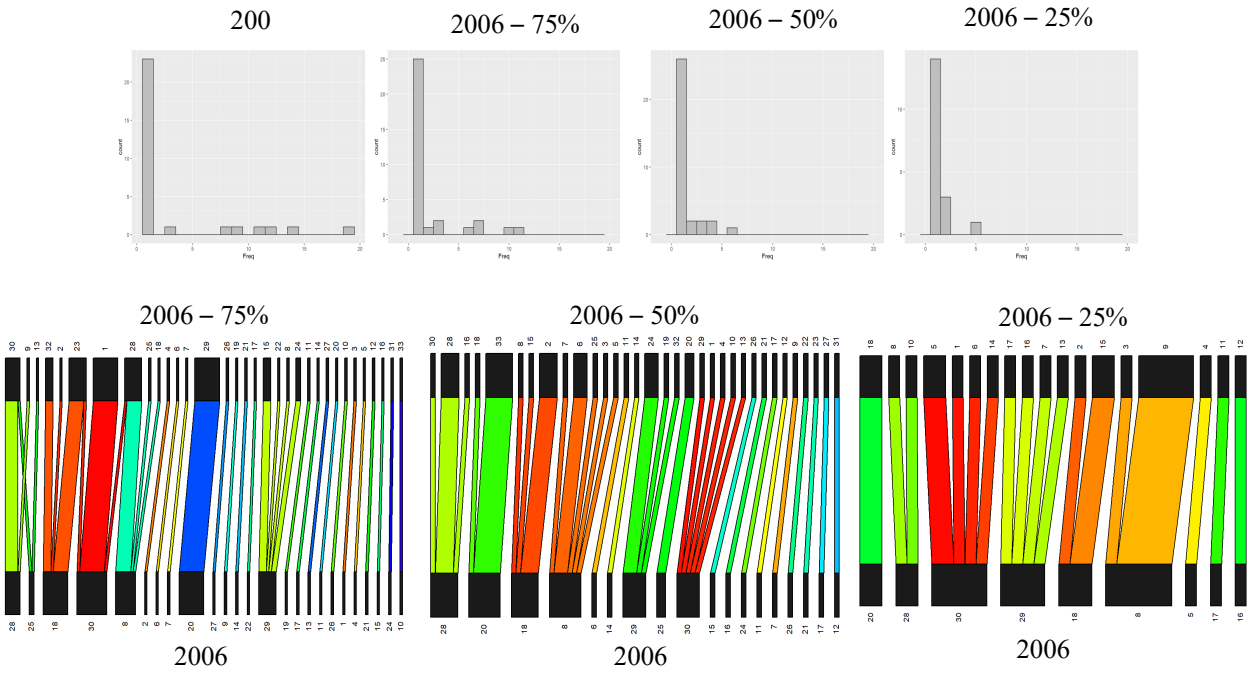


Figure 1.6.1 Bipartite networks depicting community alignment across full and subsampled datasets for the 2006 Jasper year. Squares on the top row of each bipartite network represent the communities that were detected in that subsampled dataset and the squares on the bottom row represent the communities that were detected in the full year data set. The thickness of the squares represent how many individuals are in each community. The edges connecting the rows show where the same nodes were assigned in each data set.

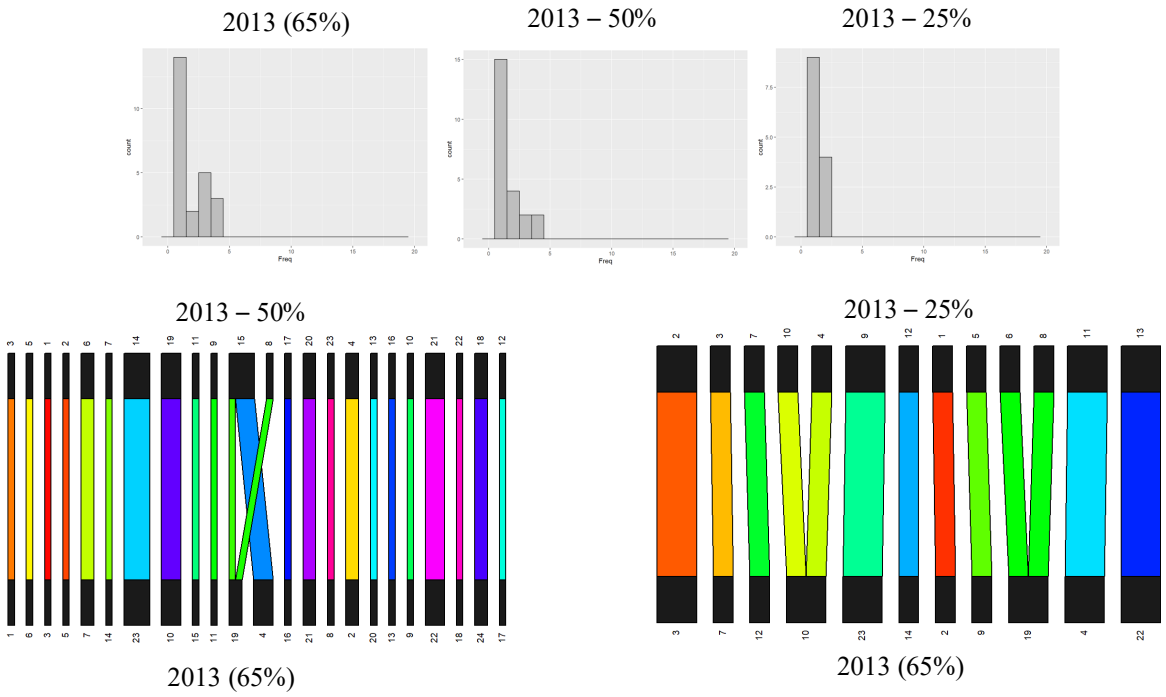


Figure 1.6.2 Bipartite networks depicting community alignment across full and subsampled datasets for the 2013 Jasper year. Squares on the top row of each bipartite network represent the communities that were detected in that subsampled dataset and the squares on the bottom row represent the communities that were detected in the full year data set. The thickness of the squares represent how many individuals are in each community. The edges connecting the rows show where the same nodes were assigned in each data set.

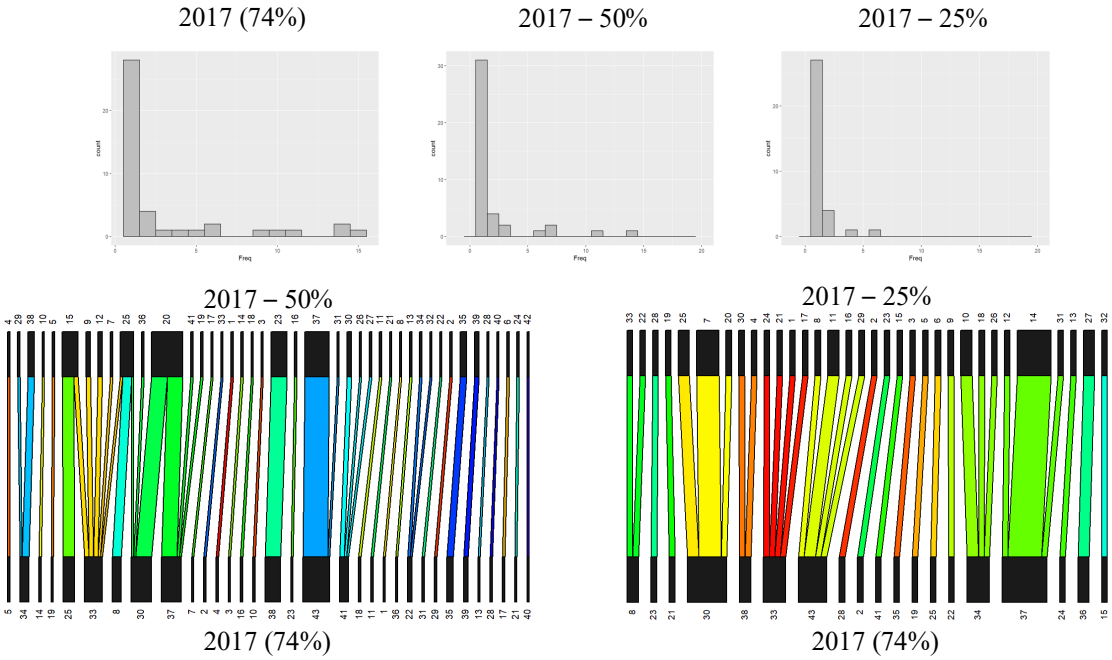


Figure 1.6.3 Bipartite networks depicting community alignment across full and subsampled datasets for the 2017 SK2 Central year. Squares on the top row of each bipartite network represent the communities that were detected in that subsampled dataset and the squares on the bottom row represent the communities that were detected in the full year data set. The thickness of the squares represent how many individuals are in each community. The edges connecting the rows show where the same nodes were assigned in each data set.

Jasper Networks

The networks were visualized both spatially and aspatially to determine if the community partitions created through the Louvain detection algorithm displayed any spatial patterns or if they lined up with the three known subpopulations of Jasper (Figure 1.7.1). As seen in figure 1.7.1, the Tonquin subpopulation was broken up into multiple communities that occupied the same spatial area; the Maligne subpopulation was detected as a single community with fairly strong connections to some individuals in the Tonquin subpopulation; and lastly, the Brazeau subpopulation was also detected as a single community. The same spatial pattern of community detection was seen in the networks created with the 2013 data set (Figure A1 in supplementary material). Although only 7 years prior, the networks created with 2006 Jasper data (Figure 1.7.2) showed a much more connected, and less fragmented overall population, with multiple communities spanning the areas of Brazeau, Tonquin, and Maligne. The assortativity coefficient calculated based on the discrete subpopulation assignment of each node was lower for the 2006 Jasper network (0.470) than for the 2013 Jasper network (0.706) demonstrating that there was more assortative mixing (nodes of the same subpopulation with more connections) for the 2013 Jasper network.

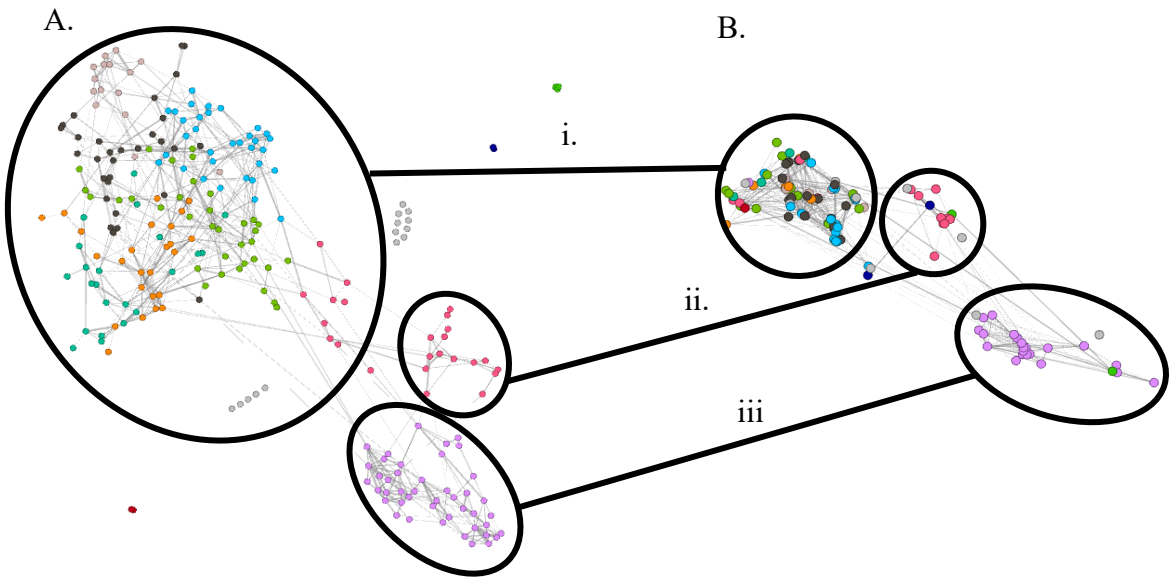


Figure 1.7.1 Individual-based genetic relatedness networks created from the full Jasper data set (2006-2016) represented (A) a-spatially and (B) spatially. Nodes are coloured according to their community assignment as determined by the Louvain community detection algorithm. Black circles are used to depict the three subpopulations: Tonquin (i), Maligne (ii), and Brazeau (iii) across both the spatial and a-spatial networks.

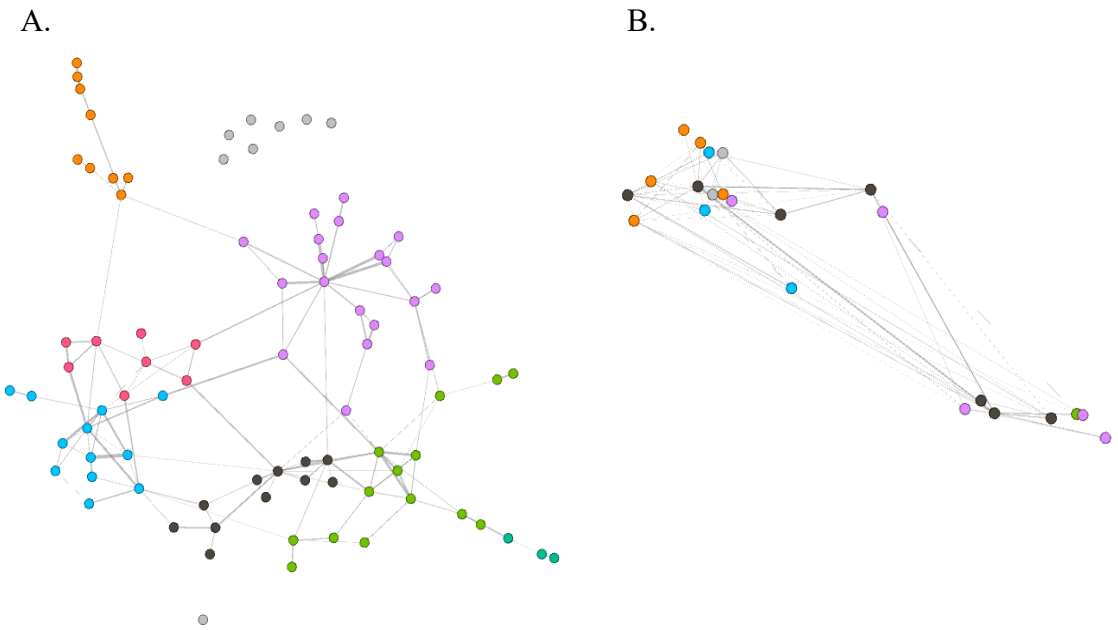


Figure 1.7.2 Individual-based genetic relatedness networks created from the 2006 Jasper data set represented (A) aspatially and (B) spatially. Nodes are coloured according to their community assignment as determined by the Louvain community detection algorithm.

SK2 Central Networks

The networks created from the full SK2 Central data set (Figure 1.8.1) were continuously connected with very few disconnected components and nodes across communities sharing connections. However, the networks still had communities detected in a spatial pattern with two communities in the west being fairly segregated (blue and black communities in Figure 1.8.1). When the data set was reduced to just the 2017 sampling year, a similar community partition pattern was shown (Figure A2 in supplementary material).

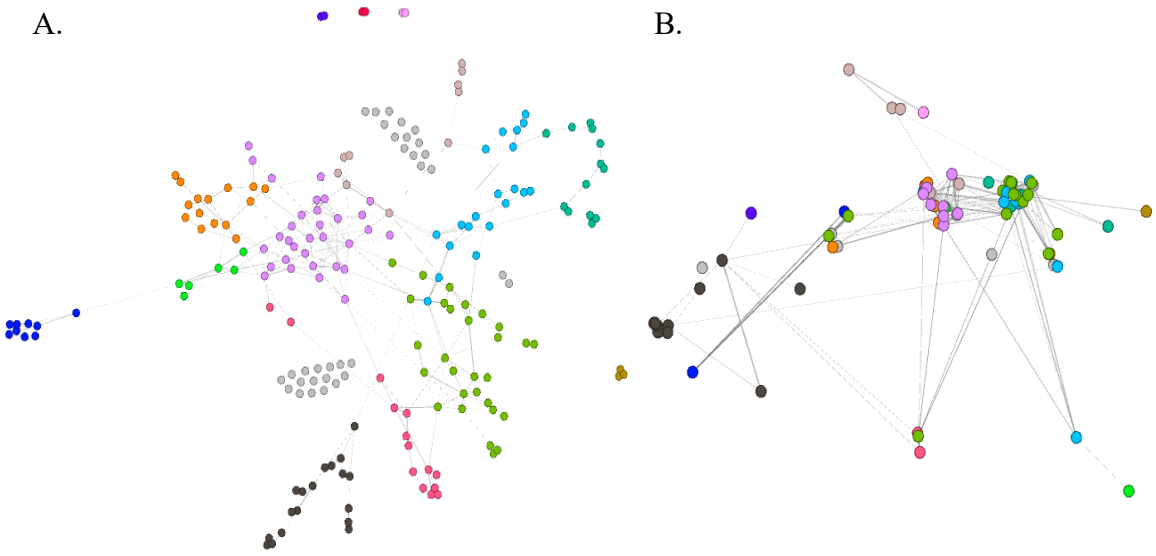


Figure 1.8.1 Individual-based genetic relatedness networks created from the full SK2 Central data set represented (A) a-spatially and (B) spatially. Nodes are coloured according to their community assignment as determined by the Louvain community detection algorithm.

Discussion

Pairwise genetic relatedness resolution – Network Edge Values

The distribution of relatedness values for the simulated pairs of individuals demonstrated little to no overlap between first-order dyads and unrelated dyads for all data sets (Figure 1.1). On closer inspection, the high proportion of first-order relationships that had relatedness values above the threshold of 0.4 and the extremely low number of unrelated individuals above that same threshold (Table 1.3) demonstrated the ability to infer accurate first-order relationships from relatedness values calculated from the 15 microsatellite loci used. This ability was not hindered by reducing the proportion of the population sampled; such robust results at low sample size were most likely due to the overall allele frequencies not changing, even when the proportion of the population that was sampled was reduced down to 25%. Although I confirmed that the resolution of the relatedness values was sufficient to differentiate first-order relationships from unrelated individuals, it was not sufficient to differentiate second order relationships from either first-order relationships or unrelated individuals (Figure 1.1). Perhaps the inclusion of additional loci would result in the resolution being sufficient to capture a higher degree of second order relationships while still maintaining confidence that few to no unrelated dyads would be captured (Foroughirad et al., 2019).

Individual-based Genetic Relatedness Network Metrics

Node Centrality Metrics

Centrality measures directly or indirectly measure how well connected a node is in the network (degree and eigenvector centrality), or how important it is to connecting otherwise disconnected components of the network (betweenness centrality). Node degree

is a basic centrality measure that represents the number of edges that connects a node to other nodes. In individual-based genetic relatedness networks built with relatedness values thresholded to represent first-order relationships, node degree represents the number of close familial relationships of a node in the given network. The distribution of node degrees for all nodes in a network can therefore represent how well connected the network of individuals are. In my analysis, the distribution of node degree centrality values for each network shifted towards zero as the proportion of the population sampled was reduced (Figure 1.2.1) with a negative correlation between the proportion of disconnected nodes and sample size (Figure 1.2.2). This was likely due to the probability of capturing both individuals of a first-order relationship dyad decreasing as the proportion of the population sampled is reduced. Given this correlation, it is important to consider sampling efforts when examining the overall connectedness of a genetic network. Similar patterns were seen with the other two centrality measures, which is to be expected since degree centrality is the most basic and direct centrality measure. Eigenvector centrality is a measure of influence of a node in a network, with its value representing the sum of its neighbors' centralities (Jones & Manseau, 2022). Therefore, eigenvector centrality in an individual-based genetic relatedness network can represent familial fitness of a node. Eigenvector centrality was also greatly affected by sample size, with the distribution shifting towards zero as sample size was reduced. Lastly, betweenness centrality is a measure of the number of shortest paths between all nodes that travel through the given node (Jones & Manseau, 2022). Betweenness centrality had a similar pattern to the other two node-based metrics and was affected by reduced sample size, with its distribution shifting towards zero as sample size was reduced.

Community Detection and Alignment

In individual-based genetic relatedness networks, communities of nodes that are more densely connected to each other than to the rest of the network are indicators of fine-scale population structure (Greenbaum et al., 2019; Greenbaum et al., 2016). When sample size was reduced, the distribution of community sizes shifted towards zero (Figure 1.3.1) indicating that the general size of communities positively correlates with the proportion of the population that is sampled. Although community size decreased, there was a general plateau of the number of communities being detected as sampling was reduced down to 50% of the estimated population size, after which point the number of communities being detected began to decrease with the sampling level (Figure 1.3.2). The general plateau of the number of communities detected despite increased sample size could be reflective of the law of diminishing returns as seen elsewhere in biological sciences (Zhang et al., 1993; Gunn et al., 2010) when solely considering the number of communities in a network. However, there still appeared to be value in increased sample size when looking at overall network topology and community structure. When community assignment was compared between the subsampled populations and the full data sets, almost all nodes were clustered in the same communities even as the sampling was reduced down to 25% (Figures 1.4.1, 1.4.2, 1.4.3). As expected, some communities broke up into smaller communities or were reduced in size but still maintained consistency in the nodes that were clustered together; this is likely due to important linking individuals being dropped from the data sets. Community alignment was consistent across both populations, and all subsampled datasets. Despite reduced sampling, the same individuals were consistently clustered together by the community detection algorithm demonstrating that even when small communities are detected in

sparsely sampled data sets, those communities can represent larger communities of individuals.

Jasper Networks

The individual-based genetic relatedness networks built with the full and 2013 subsampled data set revealed fine-scale genetic structure based on a community detection algorithm to partition the network into multiple communities (Figures 1.5.1, S1). The fine-scale genetic structure that was detected lined up with the three expected sub-populations (Tonquin, Maligne, and Brazeau) as described in McFarlane et al. (2018). Despite only 7 years prior, the network created with the 2006 subsampled dataset did not display the same strong spatial community partitions (Figure 1.5.2). The differences in the community partitions between the two subsampled datasets ranging a relatively small time scale (2006 vs. 2013) demonstrated this method's ability to detect changes to fine-scale contemporary genetic structuring.

SK2 Central Networks

The SK2 Central networks were continuously connected, with the nodes sharing connections across the entirety of the network, and very few disconnected components. The edges tended to display both long distance connections and shorter distance connections within specific areas, which lined up with previous findings of pedigree based networks of the same area (McFarlane et al., 2021). Although the networks were well connected, there was still a spatial pattern of community partitions detected in both the full and subsampled networks which could be evidence of a clinal pattern of genetic variation (Priadka et al., 2019).

Conclusions

The resolution of the pairwise relatedness values (Wang, 2002) was sufficient to differentiate first-order relationships from unrelated individuals; however it was not sufficient to differentiate second order relationships from either first-order relationships or unrelated individuals. The resolution of the relatedness values was consistent despite reduced sampling levels but it should be noted that the resolution and biologically significant threshold values will be dependent on the species, type of genetic markers used, number of loci, and the samples and populations themselves (Foroughirad et al., 2019; Pew et al., 2015). The inclusion of additional loci would be beneficial for future studies since it could result in resolution sufficient to capture a higher degree of second order relationships (Foroughirad et al., 2019).

The proportion of disconnected nodes in individual-based genetic relatedness networks increased with reduced sample size. Given this correlation, it is important to consider sampling efforts, especially when examining the overall connectivity displayed by the networks. The increased proportion of disconnected components in the networks built with reduced sample sizes restricted the ability to analyze the networks using node-based metrics. Furthermore, when analyzing the partitioning results of community detection algorithms; the general size of communities positively correlated with the proportion of the population that was sampled. Although community size decreased, there was a general consistency of the number of communities being detected as sampling was reduced. However this consistency was only true for networks built with greater than 50% of the estimated population size, after which the number of communities being detected began to decrease with the sampling level. Despite reduced sampling, the same

individuals were consistently clustered together by the community detection algorithm, although in smaller sized communities, demonstrating that even when small communities are detected in sparsely sampled data sets, those communities can represent larger communities of individuals and can still be ecologically meaningful.

Although the spatial partitioning of communities was not examined in detail in this study, the general results demonstrated that the community partitions of the individual-based networks aligned with known spatial structuring of the two caribou populations. Furthermore, the differences in the community partitions and assortative mixing between the two subsampled datasets ranging a relatively small time scale (7 years) demonstrated this method's ability to detect contemporary changes to fine-scale genetic structuring.

The results of this study highlight the importance of considering sampling efforts when using individual-based genetic relatedness networks; and that sampling efforts, although not always, can be a limiting factor depending on the type of metrics used and the questions that are trying to be answered.

References

Anderson, C. D., Epperson, B. K., Fortin, M.-J., Holderegger, R., James, P. M. A., Rosenberg, M. S., Spear, S. (2010). Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, *19*(17), 3565-3575. https://doi.org/10.1111/j.1365-294X.2010.04757.x

Baguette, M., Blanchet, S., Legrand, D., Stevens, V. M., & Turlure, C. (2013). Individual dispersal, landscape connectivity and ecological networks. *Biological Reviews*, *88*(2), 310-326. https://doi.org/10.1111/brv.12000

Bradburd, G. S., & Ralph, P. L. (2019). Spatial Population Genetics: It's About Time. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 427-449. https://doi.org/10.1146/annurev-ecolsys-110316-022659

COSEWIC. (2014). COSEWIC assessment and status report on the *Caribou Rangifer tarandus*, Northern Mountain population, Central Mountain population and Southern Mountain population in Canada. Ottawa. www.registrelep-sararegistry.gc.ca/default_e.cfm

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* 1695. https://igraph.org.

De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011). Generalized Louvain method for community detection in large networks. *11th International Conference on Intelligent Systems Design and Applications, 2011.* https://arxiv.org/abs/1108.1502

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimatorv2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209-214. https://doi.org/10.1111/1755-0998.12157

Dormann, C. F., Fründ, J., Blüthgen, N., & Gruber, B. (2009). Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks. *The Open Ecology Journal, 2: 7-24.* http://dx.doi.org/10.2174/1874213000902010007

Draheim, H. M., Moore, J. A., Etter, D., Winterstein, S. R., & Scribner, K. T. (2016). Detecting black bear source-sink dynamics using individual-based genetic graphs. *Proceedings of the Royal Society B-Biological Sciences*, *283*(1835), 9, Article 20161002. https://doi.org/10.1098/rspb.2016.1002

Dyer, R. J., & Nason, J. D. (2004). Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, *13*(7), 1713-1727. https://doi.org/10.1111/j.1365-294X.2004.02177.x

Escoda, L., Fernandez-Gonzalez, A., & Castresana, J. (2019). Quantitative analysis of connectivity in populations of a semi-aquatic mammal using kinship categories and network assortativity. *Molecular Ecology Resources*, *19*(2), 310-326. https://doi.org/10.1111/1755-0998.12967

Escoda, L., Gonzalez-Esteban, J., Gomez, A., & Castresana, J. (2017). Using relatedness networks to infer contemporary dispersal: Application to the endangered mammal Galemys pyrenaicus. *Molecular Ecology*, *26*(13), 3343-3357. https://doi.org/10.1111/mec.14133

Evans, J. S., & Cushman, S. A. (2009). Gradient modeling of conifer species using random forests. *Landscape Ecology*, *24*(5), 673-683. https://doi.org/10.1007/s10980-009-9341-0

Fischer, J., & Lindenmayer, D. B. (2007). Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, *16*(3), 265-280. https://doi.org/10.1111/j.1466-8238.2007.00287.x

Foroughirad, V., Levengood, A. L., Mann, J., & Frère, C. H. (2019). Quality and quantity of genetic relatedness data affect the analysis of social structure. *Molecular Ecology Resources*, *19*(5), 1181-1194. https://doi.org/10.1111/1755-0998.13028

Galpern, P., Manseau, M., & Fall, A. (2011). Patch-based graphs of landscape connectivity: A guide to construction, analysis and application for conservation. *Biological Conservation*, *144*(1), 44-55. https://doi.org/10.1016/j.biocon.2010.09.002

Garroway, C. J., Bowman, J., Carr, D., & Wilson, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, *1*(4), 620-630. https://doi.org/10.1111/j.1752-4571.2008.00047.x

Greenbaum, G., Rubin, A., Templeton, A. R., & Rosenberg, N. A. (2019). Network-based hierarchical population structure analysis for large genomic datasets. *Genome Research, 29.* Cold Spring Harbor Laboratory. http://genome.cshlp.org/

Greenbaum, G., Templeton, A. R., & Bar-David, S. (2016). Inference and Analysis of Population Structure Using Genetic Data and Network Theory. *Genetics*, *202*(4), 1299-1312. https://doi.org/10.1534/genetics.115.182626

Gunn I. D. M., M. O'Hare, L. Carvalho, D. B. Roy, P. Rothery, and A. M. Darwell. (2010). Assessing the condition of lake habitats: a test of methods for surveying aquatic macrophyte communities. *Hydrobiologia*, 656:87-97.

Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetical Research*, *38*(3), 209-216. https://doi.org/10.1017/s0016672300020553

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

Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKelvey, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, *19*(19), 4179-4191. https://doi.org/10.1111/j.1365-294X.2010.04808.x

Leroy, G., Carroll, E. L., Bruford, M. W., Dewoody, J. A., Strand, A., Waits, L., & Wang, J. (2018). Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evolutionary Applications*, *11*(7), 1066-1083. https://doi.org/10.1111/eva.12564

Manseau, M., Arnason, N., McFarlane, S., Wilson, P., & Pittoello, G. (2019). Population Trend Analysis for Boreal Caribou in SK2 Central Using Non-Invasive Capture-Recapture Analysis (2007-2019). *Unpublished manuscript*.

McFarlane, S., Manseau, M., Flasko, A., Horn, R. L., Arnason, N., Neufeld, L., Bradley, M., Wilson, P. (2018). Genetic influences on male and female variance in reproductive success and implications for the recovery of severely endangered mountain caribou. *Global Ecology and Conservation*, *16*, 10, Article e00451. https://doi.org/10.1016/j.gecco.2018.e00451

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

Mimura, M., Yahara, T., Faith, D. P., Vázquez-Domínguez, E., Colautti, R. I., Araki, H., Hendry, A. P. (2017). Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications*, *10*(2), 121-139. https://doi.org/10.1111/eva.12436

Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E*, *67*(2). https://doi.org/10.1103/physreve.67.026126

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577-8582. https://doi.org/10.1073/pnas.0601602103

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, *28*(19), 2537-2539. https://doi.org/10.1093/bioinformatics/bts460

Pew, J., Muir, P. H., Wang, J., & Frasier, T. R. (2015). related: an R package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources*, *15*(3), 557-561. https://doi.org/10.1111/1755-0998.12323

Priadka, P., Manseau, M., Trottier, T., Hervieux, D., Galpern, P., Mcloughlin, P. D., & Wilson, P. J. (2019). Partitioning drivers of spatial genetic variation for a continuously distributed population of boreal caribou: Implications for management unit delineation. *Ecology and Evolution*, *9*(1), 141-153. https://doi.org/10.1002/ece3.4682

Rollins, L. A., Browning, L. E., Holleley, C. E., Savage, J. L., Russell, A. F., & Griffith, S. C. (2012). Building genetic networks using relatedness information: a novel approach for the estimation of dispersal and characterization of group structure in social animals. *Molecular Ecology*, *21*(7), 1727-1740. https://doi.org/10.1111/j.1365-294X.2012.05492.x

Segelbacher, G., Cushman, S. A., Epperson, B. K., Fortin, M. J., Francois, O., Hardy, O. J., Manel, S. (2010). Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics*, *11*(2), 375-385. https://doi.org/10.1007/s10592-009-0044-5

Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., & Waits, L. P. (2010). Landscape genetics: where are we now? *Molecular Ecology*, *19*(17), 3496-3514. https://doi.org/10.1111/j.1365-294x.2010.04691.x

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

Van Dyck, H., & Baguette, M. (2005). Dispersal behaviour in fragmented landscapes: Routine or special movements? *Basic and Applied Ecology*, *6*(6), 535-545. https://doi.org/10.1016/j.baae.2005.03.005

Wang, J. (2002). An Estimator for Pairwise Relatedness Using Molecular Markers. *Genetics*, *160*(3), 1203-1215. https://doi.org/10.1093/genetics/160.3.1203

Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci*. *Conservation Genetics*, *7*(2), 167-184. https://doi.org/10.1007/s10592-005-9100-y

Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, *3*(3), 244-262. https://doi.org/10.1111/j.1752-4571.2009.00104.x

Zhang H., C. T. Haan, and D. L. Nofziger. (1993). An approach to estimating uncertainties in modeling transport of solutes through soils. *Journal of Contaminant Hydrology*, 12**:**35-50. https://doi.org/10.1016/0169-7722(93)90014-J

Appendix A

Table A1. Summary statistics for the full Jasper data set (2006-2016). Number of Alleles (Na), Number of effective alleles (Ne), Information Index (I), Observed Heterozygosity (Ho), Expected Heterozygosity (He), Unbiased Expected Heterozygosity, and Inbreeding Coefficient (F_{IS}). Summary statistics are recorded for each locus with the mean over all loci and standard error (SE) reported bellow.

LOCUS	Na	Ne	I	Ho	He	uHe	F_{IS}
BM848	7.000	4.303	1.540	0.704	0.768	0.769	0.083
BM888	11.000	4.310	1.698	0.731	0.768	0.770	0.048
FCB193	7.000	3.123	1.392	0.606	0.680	0.681	0.108
MAP2C	9.000	4.568	1.683	0.709	0.781	0.783	0.092
NVHRT16	8.000	4.360	1.586	0.694	0.771	0.772	0.099
OHEQ	13.000	6.246	1.995	0.828	0.840	0.842	0.014
RT1	8.000	4.018	1.622	0.770	0.751	0.753	-0.026
RT5	10.000	4.737	1.715	0.723	0.789	0.791	0.083
RT6	7.000	2.332	1.001	0.530	0.571	0.572	0.073
RT7	9.000	3.400	1.445	0.724	0.706	0.707	-0.026
RT9	10.000	4.919	1.770	0.776	0.797	0.798	0.027
RT13	10.000	6.100	1.926	0.797	0.836	0.838	0.046
RT24	7.000	3.290	1.440	0.677	0.696	0.698	0.028
RT27	8.000	6.325	1.928	0.853	0.842	0.844	-0.013
RT30	9.000	6.064	1.871	0.842	0.835	0.837	-0.008
MEAN	8.867	4.540	1.641	0.731	0.762	0.764	0.042
SE	0.446	0.318	0.067	0.022	0.019	0.019	0.012

Table A2. Summary statistics for the 2006 Jasper data set. Number of Alleles (Na), Number of effective alleles (Ne), Information Index (I), Observed Heterozygosity (Ho), Expected Heterozygosity (He), Unbiased Expected Heterozygosity, and Inbreeding Coefficient (F_{IS}). Summary statistics are recorded for each locus with the mean over all loci and standard error (SE) reported bellow.

LOCUS	Na	Ne	I	Ho	He	uHe	F_{IS}
BM848	7.000	4.422	1.583	0.831	0.774	0.779	-0.074
BM888	8.000	4.100	1.646	0.713	0.756	0.760	0.057
FCB193	7.000	3.016	1.371	0.630	0.668	0.672	0.057
MAP2C	6.000	4.582	1.640	0.716	0.782	0.786	0.084
NVHRT16	8.000	4.390	1.589	0.705	0.772	0.776	0.087
OHEQ	12.000	6.281	2.028	0.813	0.841	0.846	0.034
RT1	8.000	4.556	1.710	0.804	0.781	0.785	-0.030
RT5	8.000	4.464	1.639	0.656	0.776	0.780	0.155
RT6	5.000	2.331	0.975	0.538	0.571	0.574	0.057
RT7	9.000	3.160	1.375	0.742	0.683	0.687	-0.085
RT9	8.000	4.792	1.754	0.779	0.791	0.795	0.016
RT13	8.000	5.971	1.897	0.800	0.833	0.837	0.039
RT24	6.000	3.291	1.435	0.642	0.696	0.700	0.078
RT27	8.000	6.459	1.941	0.867	0.845	0.850	-0.025
RT30	8.000	5.756	1.818	0.828	0.826	0.831	-0.002
MEAN	7.733	4.505	1.627	0.738	0.760	0.764	0.030
SE	0.408	0.318	0.069	0.024	0.020	0.020	0.017

Table A3. Summary statistics for the 2013 Jasper data set. Number of Alleles (Na), Number of effective alleles (Ne), Information Index (I), Observed Heterozygosity (Ho), Expected Heterozygosity (He), Unbiased Expected Heterozygosity, and Inbreeding Coefficient (F_{IS}). Summary statistics are recorded for each locus with the mean over all loci and standard error (SE) reported bellow.

LOCUS	Na	Ne	I	Ho	He	uHe	F_{IS}
BM848	5.000	4.363	1.523	0.579	0.771	0.781	0.249
BM888	6.000	4.226	1.592	0.837	0.763	0.772	-0.097
FCB193	6.000	2.882	1.310	0.649	0.653	0.662	0.007
MAP2C	6.000	4.346	1.579	0.615	0.770	0.780	0.201
NVHRT16	5.000	4.075	1.501	0.707	0.755	0.764	0.063
OHEQ	8.000	6.108	1.923	0.897	0.836	0.847	-0.073
RT1	7.000	3.988	1.606	0.795	0.749	0.758	-0.062
RT5	6.000	3.917	1.511	0.767	0.745	0.753	-0.031
RT6	4.000	2.234	0.914	0.571	0.552	0.559	-0.034
RT7	6.000	3.509	1.422	0.744	0.715	0.723	-0.041
RT9	7.000	4.671	1.690	0.778	0.786	0.795	0.010
RT13	8.000	6.133	1.899	0.821	0.837	0.848	0.020
RT24	5.000	3.252	1.388	0.628	0.693	0.701	0.093
RT27	8.000	6.014	1.901	0.854	0.834	0.844	-0.024
RT30	7.000	5.839	1.811	0.850	0.829	0.839	-0.026
MEAN	6.267	4.371	1.571	0.740	0.752	0.762	0.017
SE	0.316	0.311	0.068	0.028	0.020	0.020	0.025

Table A4. Summary statistics for the full SK2 Central data set (2017-2019). Number of Alleles (Na), Number of effective alleles (Ne), Information Index (I), Observed Heterozygosity (Ho), Expected Heterozygosity (He), Unbiased Expected Heterozygosity, and Inbreeding Coefficient (F_{IS}). Summary statistics are recorded for each locus with the mean over all loci and standard error (SE) reported bellow.

LOCUS	Na	Ne	I	Ho	He	uHe	F_{IS}
BM848	8.000	3.776	1.561	0.709	0.735	0.737	0.035
BM888	19.000	5.079	2.059	0.842	0.803	0.805	-0.048
FCB193	11.000	7.215	2.105	0.859	0.861	0.864	0.003
MAP2C	10.000	4.900	1.840	0.772	0.796	0.798	0.029
NVHRT16	8.000	2.686	1.402	0.634	0.628	0.629	-0.010
OHEQ	14.000	5.621	2.040	0.798	0.822	0.824	0.029
RT1	8.000	3.925	1.603	0.680	0.745	0.747	0.087
RT5	10.000	4.301	1.823	0.795	0.768	0.769	-0.036
RT6	8.000	4.339	1.722	0.735	0.770	0.771	0.045
RT7	10.000	3.116	1.470	0.693	0.679	0.681	-0.021
RT9	10.000	3.142	1.432	0.686	0.682	0.683	-0.006
RT13	13.000	5.492	1.986	0.775	0.818	0.820	0.053
RT24	15.000	3.867	1.816	0.698	0.741	0.743	0.058
RT27	8.000	4.695	1.695	0.756	0.787	0.789	0.039
RT30	10.000	2.419	1.332	0.571	0.587	0.588	0.027
MEAN	10.800	4.305	1.726	0.734	0.748	0.750	0.019
SE	0.823	0.326	0.066	0.020	0.020	0.020	0.010

Table A5. Summary statistics for the 2017 SK2 Central data set. Number of Alleles (Na), Number of effective alleles (Ne), Information Index (I), Observed Heterozygosity (Ho), Expected Heterozygosity (He), Unbiased Expected Heterozygosity, and Inbreeding Coefficient (F_{IS}). Summary statistics are recorded for each locus with the mean over all loci and standard error (SE) reported bellow.

LOCUS	Na	Ne	I	Ho	He	uHe	F_{IS}
BM848	8.000	3.655	1.542	0.698	0.726	0.729	0.040
BM888	15.000	4.778	1.997	0.824	0.791	0.794	-0.042
FCB193	10.000	7.391	2.096	0.873	0.865	0.868	-0.010
MAP2C	10.000	4.670	1.785	0.750	0.786	0.789	0.046
NVHRT16	8.000	2.842	1.456	0.656	0.648	0.651	-0.012
OHEQ	14.000	5.615	2.024	0.814	0.822	0.825	0.010
RT1	8.000	4.094	1.638	0.685	0.756	0.759	0.093
RT5	10.000	4.372	1.809	0.825	0.771	0.774	-0.070
RT6	8.000	3.790	1.625	0.696	0.736	0.739	0.055
RT7	8.000	2.598	1.305	0.619	0.615	0.617	-0.007
RT9	8.000	3.016	1.366	0.672	0.668	0.671	-0.005
RT13	12.000	5.540	1.986	0.739	0.820	0.823	0.098
RT24	12.000	3.772	1.767	0.729	0.735	0.738	0.008
RT27	8.000	4.521	1.659	0.773	0.779	0.782	0.007
RT30	9.000	2.279	1.276	0.552	0.561	0.564	0.017
MEAN	9.867	4.195	1.689	0.727	0.739	0.742	0.015
SE	0.608	0.344	0.069	0.022	0.022	0.022	0.012

Table A6. Pairwise F_{ST} matrices between the three subpopulations of Jasper (Brazeau, Maligne, And Tonquin) for each of the data sets (Jasper 2013, Jasper 2006, And Jasper Full).

Jasper 2013			
	Brazeau	**Maligne**	**Tonquin**
Brazeau	0.000		
Maligne	0.100	0.000	
Tonquin	0.078	0.071	0.000
Jasper 2006			
	Brazeau	**Maligne**	**Tonquin**
Brazeau	0.000		
Maligne	0.048	0.000	
Tonquin	0.056	0.041	0.000
Jasper Full (2006-2016)			
	Brazeau	**Maligne**	**Tonquin**
Brazeau	0.000		
Maligne	0.042	0.000	
Tonquin	0.052	0.045	0.000

Table A7 Table displays the number of individuals in each data set, the estimated abundance (N) for the respective year for each data set (calculated from past CMR studies), the proportion (%) of the estimated abundance that is represented in the dataset, and the estimated effective population size (Ne) of each data set calculated using a linkage disequilibrium method in the program NeEstimator.

Data set	# Individuals in dataset	Estimated Abundance (N)	% Estimated Abundance (N)	Estimated Effective Population Size (Ne)	Lower 95% Confidence Interval (Ne)	Upper 95% Confidence Interval (Ne)
Jasper						
Complete	250	N/A	N/A	**16.4**	15.5	17.2
2006	98	98	100	**16.3**	15	17.7
2013	45	69	65	**18.9**	16.1	22.4
SK2 Central						
Complete	209	N/A	N/A	**84.5**	78.8	90.7
2017	133	181	73	**80.8**	73.2	89.6

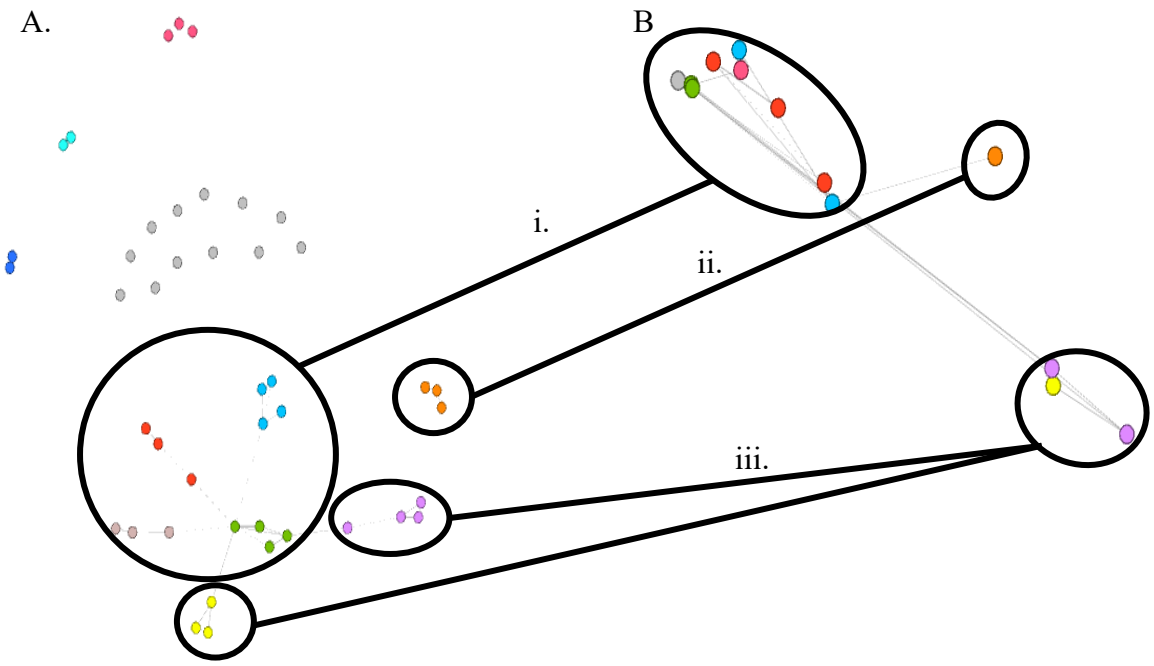


Figure A1 Individual-based genetic relatedness networks created from the 2013 Jasper data set represented (A) a-spatially and (B) spatially. Nodes are coloured according to their community assignment as determined by the Louvain community detection algorithm. Black circles are used to depict the three subpopulations: Tonquin (i), Maligne (ii), and Brazeau (iii) across both the spatial and a-spatial networks.

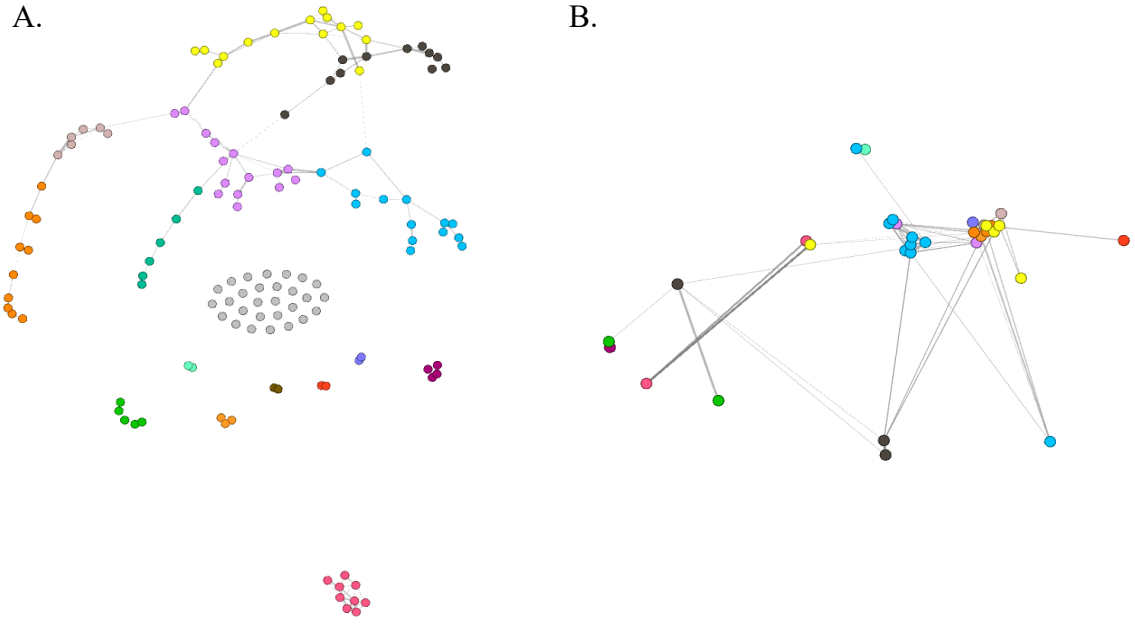


Figure A2 Individual-based genetic relatedness networks created from the 2017 SK2 Central data set represented (A) a-spatially and (B) spatially. Nodes are coloured according to their community assignment as determined by the Louvain community detection algorithm.

Chapter 2

Population and individual-based genetic networks to detect hierarchical genetic population structure and patterns of movement and connectivity of caribou in western Canada

Cory Fournier

ABSTRACT

A combination of genetic network analysis methods were used to investigate genetic structure and connectivity of caribou across different ecotypes at multiple spatial scales. Population-based genetic networks were constructed from a large dataset that spanned western Canada encompassing the Northwest Territories, the eastern side of the Yukon, northern Alberta, and a small portion of western Saskatchewan. A community detection algorithm was used to partition the network at multiple spatial scales which revealed patterns of hierarchical genetic structure. Patterns of gene flow were highlighted through the population-based genetic network topology and through node betweenness centrality values. Additionally, individual-based networks were constructed with a subset of samples from the Mackenzie Mountains Region of the Northwest Territories which revealed patterns of long-distance movements and high connectivity across the region. Using a combination of population-based and individual-based genetic networks can reveal population structure at multiple scales and allow for better interpretation of the fine-scale movements that lead to such structure, including highlighting important areas that facilitate connectivity.

Introduction

Understand of the functional connectivity and dispersal of species is essential to effectively manage and ensure their long-term viability (Baguette et al., 2013; Leroy et al., 2018; Van Dyck & Baguette, 2005). This is particularly true with Increasing landscape disturbance, fragmentation and isolation caused by anthropogenic activities and climate change (Fischer & Lindenmayer, 2007; Leroy et al., 2018; Mimura et al., 2017).

Functional connectivity at the metapopulation level relates to the long-term persistence of local populations (Hanski & Ovaskainen, 2000). Fragmentation of continuous populations of caribou can result in genetic drift and the reduction and loss of genetic diversity (Weckworth 2012, Courtois 2003). The connectivity of continuous populations and avoidance of isolated populations are important for the preservation of genetic diversity and the resulting ability to adapt to changing landscapes and environments (Courtois 2003, McLoughlin 2004).

Genetic Network Analysis

There are many approaches to investigate genetic variation, distribution, and movement of species across space and time. Such knowledge has the potential for great management and conservation implications (Bradburd & Ralph, 2019). In this study, I opt for genetic network methods of analysis which have been developed to specifically analyse genetic variation across the landscape, and to answer questions of movement and connectivity with great flexibility and relatively few assumptions (Dyer & Nason, 2004; Jones & Manseau, 2022). Genetic networks are represented as a series of nodes (populations or individuals) connected by edges which represent genetic relationships or

similarity between the nodes (Jones & Manseau, 2022). Genetic networks can be constructed at the individual or population level.

In individual-based genetic relatedness networks, nodes are individuals, and edges between the nodes are relatedness values as calculated by a pairwise genetic relatedness metric such as the Wang (2002) estimator (Jones & Manseau, 2022). Individual-based genetic relatedness networks are described in detail in the previous chapter. Another form of an individual-based genetic network is a pedigree network, where the pedigree of a given set of individuals is determined through direct observation or by full-pedigree likelihood estimator programs such as COLONY (Jones & Wang, 2010). Pedigree networks present individuals as nodes, and specific relationship types (parent/offspring, full sibling, half siblings, grandparents, cousin etc.) as edges between the nodes.

In contrast, population-based genetic networks are networks that have populations or clusters of individuals sampled in proximity to one-another as nodes in the networks; such networks provide pairwise genetic distances such as Proportion of Shared Alleles (Dps; Murphy et al., 2015), Fst (Wright, 1965) or Euclidean genetic distance (Excoffier et al., 1992) as edges between the nodes. Population-based genetic networks, much like the individual-based genetic relatedness networks, begin as fully saturated; a pairwise genetic distance value is calculated for each dyad in the network. To create meaningful networks, many population-based genetic networks need to be pruned to an edge set that allows for meaningful analysis (Dyer & Nason, 2004; Jones & Manseau, 2022; Savary et al., 2021a). Pruning of population-based genetic networks can be done in many ways, depending on the research objectives and questions, as well as the relationship between landscape distance and genetic distance of the dataset (Savary et al., 2021a). When a dataset shows a correlation between geographic and genetic distance followed by a plateau of genetic

distance which is described as a type IV isolation by distance (IBD) pattern (Van Strien et al., 2015) and the exact maximum dispersal distance of the study species is not known, it is recommended to prune population-based genetic networks using the Conditional Independence Principle method as implemented in *Popgraph* (Dyer et al., 2004; Savary et al., 2021a).

Node-based Metrics

Once networks are constructed, there are various ways to analyse them to better understand the topology of the network, how the network is connected, and how information (in the case of genetic networks: genetic information) flows through the network. Node-metrics are a very simple and useful tool to analyse such information (Newman, 2010). There are many different node-based metrics, and as network theory expands, the application of such metrics to answer real-world questions is rapidly evolving (Jones & Manseau, 2022). Node-based metrics such as degree centrality, eigenvector centrality, and betweenness centrality describe the connectedness of nodes within the network; in a genetic network framework such metrics can be used to infer fitness and demographic parameters (McFarlane et al., 2021), highlight key areas of gene flow (Garroway et al., 2008), and indicate key populations to metapopulation structure that act as bridges between otherwise disconnected components (Rozenfeld et al., 2008).

Network-wide Metrics

Where node-based metrics compare nodes amongst each other within the same network, network-wide metrics are calculated for the entirety of the network and allow for better comparison across networks. Such network metrics include Network Density – ratio of the number of edges in the network compared to the number of possible edges

indicating how connected the network is as a whole (Farine & Whitehead, 2015), Assortativity - the tendency for nodes to connect to other nodes that share a specific attribute (Newman, 2003), and Clustering Coefficient – the tendency of nodes to cluster together (Bastille-Rousseau et al., 2018).

Network Partitioning and Community Detection

Networks are often partitioned into different components. One common way to do this is by modularity optimization-based community detection algorithms such as the Louvain community detection algorithm (De Meo et al., 2011). Such community detection algorithms aim to partition the network into communities that are composed of nodes that are more connected to nodes in the same community than to nodes in other communities (De Meo et al., 2011; Newman, 2006). In genetic networks, community detection can be used to detect individuals or populations that share stronger genetic connections to each other, therefore alluding to genetic structure (Greenbaum et al., 2016). When community detection is run at multiple scales on the same dataset, it can additionally be used to infer hierarchical genetic structure (Greenbaum et al., 2019).

Northern Mountain, Boreal, and Barren-Ground Caribou

There are three different Designatable Units (DUs) of caribou in my study area that spans the Northwest Territories, Yukon, northern Alberta, and northern Saskatchewan: Northern Mountain DU, Barren-ground DU, and Boreal DU (COSEWIC, 2011).

Northern Mountain caribou require a vast range extent in order to avoid predators and other ungulates, as well as to opportunistically shift their habitat use in response to environmental and anthropogenic changes to the landscape (such as forest fires, pests,

precipitation, natural resource extraction, recreation and infrastructure) (Bergerud et al., 1984; COSEWIC, 2014; Environment Canada, 2012a; Polfus et al., 2011). Many northern mountain caribou migrate altitudinally in spring, utilizing high elevation alpine tundra plateaus and upper subalpine areas for their calving and rearing habitat and moving to lower subalpine and coniferous areas for the winter. The reason for the seasonal migratory behavior is for increased foraging opportunities and security, thus making it critical to ensure caribou continue to have the ability to move between varying ranges (Environment Canada, 2012a). In the spring and summer months, these mountain caribou move to higher alpine areas and frequently use late-lying snow patches as a relief from heat stress and insect harassment (Ion & Kershaw, 1989). Additionally, female caribou are noted to disperse in high elevation areas during calving to avoid predation (Bergerud et al., 1984). The main driver of wintering habitat selection is shallow to moderate snow depths (Bergerud, 1978; Heard & Vagt, 1998). Wintering ranges are often found where a mountain ridge or divide creates an area of reduced precipitation with plentiful lichen ground cover to forage on (Environment Canada, 2012a), becoming more confined and focused to lower elevation forests in late winter when temperatures decrease and snow depth increases (Gullickson & Manseau, 2000). A large portion of Northern Mountain Caribou habitat is located in the Mackenzie Mountains Region, referred to as Nío Nę P'ęnę́ by Shúhta (Mountain) Dene. Shúhta (Mountain) Dene and Métis people have a long, interconnected history with the land and biodiversity of the Mackenzie Mountains Region, with which comes an immense and in-depth knowledge based within their own worldviews (Andrews, MacKay, & Andrew, 2012; Andrews, MacKay, Andrew, et al., 2012). With increased concerns about the sustainability of shúhta goʔepę́ (mountain caribou) and Dene/Métis ts'įlį (ways of life) the Nío Nę P'ęnę́ Begháré Shúhta ʔepę́

Nareʔá – Trails of the Mountain Caribou Plan (2019) was drafted. The plan was written and is intended to be implemented through a collaborative, community-led approach. The plan was prepared for the ʔehdzo Got'ı̨nę Gots'é Nákedı (Sahtú Renewable Resources Board) and covers much of the shúhta goʔepǫ́ (mountain caribou) range in the Mackenzie Mountains.

Barren-ground caribou reside in the subarctic tundra, ranging from the Mackenzie Delta to the eastern coast of mainland Nunavut (COSEWIC, 2011). Barren-ground caribou are migratory, aggregating during calving, and migrate long distances between the boreal forest where they winter and the tundra for calving. The currently defined Barren-ground populations in the study area (Bluenose East, Bluenose West, and Bathurst) present overlapping winter ranges (Nagy et al., 2011), and there is evidence of variation of such ranges and behaviour over time and among populations (COSEWIC, 2011).

Boreal caribou span from the northeast corner of the Yukon east to Labrador, and south to Lake Superior. In contrast to the Northern Mountain and Barren-ground DUs, Boreal caribou are considered sedentary, with overlapping summer and winter ranges (Ferguson & Elkie, 2004). Females tend to display solitary behavior from pre-calving to late summer as a form of predator avoidance (Bergerud, 1996) and have high fidelity to general calving areas (Ferguson & Elkie, 2004). Boreal caribou require large areas of undisturbed habitat composed of mature, old-growth coniferous forest, lichens, muskegs, peatlands, and upland or hilly areas in order to disperse during unfavourable conditions, and to maintain low densities as a predator avoidance strategy (Environment Canada, 2012b). Boreal caribou are delineated and managed as 51 distinct "local populations" determined through telemetry and observational data (Environment Canada, 2011).

Caribou in the northern extent of the boreal range of central NWT co-occur with northern mountain and barren-ground caribou and have a different evolutionary history than boreal caribou from regions found further south (Polfus et al., 2017). Dene knowledge holders refer to boreal woodland caribou in central NWT as tǫdzı and distinguish them from mountain caribou (shúhta ʔepę́) and barren-ground caribou (ʔekwę́) based on behavior, habitat selection, and morphology despite significant range overlap (Polfus et al., 2016).

In this chapter I used a combination of population-based genetic networks, individual-based genetic relatedness networks, and pedigree networks to uncover spatial patterns of population genetic structure and movement of caribou spanning the Northern Mountain, Boreal, and Barren-ground Designatable units, with specific focus on the Mackenzie Mountains Region (Northern Mountain DU) of the Northwest Territories. Furthermore, to investigate areas of importance to connectivity and compare gene flow across multiple populations, I analysed the networks using node-based and network-based metrics at multiple scales. Finally, I compared these genetic results to recently published telemetry research that also attempted to delineate caribou communities using network analysis.

Methods

Sample Collection and Genotyping

A total of 1681 caribou individual genotypes from across western Canada were compiled for this study. The study area spans 3 caribou Designatable Units (Northern Mountain, Boreal, and Barren-ground) as defined by COSEWIC (2011). Samples were included from throughout the Northwest Territories, the east side of the Yukon, northern Alberta, and a small area of Northwestern Saskatchewan (Figure 2.1 for map of study area and sampling locations). Samples consisted of fecal, blood and tissue and were collected as part of a large-scale genetic sampling program over the span of 16 years (2004-2020) with a small number of older samples from the Bluenose East and Bathurst herd (Table B1). All samples were amplified and analyzed at 9 microsatellite loci (BM848, BM888, MAP2C, RT24, RT30, RT5, RT6, RT7, RT9; Bishop et al., 1994; Cronin et al., 2005; McLoughlin et al., 2004; Wilson et al., 1997) and the sex was determined by utilizing Zfy/Zfx primers according to protocols outlined in Ball et al.(2007) and Klutsch et al. (2016).

Population-based Genetic Network Analysis

All sampling locations were grouped spatially at a 100 km scale resulting in 50 individual clusters of sampling locations (Figure 2.1). The number of individuals included in each cluster ranged from 9-127 with an average of 34 individuals per location. The centroids of the resulting clusters were used in the population-based analysis as node locations in the spatial networks. Population genetic summary Statistics (Na, Ne, Ho, He, Fis) were calculated using the program GenAlEx 6.5 (Peakall & Smouse, 2012) for all nodes.

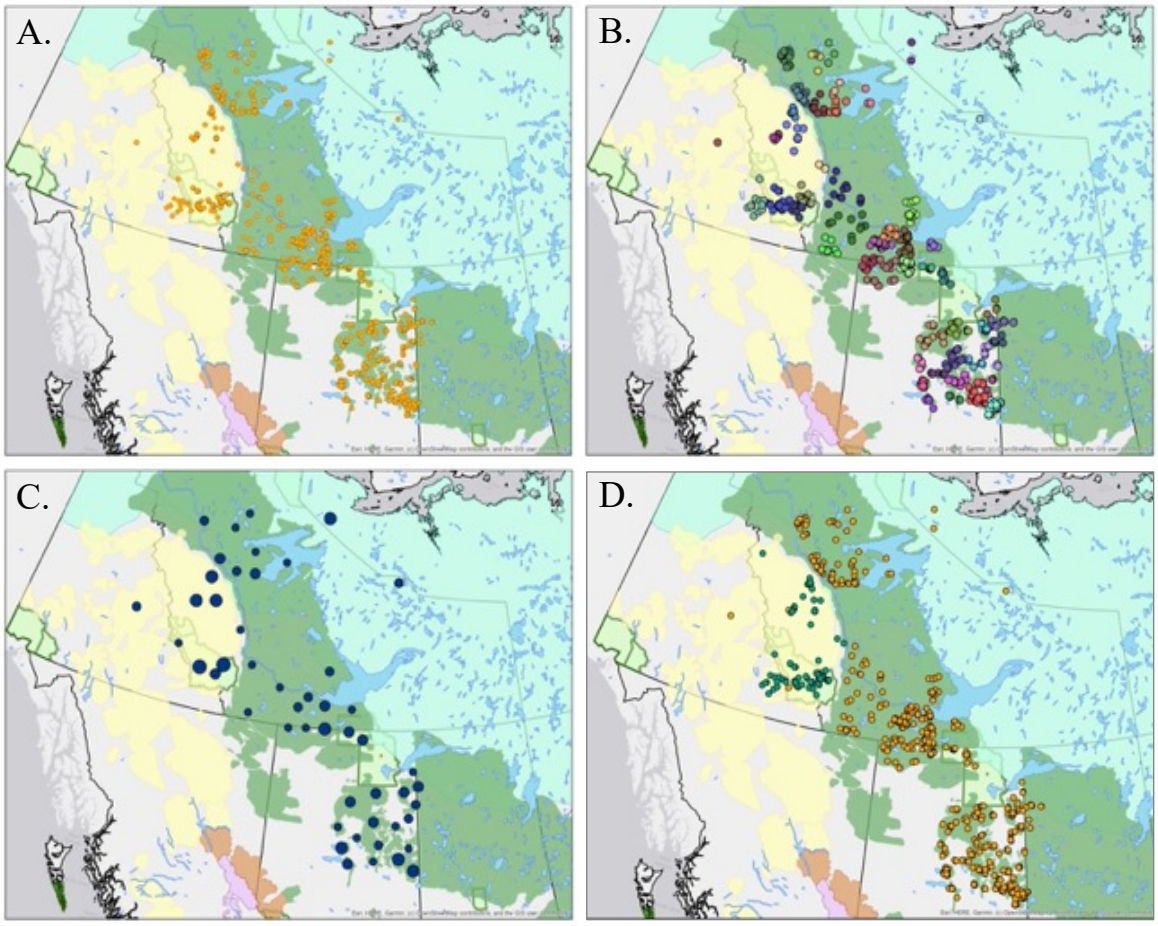


Figure 2.1. Maps of study area in western Canada spanning three Designatable Units (Northern Mountain DU: yellow, Boreal DU: dark green, Barren-ground DU: light green). Map A depicts the location of 1681 samples, Map B depicts the sampling locations clustered at a 100km scale, Map C depicts the centroid locations of each cluster with node size representing sample size for each cluster. Map D depicts the Mackenzie Mountains samples that were genotyped at 15 loci and used for the individual-based networks (green).

Population-based genetic networks were constructed, using the sampling clusters described above as nodes, and Euclidean pairwise genetic distances as calculated in Dyer & Nason (2004) as edges between the nodes. Since genetic networks begin fully saturated, meaning there is an edge between every pair of nodes, I pruned the network using the Conditional Independence Principle pruning method as outlined in Dyer & Nason (2004) which sets out to prune the graph down to the smallest edge set that sufficiently describes the among-population genetic covariance structure. Pruning using the Conditional Independence Principle method is recommended when you do not know the maximum single generation dispersal distance for the study species and the data demonstrate a type IV pattern of Isolation by Distance (Savary et al., 2021b) as these data showed (Figure B1, Appendix B). Population-based genetic networks were constructed using the R package *Graph4lg* (Savary et al., 2021b) which implements functions from the R packages *Popgraph* (Dyer & Nason, 2004) and *igraph* (Csardi & Nepusz, 2006) to construct and analyse genetic and landscape networks. Once the network was created and pruned, I ran the Louvain community detection algorithm which detects communities of nodes that are more connected to nodes in their community, than to other nodes in the network based on modularity optimization (De Meo et al., 2011). Node-based network metrics (betweenness centrality, degree centrality, closeness centrality, and mean inverse edge weight) were calculated for each node in the network using the R package *Graph4lg* (Savary et al., 2021b). The network was exported to shapefiles and was mapped using ArcMap GIS software. Box plots were created to depict the distribution, median, and mean for all node-based metrics and summary statistics, grouped by the first-order communities.

For each of the communities detected in the first-order population-based network, second order networks were created using the same methods. To test for fine-scale/ hierarchical genetic structure, the Louvain community detection algorithm was run on each of the second order networks and second order communities were detected. Node metrics (betweenness centrality, degree centrality, closeness centrality, and mean inverse edge weight) were calculated again for the second order networks. The second order population-based genetic networks, second order community detection, and network metrics were all built and analysed using the *Graph4lg* R package (Savary et al., 2021b). The networks were exported to shapefiles and were mapped using ArcMap GIS software.

Individual-based Genetic Network Analysis

In order to investigate fine-scale movement and dispersal in the Mackenzie Mountains Region at the individual level, I built individual-based genetic relatedness and pedigree networks using methods investigated and outlined in the previous chapter. To do this, a subset of 422 samples from the Mackenzie Mountains Region that had been further genotyped at an additional 6 loci for a total of 15 loci (BM4513, BM6506, BM848, BM888, BMS1788, FCB193, MAP2c, NVHRT16, OHEQ, RT1, RT13, RT24, RT27, RT30, RT5, RT6, RT7, RT9; (Bishop et al., 1994; Cronin et al., 2005; McLoughlin et al., 2004; Wilson et al., 1997) were used (Figure 2.1D).

Individual-based Genetic Relatedness Networks

Pairwise genetic relatedness values (Wang, 2002) were calculated for each dyad in the dataset using the R package, Related (Pew et al., 2015). Using the Related R Package, I then simulated 100 pairs of individuals of four known relationship types (Parent-Offspring, Full Sibling, Half Sibling, and Unrelated) based on the allele frequencies of the

data set. The simulated dyads with known relationships were then used to compare the degree of resolution between various kinship levels that was to be expected. The distributions of relatedness values for the four relationship types were analyzed by creating density plots. The proportion of 1st order, 2nd order, and unrelated pairs of individuals that had relatedness values above and below a specified threshold value were calculated.

I created individual-based genetic relatedness networks using individual caribou as nodes and the pairwise genetic relatedness values (Wang, 2002) as the edges between the nodes. The networks were created with the R package igraph (Csardi & Nepusz, 2006). The networks were thresholded at 0.375 in order to capture a high proportion of 1st order relationships (parent-offspring and full-siblings) while capturing as few unrelated individuals as possible. The details of this threshold value are discussed in the *Individual-based Pairwise Genetic Relatedness Results* section of this chapter. Node metrics (degree centrality and eigenvector centrality) were calculated for each node using the igraph R package (Csardi & Nepusz, 2006). The proportion of disconnected nodes (nodes with a degree of zero) was calculated.

Individual-based Pedigree Networks

I assigned parentage and full sibling relationships of individual caribou using COLONY (Jones & Wang, 2010) which is a computer program that implements full-pedigree likelihood methods to infer relationships of multiple levels in a given multi-locus genotype dataset. I extracted full siblings and parentage from the colony results, and further inferred grandparent, aunt/uncle, and cousin relationships. Pedigree networks were creating using individual caribou as nodes and relationship types inferred from COLONY

as edges between the nodes; networks were constructed using the r package igraph (Csardi & Nepusz, 2006). Node metrics (degree centrality, eigenvector centrality, and betweenness) were calculated for each node using the igraph R package (Csardi & Nepusz, 2006). The proportion of disconnected nodes (nodes with a degree of zero) was calculated.

For both individual-based relatedness networks and individual-based pedigree networks, I created density plots to show the distribution of degree centrality and eigenvector centrality measures for all the nodes in the network. The Louvain community detection algorithm (De Meo et al., 2011) was run on the network to partition the network into communities that contain nodes that are more connected to other nodes in the community than to nodes in the rest of the network. The networks were converted into shapefiles, and imported into ArcMap, where they were mapped. Geographic distances of edges were calculated in the ArcMap software, mean and median distances were calculated, and a histogram was created to depict the distribution of the geographic length of all edges in the network.

Results

Population-based Genetic Networks

Sampling sites were clustered at a 100 km scale, and samples were partitioned into 50 separate clusters, with resulting sample sizes ranging from 9-127 individuals. The 50 clusters were used as nodes in the population-based networks (Figure 2.1C) and population genetic summary statistics were calculated for each (Table 2.4). The resulting population-based genetic network created with edges as Euclidean genetic distances and pruned with the Conditional Independence Principle methods had a resulting 71 edges in the network (Figure 2.2).

First-Order Community Structuring

The Louvain community detection algorithm detected 6 communities: one community made up of 9 nodes from the Northern Mountain populations (henceforth referred to as Northern Mountain Community), one community of 7 nodes aligning with the Barren-ground DU (henceforth referred to as Barren Ground Community), and four communities aligning with the Boreal DU. For the four Boreal communities, one community consisted of 4 nodes along the Mackenzie River corridor in the northern stretch of the Boreal forest (henceforth referred to has Mackenzie River Community), another large community composed of 12 nodes in the Great Slave Lake area (henceforth referred to as Great Slave Lake Community), a small community of only 3 nodes just south of Lake Athabasca (henceforth referred to as Lake Athabasca Community), and lastly a large community made up of 15 nodes from the southernmost extent of the study area in Alberta (henceforth referred to as Alberta community). All the first-order communities were split into separate components – there were no edges connecting the

communities, other than the two northern most boreal communities (Mackenzie River and Great Slave Lake) which had two nodes connected by an edge southeast of Nahanni National Park Reserve.

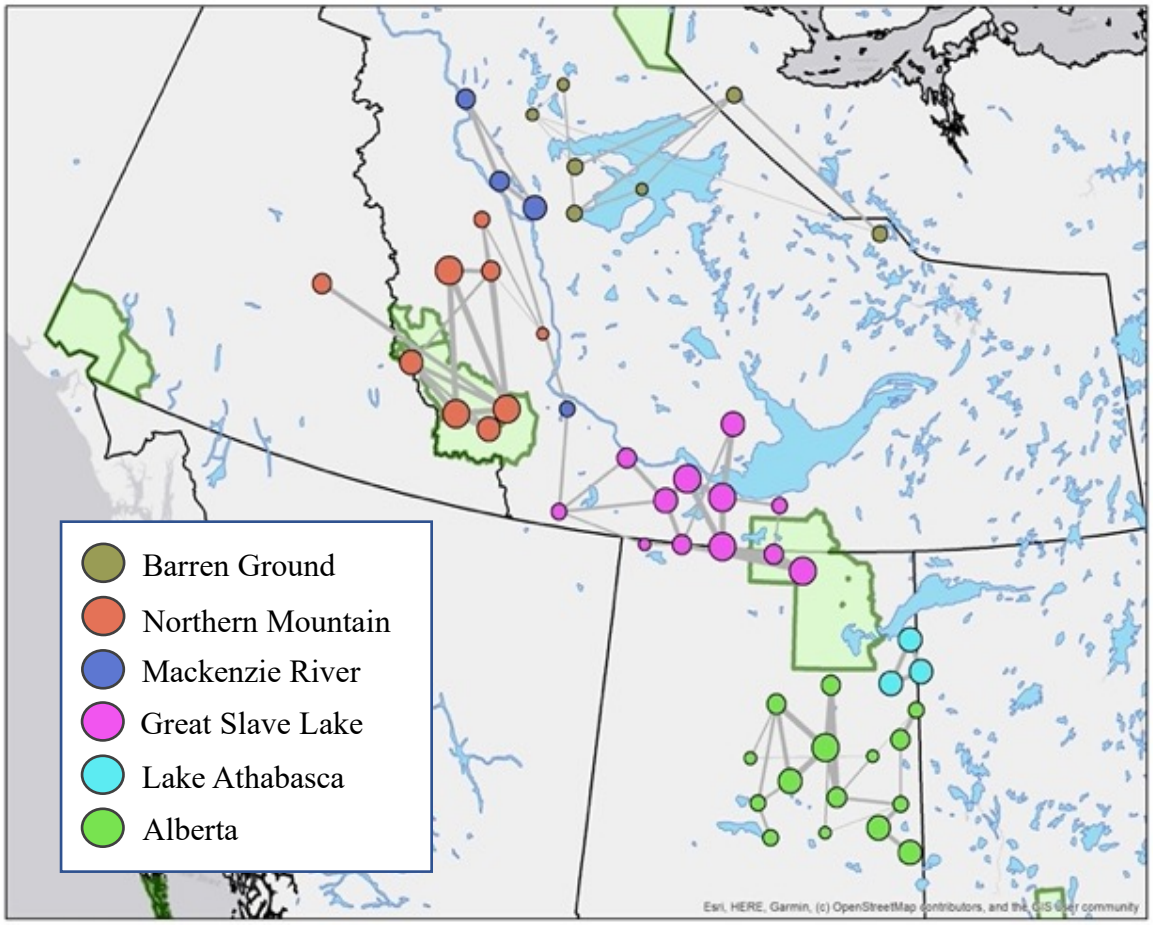


Figure 2.2. Map of population-based genetic network built with 100km clusters of samples as nodes and Euclidean genetic distances as edges between the nodes. The network was pruned using the Conditional Independence Principle pruning method. Louvain community detection algorithm was used to partition the graph into first-order communities (node colours). Node size represents the mean inverse edge weight (MIW) of each respective node while edge thickness represents the inverse edge weight of each respective edge.

Second Order Community Structuring

After creating second order networks for each of the six first-order communities and re-running the Louvain community detection algorithm on each second order network, second order community structuring was detected for each of the six communities. Three second order communities were detected in the Barren-ground community: two communities of three nodes and one community of a single disconnected node (Figure 2.3.1). The communities were all disconnected components and had spatial overlap. The Northern Mountain community had three second order communities detected, all composed of three nodes, and all connected to one another making one large network component (Figure 2.3.2). The Mackenzie River community was split into two second order communities made of up two nodes each, with both communities highly connected to one another (Figure 2.3.3). The Great Slave Lake communities were partitioned into six second order communities ranging from one to three nodes in size, with all but two communities in the west forming their own separate disconnected components (Figure 2.3.4). The Lake Athabasca community had no second order structure detected, with all three nodes forming a single community (Figure 2.3.5). Lastly, the Alberta community was partitioned into five second order communities, with node sizes ranging from one node to five nodes, and all second order communities disconnected from one another (Figure 2.3.6).

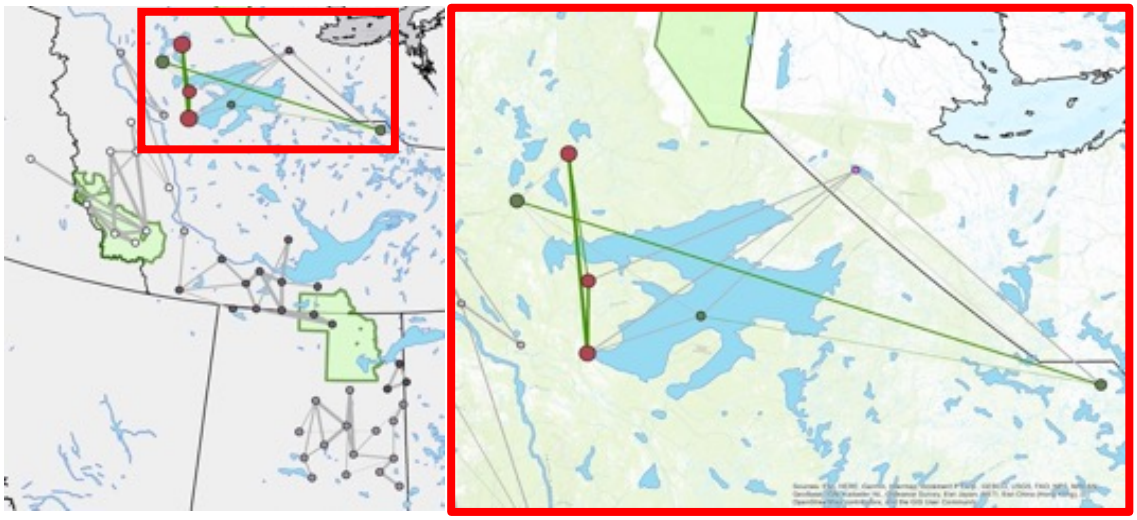


Figure 2.3.1. Second order community structuring within the Barren Ground community. Community assignment is represented by node colour. Green edges are edges in the second order network, while grey edges are edges from the first-order network. Node size represents degree, edge thickness represents inverse edge weight.

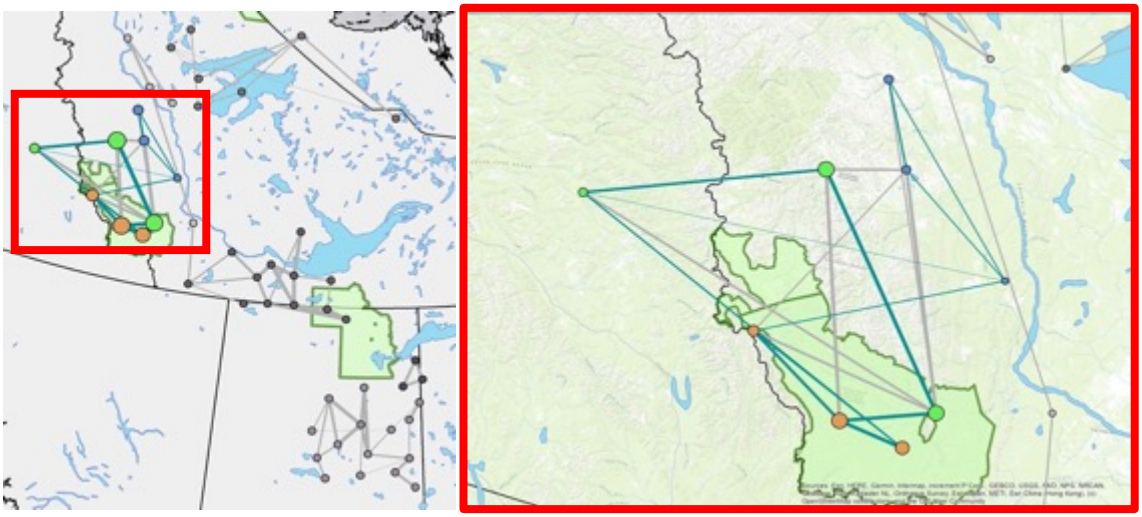


Figure 2.3.2. Second order community structuring within the Northern Mountain community. Community assignment is represented by node colour. Teal edges are edges in the second order network, while grey edges are edges from the first-order network. Node size represents degree, edge thickness represents inverse edge weight.

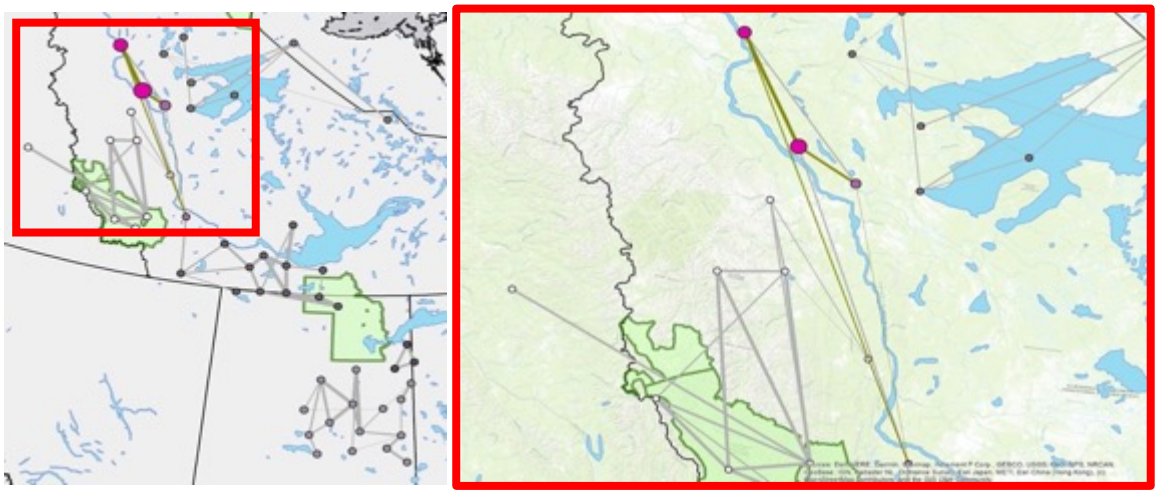


Figure 2.3.3. Second order community structuring within the Mackenzie River community. Community assignment is represented by node colour. Dark yellow edges are edges in the second order network, while grey edges are edges from the first-order network. Node size represents degree, edge thickness represents inverse edge weight.

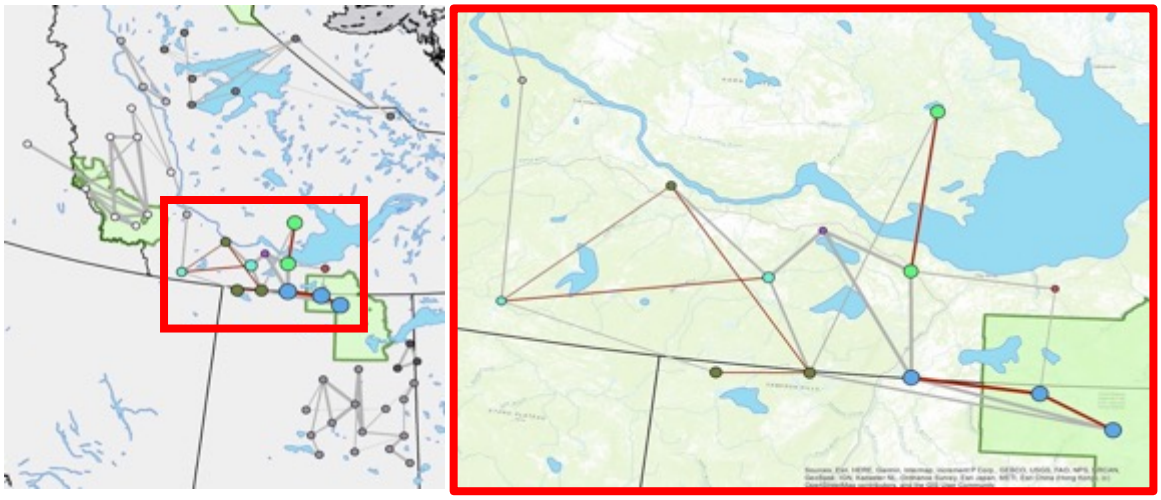


Figure 2.3.4. Second order community structuring within the Great Slave Lake community. Community assignment is represented by node colour. Dark red edges are edges in the second order network, while grey edges are edges from the first-order network. Node size represents degree, edge thickness represents inverse edge weight.

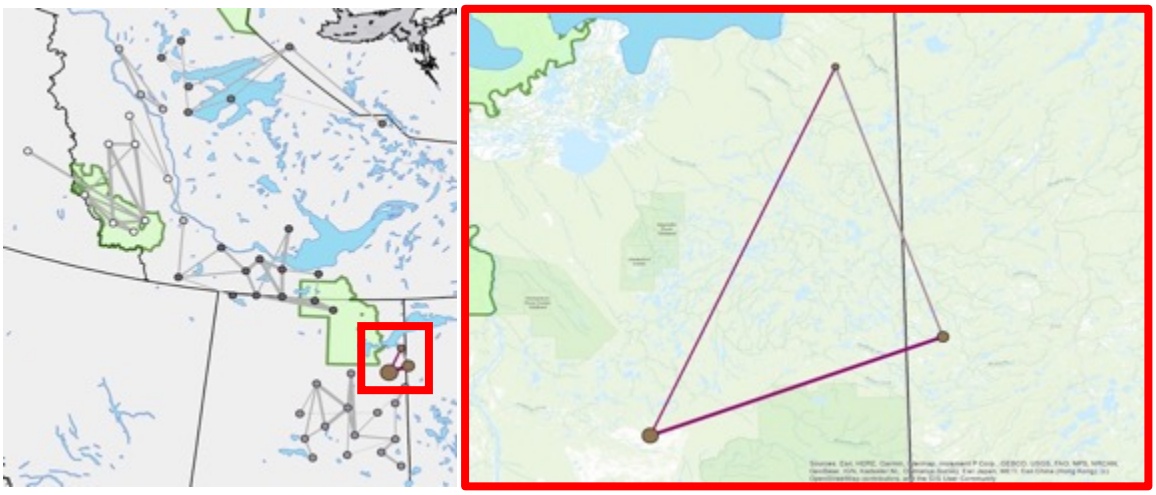


Figure 2.3.5. Second order community structuring within the Lake Athabasca community. Dark purple edges are edges in the second order network, while grey edges are edges from the first-order network. Node size represents degree, edge thickness represents inverse edge weight.

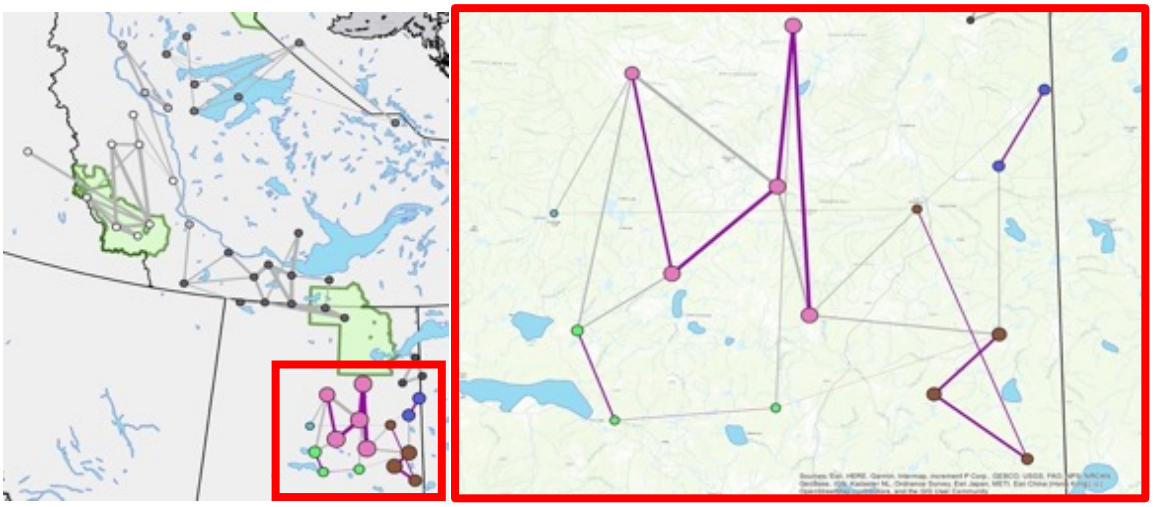


Figure 2.3.6. Second order community structuring within the Alberta community. Dark purple edges are edges in the second order network, while grey edges are edges from the first-order network. Node size represents degree, edge thickness represents inverse edge weight.

Node-based Metrics and Summary Statistics

Node-based metrics and population genetic summary statistics were calculated for all 50 nodes of the first-order population-based genetic network (Table 2.3 & 2.5). The distribution and means for all node metrics and population genetic summary statistics grouped by first order community were shown using box plots (Figure 2.4 & 2.6).

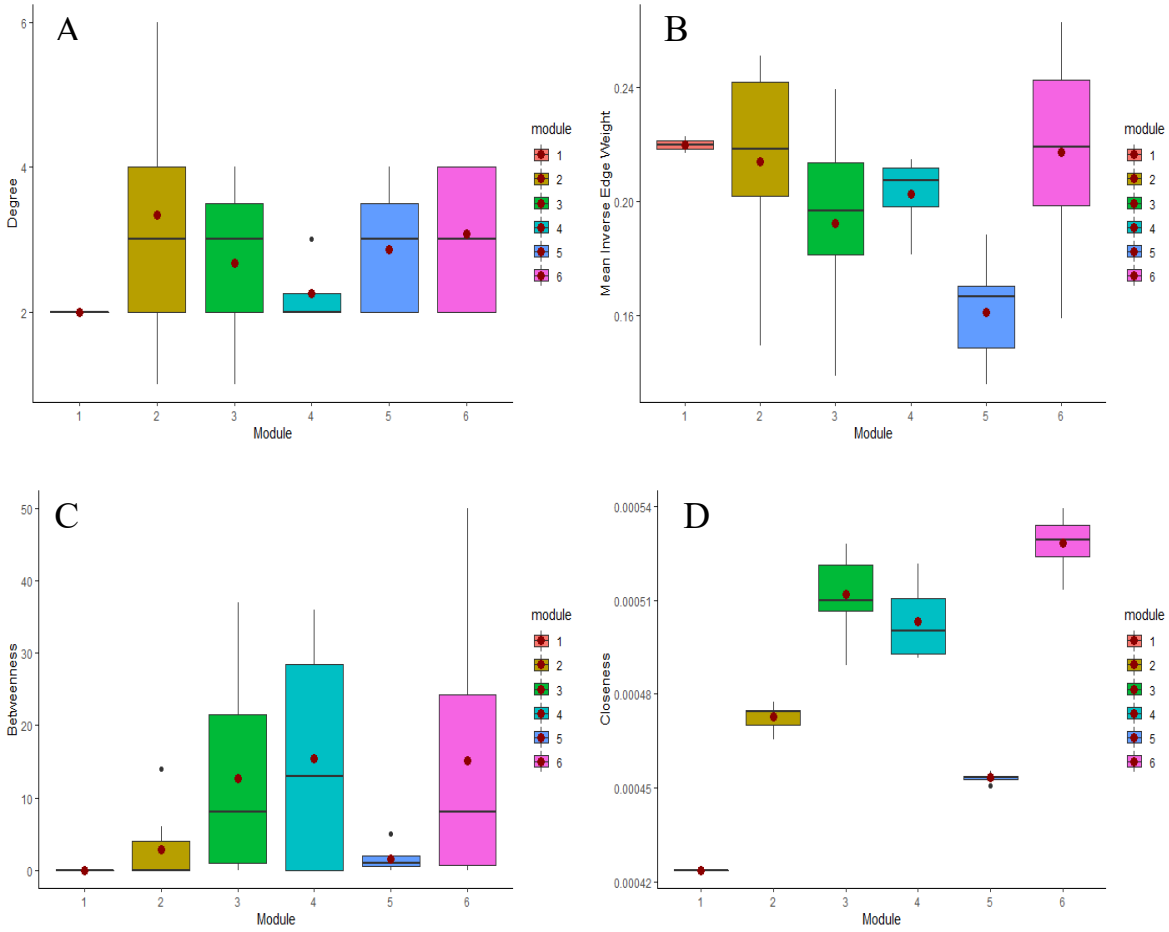


Figure 2.4. Box plots depicting distribution, median (black lines), and mean (red point) of node-based metrics calculated from the first-order population-based genetic network and grouped by the first-order communities. Module 1(red): Lake Athabasca, Module 2 (yellow): Northern Mountain, Module 3 (green): Alberta, Module 4 (teal): Mackenzie River, Module 5 (blue): Baren-Ground, Module 6 (pink): Great Slave Lake. (A) Degree, (B) Mean Inverse Edge Weight, (C) Betweenness, and (D) Closeness.

Degree centrality values were similar across all communities (degree values ranging from 2-4) except Lake Athabasca and Mackenzie River which both had a small number of nodes in each community (n=3 and n=4 respectively) and therefore had reduced degree centrality values. The mean inverse edge weight (miw) values were similar for all communities except Barren-ground which displayed much lower miw values than the remaining communities. The miw values for the two smaller communities (Lake Athabasca and Mackenzie River) spanned a much smaller distribution. Betweenness centrality values were very low, with a narrow distribution for the Northern Mountain and Barren-ground communities (means of 2.9 and 1.6 respectively); in contrast, the means of the boreal communities (Alberta, Great Slave Lake, and Mackenzie River) were greater (12.7, 15.1, and 15.5 respectively), and their distributions much larger (all with betweenness values ranging from 0 to over 20), indicating that there were more nodes with high betweenness values in the boreal communities. The exception to this was for the nodes in the Lake Athabasca community which all had betweenness values of zero due to the three nodes in the small community being all connected to one another (Figure 2.5.1). Lastly, closeness centrality values displayed a similar pattern to betweenness centrality, with the three boreal communities (Alberta, Great Slave Lake, and Mackenzie River) displaying much greater closeness values than the Northern Mountain and Barren-ground communities. Again, the exception was the small Lake Athabasca community that was composed of only three nodes (Figure 2.5.2).

Table 2.1. Node-based metric means grouped by first-order communities. Metrics include degree centrality, mean inverse edge weight, betweenness centrality, and closeness centrality.

Community (module)	Degree	MIW	Betweenness	Closeness
Lake Athabasca (1)	2.0000	0.2198	0.0000	0.0004239
Northern Mountain (2)	3.3333	0.2140	2.8889	0.0004728
Alberta (3)	2.6667	0.1923	12.7333	0.0005119
Mackenzie River (4)	2.2500	0.2027	15.5000	0.0005033
Baren-Ground (5)	2.8571	0.1614	1.5714	0.0004533
Great Slave Lake (6)	3.0833	0.2174	15.1667	0.0005281

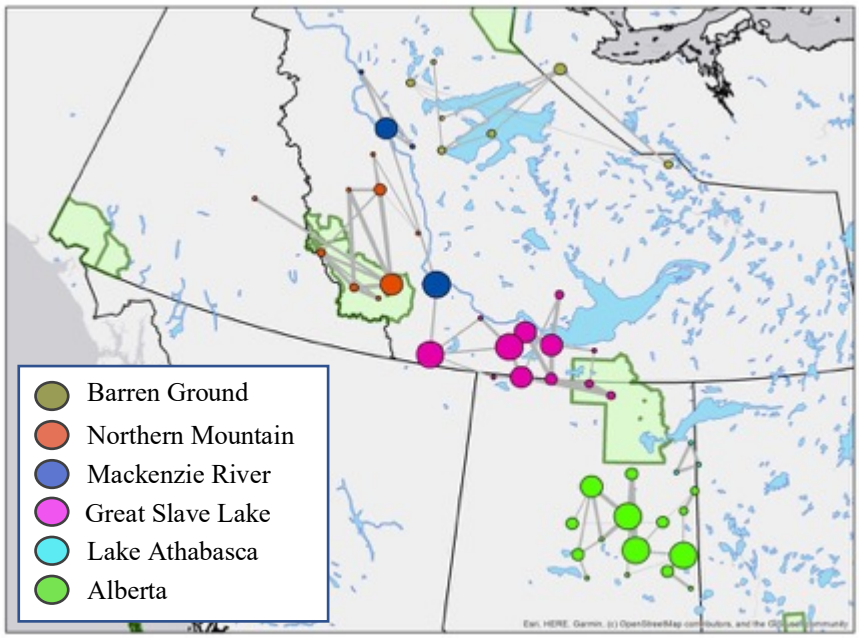


Figure 2.5.1. Population-based network with node sizes representing betweenness centrality measures and node colours representing community assignment.

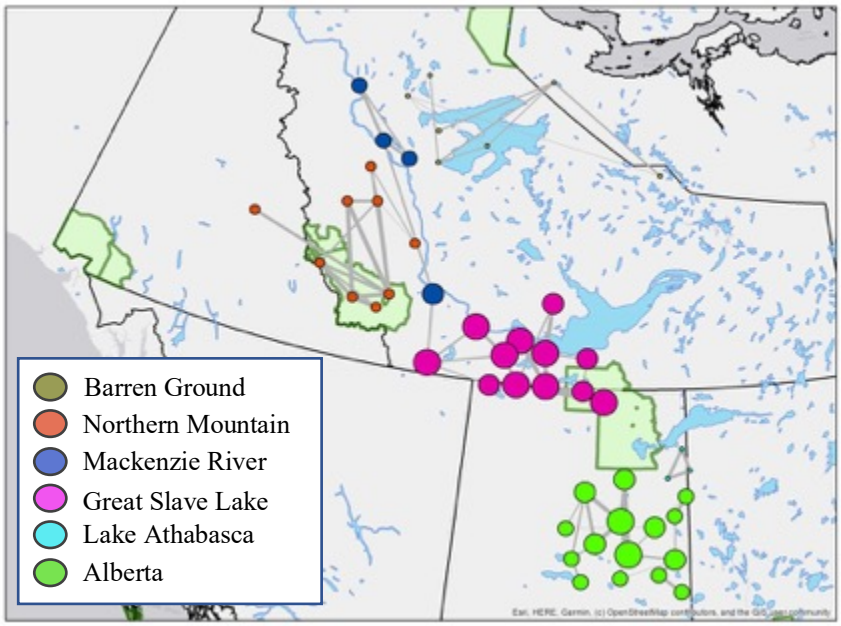


Figure 2.5.2. Population-based network with node sizes representing closeness centrality measures and node colours representing community assignment.

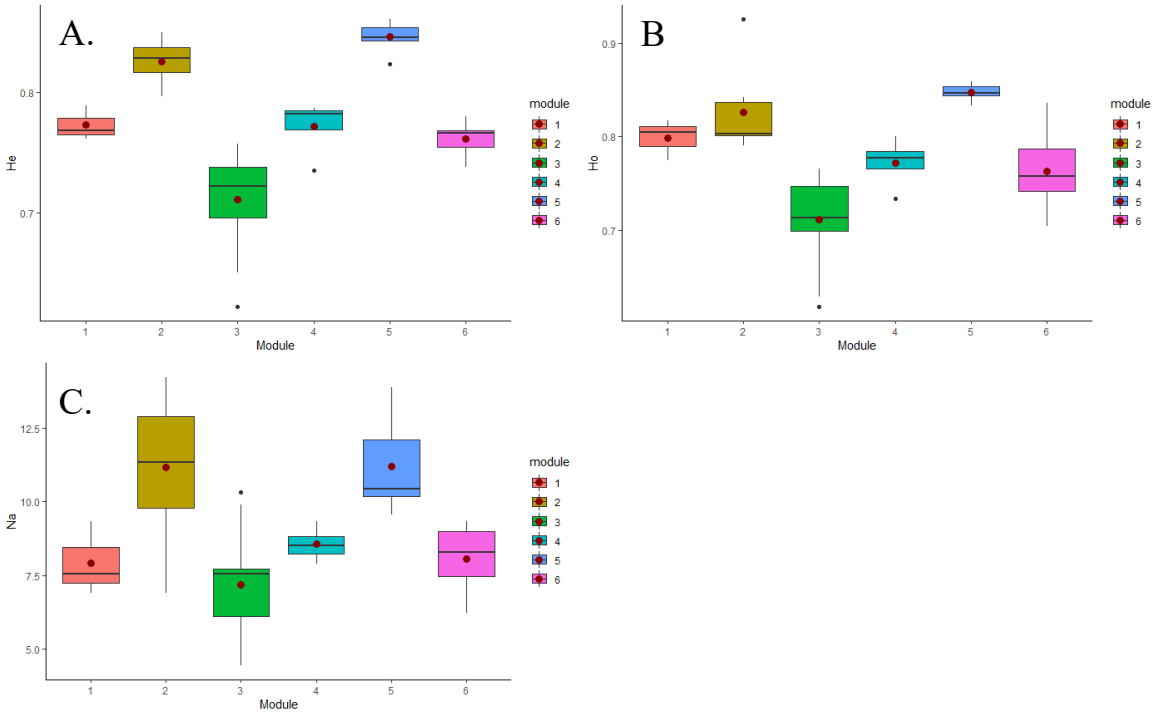


Figure 2.6. Box plots depicting distribution, median (black lines), and mean (red point) of population genetic summary statistics (He, Ho, and Na) calculated in *GenAlEx* and grouped by the first-order communities from the population-based genetic networks. Module 1(red): Lake Athabasca, Module 2 (yellow): Northern Mountain, Module 3 (green): Alberta, Module 4 (teal): Mackenzie River, Module 5 (blue): Baren-Ground, Module 6 (pink): Great Slave Lake. (A) Expected Heterozygosity, (B) Observed Heterozygosity, and (C) Number of Different Alleles.

Expected and observed heterozygosity were highest for the Northern Mountain and Barren-ground communities, with all the boreal communities having lower values, with the lowest values in the southern most boreal community (Alberta) as seen in Figure 2.5 (A and B). The Number of Different Alleles (Na) followed a similar pattern, with the most pronounced differences being the Northern Mountain and Barren-ground community having higher values, and the Boreal communities all having lower Numbers of Different Alleles(Na). Amongst the four boreal communities, the Number of Different Alleles (Na) decreased the further south the communities were situated.

Table 2.2. Means of population genetic summary statistics (He, Ho, and Na) calculated in *GenAlEx* and grouped by the first-order communities from the population-based genetic networks.

Community (module)	Mean He	Mean Ho	Mean Na
Lake Athabasca (1)	0.77	0.80	7.93
Northern Mountain (2)	0.83	0.83	11.16
Alberta (3)	0.71	0.71	7.19
Mackenzie River (4)	0.77	0.77	8.56
Baren-Ground (5)	0.85	0.85	11.21
Great Slave Lake (6)	0.76	0.76	8.06

Table 2.3. Node-based metrics: Sample Size (n), Community Assignment (Module), Degree Centrality, Closeness Centrality, Betweenness Centrality, Mean Inverse Edge Weight (MIW), Sum of Inverse Edge Weight (SIW).

NODE	n	Module	Degree	Closeness	Betweenness	MIW	SIW
13	14	1	2	0.00042	0	0.22	0.43
14	44	1	2	0.00042	0	0.22	0.44
15	39	1	2	0.00042	0	0.22	0.45
24	9	2	2	0.00047	0	0.15	0.30
25	127	2	6	0.00048	14	0.24	1.45
26	40	2	2	0.00047	0	0.22	0.45
27	113	2	4	0.00047	4	0.25	1.00
28	47	2	3	0.00047	0	0.18	0.54
29	56	2	3	0.00047	0	0.25	0.75
30	17	2	4	0.00047	2	0.22	0.87
31	49	2	5	0.00048	6	0.20	1.01
61	26	2	1	0.00047	0	0.21	0.21
1	30	3	2	0.00051	4	0.20	0.39
10	44	3	3	0.00052	8	0.21	0.63
11	20	3	2	0.00051	6	0.14	0.28
12	27	3	2	0.00050	2	0.19	0.38
2	20	3	4	0.00052	33	0.18	0.71
21	38	3	1	0.00049	0	0.19	0.19
22	65	3	3	0.00051	13	0.19	0.56
3	30	3	4	0.00053	37	0.21	0.82
4	54	3	2	0.00051	13	0.22	0.44
5	17	3	3	0.00051	9	0.14	0.42
52	50	3	1	0.00049	0	0.23	0.23
6	18	3	2	0.00051	0	0.14	0.28
7	20	3	3	0.00052	0	0.22	0.65
8	40	3	4	0.00053	36	0.24	0.96
9	37	3	4	0.00052	30	0.21	0.83
32	15	4	2	0.00052	36	0.18	0.36
44	33	4	3	0.00051	26	0.21	0.63
45	28	4	2	0.00049	0	0.20	0.41
46	29	4	2	0.00049	0	0.21	0.43
43	16	5	3	0.00045	1	0.16	0.47
48	19	5	2	0.00045	0	0.14	0.28
49	40	5	3	0.00045	2	0.17	0.51
50	56	5	4	0.00046	5	0.19	0.75

51	26	5	2	0.00045	1	0.17	0.34
53	16	5	2	0.00045	0	0.17	0.33
54	16	5	4	0.00045	2	0.14	0.54
37	17	6	2	0.00051	0	0.18	0.37
38	34	6	4	0.00053	15	0.25	0.99
39	47	6	4	0.00053	12	0.26	1.04
40	36	6	3	0.00052	1	0.21	0.64
41	34	6	3	0.00053	4	0.24	0.72
76	22	6	3	0.00054	28	0.26	0.79
77	15	6	2	0.00052	0	0.16	0.32
78	24	6	4	0.00054	50	0.23	0.91
79	16	6	4	0.00053	23	0.20	0.82
82	11	6	2	0.00053	0	0.21	0.41
83	10	6	4	0.00053	45	0.18	0.71
85	30	6	2	0.00052	4	0.23	0.45

Table 2.4. Sample Size (n), Community Assignment (Module) and Population genetic summary statistics (mean over all loci) for each node: Number of Different Alleles (Na), number of effective alleles (Ne), Shannon's information index (I), Observed Heterozygosity (Ho), Expected Heterozygosity (He), Unbiased Expected Heterozygosity (uHe), and Fixation Index.

Node	n	Module	Na	Ne	I	Ho	He	uHe	F
13	14	1	14.00	6.89	4.54	1.65	0.82	0.77	0.80
14	44	1	42.22	9.33	4.97	1.79	0.78	0.79	0.80
15	39	1	38.78	7.56	4.48	1.64	0.81	0.76	0.77
24	9	2	8.44	6.89	5.11	1.74	0.93	0.80	0.85
25	127	2	123.33	14.22	6.35	2.08	0.83	0.84	0.84
26	40	2	38.89	10.78	5.30	1.92	0.80	0.81	0.82
27	113	2	105.44	13.44	5.96	2.02	0.80	0.83	0.83
28	47	2	46.11	12.89	7.02	2.14	0.84	0.85	0.86
29	56	2	55.11	12.11	6.22	2.03	0.84	0.83	0.84
30	17	2	16.89	9.00	5.71	1.90	0.80	0.82	0.84
31	49	2	46.78	11.33	6.34	2.02	0.79	0.84	0.85
61	26	2	25.78	9.78	6.00	1.96	0.80	0.83	0.84
1	30	3	27.56	7.67	4.03	1.57	0.71	0.73	0.74
10	44	3	36.78	7.44	4.16	1.59	0.74	0.73	0.74
11	20	3	19.22	6.44	3.74	1.50	0.70	0.71	0.73
12	27	3	26.22	7.56	4.11	1.57	0.72	0.72	0.74
2	20	3	18.11	6.78	3.82	1.51	0.71	0.71	0.73
21	38	3	37.11	4.44	3.01	1.17	0.62	0.62	0.63
22	65	3	64.22	5.22	3.26	1.29	0.71	0.68	0.68
3	30	3	24.11	7.56	4.01	1.55	0.70	0.71	0.73
4	54	3	51.11	9.89	4.26	1.70	0.77	0.75	0.76
5	17	3	16.00	5.78	3.55	1.39	0.66	0.68	0.70
52	50	3	49.44	10.33	4.24	1.71	0.75	0.74	0.75
6	18	3	15.56	4.89	2.96	1.26	0.63	0.65	0.68
7	20	3	19.33	7.56	4.17	1.61	0.76	0.74	0.76
8	40	3	37.00	8.44	4.02	1.63	0.76	0.73	0.74
9	37	3	35.44	7.78	4.36	1.67	0.73	0.76	0.77
32	15	4	14.78	7.89	4.99	1.77	0.78	0.79	0.82
44	33	4	32.33	9.33	5.14	1.83	0.80	0.78	0.80
45	28	4	27.00	8.67	4.05	1.65	0.73	0.74	0.75
46	29	4	28.22	8.33	4.86	1.76	0.78	0.78	0.79
43	16	5	15.78	10.44	6.65	2.06	0.86	0.84	0.87
48	19	5	18.56	10.33	6.52	2.06	0.84	0.84	0.87
49	40	5	38.56	12.33	7.56	2.18	0.86	0.86	0.87

50	56	5	55.33	13.89	7.41	2.20	0.85	0.85	0.86
51	26	5	25.44	11.89	6.93	2.14	0.85	0.85	0.87
53	16	5	15.56	9.56	6.00	1.96	0.85	0.82	0.85
54	16	5	15.11	10.00	6.49	2.05	0.83	0.85	0.87
37	17	6	16.44	6.89	4.22	1.61	0.80	0.74	0.76
38	34	6	33.11	9.33	4.50	1.75	0.72	0.77	0.78
39	47	6	46.67	9.11	4.42	1.73	0.74	0.76	0.77
40	36	6	34.44	8.22	4.22	1.66	0.70	0.74	0.75
41	34	6	33.33	9.00	4.65	1.77	0.77	0.78	0.79
76	22	6	21.33	8.33	4.84	1.75	0.75	0.77	0.79
77	15	6	14.56	8.22	4.94	1.78	0.80	0.78	0.81
78	24	6	24.00	8.33	4.70	1.74	0.76	0.77	0.78
79	16	6	15.67	7.67	4.52	1.70	0.75	0.77	0.79
82	11	6	10.89	6.22	4.44	1.61	0.84	0.77	0.80
83	10	6	9.78	6.33	4.14	1.57	0.78	0.74	0.78
85	30	6	29.44	9.00	4.55	1.75	0.74	0.77	0.78

Individual-based Genetic Networks

Relatedness Networks

The 100 simulated pairs of individuals for each level of relatedness (parent/offspring, full sibling, half sibling, and unrelated) were plotted on a density graph to visualize the distribution of the respective relatedness values for each relationship category (Figure 2.7). Little overlap was shown between first-order (parent/offspring and full siblings) and unrelated relationships; however second order relationships (half siblings) demonstrated a large amount of overlap of both first-order and unrelated relatedness values.

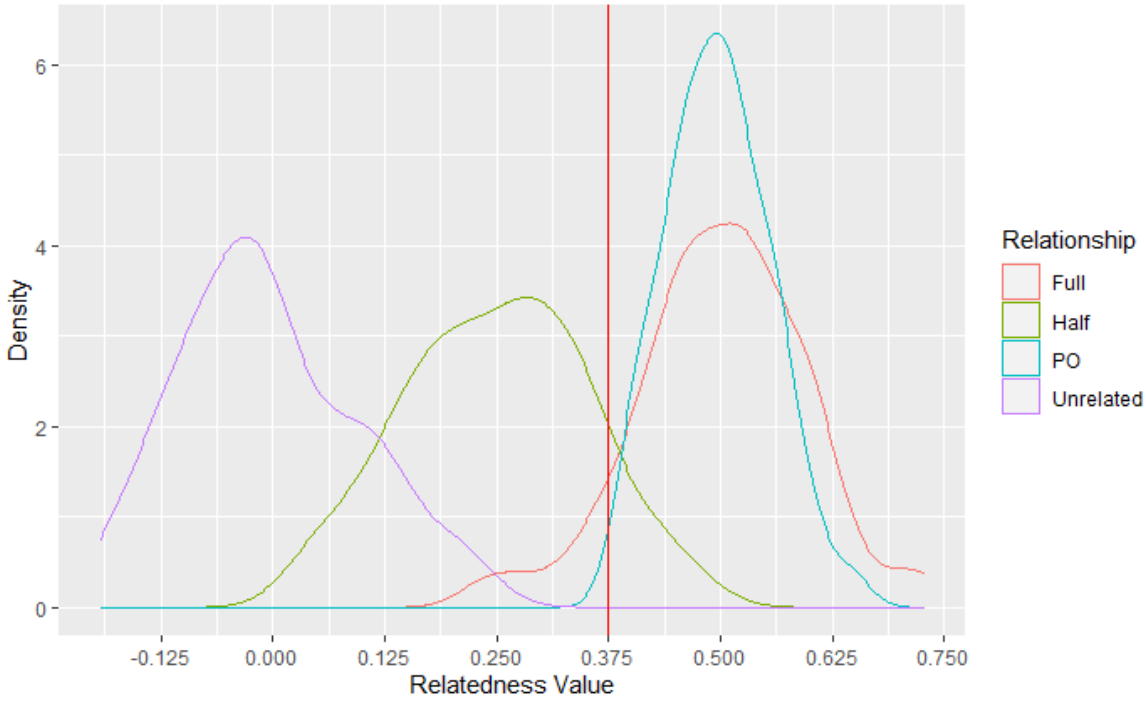


Figure 2.7. Density plot of the calculated relatedness values (Wang, 2002) for the four simulated relationship categories (PO: Parent/Offspring, Half: Half-sibling, Full: Full-sibling, and Unrelated). 100 pairs of each relationship category were simulated. Simulations were conducted and density plot was created using the R package Related. Red vertical line represents threshold value of 0.375.

Table 2.5. Proportion of pairs of individuals from each relationship type that fall below and above the 0.375 threshold for the Mackenzie Mountain dataset genotyped at 15 loci.

Relationship	Frequency (%) < 0.375	Frequency (%) >= 0.375
PO	0	100
Full	9	91
Half	90	10
Unrelated	100	0

The exact proportion of pairs of individuals from each relationship type that were above and below the 0.375 threshold for the 15 loci Mackenzie Mountains data set was calculated (Table 2.5). All parent-Offspring relationships (100%) were captured above 0.375, full sibling relationships were captured at a rate of 91%, 10% of half sibling relationships were above the 0.375 threshold, and no unrelated pairs of individuals fell above the 0.375 threshold.

The Individual-based relatedness network pruned at a relatedness value of 0.375 resulted in 120 edges detected in the Mackenzie Mountains Region. The mean and median edge distances were 72 km and 6 km respectively with a maximum distance of 380 km (Figure 2.8). There were a high number of edges connecting animals in the wintering range near the Prairie Creek Mine area to both the Nahanni complex animals to the West, and the Redstone animals to the North. The network was very disconnected, with a high proportion of nodes that were disconnected components. All node metrics skewed towards zero displaying minimal information, resembling networks built with less than 25% of the total population sampled as shown in the previous chapter.

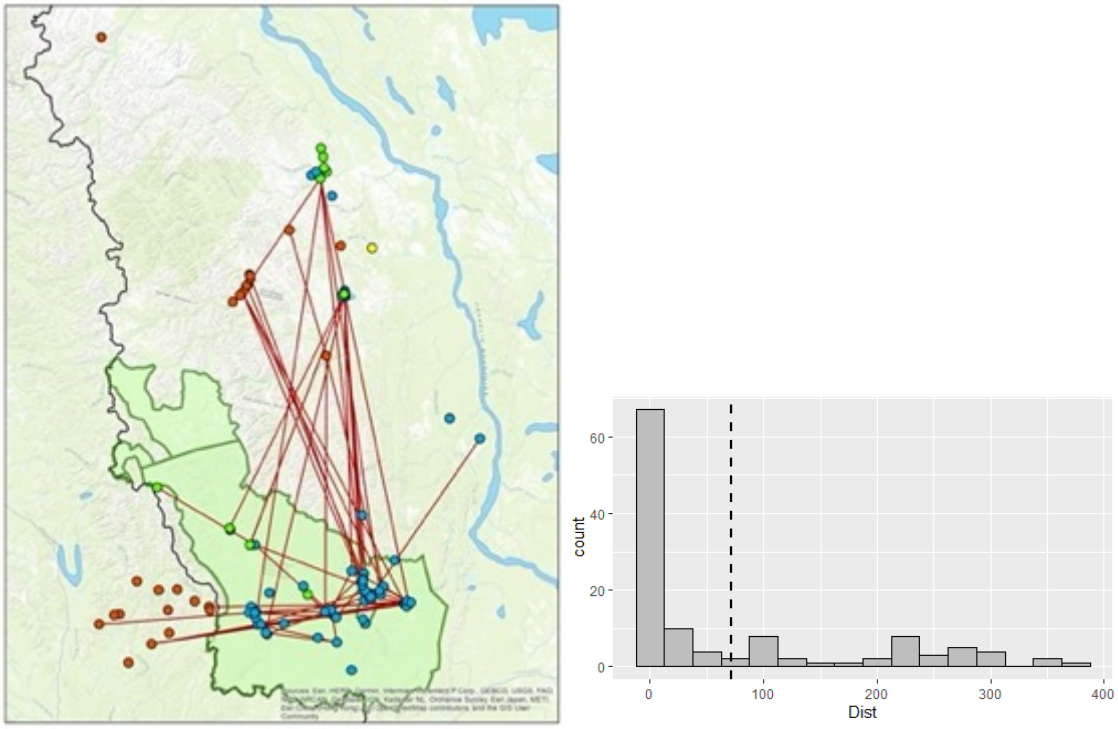


Figure 2.8. Individual-based relatedness network in the Mackenzie Mountains Region built with relatedness values (Wang, 2002) as edges, and individual caribou as nodes. Edges were pruned at a 0.375. Nodes are colored according to season: winter (blue), fall (orange), spring (green), summer (yellow). Histogram depicts distribution of edge distance (km) calculated in ArcMap GIS software.

Pedigree Networks

The individual-based pedigree network in the Mackenzie Mountains Region resulted in 172 edges detected, 81 which were first-order relationships (Parent/Offspring and Full Siblings) and 91 of which were second order relationships (Grandparent, Aunt/Uncle, Cousin). The mean and median distances of the edges were 71km and 5 km respectively with a maximum distance of 538 km (Figure 2.9). Similar to the individual-based genetic relatedness networks previously shown, the pedigree network resulted in many edges connecting the Prairie Creek mine wintering area with Nahanni complex animals in the West and Redstone animals in the North. Again, the network was very disconnected, with a high proportion of nodes that were disconnected components. All node metrics skewed towards zero displaying minimal information, resembling networks built with less than 25% of the total population sampled as shown in the previous chapter.

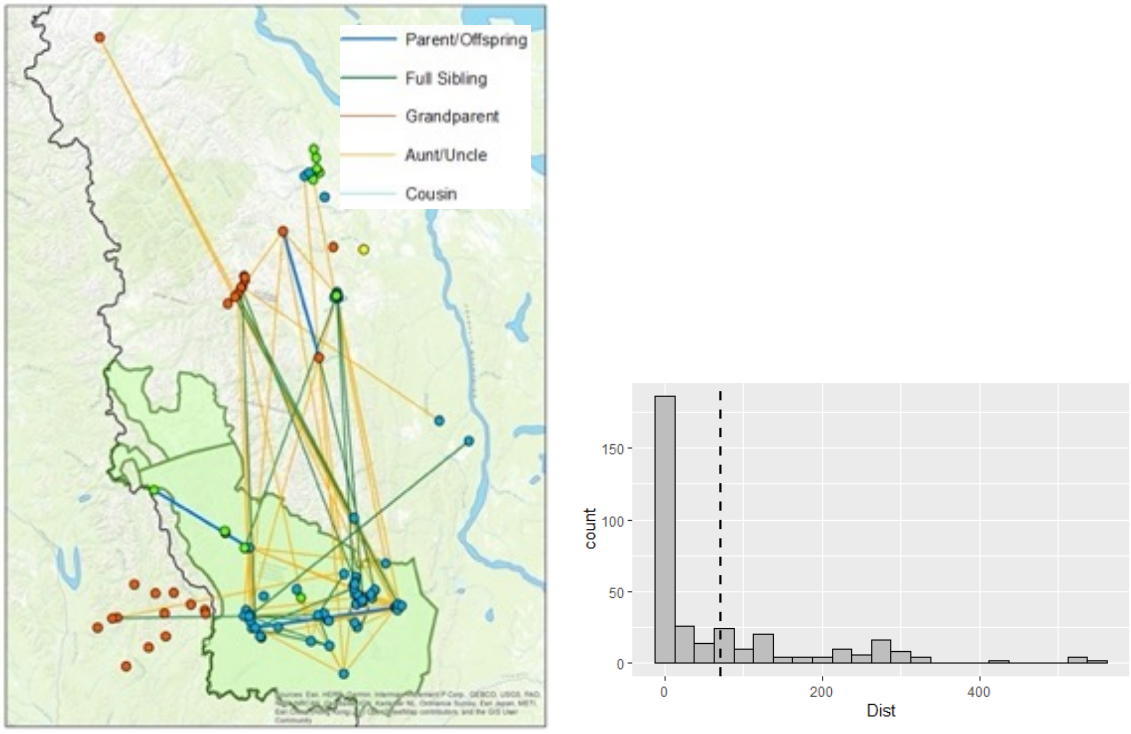


Figure 2.9. Individual-based pedigree network in the Mackenzie Mountains Region built with relationships inferred in COLONY as edges, and individual caribou as nodes. Nodes are colored according to season: winter (blue), fall (orange), spring (green), summer (yellow). Histogram depicts distribution of edge distance (km) calculated in ArcMap GIS software.

Discussion

Population-based Genetic Networks

First-Order Community Detection

The first-order population-based genetic network partitioned into the resulting 6 communities placed nodes within the three Designatable Units into separate communities, each of which were in separate disconnected components of the network. The nodes of different Designatable Units being in different components of the network, with no shared edges between them, demonstrated that there is little to no genetic exchange between the different components of the network representing the three Designatable Units detected using the Euclidean genetic distance metric, at the scale of the first-order network.

The Northern Mountain community aligned well with the boundaries of the Northern Mountain Designatable Unit, including the Redstone range to the north, the Nahanni range and the Coal River range (Environment Canada, 2012a). Many strong edges connecting the nodes within the boundaries of the Northern Mountain range with nodes adjacent to the current range boundary (specifically in the Prairie Creek mine area and further north on the west side of the Mackenzie River) displayed evidence of connectivity and gene flow reaching past the current Northern Mountain range. The Northern Mountain community extending east towards the Mackenzie River but having no edges cross over the river, east into the Boreal Designatable Unit suggested a significant role of the Mackenzie River as a driver of the structure of Northern Mountain caribou and the other Designatable units. The Barren-ground community also aligned well with the boundaries of the Barren-ground Designatable Unit in the study area

(specifically the ranges of Bluenose East and West), and had no edges connecting it to the rest of the network at this scale.

In contrast to the Northern Mountain and Barren-ground communities, there were four communities detected that spanned the range of the Boreal Designatable Unit (Mackenzie River, Great Slave Lake, Lake Athabasca, and Alberta). The first-order communities within the Boreal DU each comprised multiple local populations defined by the 2019 federal recovery strategy (Environment Canada, 2019) with the exception of the Mackenzie River community, which was entirely within the large Northwest Territories local population. The communities did however, corresponded well with the Taiga Plain, Boreal Plain, and Boreal Shield ecozone ranges (Environment and Climate Change Canada, 2019). The Mackenzie River and Great Slave Lake communities formed one large component in the network demonstrating evidence of genetic connectivity amongst local populations that reaches south to the Peace River. This genetic connectivity and structuring aligned with the two "discrete metapopulations" of Boreal Caribou described by McLaughlin et al. (2004) that are bisected by the Peace River. It should also be noted that the lack of edges between the Great Slake Lake community and the Alberta community further to the south could also be partially driven by the large spatial distance between sampling sites of those two communities (Figure 2.1) or by the phylogenetic signatures and evolutionary history shaped by past glacial cycles (Polfus et al., 2017; Taylor et al., 2021).

Second Order Community Structure

The second order networks built for each first-order community, and the resulting community detection run on those networks, further partitioned the nodes in each network

into different communities and components. These second order communities and components represent finer scale genetic structuring within the larger continuous populations.

There were three fine-scale second order communities detected in the Northern Mountain community. The first of these communities, in the south (orange nodes in Figure 2.3.2) aligned well with the Nahanni Herd Complex (Environment Canada, 2012a). The other two communities (green and blue nodes in Figure 2.3.2) aligned well with the Redstone Herd to the north and appears to show connectivity with the Tay River Herd (western most node in the Yukon) although sampling in the Tay River Herd range is currently limited. Interestingly, although the populations of caribou across the Yukon-Northwest Territories border are considered two separate herds (Tay River and Redstone) by Environment Canada(2012a), Indigenous local knowledge points towards more connected populations, facilitated by important areas such as K'á Té , which is thought to bring together at least five groups of caribou from both sides of the border (*Nío Nę P'ęnę́ Begháré Shúhta Goɔepé Narehɔ́á – Trails of the Mountain Caribou Plan*, 2019). Although the first-order Northern Mountain community could be partitioned into three different second order communities, all the communities formed one connected component, with many edges connecting the communities to one another. One node of particular interest was the node near the Prairie Creek Mine area (Figure 2.3.2: green node on the eastern side of Nahanni National Park Reserve); this node was composed of samples collected during the winter months, and appeared to form strong edges with nodes in both the Redstone Herd and the Nahanni Complex alluding to a wintering area shared by both the Redstone and Nahanni Herds. Local Indigenous knowledge additionally includes stories of caribou moving seasonally between the Drum Lake area,

down towards the northern border of Nahanni National Park Reserve (Prairie Creek Mine area) (L. Andrew, personal communication, January 25, 2022). Further evidence of this area being critical to the connectivity of the two groups of animals was its relatively high betweenness centrality value compared to the rest of the nodes in the Northern Mountain community (Figure 2.5.1). Betweenness centrality is a measure of the number of shortest paths between all nodes that travel through the given node, meaning nodes with high betweenness are important to the connectivity of the network and act as bridges between otherwise disconnected components of the network (Jones & Manseau, 2022). The betweenness value of the node in the Prairie Creek Mine area was 14 (Table 2.3: Node 25) while the mean betweenness for all nodes in the Northern Mountain community was only 2.89 (Table 2.1).

There were three fine-scale second order communities detected in the Barren-ground second order network (Figure 2.3.1). The communities formed three disconnected components, one community of three nodes (red: Figure 2.3.1), one community of a solitary node (pink: Figure 2.3.1), and another community of three nodes that had spatial overlap with the other communities (green: Figure 2.3.1). The spatial overlap of communities was to be expected given the migratory and variable behavior of Barren-ground caribou (COSEWIC, 2011; Nagy et al., 2011). These communities comprising disconnected components may be a result of the large time span and seasonality in which the Barren-ground samples were collected (spanning 29 years), especially given that migratory caribou are known to change behavior and seasonal movements (Bergerud et al., 2008).

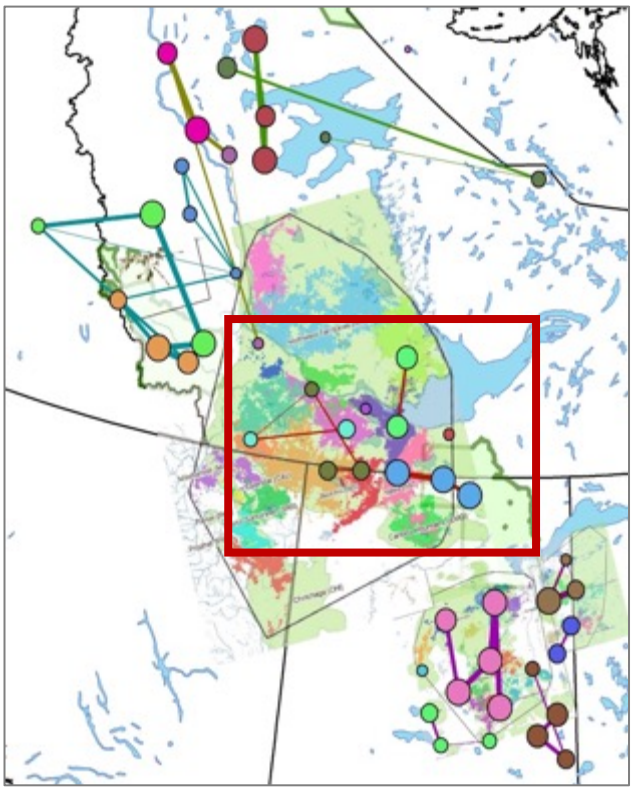
There was fine-scale structure detected within the two first-order communities in the Boreal range which aligned with known population structure uncovered from recent

telemetry studies (Wilson et al., 2020 & 2022). The Mackenzie River community was partitioned into two well connected second order communities (Figure 2.3.3). South of the Mackenzie River community was the Great Slave Lake community which was partitioned into six second order communities making up 5 different disconnected components (Figure 2.3.4). In the west there were two spatially overlapping communities (dark green and teal: Figure 2.3.4) which made up one connected component, in the northeast two nodes made up a community (lime green: Figure 2.3.4) that transverse the western arm of Great Slave Lake, to the south was a community made up of three nodes (blue: Figure 2.3.4), and lastly there were two disconnected solitary nodes (purple and pink: Figure 2.3.4). The second order communities of the Great Slake Lake area aligned well with a recent telemetry study (based on collared females) that revealed nested population structure (Wilson et al. 2022) (Figure 2.10) which showed multiple overlapping ranges in the east, two overlapping ranges that transverse the western arm of Great Slave Lake, and a larger group in the south that had no overlap with any other groups.

South of the Peace River there were two more first-order communities in the Boreal Designatable unit range that also aligned with known nested structuring (Wilson et al., 2022): the Alberta community and the Lake Athabasca community. The Lake Athabasca community had no apparent second-order community structure, with the three nodes belonging to one connected community (Figure 2.3.5). Lasty, the Alberta community was partitioned into six different communities making up six different disconnected components (Figure 2.3.6) which aligned very well with the first-order grouping from the recent telemetry results by Wilson et al. (2022) (Figure 2.11). The first and largest community was composed of 5 nodes (pink: Figure 2.3.6) and aligned with

the largest telemetry group (yellow polygon: Figure 2.11B). The second community was composed of four nodes (brown: Figure 2.3.6) and aligned with a cluster of groups in the Wilson et al. (2022) telemetry results that had overlapping spatial use which could result in gene flow. The remaining three communities comprised three nodes (green: Figure 2.3.6), two nodes (dark blue: Figure 2.3.6), and one node (light blue: Figure 2.3.6), all of which aligned with different first-order telemetry groups. The correspondence of population structure based on genetics (Figures 2.3.4-2.3.6) and movements (Wilson et al., 2022) corroborates the ability to detect broad and fine-scale genetic structure using population-genetic networks. Using a combination of telemetry and genetic network results in an area can help highlight differences between overlapping range use and genetic exchange, as well as aid in confirming isolated populations that have differing range uses and share little genetic exchange.

A.



B.

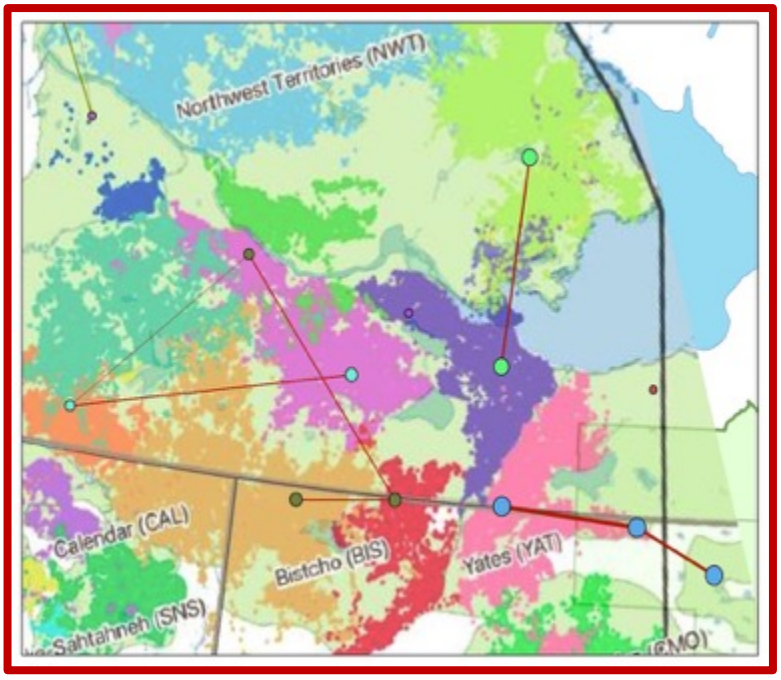
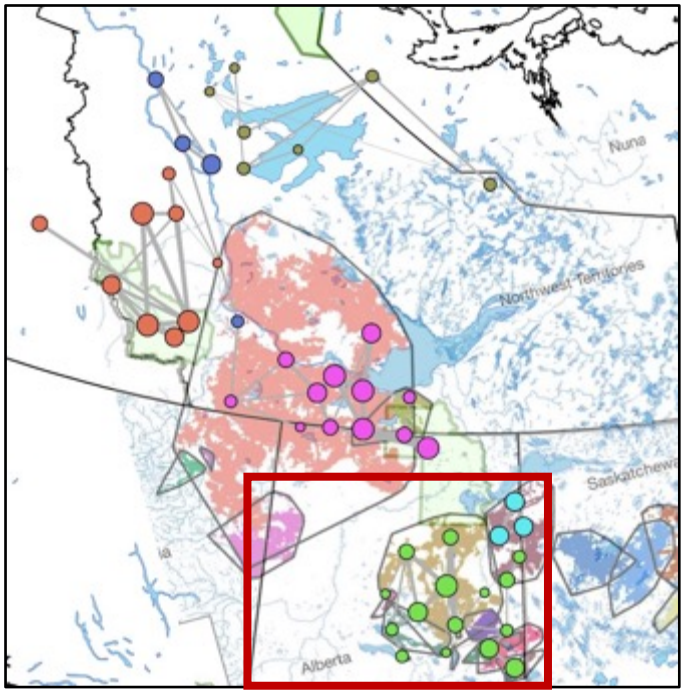


Figure 2.10. Second order genetic communities detected from population-based genetic networks overlayed onto second order groups detected in recent telemetry study conducted by Wilson et al. (2022): *Nested Population Structure of Threatened Boreal Caribou Revealed by Spatial Structuring. Ecological Modeling, in review.*

A.



B.

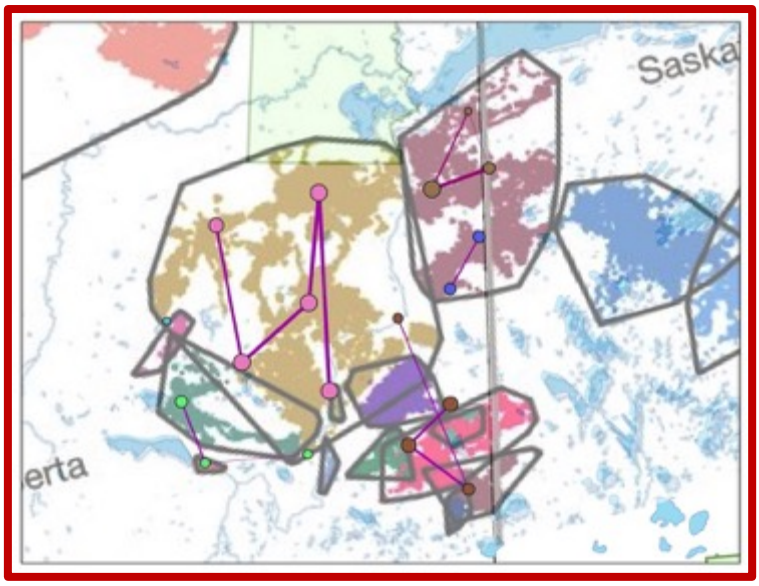


Figure 2.11. First-order (A.) and second order (B.) genetic communities detected from population-based genetic networks overlayed onto first-order groups detected in recent telemetry study conducted by Wilson et al. (2022): *Nested Population Structure of Threatened Boreal Caribou Revealed by Spatial Structuring. Ecological Modeling, in review.*

Node-based Metrics and Summary Statistics

Node-based metrics are susceptible to influence by the inherent topology of the network at hand (Opsahl et al., 2010), which is an important consideration when making inferences based on node-based metrics alone. Two metrics that appeared to differ amongst communities based on the topology of the network were degree centrality and closeness centrality. Degree centrality appeared to be strongly affected by the size of the community, with nodes in larger communities having larger degrees (Northern Mountain, Barren-Ground, Great Slave Lake, and Alberta) and nodes in smaller communities (Lake Athabasca and Mackenzie River) having smaller degrees (Figure 2.4A & Table 2.1). Closeness centrality demonstrated a similar pattern but was also driven by the size of the network component of which the community was a part, this was because closeness centrality is calculated as the reciprocal sum of the shortest path between the given node and all other nodes (Jones & Manseau, 2022) and is therefore greatly affected by the network being partitioned into differing components. Communities in large components of the network (Mackenzie River, Great Slave Lake, and Alberta) had nodes with larger closeness centralities, in contrast to communities that are part of smaller components (Lake Athabasca, Barren-Ground, and Northern Mountain) which had smaller closeness centralities (Figure 4D and Figure 2.5.2). The mean inverse edge weight of nodes in the different communities appeared to be fairly consistent, apart from the Barren-Ground community which had nodes with lower mean inverse edge weights (Figure 2.4B). The mean inverse edge weight is correlated with how similar a node is to other nodes to which it is connected, and is correlated with the numbers of migrants and immigrants in a population (Koen et al., 2016; Savary et al., 2021b). Therefore, the reduced mean inverse edge weights of nodes in the Barren-Ground community could potentially be a result of

the increased time span the samples were collected over. These results again point to the importance of sampling design, especially when sampling migratory animals that have variable behavior.

Betweenness centrality is the node-metric of most interest in this study. Betweenness centrality is a measure of the number of shortest paths between all nodes that travel through the given node (Jones & Manseau, 2022), and indicates key populations to metapopulation structure that act as bridges between otherwise disconnected components (Rozenfeld et al., 2008). In the context of this study, nodes with high betweenness centrality indicate groups of animals that facilitate gene flow between different and otherwise relatively disconnected groups of animals. There were different patterns of betweenness centrality values between communities that represented populations that are migratory (Northern Mountain and Barren-Ground) and communities that were part of the more sedentary Boreal Designatable Unit (Mackenzie River, Great Slave Lake, and Alberta). The Northern Mountain and Barren-Ground communities had nearly all nodes with very small betweenness centrality, whereas nodes in the communities making up the Boreal DU had a larger distribution of betweenness centrality values, with many nodes having had high values (Figure 2.4C and Figure 2.5.1); see Appendix B for permutation results that tested the significance of this pattern. The exception to this pattern was the Lake Athabasca community which had very low betweenness centrality values for all nodes; however this was due to the very small size of the community which was only composed of 3 nodes that were all connected to one another. The small community size and low betweenness values of the Lake Athabasca community could also be attributed to the community being adjacent of the edge of the study area. The higher betweenness centrality values of nodes in the Boreal DU could be

attributed to the more sedentary behavior of Boreal caribou. Boreal caribou, which move less on the landscape (Bergerud et al., 2008; Pond et al., 2016), display stronger patterns of Isolation By Distance (Priadka et al., 2019), which would result in nodes acting as bridges to gene flow between otherwise disconnected nodes (higher betweenness). In contrast, animals that are migratory and move long distances on undisturbed landscapes such as Northern Mountain and Barren-Ground caribou (COSEWIC, 2011) would be expected to have fewer populations that act as sole/key nodes to gene flow since their connections would be more evenly distributed in their respective component of the network.

Expected heterozygosity (He), observed heterozygosity (Ho), and number of different alleles (Na) values were all highest for caribou in the Barren-ground and Northern Mountain communities, with the four communities in the Boreal range having nodes with lower values for He, Ho, and Na (Table 2.2). The southernmost community (Alberta) had the lowest values of He, Ho, and Na. Although not a focus of this study, a general north to south pattern was displayed for He, Ho, and Na, potentially alluding to a general pattern of reduced genetic diversity for southern communities in the Boreal DU (Thompson et al., 2019). Further testing is required to quantify a relationship between measures of genetic diversity, and spatial proximity to the southern periphery of the study area.

Individual-based Genetic Networks

Both the individual-based genetic relatedness networks and the individual-based pedigree networks for the Mackenzie Mountains Region depicted a picture of a highly connected region with a mix of short and long distance migratory and dispersal

movements. Edges connecting individuals that were sampled in different seasons could be representative of migratory movements, since caribou in this region are known to seasonally move long distances (Environment Canada, 2012a). In contrast, when individuals that are closely related (such as parents and their offspring) are sampled in the same season but in two different locations, it can be inferred that a dispersal event or a change in seasonal habitat use occurred at some time in the recent past, although the direct path between the two locations is unknown. The majority of edges in both individual-based networks were between nodes that were geographically close to one another, with half of all edges between nodes less than 6 km away from one another. Both networks had edges that spanned large distances, with the individual-based relatedness network having close relationships spanning up to 380 km apart, and the pedigree network having relationships spanning up to 538 km apart. The amplitude of long-distance edges in both of these individual-based networks despite the relatively low proportion of the population that was sampled demonstrates the overall connectivity of the Mackenzie Mountains region which appears to facilitate long distance migratory and dispersal movements.

The value of individual-based genetic networks to highlight fine-scale, contemporary movements is further supported by local indigenous knowledge. During collaborative analysis of these results, a Mountain Dene Expert stated that historically, there wasn't much north/south movement north of Drum [Wrigley] Lake; however in recent years there has been an increase in the number of tracks crossing the Keele River (L. Andrew, personal correspondence, January 25, 2022); these movements appeared to be represented in the individual-based networks. Similar to the population-based networks, the Prairie Creek Mine area in the southeastern region of Nahanni National

Park Reserve contained many individuals that had edges connecting them to caribou north in the Redstone range and west in the Nahanni Complex range, demonstrating that this area is used as a wintering area by both Redstone and Nahanni animals. This area of overlapping boundaries of Nahanni and Redstone ranges could facilitate the gene flow that appeared in the population-based networks, with strong edges between the node in that region, and nodes in both the Redstone communities to the north and the Nahanni community to the west. Furthermore, this pattern of connectivity amongst the Redstone and Nahanni communities being facilitated by the Prairie Creek area was also supported by the elevated betweenness centrality of the node in this region, as discussed above.

Conclusion

Using a combination of population-based and individual-based genetic networks allows for investigation of connectivity and movement of caribou at multiple spatial and temporal scales. Population-based genetic networks are a useful tool to detect hierarchical genetic structure, as shown through the broad scale structure of caribou at the Designatable Unit level, as well as the finer scale structure of caribou more akin to the local population/herd level which is often investigated with telemetry. These various spatial scales coincide with multiple temporal scales, with the first-order communities representing a more evolutionary genetic signature and the second-order communities moving towards a more contemporary genetic structure. Node-based metrics such as betweenness centrality can highlight key areas and populations that facilitate movement and connectivity between different groups of animals. When using node-based metrics, it is also important to remember that they can vary with the topology of the network, as was shown in my analysis with degree centrality, closeness centrality, and mean inverse edge

weight measures. When sufficient genetic resolution is available, using individual-based networks in conjunction with population-based measures is useful for uncovering movement at a finer spatial and temporal scale. Individual-based genetic networks provide a much more contemporary snapshot of first order relationships across the landscape – adding yet another temporal scale to these analyses. Individual-based networks, although limited by the degree of sampling, can aid in demonstrating an area's ability to facilitate different types of movement. In study areas that hold significant biological as well as cultural value such as the Mackenzie Mountains Region, incorporating indigenous and local knowledge can greatly aid in interpreting the results and in increasing their biological relevance and acceptance (Gavin et al., 2015; Polfus et al., 2016). In the Mackenzie Mountains, increasing sampling and expanding the individual-based networks in continued collaboration with Indigenous knowledge holders and collaborators will further our biological understanding and inform stewardship efforts.

References

Andrews, T. D., MacKay, G., & Andrew, L. (2012). Archaeological Investigations of Alpine Ice Patches in the Selwyn Mountains, Northwest Territories, Canada. *Arctic*, *65*, 1-21.

Andrews, T. D., MacKay, G., Andrew, L., Stephenson, W., Barker, A., Alix, C., & Shuhtagot'ine Elders, T. (2012). Alpine Ice Patches and Shuhtagot'ine Land Use in the Mackenzie and Selwyn Mountains, Northwest Territories, Canada. *Arctic*, *65*, 22-42.

Baguette, M., Blanchet, S., Legrand, D., Stevens, V. M., & Turlure, C. (2013). Individual dispersal, landscape connectivity and ecological networks. *Biological Reviews*, *88*(2), 310-326. https://doi.org/10.1111/brv.12000

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

Bastille-Rousseau, G., Douglas-Hamilton, I., Blake, S., Northrup, J. M., & Wittemyer, G. (2018). Applying network theory to animal movements to identify properties of landscape space use. *Ecological Applications*, *28*(3), 854-864. https://doi.org/10.1002/eap.1697

Bergerud, A. T. (1978). *The Status and Management of Caribou in British Columbia.* Victoria: B.C. Fish and Wildlife Branch Report.

Bergerud, A. T. (1996). Evolving perspectives on caribou population dynamics, have we got it right yet? *Rangifer*, *16*(4), 95. https://doi.org/10.7557/2.16.4.1225

Bergerud, A. T., Butler, H. E., & Miller, D. R. (1984). Antipredator tactics of calving caribou - dispersion in mountains. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, *62*(8), 1566-1575. https://doi.org/10.1139/z84-229

Bergerud, A. T., Luttich, S. N., & Camps, L. (2008). The return of caribou to Ungava. Montreal, QC: McGill-Queen's University Press.

Bishop, M. D., Kappes, S. M., Keele, J. W., Stone, R. T., Sunden, S. L. F., Hawkins, G. A., Beattie, C. W. (1994). A Genetic-Linkage Map For Cattle. *Genetics*, *136*(2), 619-639.

Bradburd, G. S., & Ralph, P. L. (2019). Spatial population genetics: it's about time. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 427-449. https://doi.org/10.1146/annurev-ecolsys-110316-022659

COSEWIC. (2011). Designatable Units for caribou (*Rangifer tarandus*) in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. 88pp.

COSEWIC. (2014). COSEWIC assessment and status report on the *Caribou Rangifer tarandus*, Northern Mountain population, Central Mountain population and Southern Mountain population in Canada. Ottawa. Retrieved from www.registrelep-sararegistry.gc.ca/default_e.cfm

Cronin, M. R. A., MacNeil, M. D., & Patton, J. C. (2005). Variation in mitochondrial DNA and microsatellite DNA in caribou (*Rangifer tarandus*) in North America. *Journal of Mammalogy*, *86*(3), 495-505. https://doi.org/10.1644/1545-1542(2005)86[495:vimdam]2.0.co;2

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems,* 1695. https://igraph.org

De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011). Generalized Louvain method for community detection in large networks. *11th International Conference on Intelligent Systems Design and Applications, 2011.* https://arxiv.org/abs/1108.1502

Dyer, R. J., & Nason, J. D. (2004). Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, *13*(7), 1713-1727. https://doi.org/10.1111/j.1365-294x.2004.02177.x

Environment Canada. (2011). Scientific assessment to support the identification of critical habitat for woodland caribou (*Rangifer tarandus caribou*), boreal population in Canada. Ottawa, ON. 101pp. plus appendices. Retrieved from: https://www.registrelep-sararegistry.gc.ca/

Environment Canada. (2012a). Management plan for the Northern Mountain Population of Woodland Caribou (*Rangifer tarandus caribou*) in Canada. Ottawa Retrieved from https://wildlife-species.canada.ca/species-risk-registry

Environment Canada. (2012b). Recovery strategy for the woodland caribou *(Rangifer tarandus caribou)* boreal population, in Canada. Species at Risk Act Recovery Strategy Series. Environment Canada, Ottawa. vi + 55 pp. Retrieved from: https://www.registrelep-sararegistry.gc.ca/

Environment and Climate Change Canada. (2019). Amended Recovery Strategy for the Woodland Caribou (Rangifer tarandus caribou), Boreal population, in Canada [Proposed]. *Species at Risk Act* Recovery Strategy Series. Environment and Climate Change Canada, Ottawa. Retrieved from: https://www.registrelep-sararegistry.gc.ca/

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479-491. https://doi.org/10.1093/genetics/131.2.479

Farine, D. R., & Whitehead, H. (2015). Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology*, *84*(5), 1144-1163. https://doi.org/10.1111/1365-2656.12418

Ferguson, S. H., & Elkie, P. C. (2004). Seasonal movement patterns of woodland caribou (*Rangifer tarandus caribou*). *Journal of Zoology*, *262*(2), 125-134. https://doi.org/10.1017/s0952836903004552

Fischer, J., & Lindenmayer, D. B. (2007). Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, *16*(3), 265-280. https://doi.org/10.1111/j.1466-8238.2007.00287.x

Garroway, C. J., Bowman, J., Carr, D., & Wilson, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, *1*(4), 620-630. https://doi.org/10.1111/j.1752-4571.2008.00047.x

Gavin, M. C., McCarter, J., Mead, A., Berkes, F., Stepp, J. R., Peterson, D., & Tang, R. F. (2015). Defining biocultural approaches to conservation. *Trends in Ecology & Evolution*, *30*(3), 140-145. https://doi.org/10.1016/j.tree.2014.12.005

Greenbaum, G., Rubin, A., Templeton, A. R., & Rosenberg, N. A. (2019). Network-based hierarchical population structure analysis for large genomic datasets. *Genome Research, 29.* Cold Spring Harbor Laboratory. http://genome.cshlp.org/

Greenbaum, G., Templeton, A. R., & Bar-David, S. (2016). Inference and analysis of population structure using genetic data and network theory. *Genetics*, *202*(4), 1299-1312. https://doi.org/10.1534/genetics.115.182626

Gullickson, D., & Manseau, M. (2000). *South Nahanni woodland caribou herd seasonal range use and demography*. Parks Canada Agency.

Hanski, I., & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, *404*(6779), 755-758. https://doi.org/10.1038/35008063

Heard, D. C., & Vagt, K. L. (1998). Caribou in British Columbia: A 1996 status report. *Rangifer*, *18*(10), 7. https://doi.org/10.7557/2.18.5.1548

Ion, P. G., & Kershaw, G. P. (1989). The Selection Of Snowpatches As Relief Habitat By Woodland Caribou (Rangifer-Tarandus-Caribou), Macmillan Pass, Selwyn Mackenzie Mountains, Nwt Canada. *Arctic and Alpine Research*, *21*(2), 203-211. https://doi.org/10.2307/1551633

Jones, O. R., & Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular ecology resources*, *10*(3), 551-555. https://doi.org/10.1111/j.1755-0998.2009.02787.x

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

Klutsch, C. F. C., Manseau, M., Trim, V., Polfus, J., & Wilson, P. J. (2016). The eastern migratory caribou: the role of genetic introgression in ecotype evolution. *Royal Society Open Science*, *3*(2), 13, Article 150469. https://doi.org/10.1098/rsos.150469

Koen, E. L., Bowman, J., & Wilson, P. J. (2016). Node-based measures of connectivity in genetic networks. *Molecular Ecology Resources*, *16*(1), 69-79. https://doi.org/10.1111/1755-0998.12423

Leroy, G., Carroll, E. L., Bruford, M. W., Dewoody, J. A., Strand, A., Waits, L., & Wang, J. (2018). Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evolutionary Applications*, *11*(7), 1066-1083. https://doi.org/10.1111/eva.12564

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

McLoughlin, P. D., Paetkau, D., Duda, M., & Boutin, S. (2004). Genetic diversity and relatedness of boreal caribou populations in western Canada. *Biological Conservation*, *118*(5), 593-598. https://doi.org/10.1016/j.biocon.2003.10.008

Mimura, M., Yahara, T., Faith, D. P., Vázquez-Domínguez, E., Colautti, R. I., Araki, H., Hendry, A. P. (2017). Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications*, *10*(2), 121-139. https://doi.org/10.1111/eva.12436

Murphy, M., Dyer, R., & Cushman, S. A. (2015). Graph theory and network models in landscape genetics. *Landscape Genetics* (pp. 165-180). https://doi.org/https://doi.org/10.1002/9781118525258.ch10

Nagy, J. A., Johnson, D. L., Larter, N. C., Campbell, M. W., Derocher, A. E., Kelly, A., . . . Croft, B. (2011). Subpopulation structure of caribou (*Rangifer tarandus L.)* in arctic and subarctic Canada. *Ecological Applications*, *21*(6), 2334-2348. https://doi.org/10.1890/10-1410.1

Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E*, *67*(2). https://doi.org/10.1103/physreve.67.026126

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577-8582. https://doi.org/10.1073/pnas.0601602103

Newman, M. E. J. (2010). *Networks: an introduction*. Oxford University Press.

Nío Nę P'ęnę́ Workng Group. *(2019). Nío Nę P'ęnę́ Begháré Shúhta Goɔepé Narehɔá – Trails of the Mountain Caribou Plan*. Tulita, NT.

Opsahl, T., Agneessens, F., & Skoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245-251.

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics*, *28*(19), 2537-2539. https://doi.org/10.1093/bioinformatics/bts460

Pew, J., Muir, P. H., Wang, J., & Frasier, T. R. (2015). related: an R package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources*, *15*(3), 557-561. https://doi.org/10.1111/1755-0998.12323

Polfus, J. L., Hebblewhite, M., & Heinemeyer, K. (2011). Identifying indirect habitat loss and avoidance of human infrastructure by northern mountain woodland caribou. *Biological Conservation*, *144*(11), 2637-2646. https://doi.org/10.1016/j.biocon.2011.07.023

Polfus, J. L., Manseau, M., Klütsch, C. F. C., Simmons, D., & Wilson, P. J. (2017). Ancient diversification in glacial refugia leads to intraspecific diversity in a Holarctic mammal. *Journal of Biogeography*, *44*(2), 386-396. https://doi.org/10.1111/jbi.12918

Polfus, J. L., Manseau, M., Simmons, D., Neyelle, M., Bayha, W., Andrew, F., Andrew, L., Klütsch, C., Rice, K., Wilson, P. (2016). Leghagotsenete (learning together): the importance of indigenous perspectives in the identification of biological variation. *Ecology and Society*, *21*(2), 35, Article 18. https://doi.org/10.5751/es-08284-210218

Pond B. A., G. S. Brown, K. S. Wilson, and J. A. Schaefer. 2016. Drawing lines: Spatial behaviours reveal two ecotypes of woodland caribou. *Biological Conservation*, 194**:**139-148. https://doi.org/10.1016/j.biocon.2015.12.005

Rozenfeld, A. F., Arnaud-Haond, S., Hernández-García, E., Eguíluz, V. M., Serrão, E. A., & Duarte, C. M. (2008). Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences*, *105*(48), 18824-18829. https://doi.org/doi:10.1073/pnas.0805571105

Savary, P., Foltête, J. C., Moal, H., Vuidel, G., & Garnier, S. (2021a). Analysing landscape effects on dispersal networks and gene flow with genetic graphs. *Molecular Ecology Resources*, *21*(4), 1167-1185. https://doi.org/10.1111/1755-0998.13333

Savary, P., Foltête, J. C., Moal, H., Vuidel, G., & Garnier, S. (2021b). graph4lg: A package for constructing and analysing graphs for landscape genetics in R. *Methods in Ecology and Evolution*, *12*(3), 539-547. https://doi.org/10.1111/2041-210x.13530

Taylor, R. S., Manseau, M., Klütsch, C. F. C., Polfus, J. L., Steedman, A., Hervieux, D., Kelly, A., Larter, N., Gamberg, M., Schwantje, H., Wilson, P. J. (2021).

Population dynamics of caribou shaped by glacial cycles before the last glacial maximum. *Molecular Ecology*, *30*(23), 6121-6143. https://doi.org/10.1111/mec.16166

Van Dyck, H., & Baguette, M. (2005). Dispersal behaviour in fragmented landscapes: Routine or special movements? *Basic and Applied Ecology*, *6*(6), 535-545. https://doi.org/10.1016/j.baae.2005.03.005

Van Strien, M. J., Holderegger, R., & Van Heck, H. J. (2015). Isolation-by-distance in landscapes : Considerations for landscape genetics. *Heredity*, *114*(1), 27-37. https://doi.org/10.1038/hdy.2014.62

Wang, J. (2002). An Estimator for Pairwise Relatedness Using Molecular Markers. *Genetics*, *160*(3), 1203-1215. https://doi.org/10.1093/genetics/160.3.1203

Wilson, G. A., Strobeck, C., Wu, L., & Coffin, J. W. (1997). Characterization of microsatellite loci in caribou Rangifer tarandus, and their use in other artiodactyls. *Molecular Ecology*, *6*(7), 697-699. https://doi.org/10.1046/j.1365-294X.1997.00237.x

Wilson, S. F., Crosina, W., Dzus, E., Hervieux, D., McLoughlin, P.D., Trout, L.M., Nudds, T. D. (2022). Nested Population Structure of Threatened Boreal Caribou Revealed by Spatial Structuring. *Ecological Modeling, in review.*

Wilson, S. F., Sutherland, G., Larter, N., Kelly, A., McLaren, A., Hodson, J., Hegel, T., Steenweg, R., Hervieux, D., & Nudds, T. (2020). Spatial structure of boreal woodland caribou populations in northwest Canada. Rangifer, 40(1), 1–14. https://doi.org/10.7557/2.40.1.4902

Wright, S. (1965). The Interpretation Of Population Structure By F-Statistics With Special Regard To Systems Of Mating. *Evolution*, *19*(3), 395-420. https://doi.org/10.1111/j.1558-5646.1965.tb01731.x

Appendix B

Table B1. Sample collection details (province, Designatable Unit range, local populations or local areas where the samples were collected, years that the samples were collected, and the season the samples were collected) for all samples that were used within the genetic network analyses in Chapter 2.

Number Of Samples	Province	Designatable Unit	Local Populations/ Areas	Years	Season Collected
639	AB	Boreal	Cameron Hills, Cold lake, ESAR, Nipisi, Pine Point, Buffalo Lake, Red Earth, Richardson, Slave Lake, WSAR,	2005-2019	Winter
383	NT	Boreal/ Barren-ground	Cameron Hills, Decho North, Decho South, Hay River Lowlands, Mackenzie, Pine Point, Buffalo Lake, Sahtu, Wood Buffalo	2004-2016	Winter
175	NT	Barren-ground	Bathurst, Blunose	1991-2015	Winter
358	NT	Northern Mountain	Redstone, Nahanni	2012-2020	Winter
91	NT	Northern Mountain	Redstone, Nahanni	2013-2020	Fall
9	NT	Northern Mountain	Redstone, Nahanni	2003-2013	Spring/summer
26	YT	Northern Mountain	Pelly, Tay	2002-2010	Fall

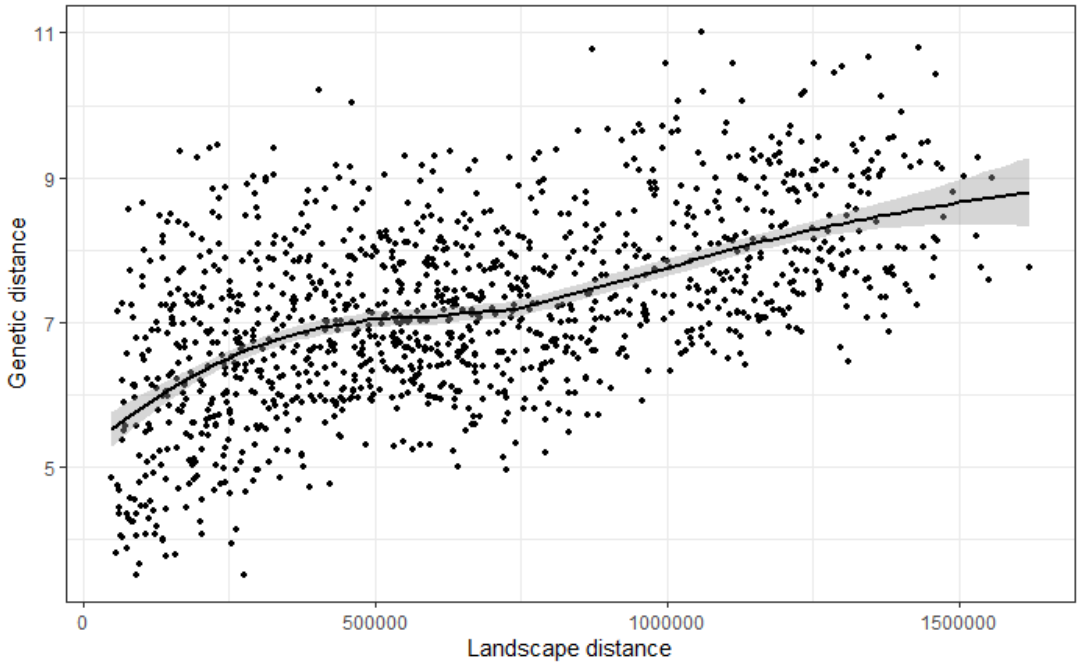


Figure B1. Scatter plot depicting Isolation By Distance (IBD) pattern of all 1681 samples genotyped at 10 loci that were used in the population-based genetic network. Euclidean distance is on the y axis and landscape distance (m) is on the x axis. A type IV pattern of IBD is shown as described by Hutchison and Templeton (1999).

Network permutations

Network permutations were run to test the statistical significance of the difference between node-based metric values between communities. A total of 1000 permutations were run in which the number of edges and edge values were kept constant, but the nodes that the edges connected were swapped. The community assignment for each node was also kept consistent. Statistical significance was found for the difference of betweenness values for the Northern Mountain and Boreal Communities. The Northern Mountain community was found to have significantly lower betweenness than expected (p=0.034), while the Barren-ground community was trending towards a lower betweenness than expected (p=0.064). See Figure B2 for visualization of the distribution of mean betweenness of the permutations and the actual mean betweenness for each community.

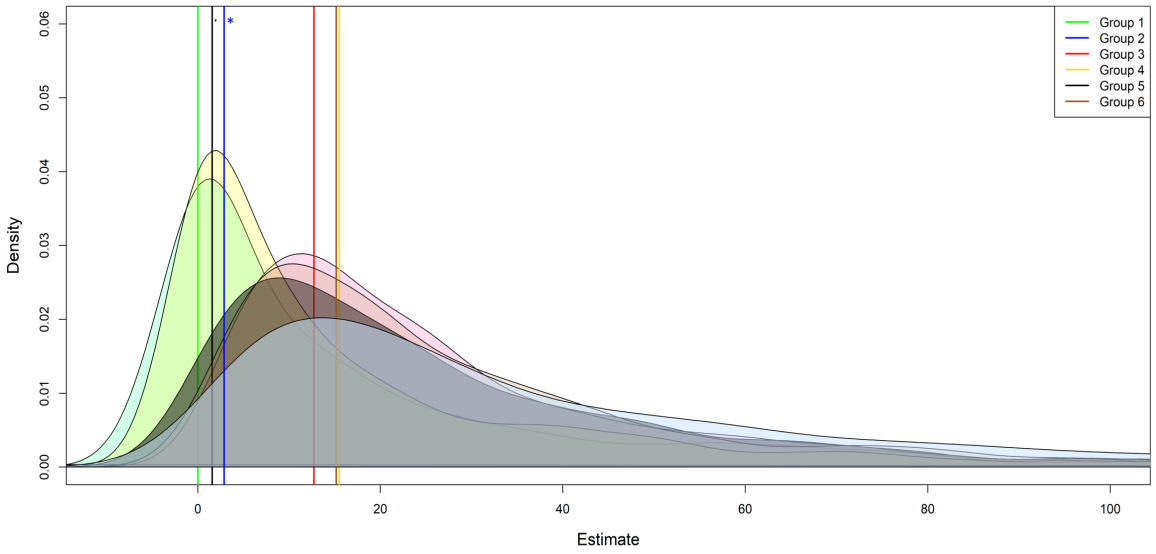


Figure B2. Density plots depicting the distribution of mean betweenness for each community in the 1000 permutations. Solid vertical line represents the actual mean betweenness for each community. Group 2 (Mackenzie Mountains) showed significantly lower betweenness values (P=0.034), whereas Group 5 (Baren-ground) was trending towards a lower betweenness than expected (P=0.064).

Table B2. Network wide metrics for both the individual-based networks

	Relatedness Network	**Pedigree Network**
Density	0.00116183	0.00333059
Global Clustering Coefficient	0.30357143	0.34375
Average Clustering Coefficient	0.36622935	0.51542988
Average Path Length	3.38751814	1.88411215
Disconnected Components	306	298
Modularity Score	0.9025	0.82258991
Assortativity	0.95586937	0.90414445

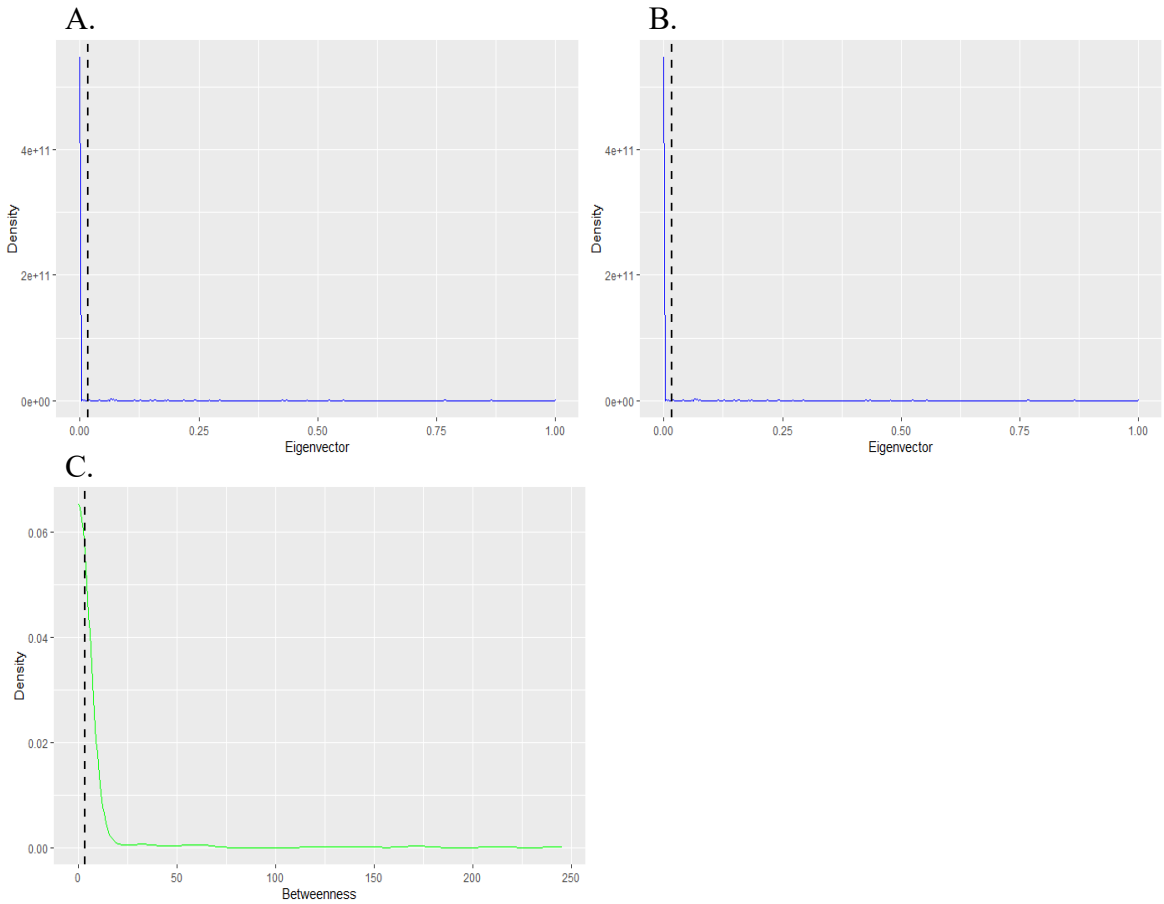


Figure B3. Density plots depicting the distribution of node-based metric values for (A) Degree Centrality, (B) Eigenvector Centrality, and (C) Betweenness Centrality for the individual-based relatedness network. Highly skewed towards 0 for all values, indicating a highly disconnected network due to a small proportion of the population being sampled.

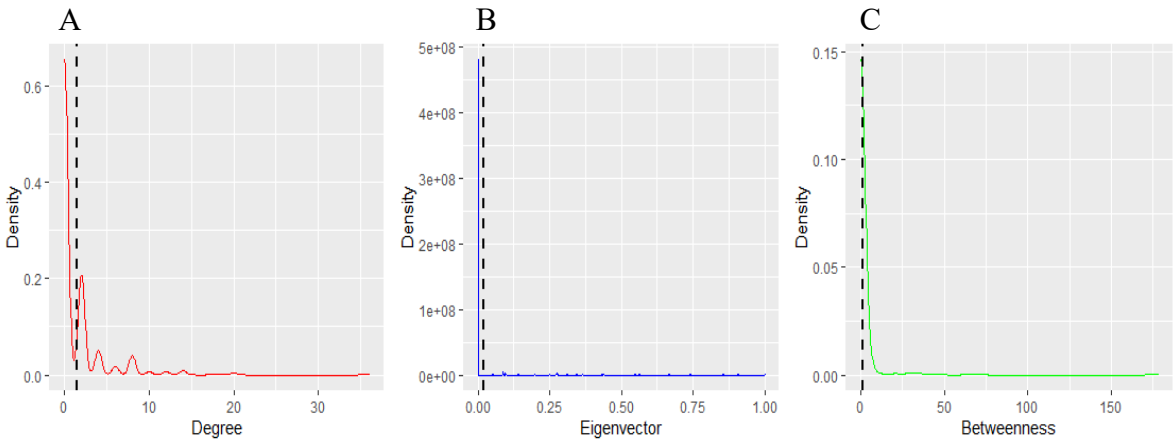


Figure B4. Density plots depicting the distribution of node-based metric values for (A) Degree Centrality, (B) Eigenvector Centrality, and (C) Betweenness Centrality for the individual-based pedigree network. Highly skewed towards 0 for all values, indicating a highly disconnected network due to a small proportion of the population being sampled.

General Conclusion

In remote areas where sampling is costly and logistically difficult and non-invasive sampling is preferred, using a combination of population-based and individual-based genetic networks allows for investigation of connectivity and movement of caribou at multiple scales.

The results of the first chapter of this study highlighted the importance of considering sampling efforts when using individual-based genetic relatedness networks; and that sampling efforts, although not always, can be a limiting factor depending on the type of metrics used and the questions that are trying to be answered. In addition to sampling efforts, genetic resolution is an important factor to consider when constructing individual-based genetic networks. The resolution of the pairwise relatedness values (Wang, 2002) estimator for caribou genotyped at 15 loci is sufficient to differentiate first-order relationships from unrelated individuals; however it was not sufficient to differentiate second order relationships from either first-order relationships or unrelated individuals. The inclusion of additional loci would be beneficial for future studies since it could result in genetic resolution sufficient to capture a higher degree of second order relationships (Foroughirad et al., 2019). Such additional loci could help offset the relatively low proportion of the population that is sampled in remote areas such as the Mackenzie Moutnains Region.

The second chapter uncovered hierarchical genetic structure of caribou in the study area. First-order structuring detected by population networks had concordance with COSEWIC's Designatable units (COSEWIC, 2011) and Indigenous knowledge (Polfus et al., 2016), showing discrete components of the network for Boreal caribou (tǫdzı),

Barren-ground caribou (ʔekwę́), and Northern Mountain caribou (shúhta ʔepę́). Both first and second order structure detected by the population networks aligned with recent telemetry work on boreal caribou in the Great Slave Lake area and northern Alberta (Wilson et al., 2022), as well as herd delineations of Northern Mountain Caribou in the Mackenzie Mountains Region (Environment Canada, 2012). Incorporating individual-based genetic relatedness network analysis supported the population-based networks in depicting the Mackenzie Mountains Region as a highly connected region, with long distance movements, and overlapping wintering regions of the Redstone herd and Nahanni Herd Complex.

These results highlight the ability to detect accurate hierarchical genetic structure using population genetic network analysis on non-invasively collected samples. Furthermore, using a combination of population-based and individual-based genetic network analysis methods proves to be useful when investigating population structure, connectivity, and movement of species, specifically when sampling levels and genetic resolution are not sufficient for individual-based genetic network analysis alone. With the expansion of sampling efforts and incorporation of additional samples with increased genetic resolution in the Mackenzie Mountains Region, expanding upon these genetic network results in continued collaboration with Indigenous knowledge holders and collaborators will further strengthen future results and inform stewardship efforts.

References

COSEWIC. (2011). Designatable Units for caribou (*Rangifer tarandus*) in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. 88pp.

Environment Canada. (2012). Management Plan for the Northern Mountain Population of Woodland Caribou (*Rangifer tarandus caribou*) in Canada. Ottawa, ON. Retrieved from: https://wildlife-species.canada.ca/species-risk-registry

Foroughirad, V., Levengood, A. L., Mann, J., & Frère, C. H. (2019). Quality and quantity of genetic relatedness data affect the analysis of social structure. *Molecular Ecology Resources*, *19*(5), 1181-1194. https://doi.org/10.1111/1755-0998.13028

Polfus, J. L., Manseau, M., Simmons, D., Neyelle, M., Bayha, W., Andrew, F., Andrew, L., Klütsch, C., Rice, K., Wilson, P. (2016). Leghagotsenete (learning together): the importance of indigenous perspectives in the identification of biological variation. *Ecology and Society*, *21*(2), 35, Article 18. https://doi.org/10.5751/es-08284-210218

Wang, J. (2002). An estimator for pairwise relatedness using molecular markers. *Genetics*, *160*(3), 1203-1215. https://doi.org/10.1093/genetics/160.3.1203

Wilson, S. F., Crosina, W., Dzus, E., Hervieux, D., McLoughlin, P.D., Trout, L.M., Nudds, T. D. (2022). Nested population structure of threatened boreal caribou revealed by spatial structuring. *Ecological Modeling, in review.*