ASSESSING MOLECULAR AND ECOLOGICAL DIFFERENTIATION IN WILD CARNIVORES

A Thesis Submitted to the Committee on Graduate Studies

In Partial Fulfillment of the Requirements for the Degree of Master of Science

in the Faculty of Arts and Science

Trent University

Peterborough, Ontario, Canada

Environmental and Life Sciences M.Sc. Graduate Program

January 2019

ABSTRACT

Assessing molecular and ecological differentiation in wild carnivores

Justin Brian Johnson

Wild populations are notoriously difficult to study due to confounding stochastic variables. This thesis tackles two components of investigating wild populations. The first examines the use of niche modeling to quantify macro-scale predator-prey relationships in canid populations across eastern North America, while the second examines range-wide molecular structure in Canada lynx. The goal of the first chapter is to quantify niche characteristics in a *Canis* hybrid zone of *C. lupus, C. lycaon,* and *C. latrans* to better understand the ecological differentiation of these species, and to assess the impacts of incorporating biotic interactions into species distribution models. The goal of the second chapter is to determine if DNA methylation, an epigenetic marker that modifies the structure of DNA, can be used to differentiate populations, and might be a signature of local adaptation. Our results indicated that canids across the hybrid zone in eastern North America exhibit low levels of genetic and ecological differentiation, and that the importance of biotic interactions are largely lost at large spatial scales. We also identified cryptic structure in methylation patterns in Canada lynx populations, which suggest signatures of local adaptation, and indicate the utility of DNA methylation as a marker for investigating adaptive divergence.

Keywords: ecological genetics, species distribution models, DNA methylation, landscape genomics, genotyping, population genetics, epigenetics

ACKNOWLEDGEMENTS

This thesis is the product of a multitude of contributors, the core of which come from the research groups of Professor Dennis Murray and Professor Aaron Shafer. Dr. Murray offered me the initial position to investigate niche space in wild canids and provided me with the freedom to pursue tangential projects and expand my academic expertise. One of these tangential projects became a second MSc chapter and provided the foundation for a strong and fruitful academic relationship with Dr. Shafer. I could not ask for a better supervisor dynamic, where I have been provided with the freedom to explore and the critical eye and wisdom to keep me from getting too lost in the weeds.

The members of the Murray and Shafer research groups have provided unequivocal support whether it be through experimental advice (Catarina Ferreira, Sibelle Torres Vilaça), statistical help (Elizabeth Kierepka, Lynne Beaty, Angela Eads, Daria Martchenko), brain-storming and sound-boarding (Melanie Boudreau, Peter Mills, Meghan Britt, Jessica Breen) or maintaining overall sanity (Jasper Leavitt, Patrick Heney, Lindsey Bargelt). I would also like to thank Dr. Joe Northrup from the MNRF for friendly paper reviews, and invaluable statistical advice.

This project is specifically made possible due to the previous work of several wolf researchers including Dr. Paul Wilson, Dr. Linda Rutledge, Tyler Wheeldon, Dr. John Benson, Dr. Brent Patterson, and Josee-Anne Otis. Their work in genotyping and organizing samples paved the way for a large-scale ecological genetic study on our canids. I received a good deal of assistance and advice for

the laboratory work of this thesis from the NRDPFC at Trent University, primarily Matthew Harnden and Nguyen Thi Xuan Nguyen, who patiently helped me understand the intricacies of canid genotyping at Trent. My second chapter would not exist without the patience, advice, and reagents of Niels Wagemaker from Radboud University, and the sequencing help from Sergio Pereira and TCAG at SickKids Hospital in Toronto.

I would like to thank the various administrators who allowed us entry to the fur auctions to collect samples (Vittorio Villacis – NAFA, Howard Noseworthy – FHA), as well as the individuals that had previously coordinated with Josee for sample collection (Pierre-Yves Collin, Joseph Bopp, Andy MacDuff, Kelly Leavesley). I would also like to thank WCS Canada and the W. Garfield Weston Foundation for their monetary contribution for genotyping, and the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs program. A special thanks to the EnLS program for providing an office and administrative assistance, and to Dr. James Conolly for his input on early project directions which helped lead to two unique chapters.

Above all, I would like to thank my family for their unyielding support, who although my research may never make complete sense to them, understand my passion for it and support me entirely. I especially want to thank Alice Pintaric, who continually endures my abstract rants on DNA and genomics and who is my absolute rock, who I owe a great deal due to the time investment I have spent into the following pages and the excessive amount of my life that I spend thinking about research.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2.

CHAPTER 3.

LIST OF FIGURES

CHAPTER 2.

SPECIAL NOTE

Chapters I and IV of the thesis are written in the singular '*I*' form, as these chapters are primarily my thoughts and reflections regarding the research outlined in chapters II and III. Chapters II and III are written in the plural '*we/our*' form, as these chapters are necessarily collaborative and involved the thoughts and actions of all those involved.

CHAPTER 1

GENERAL INTRODUCTION

GENERAL INTRODUCTION

The system: large scale patterning in high-latitudinal carnivores

Evolutionary processes affect all wild populations and are responsible for the diversity of organisms that exist on earth (Darwin, 1859). Although laboratory experiments in population genetics have revealed invaluable knowledge in evolutionary processes like the neutral model of evolution (Kimura, 1977), a majority of these findings surround a few model organisms, notably *Drosophila melanogaster* (Partridge et al., 1994) or *Arabidopsis thaliana* (The *Arabidopsis* Genome Initiative, 2000). Insights from these studies are invaluable for understanding heritability and population-level processes in molecular genetics, but empirical studies in wild populations are required to test theoretical hypotheses and integrate evolutionary theory into applied field research in adaptation. The methodology and theoretical framework generated from research in model organisms has subsequently provided the foundation for evolutionary research into non-model organisms and wild populations.

The Canada lynx (*Lynx canadensis*) and lupine canids (*C. lupus, C. lycaon, C. latrans*) exist across high latitudinal North American ecosystems and provide a fascinating system to examine gene flow, ecological speciation, and hybridization. Both groups have extensive dispersal capabilities and low overall levels of genetic differentiation due to high rates of gene flow across the continent (Roy et al., 1994; Row et al., 2012). Both groups have empirical evidence for genetic clustering due to climatic or environmental gradients (Geffen et al., 2004; Stenseth et al., 2004), and thus both provide useful systems to

examine macro-scale relationships between wild populations and their environment. Specifically, these systems provide a system to examine the genetic and epigenetic signatures of local adaptation (e.g. Canada lynx) and the framework for assessing the ability to infer macro-scale ecological characteristics of subtle genetic population structure across the landscape (e.g. North American lupine canids). Furthermore, both systems serve as opportunities for investigating a particular aspect of the ecological speciation continuum. The geographically-isolated Newfoundland lynx provide a natural system to examine the effects of drift and selection on a population that was likely isolated less than 10,000 years ago and might be in the early stages of ecological speciation. Alternatively, the canid system offers a glimpse into genotype-environment relationships at macro-scales, allowing for an ecological examination of putative species boundaries and the effects of hybridization on environmental associations to determine the ecological signatures of divergence.

Canids: hybridization and the ecological niche of genetic intermediates

The ancestral story of North American canids is riddled with conflicting perspectives (Roy et al., 1994; Wilson et al., 2000). All perspectives recognize that a large canid, the Holarctic gray wolf (*Canis lupus*), and a smaller canid known as the Coyote that is endemic to the American southwest (*C. latrans*), exist as distinct species across the North American continent. Previous estimates of lineage divergence using mitochondrial DNA estimated the C. *lupus* and C. *latrans* split at 1 million years (Wilson et al., 2000). However, whole-genome resequencing of 28 canids across North America has identified a divergence time of $T = 0.38\ N$ (with

current N_e of 45,000), equivalent to a ~50,000 year split (VonHoldt et al., 2016). Notably, a generation time of 3 years was used for this analysis, which is lower than expected gray wolf estimates (Mech et al., 2016), so the time split is most likely slightly earlier. Similarly, genetic differentiation, even between Eurasian and North American gray wolves, is notably low (F_{ST} = 0.099; VonHoldt et al. 2016a). Contemporary disagreements have primarily arisen out of identifying the ancestral and taxonomic status of an intermediate sized canid within the Great Lakes region, termed the eastern wolf (*C. lycaon*).

Although the advancement of molecular techniques and sequencing technology would intuitively aid in clarifying genetic subdivisions and ancestry, the integration of genomic data into *Canis* has seemed to further divide initial perspectives on the ancestry of the canid that exists in the Great Lakes region of eastern North America (Rutledge et al., 2010; vonHoldt et al., 2011; Rutledge et al., 2012, 2015; VonHoldt et al., 2016). Notably, although hybrids from unadmixed gray wolves and coyotes are not documented to occur naturally in the wild (Mech, 2011), their offspring are viable and fertile (Mech et al., 2014, 2017). Eastern wolves are known to readily hybridize with coyotes and gray wolves in sympatry (Kolenosky, 1971), and thus can facilitate genetic transfer between the two species (Rutledge et al., 2010).

One perspective on *Canis* in eastern North America, an area for our purposes delineated from Manitoba and Minnesota east through New England and Québec, is the historical occupation by a mid-sized canid (*C. lycaon*) (Nowak, 2002). This mid-sized canid has experienced contemporary (~ 200 years)

introgression from encroaching coyote (*C. latrans*) populations from their southwest range, as historical extirpation of the larger canids across the continent left vacant niche space for a canid-like predator (Lehman et al., 1991). This introgression has morphological and ecological consequences, where hybridization between coyotes and wolves has been shown to cause a shift in niche space in the expanding coyote populations (Thornton & Murray, 2014), with hybrids even postulated to exhibit intermediate niches between progenitors (Otis et al., 2017). Thus, understanding macro-scale ecological signatures from these groups can help clarify species boundaries, and assist in determining the unique ecological requirements for each group to inform conservation measures.

Despite the disagreements over the ancestry and taxonomy of *Canis* across eastern North America, the contemporary ecological role of apex predator in these ecosystems still falls largely on its lupine canid-like organisms. Thus, if specific canid groups have a unique ecological relationship with their prey, then the persistence of a specific canid group on the landscape might need to be prioritized to maintain functional ecosystems with a predator that can maintain prey populations. Despite the taxonomic ambiguity surrounding this system, the ecological relationships between predator and prey within the genetic subdivisions of canids is critical to understanding ecosystem function (i.e., prey population dynamics) and identifying habitat that is required to maintain adaptive potential.

To investigate macro-scale predator-prey relationships in this system, I first subdivided my *Canis* samples into genetic groups using neutral autosomal

microsatellite genotyping and Bayesian clustering as outlined in contemporary *Canis* genotyping methods (Wheeldon & White, 2009; Rutledge et al., 2010). I then used these georeferenced and genotyped canid samples to create species distribution models, or niche models, to identify macro-scale relationships between occurrence locations and environmental predictors for each canid group.

Species distribution models have been used in a variety of contexts, notably to predict suitable habitat for reintroductions (Malone et al., 2018), examine species-environment relationships (Barbet-Massin & Jetz, 2014), or to predict distributional shifts due to changing climates (Bay et al., 2018). A contentious issue in the modelling community is the importance of incorporating biotic environmental predictors into distribution models (Araújo & Luoto, 2007; Araújo & Rozenfeld, 2014; De Araújo et al., 2014; Fraterrigo et al., 2014; Anderson, 2017). Research on the importance of competition in predicting species distributions has been done using simulations (Godsoe et al., 2016), which revealed that it is ultimately difficult to disentangle the biotic and abiotic factors controlling species distributions. However, some empirical studies have suggested that biotic interactions are important for shaping and predicting species distributions, even at macro-scales (Heikkinen et al., 2007; Aragón & Sánchez-Fernández, 2013; De Araújo et al., 2014).

Larger canids (i.e. *C. lycaon, C. lupus*) are highly dependent on their ungulate prey, where moose are frequently their primary food source (Bergerud et al., 1983; Peterson et al., 1984). Smaller canids (i.e. *C. latrans*) tend to specialize on

smaller prey such as rodents and even occasionally eating insects (Dumond et al., 2001). However, coyotes are documented to have a regulatory effect on white-tailed deer (*Odocoileus virginianus*) populations (Kilgo et al., 2010), and have even been seen taking down adult moose (Benson & Patterson, 2013a). The dependency of a predator on its prey facilitated our assessment herein, where I determined predator-prey associations across canid groups by creating SDMs for white-tailed deer and moose, and then integrating these variables into our canid models. We expected to find a gradient in the strength and direction of the response, with gray wolves exhibiting the largest increase in model fit when including moose habitat variables, while moose exhibited the smallest increase in model fit for coyotes. Overall, I sought to examine macro-scale genetic and ecological differentiation in canids across eastern North America, while also assessing the impacts of incorporating biotic interactions in species distribution models.

The main objectives of this chapter were thus to i) determine genetic structure and macro-scale ecological differentiation among canids in the admixture zone, and to ii) determine if prey variables (biotic interactions) are disproportionately important for predicting canid predator distributions, with a prediction that we see a gradient in the strength of the response dependent on prey and predator size. These results will therefore inform the ongoing *Canis* debate by comparing macro-scale ecological determinants of canid distributions in a hybrid zone, while providing an empirical study that assesses the effects of incorporating biotic interactions into ecological niche models.

The Canada lynx: climatically-associated population structure

Increased relative fitness of one conspecific population to another in a specific environment due to its phenotype is the basis of evolutionary local adaptation (Levene, 1953). Evolutionary processes like mutation (Schluter, 2009), drift (Uyeda et al., 2009), natural selection (Rundle et al., 2000), and ecological processes like migration and sexual selection (Lande, 1981) are the primary drivers of reproductive isolation, divergence, and ecological speciation (Turelli et al., 2001; Wolf et al., 2010; Shafer & Wolf, 2013). Divergence, however, is difficult to detect in systems experiencing high rates of migration (although speciation can still occur, see Nosil 2008, Feder et al. 2012), as the homogenizing effects of gene flow reduce genetic diversity outside of barrier loci, or loci that putatively play a role in reproductive isolation (Wolf & Ellegren, 2017). Consequently, detecting the underlying genetic differences between populations can be informative for understanding the trajectories and histories of selection.

Although foundational studies in divergence quantified differentiation at putatively neutral markers (Rundle et al., 2000), high-throughput genomic data has allowed for investigations into genome-wide variation that is indicative of divergence. Notably, genomic data has increased our understanding of divergence in the face of gene flow, and has weakened the foundations of Kimura's (Kimura, 1977) theory of neutral evolution, particularly the role of linked selection in divergence (e.g., 'genetic draft' (i.e. divergence hitch-hiking, genome hitch-hiking), background selection; Aeschbacher et al. 2017, Ellegren and Wolf 2017, Kern and Hahn 2018).

Interactions between the environment and phenotype have been noticed since the early twentieth century, as early as Thomas Morgan's identification of temperature regulation in the development of *Drosophila* (Markow, 2015). Research and debate continued regarding gene x environment interactions (GxE), particularly with Huxley's research group including Hogden, who found environmental associations with development in *Xenopus* (Tabery, 2008). These investigations lead to more thorough inquiries into environmentally induced phenotypic plasticity, and lead to the eventual development of epigenetics as a discipline, in which modifications to DNA can cause pronounced changes in its expression (Griffith & Mahler, 1969). One of these epigenetic modifications, DNA methylation, can cause modifications to gene expression, and could be an environmentally induced molecular mechanism underlying local adaptation and speciation (Smith et al., 2016). Our research thus seeks to investigate the patterns of DNA methylation underlying population structure across a species that experiences high rates of gene flow, where the results could inform disciplines ranging from ecological speciation to island biogeography, due to the potential integration of DNA methylation into the working model of speciation (via linked selection, Feder et al. 2012).

The Canada lynx is a highly mobile mid-sized felid that has remarkably low interpopulation genetic differentiation in mainland North America (Schwartz et al., 2002). Despite sampling individuals across 3,100 km, Schwartz *et al. (2002),* found the highest pairwise F_{ST} between populations to be 0.051 across the western range spanning from Montana to Alaska (Schwartz et al., 2002). These

results were corroborated by Rueness *et al. (2003),* who found a low degree of relative differentiation between east, mid- and northern continental populations, with the greatest genetic divisions between Northwestern and BC populations (pairwise F_{ST} 0.024; Rueness et al. 2003). This research was expanded upon by Row *et al (2012).*, who found slightly similar trends in pairwise F_{ST} between mainland populations, while broadening their investigation to the island of Newfoundland, which exhibited starkly higher levels of differentiation to the mainland (pairwise F_{ST} 0.180 – 0.203; Row et al. 2012). Previous research found that subtle genetic clustering in mainland Canada lynx is due to climatically-associated divides, postulated to be due to winter snow conditions (Stenseth et al. 2004, Row et al. 2014). Furthermore, the geographically isolated individuals of Newfoundland have evidence of diminished morphological size (Van Zyll De Jong, 1975; Khidas et al., 2013), which suggests that genetic or epigenetic variants might be underlying the morphological divergence seen in this insular population.

The goal of the second chapter is thus to quantify genome-wide genetic and epigenetic differentiation in individuals of Canada lynx across their geographic range to assess signatures of local adaptation that might be associated with environmental conditions. I predicted, consistent with previous research in Canada lynx, that mainland populations would be largely genetically undifferentiated. However, I also predicted that we would find unique patterns of DNA methylation in our geographic populations of Alaska and Newfoundland compared to mainland Québec and Manitoba individuals, likely due to

environmentally induced local adaptation via epigenetic regulation. Provided that these patterns of methylation impart a fitness benefit and are subsequently under selection, the underlying genetic sequence could become conserved in adapted populations, and could be a precursor leading to adaptive divergence. Consequently, DNA methylation could thus be integrated into the current working model of speciation, even over putatively neutral regions, due to linked selection (Feder et al., 2012).

CHAPTER 2

REGIONAL NICHE DYNAMICS IN A WILD CANID ADMIXTURE ZONE REVEAL ECOLOGICAL CONGRUENCE AND IMPERCEPTIBLE EFFECTS OF BIOTIC INTERACTIONS

ABSTRACT

Species' distributions are governed by environmental conditions that allow for their persistence. While climate primarily dictates occupancy of thermally-sensitive species, distribution of primary resources (i.e., food) should be critical in shaping occupancy of mammalian predators that have a strong relationship with prey. It follows that large-scale predator distributions should correspond to relative dependency on specific prey types, but it is less clear whether prey-related spatial differentiation is expressed across closely-related co-existing predator groups. We assessed the impact of integrating prey habitat features into distribution models of canid predators (*Canis lupus*, *C. lycaon*, *C. latrans* and their hybrids) within an admixture zone in northeastern North America, and predicted that canid distributions would show associations between large (moose) and medium-sized (white-tailed deer) prey that correspond to a genetic gradient across the zone. We used a novel two-pronged approach involving: 1) performance-based comparison of models with only environmental variables vs. those including prey distributions; and 2) permutation-based comparison that standardized number of predictors across model comparisons. Our results indicate low overall genetic differentiation among canids across the admixture zone, with genetic structuring being largely consistent with spatial distribution patterns. We found that at regional scales, canid groups had negligible functional differentiation in terms of habitat niche characteristics and most groups were more similar than expected. Furthermore, our conservative assessment of the relative importance of variables on model fit identified no consistent differential patterns of co-occurrence between the canid genetic gradient and prey

distributions. We infer high genetic and ecological congruence among canids at a regional scale in the admixture zone across eastern North America, and reveal the inadequacy of single analytical tests for examining the effects of biotic interactions. Accordingly, our investigation of niche model performance suggests that biotic interactions only weakly influence distributions of closely-related predators at regional scales.

INTRODUCTION

Organisms are distributed across the landscape in response to the distribution and abundance of biotic and abiotic resources that are important for survival and persistence (Guisan & Thuiller, 2005). While abiotic conditions are fundamental for describing total potential environmental space available to an organism, many species' distributions are constrained by the interplay of competition, predation, and other interactions between co-occurring biota (Soberón & Nakamura, 2009). Detailed insight into the complex factors underlying species' distributions can therefore inform our understanding of the mechanisms promoting species co-occurrence (Bullock et al., 2000), aid in predicting spread of invasive species (Smolik et al., 2010), or provide valuable information for reintroducing species (Malone et al., 2018). As such, disentangling the relative contribution of biotic and abiotic conditions that allow persistence is essential for understanding determinants of species' distributions and community organization.

At small spatial scales, interactions between coexisting species largely determine fine-scale distributions due to processes underlying availability of resources, such as competition (Soberón & Nakamura, 2009; Araújo & Rozenfeld, 2014). Specifically, negative interactions including competitive exclusion can dramatically influence native species' distributions (Maran & Henttonen, 1994), whereas range contractions due to shifting predator distributions may alter ecosystem assemblages due to mesopredator release (Newsome et al., 2017). The relative contribution of biotic and abiotic conditions for describing species' distributions is assumed to be scale dependent, and that

at large spatial scales, abiotic conditions supersede biotic interactions as range determinants, whereas at local scales biotic interactions should prevail (Soberón and Nakamura 2009). Nonetheless, there is growing recognition of the importance of biotic interactions across a range of spatial scales (Wisz et al., 2013), although the scale required for detecting accurate signals of species co-occurrence (i.e., a biotic interaction), likely depend on complex factors including number and strength of ecological interactions (Dormann et al., 2018). Few empirical studies have evaluated large-scale patterns of co-occurrence across closely-related species, requiring that further investigation address underlying regional patterns of co-occurrence in natural systems to provide insight into the determinants of occupancy patterns.

Most empirical studies examine large-scale biotic interactions in a single focal species (Anderson et al. 2002, Heikkinen et al. 2007, De Araújo et al. 2014, but see Aragón and Sánchez-fernández 2012), and do not test predictions within clades in systems that may have a gradient in the strength and direction of biotic interactions. Notably, empirical studies often only examine additive impacts of including additional biotic variables into their models (e.g., Heikkinen et al. 2007, De Araújo et al. 2014), but this approach is problematic because adding more variables inherently increases model fit when not considering model overparameterization (Forster 2000, Radosavljevic and Anderson 2014). It follows that such problems limit robust assessment of the relative contribution of abiotic and biotic factors on species' distributions, especially when niche differentiation is limited across the clade. Thus, there is a need for empirical

studies that examine the relative importance of biotic interactions across a gradient of closely-related species, while adjusting for potential confounds such as model complexity.

Across the landscape, predators having a known positive relationship to a specific prey type should exhibit high patterns of co-occurrence with their prey (Kissling et al., 2012), whereas those with lesser relationships to specific prey types (e.g., generalist feeders) should exhibit weaker responses due to the confounding effect of abiotic influence. The gray wolf (*Canis lupus*), is a widely-distributed wild canid that preys primarily on large ungulates like moose (*Alces alces*) and caribou (*Rangifer tarandus*) (Peterson et al., 1984; Messier, 1994, 1995). Although climatically-associated subdivisions structure gray wolf populations at a continental scale (Geffen et al., 2004), dependency on prey is a documented driver of spatial and genetic clustering across the landscape (Carmichael et al., 2001). In contrast, smaller canids such as coyotes (*C. latrans*) forage on smaller mammals and exhibit weaker associations to large prey (Paquet, 1992), and stronger links to agricultural lands and shallow snow cover (Otis et al., 2017). Other canids in the clade, including eastern wolf (*C. lycaon*) or related hybrids, are intermediate in size and should exhibit corresponding relationships with prey and snow. Thus, within the canid admixture zone in eastern North America, the genetic gradient should be expressed via disparity in the influence of large prey on distributions, leading to the expectation that larger canids will have stronger association with large ungulate prey habitat and lesser correlation to abiotic conditions.

We used species distribution models to assess ecological differentiation between canids in an admixture zone across eastern North America, while also examining regional patterns of co-occurrence with important prey types. First, we predicted that ecological differentiation between canids would reflect the genetic gradient, with the largest (*C. lupus*) and smallest (*C. latrans)* species exhibiting the most dissimilar niche space, whereas intermediately-sized canid types (*C. lycaon, C. lupus x C. lyc*aon and *C. lycaon x C. latrans* hybrids) having intermediate and potentially overlapping niches. Second, larger canids should be more strongly associated with large ungulate prey whereas smaller, generalist canids should exhibit weaker relationship to large prey and thus stronger relative relationship to abiotic factors, specifically snow.

METHODS

Study region

The study region spanned 2,268,309 km^2 across the area of contemporary *Canis* hybridization in eastern North America (Stronen et al., 2012), ranging from north-central Manitoba to the Atlantic coast of Canada, and including western Minnesota through to Maine, USA. Exact geographic boundaries were determined by point locations of our canid samples, using a minimum bounding geometry function in ArcGIS, with a 100 km buffer to account for dispersal. This large spatial scale was chosen because of the large spatial requirements and dispersal capabilities of wild canids, and that the scale included a range of environmental conditions and prey distributions that may influence their large-scale niche characteristics. In general, the region is characterized by a variety of habitats ranging from mixed deciduous forests in the south to boreal forests and lowlands in the north, interspersed with areas of development and agriculture. Climatic conditions across the range varied from -19.3 to 3.2° C in winter, with snow precipitation ranging from 22 to 434 mm. The large region is occupied by other large predators including Canada lynx (*Lynx canadensis*), bobcat (*Lynx rufus*) and black bear *(Ursus americanus*). Potential prey species for wild canids include moose (*Alces alces*) mostly in the northern and central range, white-tailed deer (*Odocoileus virginianus*) mostly in the southern and central range, and beaver (*Castor canadensis*) and smaller mammals throughout the region. Medium-sized and smaller prey like lagomorphs and small mammals occur throughout the region.

Sample distribution and processing

Spatial and genotypic data for wild canids were obtained from 1366 georeferenced pelt samples obtained from fur auction houses in Ontario, Canada (North American Fur Auctions, Inc.; Fur Harvester's Auction, Inc., see Otis et al. 2017). Locational certainty for most (approx. 85%) of our samples was < 1,000m (J. Otis, unpubl.). Some samples were used previously to address questions related to canid genetics, hybridization, and niche dynamics (see Rutledge et al. 2010, Wheeldon and Patterson 2012, Otis et al. 2017), with 3.4% of the present samples being collected to fill specific gaps in spatial distribution of canid occurrence in eastern North America. Our analysis is an extension of a previous assessment of canid niche overlap (Otis et al., 2017) by considering *C. lupus* and related hybrids in the larger complex of wild canids while also developing models that include prey habitat suitability as a potential predictor of canid occupancy. Occurrence locations for large prey species (moose, white-tailed deer) were compiled from the Global Biodiversity Information Facility (GBIF.org, 2017) and harvest records from governmental organizations (Pickles et al., 2013; Feldman et al., 2017).

We quantified canid sample DNA using ~20 mg of dried pelt (DNeasy Blood and Tissue Kit, Qiagen, Valencia, CA, USA). DNA was quantified and standardized to 2.5 ng/µl using a Qubit 3.0® fluorometer (Thermo Fisher Scientific) and twelve neutral nuclear microsatellite loci were amplified in three multiplexed reactions (Supplemental Table S1), including positive controls for both sexes (Wheeldon & White, 2009). Alleles were genotyped multiple times

and scored using two separate programs (Geneious v10, (Kearse et al., 2012); GeneMarker® v2.7), and by two observers to ensure consistency.

We identified species and putative hybrids using an ancestry coefficient (*Q*-value) matrix generated from *structure v2.3.4,* which sorts individuals into clusters that conform to Hardy-Weinberg Equilibrium (Pritchard et al., 2000). Our total genotypic dataset included 136 individuals of known ancestry that were used to evaluate genetic clustering, and 1,230 unassigned individuals. Unadmixed individuals were designated as having a *Q*-value ≥ 0.80 for any of the three genetic clusters (K), while remaining individuals were assigned respective hybrid designations based on proportionate assignments, consistent with an established genetic threshold (Rutledge et al. 2010, Wheeldon and Patterson 2012, Otis et al. 2017; additional details in Supplementary Materials). We visualized *structure* results in R (R Core Team, 2017) and further assessed overall genetic structure with a principal components analysis implemented in *adegenet* (Jombart, 2008). Microsatellite data for canids has been deposited in a DRYAD repository, along with occurrence locations for prey species.

Spatial Data Preparation and Parameterization

Environmental variables for prey distributions are available in Supplemental materials. Environmental variables for baseline canid models included: 1) snow depth; 2) snow precipitation; 3) tree cover; and 4) a human-use distance variable (DeFries et al., 2000; Mesinger et al., 2006; Adaptwest, 2015; ESA, 2017). These abiotic variables are biologically meaningful because they reflect constraints on travel (snow depth, snow precipitation) and colonization and occupancy (human

use) that should have variable influence across the wild canid clade. In addition, canid models included the habitat suitability models for: 5) white-tailed deer and 6) moose (prey model descriptions in Supplemental Information). Note that a model for beaver was developed and later dropped because of its high correlation with the deer model (Supplemental Figure S1). This methodological approach for examining co-occurrence is comparable to other studies (e.g., Barbet-Massin and Jiguet 2011, De Araújo et al. 2014), except that prey habitat models were modeled as continuous predictors to avoid complications in threshold choice and to reflect the full continuum of suitable habitats for prey across the wild canid clade.

We thinned occurrence locations based on environmental heterogeneity at three stratified rarefication distance classes for all canids (5, 15, 25 km; 20 – 69% data retention), and two rarefication distance classes for prey (50, 75 km; 12 – 36% data retention), implemented in *sdmtoolbox v2.0* (Brown et al. 2017; Supplemental Table S2, Supplemental Figure S2). All environmental variables were scaled to a resolution of 1.8 km^2 and analyzed using an equal area projection. Collinearity between explanatory variables was assessed using the variance inflation factor (VIF) as well as Pearson's correlation coefficients. We set thresholds of ≤ 6.0 (VIF) and 0.80 (Pearson's) as our ceilings to remove collinear variables (Supplemental Table S3, Supplemental Figure S3, S4).

Niche Modeling

We created cross-validated ensemble distribution models by generating a weighted average of six algorithms that use regression and non-parametric

statistical learning methods including MaxEnt v3.3.3e (Phillips et al., 2006), boosted regression tree (BRT), random forest (RF), generalized additive model (GAM), multivariate adaptive regression spline (MARS), and mixture discriminant analysis (MDA). Individual model parameterization was optimized using a canid-group level based on relevant techniques, where applicable (Supplemental Figure S5, S6), and final ensembles were created and averaged using a weighted mean based on receiver operating characteristic (ROC) area-under-the-curve (AUC_{test}) statistics, implemented in the package *sdm* (Naimi & Araújo, 2016). Model fit was evaluated using a threshold-independent metric (AUC_{test}) in addition to the threshold-dependent true skill statistic (TSS). Further details can be found in Supplemental Materials. While the extrapolation of habitable space beyond the environmental distribution of the modelled species (Owens et al., 2013) is a concern, the historical occupation of the entire extent (Leonard et al., 2005) and the extensive dispersal capabilities of canids (Geffen et al., 2004), makes such concerns unsubstantiated.

Three-fold cross validation was used to generate ensemble distribution models, and each model was run in triplicate for a total of 54 models per canid group. For canid models, the entire ensemble process was repeated 3 times to generate confidence intervals and assess stochasticity in comparative niche evaluation (i.e., niche overlap, permutation importance). We averaged relative covariate importance of each explanatory variable during model creation across all methods, and generated response curves from independent MaxEnt runs with identical parameterization. We then analyzed differential covariate importance

between canid groups using linear mixed models with permutation importance (based on AUC_{test}) as a response variable, canid group as an explanatory variable, and algorithm as a random effect.

As background environmental conditions varied between canid groups, we evaluated niche overlap by performing background similarity tests, which compared observed niche overlap against a null distribution of 100 generated niche models created for each species within the background environment of all other canid groups (Warren et al., 2010). This symmetrical analysis determined if pairwise distributions differed more than expected by chance when the background environments between groups were identical (Warren et al. 2010).

Quantifying prey importance in canid models

We quantified impacts of integrating prey distributions into predator models with a two-prong approach (Supplemental Figure S7). First, we used a standard approach by assessing model fit (e.g. AUC_{test}) between models with: i) baseline environmental variables (snow depth, snow precipitation, distance to human areas, tree cover); ii) baseline variables with the most important variable for predicting prey habitat (identified as mean annual temperature, MAT); iii) baseline variables with moose habitat and MAT; iv) baseline variables with only deer habitat; v) baseline variables with only moose habitat; and vi) baseline variables with both prey habitats (deer habitat and MAT were above the threshold for multicollinearity and were not assessed). The underlying variable predicting prey habitat was included to determine if the abiotic determinants of

prey distributions might equally influence predator distributions compared to the prey distributions themselves.

Our second approach assessed model fit between hundreds of ensemble models with various permutations of covariates. A permutation approach was chosen because while our first approach is commonly used when modeling biotic interactions in species distributions (e.g., Heikkinen et al. 2007, De Araújo et al. 2014), additive impacts of including prey habitat variables in our first approach might bias model evaluation compared to models with fewer covariates. Consequently, we permuted ensemble models for all five canid groups (N = 1,008 per group) that included various combinations of only three covariates, drawing from eight total variables, therefore testing 56 covariate combinations (Supplemental Table S4). In addition to the four baseline environmental variables, we included two prey habitat variables and two additional climatic variables, one of which was the primary environmental variable associated with prey distributions (Adaptwest, 2015). We included these climatic variables to determine if the underlying climatic variables that are associated with prey distributions might explain as much variance in model fit as the prey models themselves. The permutation approach is therefore a straightforward method to conservatively test biologically meaningful questions while the limited number of covariates in permutations allow for testing additional variables that would have been removed due to multicollinearity. Model permutations were assigned a categorical designation based on covariate composition (i.e., same designations

as the baseline approach, i - vi; Supplemental Table S4), and were evaluated using both AUC_{test} and the TSS.

Both approaches to quantify impacts of biotic interactions (i.e. baseline, permutation) were analyzed by performing linear mixed models for each canid group with model fit (AUC_{test}) as response variable and algorithm as a random effect. For both analyses, categorical covariate composition was the explanatory variable. We determined effect size of the relationship between biotic interactions and model fit by assessing variance explained by model fixed effects (marginal R^2; R^2M) compared to variance explained by the full model (biotic effect with algorithm) (conditional R^2; R^2C), implemented in MuMin (Bartoń, 2018), using the procedure outlined in Nakagawa and Schielzeth (Nakagawa & Schielzeth, 2013). We then quantified percent increase in model fit (AUC_{test}) for all canids by calculating fold change with models without prey covariates as our baseline. We repeated this process except with the TSS as a response variable and arrived at the same conclusions (Supplemental Table S5, S6).

RESULTS

Genetic differentiation

Wild canids were largely genetically unstructured across the study region. Based on our analysis of neutral microsatellite loci, we found that clustering samples into K = 3 correctly identified 91.3% of animals with known ancestry, while K = 2 left *C. lupus* and *C. lycaon* largely undifferentiated (Supplemental Figure S8). Further analyses identified relatively low separation between these genetic clusters, where the greatest genetic differentiation was between *C. lupus* and *C. latrans* (F_{ST} = 0.07; Fig. 2C). In contrast*, C. lycaon*, *C. latrans,* and hybrids were markedly less differentiated from other groups, with practically no differentiation between *C. latrans* and *C. lycaon* x *C. latrans* hybrids (F_{ST} < 0.01). Overall, PCA and F_{ST} analysis revealed a low level of structuring among wild canids occurring across the study region (Fig. 2B, 2C, Supplemental Table S7), while an unrooted phylogenetic tree (Figure 2D) visually confirmed the gradual genetic cline between canid groups.

Canid niche dynamics

Qualitatively, canid suitability models showed variable distributions ranging from *C. latrans* in the south-eastern range of the study extent, *C. lupus* in the north and western range, and *C. lycaon* in the mid-eastern extent, with suitable habitat for *C. lycaon* appearing more similar to *C. latrans* than *C. lupus* (Figure 3A). Less suitable habitat across the region existed for hybrids, with both hybrid groups qualitatively exhibiting intermediate niche space between unadmixed progenitors

(Figure 3A). Models fit data with reasonable consistency, with *C. lycaon* exhibiting the highest overall model fit, despite also showing the highest sensitivity to specific algorithmic choice compared to other groups (AUC_{test}, MDA: 0.76 ± 0.03; GAM: 0.84 ± 0.04; Supplemental Table S9).

Direct comparison of niche overlap values indicated a gradient in suitable habitats from *C. latrans* to *C. lupus*, which was further investigated by symmetrical background similarity tests that revealed more niche overlap than expected by chance for most pairwise comparisons between canid groups (Table 1, Supplemental Figure S9). Only 3 of 20 pairwise comparisons (N = 20, due to all groups being both the observed and background groups) identified significantly different niche space between groups (i.e., less overlap than expected if using the same background environments). Specifically, only *C. latrans* exhibited significant niche differentiation compared to *C. lycaon, C. lycaon x C. lupus, and C. lupus,* assuming that *C. latrans* had access to the same environmental conditions available to those species. All other pairwise comparisons identified no significantly differentiated niche space (Table 1, Supplemental Figure S9).

Prey as predictors of canid occurrence

Prey habitat, snow conditions, and distance to human areas strongly influenced canid distributions, whereas tree cover had low importance (Figure 3B). All canid models were positively associated with prey, with the strongest relationship being between *C. lycaon* and moose habitat. As predicted, larger canids responded

more favorably to moose habitat than smaller canids, with the weakest association between *C. latrans* and *C. lycaon x C. latrans* hybrids and moose habitat (Figure 3B, Supplemental Table S10). White-tailed deer habitat also strongly predicted canid distributions, with the weakest associations between *C. lycaon*, and the strongest with *C. lycaon x C. latrans* hybrids; Figure 3C). While minor differences in importance were seen between canid groups for the remaining environmental variables (Figure 3B), statistical analyses identified more variation due to algorithmic choice than between canid groups (Supplemental Table S11; $R^2C - R^2M$).

Variance explained by differences between canid groups (R^2M) was greater than variance due to algorithmic choice ($R^2C - R^2M$) only for moose and deer habitats (Supplemental Table S11), while differences between canids were smaller than differences between algorithmic choice for snow depth, snow precipitation, distance to human areas, and tree cover (Supplemental Table S11).

Do prey distributions improve canid models?

Our two-prong evaluation of biotic interactions revealed a minor increase in model fit for all canid groups, while the different analytical approaches uncovered different underlying patterns. Using our baseline approach, *C. lupus* and *C. lycaon* models slightly increased in predictive accuracy with both moose and white-tailed deer distributions, while *C. latrans* and *C. lycaon x C. latrans* hybrids models saw a slight increase with deer, and no increase with moose habitat

(Figure 4A). Therefore, there is apparently some support for the expected cline in prey habitat importance along the genetic gradient. Notably, *C. lycaon x C. lupus* hybrid models performed worse when prey habitats were included. All canid groups responded comparably to both prey habitats as to the primary environmental variable associated with prey distributions. Overall, model fit slightly increased for 4 canid groups across all covariate compositions compared to baseline, except for *C. lycaon* x *C. lupus* hybrids. Thus, prey covariates (or the climatic variables underlying prey distributions) slightly increased model performance. Prey variables (or the underlying climatic variables) explained the most variance in predictive accuracy for *C. lupus*, and the least for *C. latrans*, while *C. lycaon* and hybrids fell intermediate to all progenitors (R^2M, Table 2). The random effect of algorithm ($R^2C - R^2M$) explained more variation than covariate treatments for all groups, explaining 7.7 times more variation than prey composition for *C. latrans*.

The results of our permutation approach found notable differences between moose habitat and its confounded underlying climatic variable for *C. lycaon* (Figure 4A, 4B). *C. lupus* exhibited no increase in performance with both prey habitats, and a decrease in model fit for models that incorporated only moose (Figure 4B). No significant statistical variation in model fit was seen for permutations that included deer habitat for any canid group, whereas minor differences were seen for incorporating moose habitat for three of the five canid groups.

DISCUSSION

We found low genetic and large-scale ecological differentiation across a clade of closely-related wild canids, with the greatest divergence being between *C. lupus* and *C. latrans*, the most distal species. Our prediction that canid group occurrence should follow a gradient with large ungulate prey distribution was not supported given that biotic and abiotic determinants related comparably poorly to all canid distributions. Lack of genetic and niche differentiation in this clade is contrary to previous results (e.g., Kittle et al. 2017, Otis et al. 2017) and indicates a higher than expected similarity among canid groups, when assessed at a regional scale. We propose that our novel permutation-based approach for assessing the relative explanatory power of biotic and abiotic factors represents a more conservative and robust test than is normally used for such investigations (e.g., Heikkinen et al. 2007, De Araújo et al. 2014), and should be adopted in future studies of species distributions. Ultimately, a more judicious evaluation of how species interact with their environment will provide stronger inference when conducting large-scale assessment of species-environment interactions.

A continuum of canid genotypes

We identified low genetic differentiation between canids in the admixture zone, where genetic population structure largely reflected spatial structuring across the landscape. Recent studies in this region recognize three distinct genetic clusters among *C. lupus, C. lycaon,* and *C. latrans* (K = 3; Benson et al. 2012, Otis et al. 2017), and our *structure* results support a comparable level of differentiation. Quantitative differentiation (F_{ST}) was surprisingly low for interspecific

comparisons, and we did identify a genetic gradient across the *Canis* clade that was suggestive of a largely continuous gradient of canid genotypes across the region. This observation is not inconsistent with the observed morphological gradient in wild canids in the same region (Benson et al., 2012), and reinforces that limited differentiation between hybrids and their progenitors challenges our ability to apply standard species and niche concepts to this system.

The finding that canids in eastern North America span a subtle genetic gradient is also consistent with recent revelations from whole-genome analyses (VonHoldt et al., 2016) and SNP arrays (vonHoldt et al., 2016), and suggests that rampant admixture has substantially influenced canid population structure and differentiation in the region. While discrete species delimitation can be valuable for answering some ecological questions (e.g., Yoder et al. 2018), our PCA results using microsatellites indicate low interspecific differentiation and notable distinction only at the extremes between *C. lupus* and *C. latrans*. Thus, our genetic results provide an important example of the complexities associated with species status assessment and management in the face of introgressive hybridization (Allendorf et al., 2001; vonHoldt et al., 2018), and suggest that functional ecological differentiation may be required to adequately distinguish among ecologically-distinct groups (see Benson and Patterson 2013).

Limited regional niche differentiation

Recent studies show that both fine (Benson & Patterson, 2013b) and coarse-scale (Kittle et al., 2017; Otis et al., 2017) ecological differentiation exists across the wild canid clade in eastern North America. While research in this system

identified regional differences between canids (Otis et al., 2017), their approach appears to have failed to fully account for substantial autocorrelation between observations and used differential modelling parameters between canid groups, and likely detected only signatures of differential contemporary occupancy and not true ecological differences in niche space. Consistent parameterization of regularization coefficients (Warren & Seifert, 2011) and autocorrelation mitigation measures such as data thinning (Dormann et al. 2007, Boria et al. 2014) are essential for successful large-scale niche occupancy comparison among groups of organisms. In our system, these concerns are highlighted by the restricted distribution especially of *C. lycaon* (see Figure 1) as this distribution may, to an unknown degree, reflect historical range contraction due to human extirpation rather than the actual environmental space available to the species (Kyle et al. 2006, see also Faurby and Araújo 2018).

Alternatively, the scale of our spatial analysis could impact inferred ecological differentiation between canid groups (e.g. Razgour et al. 2018). Wolves and coyotes are not functionally equivalent ecological units, with wolves being directly responsive to density of large ungulate prey (Messier, 1994) and smaller coyotes serving as generalist mesocarnivores and using a variety of prey types (Roemer et al., 2009). While *C. latrans* can kill adult moose (Benson & Patterson, 2013a), we reasonably expected substantial differences in positive moose associations, at a minimum, between *C. lupus* and *C. latrans*. The absence of such a signal after accounting for background conditions implies that either canid groups in our region are functionally equivalent at large spatial scales, or that biotic interactions

between predators and prey are lost at the spatial extent used in the present study. Support for the former hypothesis is consistent with both our genetic and niche differentiation data, and in tandem with the short evolutionary time scale involved (VonHoldt et al., 2016), indicate that phylogenetic inertia may have maintained largely similar niche space between these genetically similar canids (Knouft et al., 2006; Eaton et al., 2008; Peixoto et al., 2017).

Inconsequential biotic interactions at regional scales

Our results indicate that the novel permutation approach provides a conservative test of variable importance in species models, and offers a straightforward and transferable context for interpreting the results. Specifically, the permutation approach helped distinguish the confounded relationship between moose distribution and underlying environmental variables associated with moose, particularly for *C. lycaon* (Fig 4B). These results were consistent with our ecological background tests and therefore superior to the standard approach for assessing variable importance by sequentially adding parameters to species distribution models. Notably, the permutation approach is dependent on number of covariates chosen for each replicate, so species with complex habitat requirements might inevitably be missing important environmental variation if few variables are included in replicates (Barry & Elith, 2006). This may be the case with *C. lupus,* which exhibited low overall model fit for all permutations, in contrast to the standard approach. Nonetheless, our permutation approach provides a more robust assessment of the relative contribution of biotic and abiotic determinants, while minimizing model complexity (Radosavljevic &

Anderson, 2014) and accounting for additive impacts of additional covariates (e.g., Heikkinen et al. 2007, De Araújo et al. 2014). We suggest that this approach is more reliable than traditional methods for testing variable importance and is easily applicable to a variety of systems and questions.

The observed negligible importance of large prey distribution on differentiation among canid groups is surprising. Especially for *C. lupus*, this result is contrary to previous findings (e.g., Bergerud et al. 1983, Peterson et al. 1984, Hayes et al. 2003, Kittle et al. 2017), and suggests that abotic factors are primary determinants of occurrence and niche space. This is consistent with empirical (Fraterrigo et al., 2014) and theoretical (Araújo & Rozenfeld, 2014) understanding of species range determinants at regional scales. This is reinforced by the similar response to prey density across the canid clade, supporting that abiotic factors are primary determinants at regional scales (Soberón & Nakamura, 2009). Despite the contrast between our results and those in other recent studies (e.g., Atauchi et al. 2018, Palacio and Girini 2018) concerning the role of prey distribution (and more generally, biotic interactions) on species occupancy patterns, we infer that this disparity may also arise from contemporary patterns of canid occupancy across North America, and the fact that distributions may be in disequilibrium due to the legacy of historical extirpation of wild canids in the region (Faurby & Araújo, 2018).

We conclude that despite a lack of large-scale genetic and niche differentiation among wild canids within the admixture zone in eastern North America, smaller-scale ecological differences between groups are well known

(Benson & Patterson, 2013b). Future research should synthesize stable isotope signatures with distributional data to precisely track the association between dietary prey and canid occurrences (e.g., Genner et al. 1999), while expanding the study extent to encompass the full extent of all progenitors. Ultimately, additional focus on the determinants of species distributions and the relative role of biotic interactions at a variety of spatial scales will help us reconcile the complex factors that underly contemporary distributions and species co-occurrence, and will assist in our understanding of what drives species persistence.

Table 1. Background similarity tests of niche differentiation for 5 wild canid groups in eastern North America. *I* statistics include asterisks to indicate significant similarity/dissimilarity in niche space, after accounting for relative occupancy across different environmental conditions. Note that this is a two-tailed test, so the result depends on which canid is being used as the observed or background group. Bold fields indicate groups that are more significantly similar than expected by chance, while underlined fields indicate groups that are more different than expected.

Background Group / *Observed Group*	*C. latrans*	*C. lycaon x C. latrans*	*C. lycaon*	*C. lycaon x C. lupus*	*C. lupus*
C. latrans	-	***0.92****	*0.76**	*0.69**	*0.62**
C. lycaon x C. latrans	*0.92*	-	*0.84*	*0.73*	*0.61*
C. lycaon	***0.76****	***0.84****	-	*0.86*	*0.69*
C. lycaon x C. lupus	***0.69****	***0.73****	*0.86*	-	***0.89****
C. lupus	*0.63*	*0.61*	*0.69*	*0.89*	-

* *p-value* < 0.05

Table 2. Change in distribution model fit (Δ AUC, *t*-statistic) by incorporating specific prey covariates for 5 wild canid groups in eastern North America. Changes are reported relative to the baseline, which excludes biotic variables and includes four environmental covariates. Comparisons include models with prey habitat as well as the primary environmental covariate predicting prey habitat (mean annual temperature, MAT). R^2M refers to variance explained by covariate composition, while R^2C – R^2M indicates variance attributed to algorithm choice.

	MAT Δ AUC (t-stat)	Deer Only Δ AUC (t-stat)	Moose Only Δ AUC (t-stat)	Moose & MAT Δ AUC (t-stat)	Both Prey Δ AUC (t-stat)	R^2M (Treatment)	R^2C-R^2M (Algorithm)
C. latrans	0.02 (4.2) ***	0.01 (2.7) **	0.00 (-0.8)	0.01 (2.2) *	0.02 (3.5) ***	0.06	0.46
C. lycaon x C. latrans	0.02 (2.3) *	0.02 (2.1) *	0.00 (-0.4)	0.02 (2.2) *	0.02 (1.9)	0.03	0.25
C. lycaon	0.05 (6.2) ***	0.02 (2.7) **	0.03 (3.9) ***	0.04 (5.2) ***	0.03 (3.1) **	0.08	0.37
C. lycaon x C. lupus	0.04 (4.2) ***	-0.01 (-0.8)	0.00 (0.2)	0.03 (3.0) **	-0.01 (-1.5)	0.09	0.32
C. lupus	0.05 (9.7) ***	0.02 (3.3) **	0.02 (4.1) ***	0.05 (8.8) ***	0.03 (6.0) ***	0.21	0.29

* *p-value* < 0.05, ** *p-value* < 0.01, *** *p-value* < 0.001

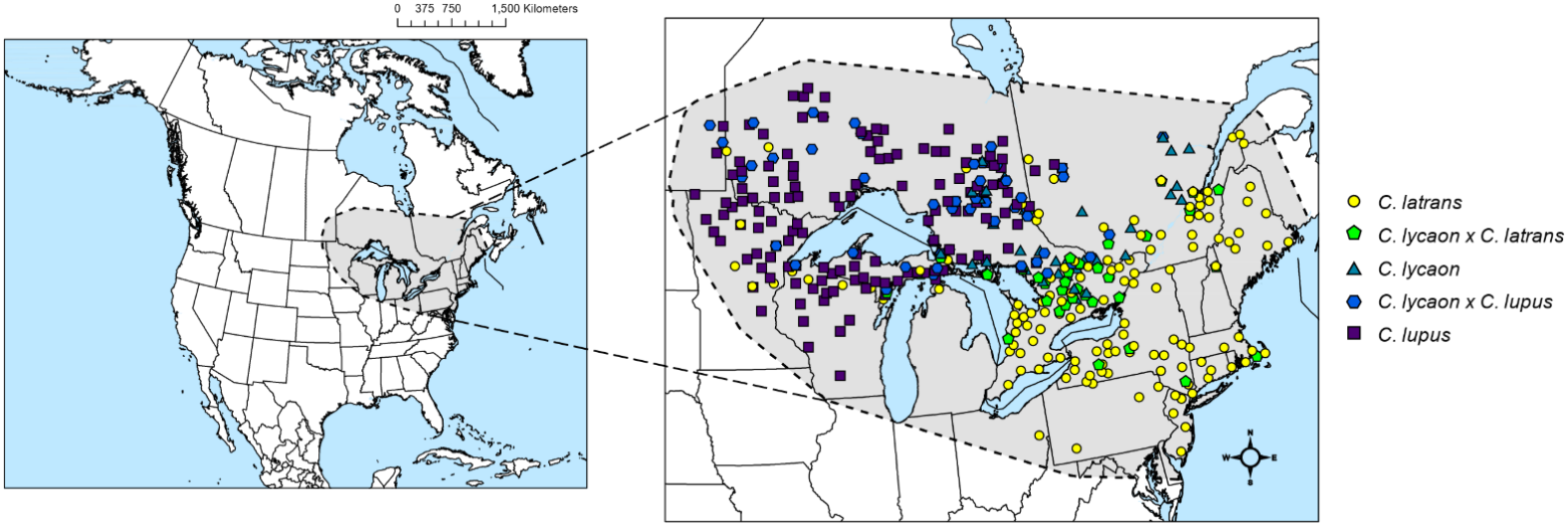


Figure 1. Study region and sample distribution for 5 wild canid groups in eastern North America. Samples depicted in the figure represent those retained after rarefaction (N = 408). Canid groups are delineated by colour, with the study extent outlined in gray.

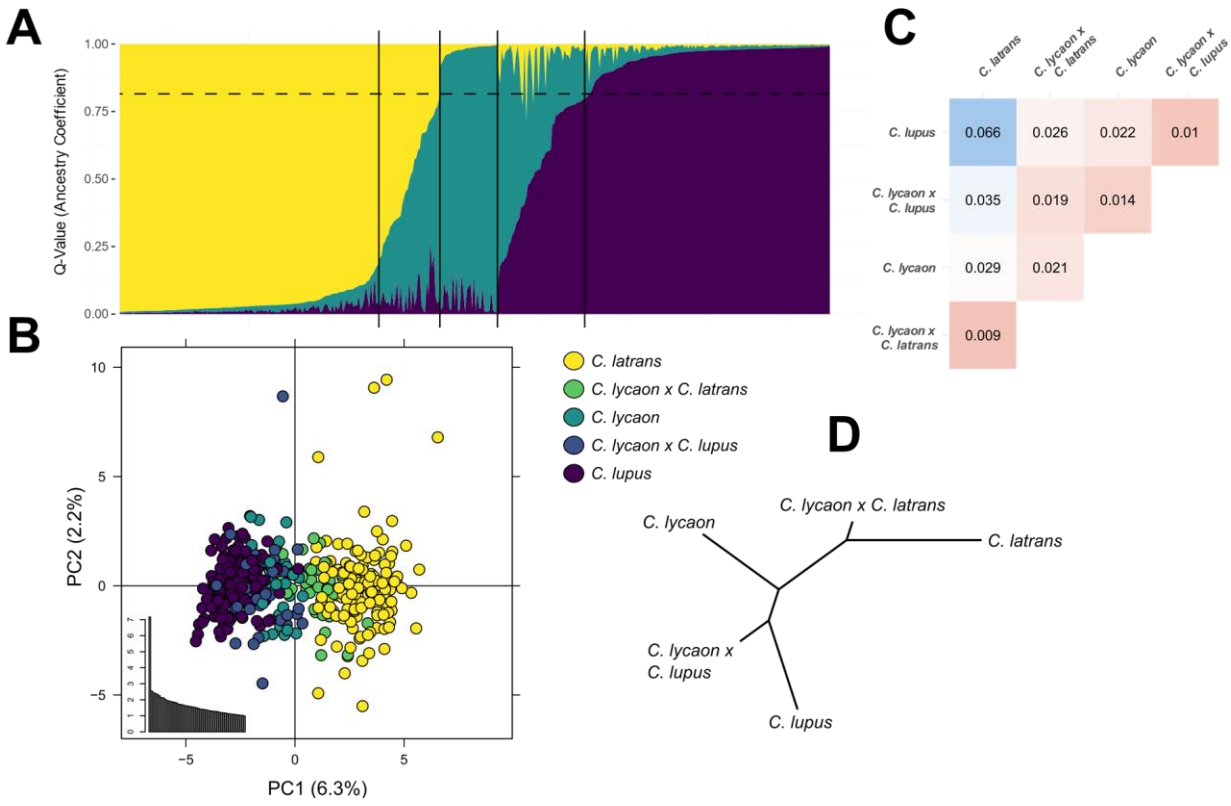


Figure 2. Genetic analyses of 408 samples for 5 wild canid groups in eastern North America. Samples were genotyped at 12 neutral microsatellite loci. **(A)** *Structure* plot for K = 3 genetic clusters, with each sample represented by a single vertical line and the three previously identified genetic groups (*C. lupus, C. lycaon, C. latrans)* delineated by colour. Y-axis indicates the *Q*-value ancestry coefficient, with the dashed line indicating the threshold for unadmixed individuals at *Q* = 0.8. **(B)** PCA plot showing genetic differentiation, with eigenvalues displayed in the bottom left corner. **(C)** Heatmap indicating pair-wise F_{ST} between canid groups. **(D)** Unrooted phylogenetic tree showing *Canis* lineage.

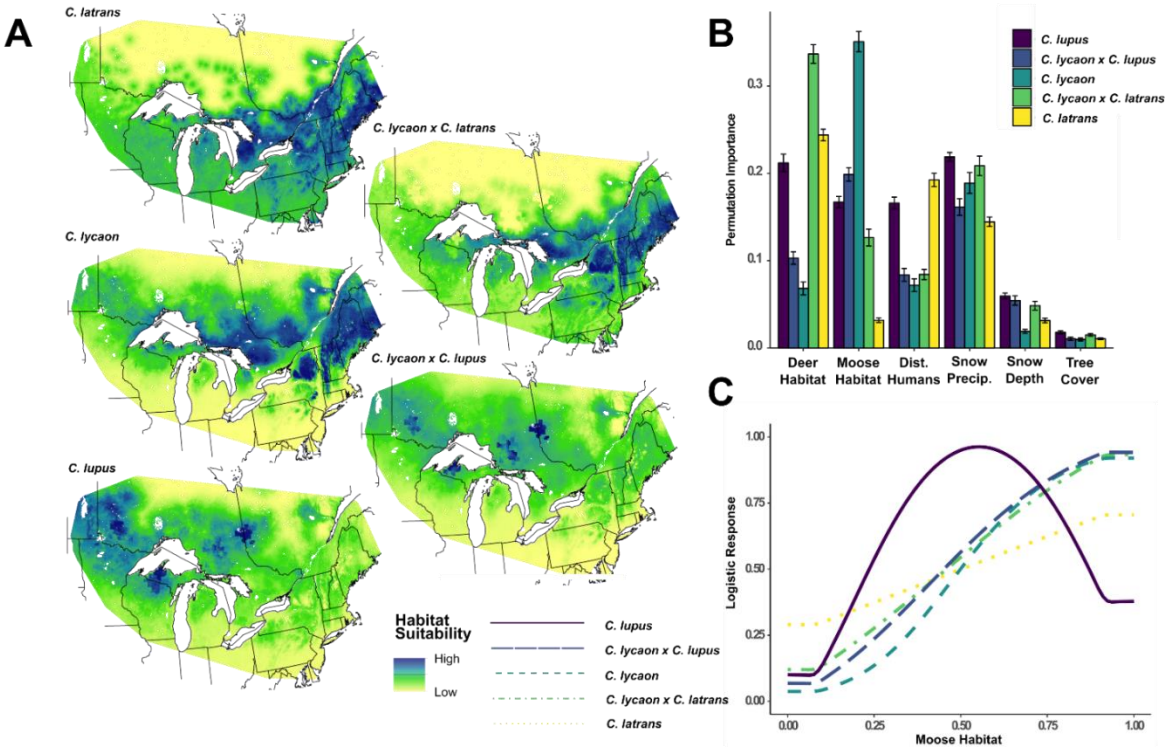


Figure 3. Ecological niche characteristics among 5 wild canid groups in eastern North America. **(A)** Ensemble species distribution models for the 5 canid groups, with unscaled habitat suitability visualized with poor habitat in yellow and better habitat in blue. **(B)** Permutation importance (calculated based on AUC_{test}, mean ± SE) for each covariate, with canid group distinguished by colour. **(C)** MaxEnt Response curves representing the relationship between each canid group and moose habitat. Moose habitat was rescaled, where 1 indicates the best, and 0 indicates the poorest, moose habitat.

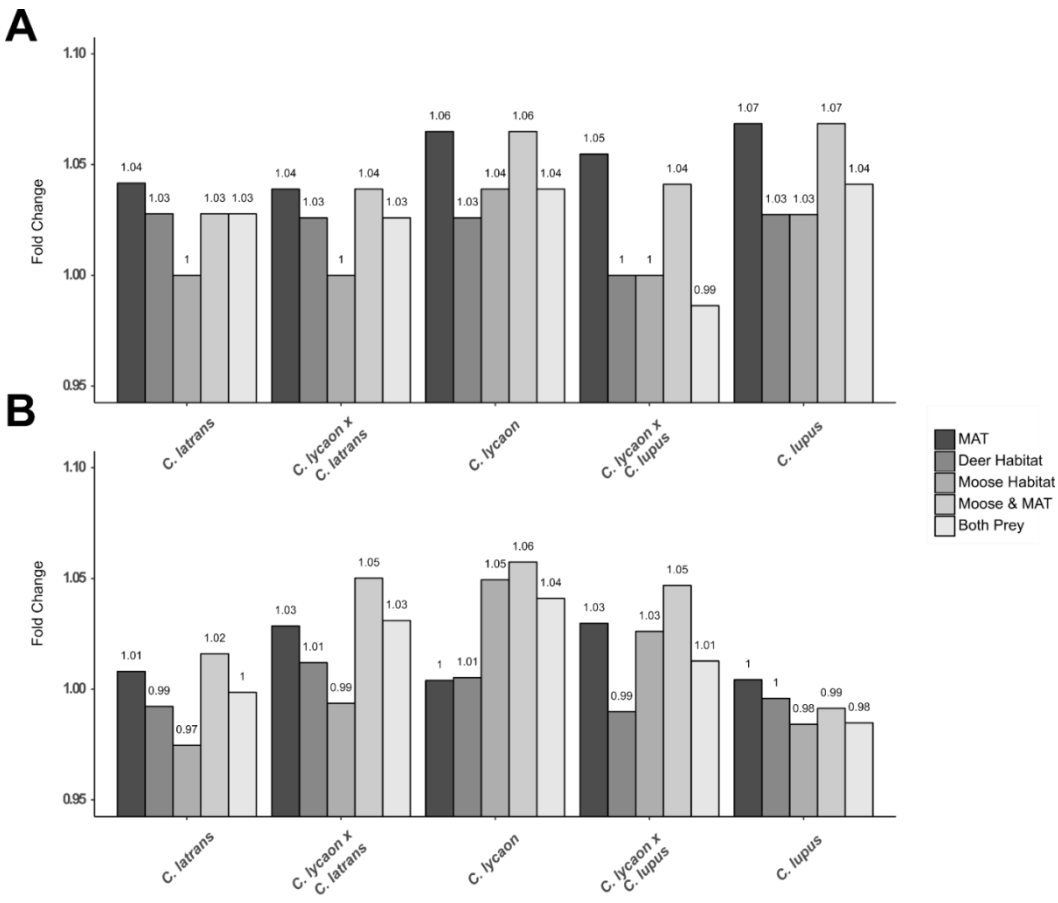


Figure 4. The importance of prey habitat variables in species distribution models for 5 wild canid groups in eastern North America. Fold changes reflect percent change from baseline covariates (e.g., snow depth, snow precipitation, distance to human areas, and tree cover). MAT refers to mean annual temperature (the primary variable predicting prey distributions). **(A)** Percent change based on the first baseline approach. **(B)** Percent change based on permutation approach.

CHAPTER 3

METHYLATION PATTERNS REVEAL CRYPTIC STRUCTURE AND A PATHWAY FOR ADAPTIVE DIVERGENCE IN A PANMICTIC CARNIVORE

ABSTRACT

Determining molecular signatures of population divergence is a fundamental component of evolutionary biology. Identifying divergence is particularly challenging between populations of highly mobile species that undergo substantial gene flow, such as the Canada lynx (Lynx canadensis), where populations are considered panmictic when only neutral genetic markers are considered. Here, we used high-throughput bisulfite sequencing to examine the environmental determinants of methylation structure across the distributional range of Canada lynx. Despite a high degree of genetic similarity inferred from SNPs among mainland populations, epigenetic structure revealed hidden levels of differentiation coincident with environmental associations, particularly in the peripheral Newfoundland and Alaskan populations. Several genes related to body-size were hypermethylated on the island of Newfoundland, providing a putative mechanism for adaptive evolution and the observed island effect on organism size. Our results indicate that epigenetic modifications, specifically DNA methylation, are powerful markers to investigate population differentiation and rapid evolutionary response.

INTRODUCTION

Investigations into divergent selection have often relied on quantifying phenotypic variation and the heritability of such traits, the latter assumed to be genetic polymorphisms transmitted via an organism's DNA (Rundle et al. 2000). Environmental conditions can be a powerful driver of adaptive divergence, where relationships are traditionally ascertained by correlating allele frequencies to environmental variation (Shafer and Wolf 2013) However, detecting adaptive divergence in species that experience high rates of gene flow is challenging due to the homogenization of genomic regions that are neutral or under weak selection (Feder et al. 2012). An aspect of adaptive molecular evolution that is undetected by standard genetic sequencing involves direct modifications to the structure of DNA. Epigenetic modifications like DNA methylation are influenced by environmental conditions, directly affect gene expression, and may be indicative of early adaptive divergence due to local adaptation (Jones and Takai 2001, Dubin et al. 2015, Artemov et al. 2017). DNA methylation could play a role in local adaptation due to its regulatory role in transcription by modifying chromatin structure, repressing transcription factors, or recruiting protein complexes that block transcriptional machinery, especially around CpG islands (Fujita et al. 2003, Lorincz et al. 2004, Maunakea et al. 2010) that are dense clusters of cytosine-guanine dinucleotides and frequently occur near the transcription start site of genes and have functional role with gene expression (Wutz and P. Barlow 1998, Han et al. 2008, Jones 2012). Consequently, DNA methylation, especially around CpG islands, could explain the molecular basis of

local adaptation in many instances due to its regulatory function on gene expression, and to date has been overlooked molecular marker of adaptive divergence and rapid evolutionary adaptation in wild populations.

Here, we assessed whether environmental variation, geographic distance, or insularity were determinants of DNA methylation structure in a free-ranging carnivore. Our study species, the Canada lynx (*Lynx canadensis*), is a mid-sized felid that is highly mobile and whose neutral genetic variation (i.e., microsatellites) exhibits low levels of genetic differentiation ($F_{ST} < 0.01$ across the mainland, with divergent island populations (Row et al. 2012, Prentice et al. 2017). Despite this low degree of overall genetic structure, two potential mechanisms might drive epigenetic divergence in Canada lynx. First, allele frequencies are correlated to climatic gradients in both population time-series and fine-scale genetic analyses (Stenseth et al. 1999, Row et al. 2012, 2014); thus, given the large distributional range of lynx, climate might play a larger role in shaping DNA methylation patterns than can be detected with microsatellites. Second, Canada lynx show a subtle cline in body size with larger individuals in Alaska (Van Zyll De Jong 1975) and smaller individuals in insular populations, including Newfoundland and Cape Breton Island (Khidas et al. 2013). Body size changes in island populations appear to be consistent with the 'Island Rule' (Khidas et al. 2013), where insular mammals are smaller in size compared to their mainland counterparts (Foster 1964, Van Valen 1973, Wayne et al. 1991, Lomolino 2005). If functional genes related to body size are repressed in geographically isolated populations due to DNA methylation, then epigenetic

modifications might be an underlying mechanism driving population divergence (Flatscher et al. 2012). Based on these two mechanisms (climate and island isolation), we formulated two predictions: 1) climatic conditions are the driving force behind spatial epigenetic structuring (Stenseth et al. 2004); and 2) patterns of DNA methylation over CpG islands and gene bodies are more correlated with environmental substructuring than methylation patterns over DNA of unknown function, analogous to patterns seen in genetic data in regions under selection compared to putatively neutral regions (Keller et al. 2016).

METHODS

Experimental Design

We investigated differential patterns of methylation between four populations of Canada lynx across North America using high-throughput bisulfite sequencing. 23 – 24 individuals were sampled from each population, spanning a wide climatic and geographical region. Sequencing effort was reduced by preparing reduced representation sequencing libraries, and genetic data (i.e. SNPs) were identified from bisulfite converted reads using conservative calling parameters. We examined population structure using Euclidean distance-based ordinations (PCoAs) and assessed environmental associations against this variation using distance-based statistical frameworks that allow multiple response variable inputs (i.e. db-RDAs). We then determined gene-specific differential methylation between populations using beta regressions.

Sample Acquisition and Reduced-Representation Bisulfite Sequencing (RRBS) Library Preparation

Georeferenced Canada lynx tissue samples were collected from fur auction houses throughout eastern North America from dried pelts (North American Fur Auctions, Fur Harvester's Auctions, Inc.). Individuals from four geographic locations were chosen for this study, spanning the longitudinal and latitudinal range of the species and consistent with previous research on the Canada lynx system (Rueness et al. 2003). Tissue was consistently taken from the same morphological location from each adult-sized pelt. Genomic DNA was isolated

using a MagneSil® (Promega Corporation) Blood Genomic Max Yield System a JANUS® workstation (PerkinElmer, Inc.) and quantified and standardized to 20 ng/µl with a Quant-iT PicoGreen® ds-DNA assay using manufacturer's instructions (Thermo Fisher Scientific). We adapted an existing reduced representation bisulfite sequencing library preparation workflow designed for multiplexed high-throughput sequencing (van Gurp et al. 2016). Genomic DNA (400 ng) for 95 Canada lynx samples was digested with NsiI and AseI restriction enzymes overnight and subsequently ligated with methylated adapters (Table S6). An individual sample of completely non-methylated lambda phage genomic DNA (200 ng; Sigma-Aldrich – D3654) with a unique barcode was included to assess bisulfite conversion efficiency. Barcoded samples were then combined into eight pools to ensure consistent reaction environments for the entire library using a QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA) following manufacturer's instructions and 10 µL of 3M NaAc was added to neutralize pH. We then performed a SPRI size selection on each pool (0.8x volume ratio) with Agencourt® AMPure® XP beads (Beckman Coulter, Inc.). Nicks between the 3' fragment overhang and the 5' non-phosphorylated adapter nucleotide were repaired with DNA polymerase I and bisulfite conversion was performed on each pool using an EZ DNA Methylation-Lightning Kit™ (Zymo Research) with a 20-minute desulphonation time.

Pools were amplified in three separate PCRs to mitigate stochastic differences in amplification and were subsequently concentrated using a QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA). The eight pools

were quantified with Qubit 3.0® (Thermo Fisher Scientific) and appropriate amounts were added to a final super-pool for equal weighting. A final magnetic bead clean-up (0.8x volume ratio) was performed to remove any adapter dimer. We checked final library concentration and fragment distribution with a Qubit® 3.0 (ThermoFisher Scientific) and Hi-Sense Bioanalyzer 2100 chip (Agilent), respectively. Paired-end 125-bp sequencing was performed on a single lane on an Illumina HiSeq 2500 platform at the Centre for Applied Genomics at the Hospital for Sick Children (Toronto, Ontario, Canada). Raw sequence data FastQ files are available on the Sequence Read Archive (ID Pending).

Bioinformatics – quality checks and SNP calling

We assessed sequencing success as well as removed adapter and low-quality reads via FastQC (Andrews 2010) and Cutadapt (Martin 2011) implemented in TrimGalore! v0.4.4 (Krueger 2012). Individuals were demultiplexed using python scripts (van Gurp et al. 2016). Paired-ends reads for Canada lynx samples were initially aligned to several genomes (*Felis catus, Homo sapiens,* Lambda Phage) using Bismark (Krueger and Andrews 2011) to assess mapping efficiency and contamination (Table S1). Paired-end reads for downstream analyses were aligned to the domestic cat genome with Bismark, using relaxed mapping parameters (score_min L,0,-0.6) (Krueger and Andrews 2011).

SNPs were called from indexed BAM files with CGmapTools (Guo et al. 2017) using a coupled Bayesian wildcard algorithm with a conservative 0.01 error rate and a static 0.001 p-value for calling variant sites, which generated variant call files (VCFs). VCF files were indexed, merged, and filtered using VCFtools

(Danecek et al. 2011) for bi-allelic loci with a sequencing depth of at least five, and were shared between at least 50% of the individuals (max-missing 0.5) and a minor allele frequency (maf) of > 0.001. Sites were further filtered by removing any variants out of Hardy-Weinberg equilibrium within populations (p-value < 0.05). A pair-wise Euclidean dissimilarity (distance) matrix was computed on the SNP data using the function *daisy* within the package *cluster* (Maechler et al. 2018) using R v3.4.2 (R Core Team 2017). This dissimilarity matrix was then summarized in a principal coordinates analysis (PCoA) using the *dudi.pco* function in *adegenet* (Jombart 2008). Missing SNP data was imputed by mean allele at a population level. Pairwise F_{ST} was calculated using StAMPP (Fig. S3) and relative F_{ST} was measured using a null model Bayesian approach implemented in BayeScan v2.0 (Foll and Gaggiotti 2008) (Fig. S3, further information in Supplemental). AMOVA between populations was computed using the *poppr.amova* function within *poppr* (Kamvar et al. 2014), and heterozygosity was assessed with *adegenet.*

Bioinformatics – DNA methylation

We identified methylated and non-methylated positions by first filtering BAM files for incomplete bisulfite conversion based on reads containing more than three methylated positions in a CHH or CHG context. In the remaining reads, methylated positions in a CpG context with a sequencing depth of at least five were extracted with Bismark (Krueger and Andrews 2011), truncating the last two bases of the forward mate-paired reads (R1) and the first two bases of the reverse mate-paired (R2) reads. Methylation polymorphisms in areas of overlap

between read pairs were extracted only once. After confirming that our non-directional library contained roughly equal reads for all possible amplified DNA strands, we proceeded to analysis.

We first generated a custom CpG island annotation track using hidden Markov models based on $CpG_{o/e}$ implemented in *makeCGI* (Wu et al. 2010). Only islands with a calculated posterior probability greater than 99.5% were retained for analysis, based on $CpG_{o/e}$. Mapped and extracted methylated sites were then imported into Seqmonk (Andrews 2007) using the generic text importer, and raw data was qualitatively visualized against the annotated domestic cat genome (felCat9.0). We analyzed DNA methylation over CpG islands and gene bodies by creating 5,000-bp running windows directly over and 25,000-bp upstream of gene bodies, combined with windows directly over CpG islands. Each window was assigned a methylated percentage score based on the overall ratio of methylated to non-methylated bases within the feature. We filtered this window-set for regions that had at least one CpG and equal representation from each population. This process was repeated for our second subset of analyses that investigated methylation patterns of unannotated regions of the genome. For this analysis, we created 5,000-bp running windows across the entire genome and removed any windows that overlapped with the initial windows over CpG islands and gene bodies by more than 1%. Data was filtered, exported, and summarized in the same way as the first analysis.

Sensitivity analyses

To determine the implications of missing data and PCoA axis retention thresholds, we performed two sensitivity analyses. The first analysis examined the effects of missing data by repeating all analyses using a subset of 10 individuals per population with the least amount of missing data (N = 40), and again with a subset of 6 individuals per population with the least amount of missing data (N = 24), across all three datasets (Fig. S1). Overall trends in explanatory effects (adjusted R^2) and qualitative inferences (PCoA clustering) were investigated and no change in inferences were determined. We examined the implications of arbitrary PCoA axis retention by repeating all analyses, but instead using different cumulative variation explained thresholds as response variables. We performed a number of db-RDAs using all axes explaining 30%, 50%, 75%, and 95% cumulative variation as response variables, and results were qualitatively similar regardless of axis retention (Fig. S1).

Quantifying environmental associations

To determine if patterns of DNA methylation and genomic variability could be explained by macro-scale climatic conditions, geographic distance, or insular divergence, we performed a distanced-based redundancy analysis on the summarized SNP and DNA methylation data. Meaningful axes explaining > 30% of the cumulative variation in the data were used as response variables in a distance-based redundancy analysis (db-RDA) conducted in *vegan* (Dixon 2003), using variables that putatively describe the environmental determinants of population structure in Canada lynx. Our covariates included a binary variable of insularity, which identified the Newfoundland population against mainland

populations and was used to describe the largely impassable barrier of the Strait of Belle Isle between Newfoundland and mainland Labrador (Koen et al. 2015). A variable of geographic distance was included which was simply the first axis of a principal coordinates analysis (PCoA) on a Euclidean distance matrix of latitude and longitude (PCo1 = 99.7% of the variation). In addition to the geographic variables, we included a biotic variable of percent tree cover (DeFries et al. 2000), a randomly generated numerical variable to assess the effect of noise, and a climate variable. We created a climate variable that summarizes winter conditions for lynx by performing a PCA to reduce multi-collinearity, which reduced annual temperature ranges (BIO7), winter precipitation (BIO19), and minimum coldest temperature (BIO6) to a single PCA axis (Fick and Hijmans 2017) (PC1 = 85.6% of the variation). All raster data had a resolution of 1 km^2, and raster values for all explanatory variables were extracted from each cell for all 95 samples for db-RDAs. Linearity was confirmed between response and explanatory variables, and multi-collinearity between explanatory variables was assessed using the VIF and any variables > 4 were removed (Table S5). Step-wise model selection using the function *ordistep* within *vegan* (Dixon 2003) was performed to isolate the best overall model using a QR decomposition technique based on p-values (Table S4). To isolate the individual explanatory power of each variable, we performed partial distanced-based redundancy analyses (p-dbRDAs) on the variables that were identified as significant in the full db-RDA.

Differentially methylated regions and gene ontology

We identified differentially methylated regions with putative functional correlates by performing beta-regressions on windows over CpG islands and gene bodies for all 95 individuals with percent methylation as the response variable and population as the explanatory variable. Beta regressions are appropriate for proportion or percentage data (Ferrari and Cribari-Neto 2004), and we set an alpha threshold at conservative levels seen in similar studies (Le Luyer et al. 2017) (p-value < 0.001). Methylation values of 0 and 1.0 were modified to 0.001 and 0.999 to allow for statistical analyses. Overall differences in percent methylation were calculated, and genes were identified as either hypermethylated or hypomethylated based on their relative degree of methylation compared to other populations. Direct overlap between our differentially methylated regions and the felcat9.0 gene annotations were extracted using Seqmonk (Andrews 2007). We then identified functional associations by searching UniProt (https://www.uniprot.org) by gene name against the database of genes in the domestic cat genome using the search term “organism: *’Felis catus* (Cat) (*Felis silvestris catus*) (9685)’”.

RESULTS

We identified differential methylation at base-pair resolution across the genome in 95 Canada lynx epidermal tissue samples from four populations (n = 23-24 per sampling area) across North America, including one insular population in Newfoundland (Fig. 1). The sampled populations have a wide geographical spread, with an average minimum distance between populations (Québec and Newfoundland) of 1,158 km and a maximum distance (Alaska and Newfoundland) of 5,520 km. Habitats around these populations present a dynamic range of environmental conditions, ranging from 32 to 432 mm of winter precipitation and a mean annual temperature range (Fick and Hijmans 2017) of -6.3 to 4.7° C. To determine associations between these environmental conditions and genome-wide patterns of DNA methylation, we created a reduced representation bisulfite sequencing (RRBS) library (full protocol in Supplemental). Paired-end sequencing on a HiSeq2500 generated a total of 210,773,612 filtered and demultiplexed reads that were aligned to the domestic cat genome (85.0% average mapping success; Table S1; *Felis catus*; NCBI felcat9.0) and variants were called using specially designed software for bisulfite converted reads (Krueger and Andrews 2011). We mapped our reads to the human (*Homo sapiens*) and lambda phage genomes to rule out contamination (2.7% and <0.1% success, respectively), and assessed bisulfite conversion efficiency by including non-methylated lambda phage DNA in the sequencing lane (Table S2). Furthermore, we quantified the temporal effects of methylation, the ramifications

of missing data, and model parameterization with sensitivity analyses, with no implications on overall inferences (Fig. S1).

Pronounced population structure revealed by methylation patterns

To examine regions with putative regulatory function, CpG islands were first bioinformatically identified *de novo* (N = 28,127) using hidden Markov models and contained an average GC content of 59.2% with a posterior probability of observed-to-expected GC content ($CpG_{o/e}$) of 1.13. Our DNA methylation analysis was broken into two subsets based on proximity to genomic features, with identical filtering parameters for each subset to maximize comparative inferences. Our subset over CpG islands and gene bodies contained 329 5,000-bp windows with 4,611 CpG positions, while the unannotated dataset contained 376 5,000-bp windows with 5,031 CpG positions (Table S3). For our purposes, gene bodies included both introns and exons, as the explicit epigenetic function has not been determined between the two, although exon methylation appears to be more conserved across organisms (Feng et al. 2010). Qualitatively, population structure between geographically peripheral populations was most pronounced in DNA methylation patterns over CpG islands and gene bodies compared to methylation patterns over regions of unknown function. The first two axes of a principal coordinates analysis (PCoA) on a Euclidean distance matrix summarizing CpG island and gene body methylation (PCoA1 = 6.2%, PCoA2 = 5.0% variation explained) distinctly cluster Alaska from the remaining mainland populations, while the mid-continental populations show no structure (Fig. 2). Similar, but less distinct patterns were seen in the ordination of methylation

patterns over unannotated regions of the genome, which showed more distinct structure between the Québec and Manitoba populations (Fig. 2). Analogous population-level trends between the datasets were further confirmed with Pearson and Spearman's rank coefficients of the first PCoA axis ($r = 0.91$; $\rho = 0.88$).

Environmental variation associated with epigenetic structure

For both methylation datasets, model selection identified three significant variables in our db-RDA analysis: geographic distance, climate and a binary variable representing insularity for the Newfoundland population (pseudo-***F*** = 16.26 - 16.27; adjusted $R^2 = 0.33$; all $p = 0.001$). Tree cover ($p = 0.13$ - 0.50) and a randomly-generated numeric variable to assess the effect of noise ($p = 0.64$ – 0.78) added no explanatory power to either model (Table S4). Collinearity was low between all retained variables (VIF = 2.09 – 3.77; Table S5). We examined the explanatory power of each variable independently using partial db-RDAs (Fig. 3), which revealed similar trends between both methylation datasets for geographic distance (pseudo-***F*** = 15.72 – 17.47; adjusted $R^2 = 0.11$ – 0.12; $p = 0.001$) and climate (pseudo-***F*** = 6.75 – 6.90; adjusted $R^2 = 0.04$; $p = 0.001$). However, within the insular Newfoundland population, methylation patterns over CpG islands and gene bodies were more strongly associated with epigenetic variation (pseudo-***F*** = 13.66; adjusted $R^2 = 0.09$; $p = 0.001$) than with methylation patterns over unannotated regions (pseudo-***F*** = 10.03; adjusted $R^2 = 0.07$; $p = 0.001$), suggesting stronger epigenetic divergence in putatively regulatory regions of the genome in insular Canada lynx (Fig. 4).

Differential methylation over genes related to morphology

We identified differential methylation directly over CpG islands and gene bodies using beta regressions (n = 329), with population as an explanatory variable (i.e. windows with > 10% overall difference in methylation and p-value < 0.001). Of the 16 gene regions identified using this criterion, 11 were significantly differentiated in the Newfoundland population (Table 1). Several differentially methylated regions over genes with associated morphological function were hypermethylated, while genes with epigenetic regulatory function were both hyper- and hypomethylated (i.e. transcription regulation, DNA-binding transcription factors; Table 1). The 11 differentially methylated regions in the insular population have a difference in methylation of 14%-40% compared to mainland populations, suggesting a pathway for morphological divergence associated with differential patterns of DNA methylation. Specifically, hypermethylation of HDAC9, and thus putative repression of its transcriptional activity, offers a mechanism underlying diminished insular body size and is consistent with research in model organisms (Chatterjee et al. 2014). Additionally, we identified three differentially methylated regions in the Alaskan population related to spectrin, carbohydrate, and ATP binding, suggesting differential expression of genes with metabolic function. Investigations into explicit levels of gene expression (i.e. RNA-seq), and chromatin accessibility (i.e. ChIP-seq) will be needed to provide additional causal evidence into the regulatory function of epigenetic mechanisms in adaptive divergence provided here.

Genetic structure driven by insular and climatic divides

We examined the relationship between neutral SNPs, environmental variation, geographic distance, and insularity by first performing a principal coordinates analysis on a Euclidean distance dissimilarity matrix of 489 genomic SNPs identified using a Bayesian wildcard algorithm with conservative calling and filtering parameters (Guo et al. 2017). Consistent with previous research on population structure in Canada lynx (Rueness et al. 2003, Row et al. 2012), our SNP data analysis indicated substantial genetic separation of the insular Newfoundland population (pair-wise F_{ST} = 0.10 – 0.13; Fig. S3). We complemented our pair-wise F_{ST} analysis with a null model approach that calculated the relative differentiation (F_{ST}) between populations against absolute zero (Foll and Gaggiotti 2008), which similarly identified mainland populations as relatively similar (F_{ST} = 0.005 – 0.024), with Newfoundland as the most segregated (F_{ST} = 0.096, Fig. S3, further information in Supplemental). Consistent with previous research (Prentice et al. 2017), gene diversity was lowest in the Newfoundland population (H_e = 0.224). An analysis of molecular variance (AMOVA) identified substantial genetic variation between populations, which explained 8.58% of the total variation in SNP data (σ = 5.84, p = 0.01). Outlier detection with Bayescan v2.0 (Foll and Gaggiotti 2008) (false discovery rate > 0.05) identified no loci under selection.

We determined relationships between neutral SNPs and environmental variables by performing a db-RDA step-wise model selection on the retained PCoA axes, and identified only climate (pseudo-***F*** = 51.13; p-value = 0.001) and

insularity (pseudo-***F*** = 26.33; p-value = 0.001) as important for explaining genetic variation (Fig. S2). Geographic distance (p-value = 0.31), tree cover (p-value = 0.95), and the randomly generated numeric variable (p-value = 0.33) were all dropped during the model step-selection. The first axis of the db-RDA was strongly representative of both climate (db-RDA1 = -0.86) and insularity (db-RDA1 = 0.99), while the second axis primarily summarized genetic variation associated with climate (db-RDA2 = 0.51). Partial db-RDAs identified a substantial amount of the net variation was explained by the mainland – Newfoundland divide (pseudo-***F*** = 26.3; adjusted R^2 = 0.15; p-value = 0.001), while climatic variation explained much of the variation seen in mainland populations (pseudo-***F*** = 13.70; adj. R^2 = 0.08; p-value = 0.001). We identified a modest to weak correlation between SNP data and unannotated region methylation (ρ = -0.62) than with CpG island and gene body methylation (ρ = -0.61) from the first axes of PCoAs using Spearman's ranking coefficients.

DISCUSSION

Our results indicate that DNA methylation detects population divergence in the face of substantial gene flow; as such, it can be a powerful marker to examine cryptic population structure in species that otherwise appear genetically panmictic. In addition, identifying the molecular basis of adaptive divergence between populations has relied on comparing allele frequencies or isolating outlier loci under selection (Beaumont and Balding 2004, Jones et al. 2012); our results suggest that epigenetic modifications increase resolution when defining population structure and are a useful marker for understanding adaptive differentiation and Island biogeography.

DNA Methylation and the Island Rule

For decades, the adaptive role of epigenetic modifications in evolution has been controversial (Ho and Saunders 1979, Jablonka and Raz 2009, Robertson and Richards 2015). Our data provide evidence that epigenetic modifications can be an informative molecular marker to investigate rapid evolutionary change. The evolutionary history of Canada lynx on Newfoundland is unclear, yet evidence suggests post-glacial colonization (10,000 years) (Row et al. 2012). The higher levels of differential methylation in the smaller Newfoundland lynx, particularly over genes with morphological significance, suggest an epigenetic pathway for rapid evolutionary responses to geographic isolation (Ho and Saunders 1979). Insular populations of mammals generally have notable phenotypic differentiation (Lomolino 2005, Meiri et al. 2005), and DNA methylation serves as a potential mechanism driving insular phenotypic divergence due to its plasticity to

environmental conditions (Lea et al. 2016). The differentiated DNA methylation patterns in the Newfoundland population suggest a molecular pathway for morphological differences between mainland and island populations, as predicted by the Island Rule. Evidence for this adaptive divergence was observed in both disparities between methylation in CpG islands/gene bodies and DNA of unknown function, and differential methylation of specific genes. We identified two notable genes (ZEB1 and HDAC9) that may underlie differences in body size between mainland and Newfoundland lynx. The hypermethylated HDAC9 gene is of particular interest because downregulation in HDAC9 is associated with diminished body mass (Chatterjee et al. 2014). Commonly, insular mammalian carnivores are smaller in size (Lomolino 2005), which is consistent with existing data on the insular population of Canada lynx (Van Zyll De Jong 1975, Khidas et al. 2013). DNA methylation remains an overlooked molecular marker of population divergence and local adaptation, and offers a putative mechanism underlying rapid evolutionary response and the Island Rule.

An overlooked marker for examining adaptive divergence

In Canada lynx, geographic distance was correlated to both methylated datasets, but not neutral SNPs. Row et al. showed a minimal barrier effect of the Rocky Mountains to Canada lynx (Row et al. 2012), but environmental variation appears to be driving unique patterns of functionally important DNA methylation at the range margins. Despite the lack of genetic structure between mainland populations (Rueness et al. 2003), we see distinct patterns of DNA methylation that differentiate Alaska from mid- and east-continental populations, and variation

that is associated with climate that is not accounted for by geographic distance. Although both DNA methylation datasets found roughly similar patterns in population structure, methylation patterns over regions of unannotated DNA identified more subtle differentiation between the Québec and Manitoba populations. These results suggest that genome-wide methylation patterns follow roughly analogous trends in population structure to genetic data, where less dramatic but more overall structure is seen in putatively neutral areas of DNA methylation compared to regions with putative regulatory function. We find this surprising due to the disparate nature of genetic and epigenetic markers and their distinct underlying mechanisms. Further investigations will be needed to assess the stability and role of DNA methylation between populations in the face of drift, recombination, and selection. With time, divergent methylation patterns should lead linked SNPs to act as if under directional selection, and thus could be factored into the working model of speciation (Feder et al. 2012) and our understanding of adaptive differentiation. The identification of cryptic epigenetic structure also has implications for conservation. For example, genetically homogenous individuals could be epigenetically adapted to local conditions (Le Luyer et al. 2017), and thus, greater care would be needed to select individuals for translocations (Murray et al. 2008). However, the plasticity of DNA methylation throughout the lifetime of an individual could allow for adaptive epigenetic responses (Lea et al. 2016).

Overall, these results demonstrate that epigenetic modifications like DNA methylation are powerful markers for investigating population differentiation in

wild organisms. We reveal the utility of integrating epigenetic assays into investigations of population structure, where traditional genetic surveys would have not detected subtle molecular differentiation between all populations. Comprehensive investigations into molecular population structure should thus quantify epigenetic differentiation before claiming uniform structure, which has far reaching implications across all population genetic research.

Table 1. Differential methylation between populations over gene regions. Differential methylation between populations over gene regions. List of differentially methylated regions that have direct overlap with annotated genes. Differentially methylated regions were identified using beta regressions with a p-value < 0.001 and when methylation levels differed by more than 10% between the identified population and the others. Hyper- and hypo-methylation is relative to the other three populations. GO annotated functions were retrieved from UniProt (www.uniprot.org).

Population	Gene	UniProt ID	P-Value	% Difference	Hyper/Hypo-Methylated	GO Annotated Function
Alaska	ADGRL2	M3XFT6_FELCA	7.58E-07	20.17	Hyper	Carbohydrate binding, G-protein coupled receptor activity
	ANK2	M3WXH8_FELCA	2.41E-05	16.10	Hyper	Spectrin binding, maintenance of cytoskeletal structure
	ATP8A1	M3WM63_FELCA	0.000701	14.99	Hyper	ATP binding, magnesium ion binding
Manitoba	NRG3		0.000216	25.43	Hypo	No Uniprot Annotation
	RWDD1	M3XFU2_FELCA	1.86E-05	17.61	Hyper	Cytoplasmic translation
Newfoundland	CDH18	M3WJ78_FELCA	2.83E-05	21.23	Hypo	Calcium ion binding
	CENPU	M3W3F1_FELCA	0.000895	13.71	Hyper	Embryonic development
	CRISPLD1	M3W2H3_FELCA	0.000337	19.82	Hyper	Face morphogenesis
	DCC	M3WHY3_FELCA	0.000399	37.36	Hyper	Spinal cord ventral commissure morphogenesis
	FAM35A	M3WZ20_FELCA	0.000641	19.62	Hyper	No Uniprot Annotation
	HDAC9	A0A2I2U9G1_FELCA	4.40E-05	39.60	Hyper	Histone deacetylase; transcription regulation; downregulation associated with diminished body mass, adaptive thermogenesis (Chatterjee et al., 2014)
	LOC101091724		0.000698	26.54	Hypo	lncRNA
	LOC109493917		0.000332	38.44	Hypo	lncRNA
	PBX3	M3WKS0_FELCA	3.21E-05	26.60	Hypo	DNA binding transcription factor activity
	TMOD2	M3X1Q9_FELCA	0.000847	29.43	Hyper	Myofibril assembly, muscle contraction, epithelial cell morphology (Weber et al., 2007)
	ZEB1	A0A2I2UAU6_FELCA	5.73E-05	37.60	Hyper	DNA-binding, transcription factor, regulation of adipose tissue mass (Saykally et al., 2009)

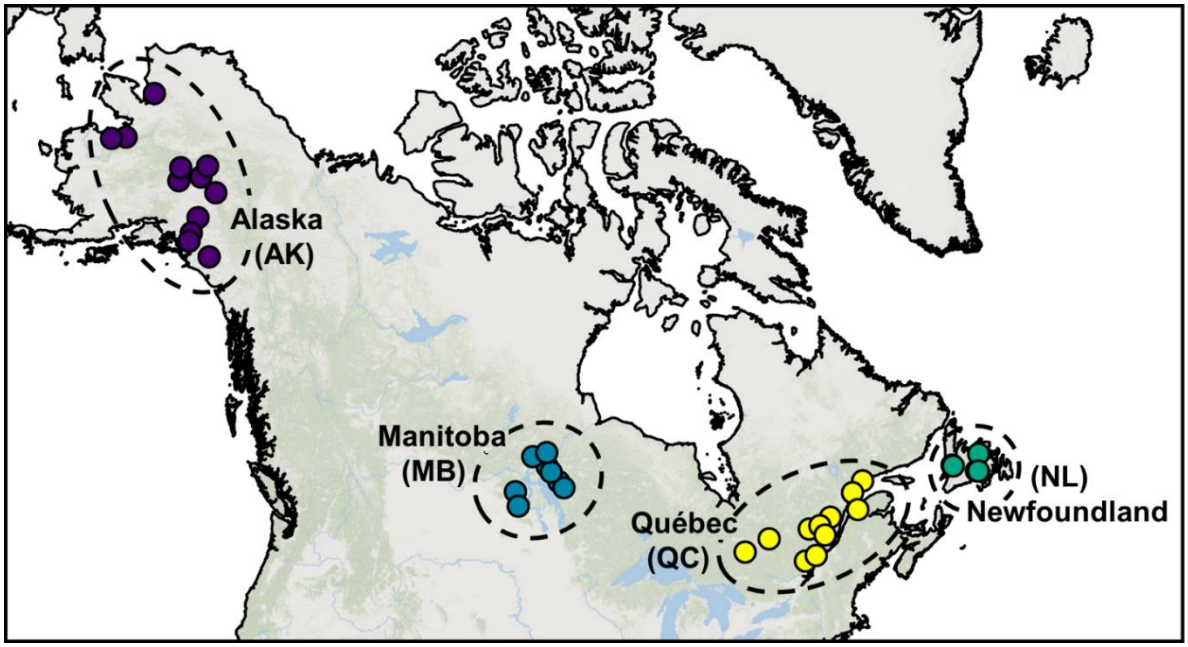


Figure 1. Sample distribution and study extent. Distribution of 95 Canada lynx (*Lynx canadensis*) samples across North America, used for high-throughput bisulfite sequencing. All four populations are delineated by colour and include 24 individuals, except Alaska (n = 23).

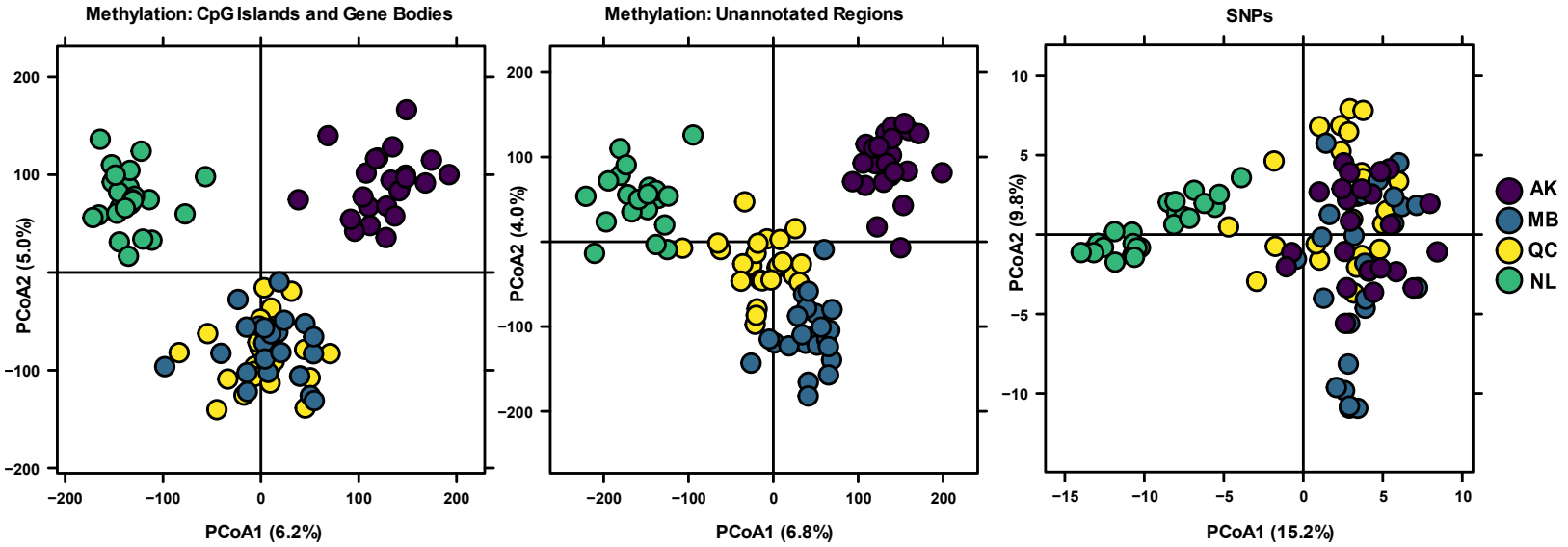


Figure 2. Epigenetic and genetic population structure. Principal coordinates analysis (PCoA) plots of variation across three molecular marker datasets, with individuals as single circles and populations delineated by colour. All molecular data was summarized with a pair-wise Euclidean dissimilarity matrix. Methylation was summarized with 5,000-bp running windows over CpG islands and gene bodies (n = 329) and over unannotated regions (n = 376). SNP variants were called from bisulfite converted reads and reflect unstructured mainland populations (n = 496).

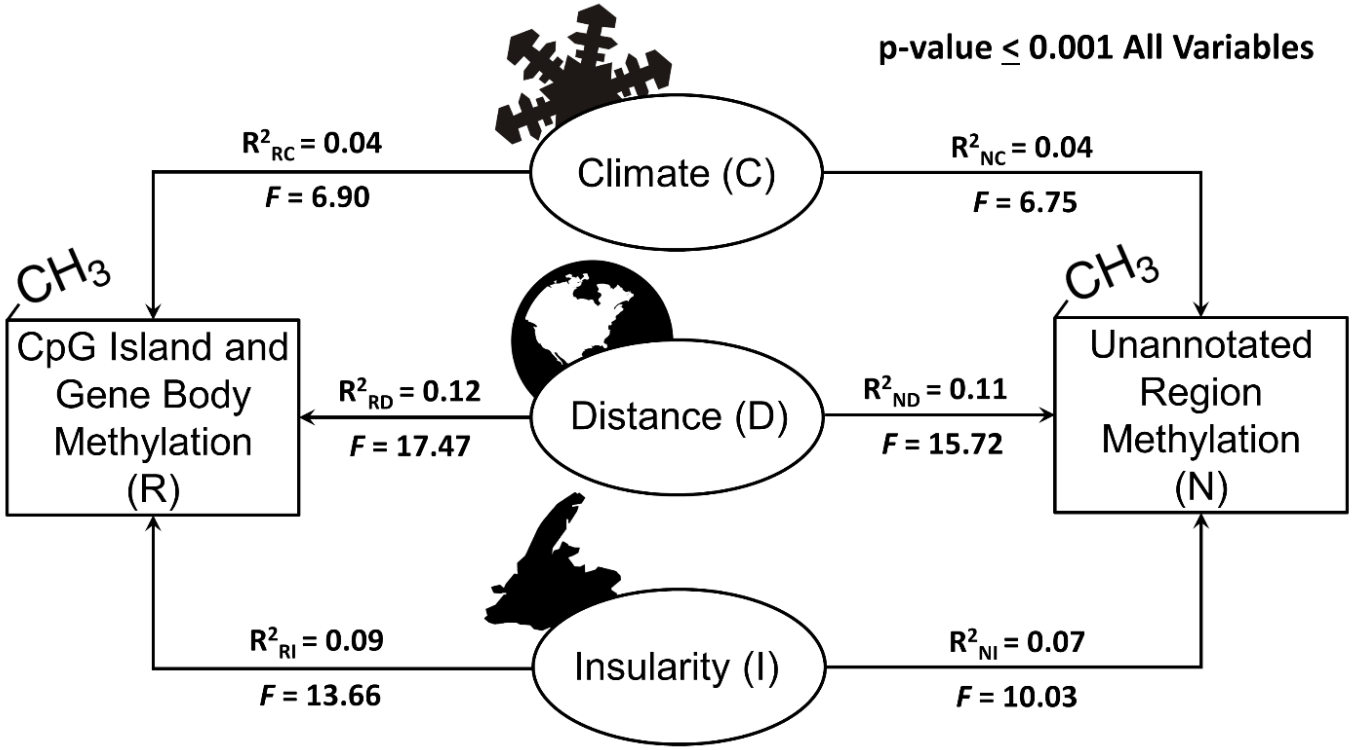


Figure 3. Importance of environmental variables in describing the epigenetic landscape. Visual depiction of partial distance-based redundancy analyses (p-db-RDAs). The effect sizes indicate the independent explanatory effects of each variable on explaining methylation patterns, subtracted from the effect of any other variable. The effect size is adjusted R^2, (adj. R^2) and the test-statistic is a pseudo-***F*** generated using QR decomposition within *vegan*.

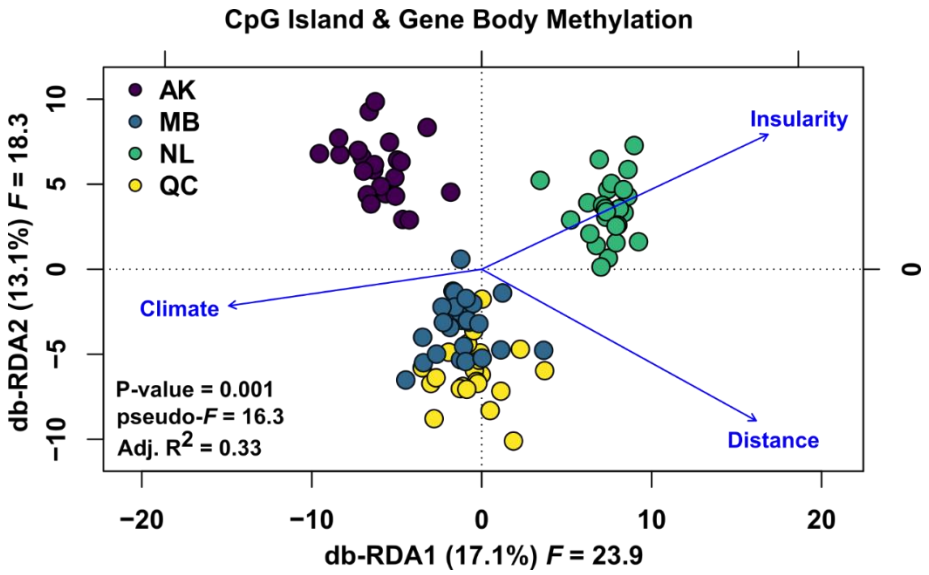


Figure 4. Associations between epigenetic data and environmental variables. Distance-based redundancy analysis (db-RDA) on DNA methylation data over CpG islands and gene bodies, with population delineated by colour. The axes of a principal coordinates analysis (PCoA) were used a response variable to determine biogeographical relationships. The explanatory variables included a distance variable (the first axis of a PCoA on latitude and longitude); insularity (a binary variable distinguishing the Newfoundland island population); and climate (the first axis of a PCA summarizing winter precipitation, annual temperature ranges, and coldest minimum temperature).

CHAPTER 4

GENERAL DISCUSSION

GENERAL DISCUSSION

Summary of Findings

My genetic data on *Canis* across eastern North America indicated a relatively unstructured hybrid-zone between canids. Contrary to my predictions, I found this unstructured pattern largely reflected even in macro-scale predator-prey relationships from niche models. I expected to see a substantial increase in model performance by incorporating moose habitat variables in my niche models for larger canids, and detected a pattern, although subtle, that suggests that biotic interactions only play a minor, and biologically trivial, role in species distribution models at regional scales. My systematic evaluation of biotic interactions using a two-pronged approach highlighted the need for critical interpretation of niche modeling outputs, while my overall investigation into canid niche dynamics identified extremely subtle ecological differentiation of canids across the hybrid zone, with convergence for affinity to most environmental variables. The results of my background similarity tests indicate that the canids within the study extent exhibit largely similar responses to environmental conditions, provided that they exist in the same background environment, reinforcing the overall unstructured patterns seen in the genetic data.

My genomic data in Canada lynx across its North American range supported a relatively panmictic, interconnected population (Rueness et al., 2003; Row et al., 2012), while my data on DNA methylation patterns revealed a hidden molecular architecture that might be indicative of local adaptation. The degree of structuring in genetic data for mainland populations is not mirrored in epigenetic

data, where individuals from Alaska harbored distinct epigenetic profiles that could be reflective of their larger body size and unique metabolic requirements for the landscape, suggested by differential methylation over genes related to metabolism. Furthermore, the hypermethylation of genes related to body mass and morphology in the Newfoundland individuals suggests that DNA methylation could be a molecular marker underlying rapid adaptive evolution, as this population is recently diverged (< 10,000 years) and exhibits unique morphological phenotypes. Overall, my results indicated that DNA methylation is a powerful and overlooked marker to investigate adaptive divergence and offered empirical results to push the discipline of ecological epigenetics.

Discussion of canid niche dynamics in a hybrid zone

The lack of a unique identity in niche space between most canids in the study extent is unexpected given previous research (Otis et al., 2017), but not entirely surprising given the low degree of genetic differentiation observed between groups. After accounting for sampling bias and parameterization, canids were largely undifferentiated for nearly all environmental variables. However, I did identify subtle associations between canid size and prey distributions, although the strength of the signal was weak and likely biologically meaningless. While I expected the differences to be consistent and substantial, my results revealed inconsistent signals between the canid genetic gradient and prey distirbutions. Consequently, my research suggests high ecological convergence among canids in the admixture zone at regional scales, and the unimportance of biotic interactions at this resolution, even in a highly dependent predator.

My results also indicated substantial variance between algorithms used to create species distributions models. Choice of algorithm affected both overall model fit (AUC_{test} / TSS) and evaluations of covariate importance, which suggests that thorough investigations into niche modeling methodology should be undertaken before assigning biological relevance. Notably, when examining differential covariate importance between canids, I found a larger effect size on algorithmic choice than between actual canid groups for most environmental predictors. This result is not completely unexpected given the lack of niche identity determined from background similarity tests, but is surprising given the initial covariate importance results that indicated unique environmental relationships for many canid groups. Specifically, the variance explained by algorithm choice was only half has important as differences between canid group when examining differential responses to deer habitat. Similarly, more variance was explained by algorithm than by canid group for the variable reflective of human disturbance.

Discussion on the epigenetic structure of Canada lynx populations

I detected a cryptic, underlying epigenetic structure in Canada lynx populations that would have been undetected with only genetic data. I find the patterns surprising, given the disparate nature of genetic and epigenetic data. Specifically, as DNA methylation is a modification to DNA that determines cell fate, identity, and expression (Kelsey et al., 2017), we would only hypothesize to detect differential methylation if it served a purpose, and thus would be unlikely to detect substantial epigenetic structure in a similar species, particularly one with such

low genetic differentiation. Our results thus indicate that DNA methylation might play a role in local adaptation to specific environments, reinforced by our data on specific differentially methylated regions over genes related to morphology and metabolism.

Our results indicate that DNA methylation can be a useful marker for detecting subtle population differentiation, and could even be worked into the current model of speciation (Feder et al., 2012). These results have far-reaching ramifications for detecting differentiation at a molecular level, and for examining early signatures of divergence in populations at an evolutionary level. Our results suggest that DNA methylation could be a marker indicative of early adaptive divergence, where differential patterns of methylation could act as initiators of genomic islands of divergence and could thus favor alleles with physical linkage to the methylated regions, and bring about the early stages of ecological speciation via linked selection (e.g. genetic draft, Via 2012).

Conclusions & Future Directions

This thesis examined the genetic and ecological determinants of wild canids across eastern North America and found only subtle differentiation at both levels. In contrast, my results indicate ecological convergence among canid groups and the negligible importance of biotic interactions at regional scales. My thesis also generated genomic and epigenomic data in a wild felid system, the first of its kind. We detected hidden epigenetic structure that was undetected by genetic data, which suggests the utility of incorporating DNA methylation into analyses of population differentiation.

Future directions in *Canis* would require detecting ancient vs. contemporary hybridization via assays of larger genomic regions and linking those patterns with fine-scale ecological data. Continent-wide sampling will likely be required to fully understand the *C. lupus, C. lycaon, C. latrans* system, incorporating both fossil records and contemporary distributions. Most importantly, fine-scale investigations of resource selection will likely be necessary to fully differentiate canids at an ecological level, as contemporary distributions exist as an artifact of historical extirpations as well as modern exclusion, so fine-scale resource use in dynamic environments will likely be needed to make the most veridical comparisons across groups. Further evaluation of molecular variation between canid groups could identify differential expression between groups that have functional correlates. Specifically, genes underlying coat morphology might be under stronger selective pressure in animals in colder environments, so both intra-group (north *latrans* vs. south *latrans*) and inter-group (*latrans* vs. *lupus*) transcriptomic and methylomic data could reveal important ecologically functional variation underlying these groups.

A future direction in the Canada lynx system would be to first develop a high-quality genome and methylome, and to identify explicitly differential patterns of methylation that are related to functional gene expression. Larger genomic regions need to be scanned across individuals in all sampling areas to evaluate signatures of selection, at which point transcriptomic data would be useful to further substantiate actual differences in expression. If differences in expression are identified, then comparative methylome analyses would provide insights into

the mechanisms underlying phenotypic, genomic, transcriptomic, and methylomic divergence in wild populations, thus informing the discipline on the molecular signatures of ecologically-based divergence and the processes underlying ecological speciation.

REFERENCES

Adaptwest (2015) Gridded current and projected climate data for North America at 1km resolution, interpolated using the ClimateNA v5.10 software (T. Wang et al., 2015). Available at adaptwest.databasin.org.

Aeschbacher, S., Selby, J.P., Willis, J.H., & Coop, G. (2017) Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences*, **114**, 7061–7066.

Allendorf, F.W., Leary, R.F., Spruell, P., & Wenburg, J.K. (2001) The problems with hybrids: Setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

Allouche, O., Tsoar, A., & Kadmon, R. (2006) Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

Anderson, R.P. (2017) When and how should biotic interactions be considered in models of species niches and distributions? *Journal of Biogeography*, **44**, 8–17.

Anderson, R.P., Peterson, A.T., & Gomez-Laverde, M. (2002) Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, **98**, 3–16.

Andrews, S. (2007) Seqmonk: A tool to visualise and analyse high throughput mapped sequence data. Available at https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/. Online.

Andrews, S. (2010) FastQC: A quality control tool for high throughput sequence data. Available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Online.

Aragón, P. & Sánchez-Fernández, D. (2013) Can we disentangle predator-prey interactions from species distributions at a macro-scale? A case study with a raptor species. *Oikos*, **122**, 64–72.

De Araújo, C.B., Marcondes-Machado, L.O., & Costa, G.C. (2014) The importance of biotic interactions in species distribution models: A test of the Eltonian noise hypothesis using parrots. *Journal of Biogeography*, **41**, 513–523.

Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.

Araújo, M.B. & Rozenfeld, A. (2014) The geographic scaling of biotic interactions. *Ecography*, **37**, 406–415.

Artemov, A. V., Mugue, N.S., Rastorguev, S.M., Zhenilo, S., Mazur, A.M.,

Tsygankova, S. V., Boulygina, E.S., Kaplun, D., Nedoluzhko, A. V., Medvedeva, Y.A., & Prokhortchouk, E.B. (2017) Genome-Wide DNA Methylation Profiling Reveals Epigenetic Adaptation of Stickleback to Marine and Freshwater Conditions. *Molecular Biology and Evolution*, **34**, 2203–2213.

Atauchi, P.J., Peterson, A.T., & Flanagan, J. (2018) Species distribution models for Peruvian plantcutter improve with consideration of biotic interactions. *Journal of Avian Biology*, **49**, 1–8.

Barbet-Massin, M. & Jetz, W. (2014) A 40-year, continent-wide, multispecies assessment of relevant climate predictors for species distribution modelling. *Diversity and Distributions*, **20**, 1285–1295.

Barbet-Massin, M. & Jiguet, F. (2011) Back from a predicted climatic extinction of an island endemic: A future for the Corsican Nuthatch. *PLoS ONE*, **6**, 1–8.

Barry, S. & Elith, J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413–423.

Bartoń, K. (2018) MuMIn : Multi-modal inference. Model selection and model averaging based on information criteria. R Package.

Bay, R.A., Harrigan, R.J., Underwood, V. Le, Gibbs, H.L., Smith, T.B., & Ruegg, K. (2018) Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, **359**, 83–86.

Beaumont, M.A. & Balding, D.J. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Benson, J. & Patterson, B. (2013a) Moose (*Alces alces*) predation by eastern coyotes (*Canis latrans*) and eastern coyote × eastern wolf (*Canis latrans* × *Canis lycaon*) hybrids. *Canadian Journal of Zoology*, **91**, 837–841.

Benson, J.F. & Patterson, B.R. (2013b) Inter-specific territoriality in a *Canis* hybrid zone: Spatial segregation between wolves, coyotes, and hybrids. *Oecologia*, **173**, 1539–1550.

Benson, J.F., Patterson, B.R., & Wheeldon, T.J. (2012) Spatial genetic and morphologic structure of wolves and coyotes in relation to environmental heterogeneity in a *Canis* hybrid zone. *Molecular Ecology*, **21**, 5934–5954.

Bergerud, A.T., Wyett, W., & Snider, B. (1983) The Role of Wolf Predation in Limiting a Moose Population. *The Journal of Wildlife Management*, **47**, 977–988.

Boria, R.A., Olson, L.E., Goodman, S.M., & Anderson, R.P. (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, **275**, 73–77.

Brown, J., Bennet, J., & French, C. (2017) SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species

distribution model analyses. *PeerJ*, **5**, 694–700.

Bullock, J.M., Edwards, R.J., Carey, P.D., & Rose, R.J. (2000) Geographical separation of two *Ulex* species at three spatial scales: does competition limit species' ranges? *Ecography*, **23**, 257–271.

Carbaugh, B., Vaughan, J.P., Bellis, E.D., & Graves, H.B. (1975) Distribution and Activity of White-Tailed Deer along an Interstate Highway. *The Journal of Wildlife Management*, **39**, 570–581.

Carmichael, L.E., Nagy, J.A., Larter, N.C., & Strobeck, C. (2001) Prey specialization may influence patterns of gene flow in wolves of the Canadian Northwest. *Molecular Ecology*, **10**, 2787–2798.

Chatterjee, T.K., Basford, J.E., Knoll, E., Tong, W.S., Blanco, V., Blomkalns, A.L., Rudich, S., Lentsch, A.B., Hui, D.Y., & Weintraub, N.L. (2014) HDAC9 knockout mice are protected from adipose tissue dysfunction and systemic metabolic disease during high-fat feeding. *Diabetes*, **63**, 176–187.

Chhatre, V.E. & Emerson, K.J. (2017) StrAuto: Automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*, **18**, 1–5.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., & Durbin, R. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life.* London.

DeFries, R.S., Hansen, M.C., Townshend, J.R.G., Janetos, A.C., & Loveland, T.R. (2000) A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, **6**, 247–254.

Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**, 927–930.

Dormann, C.F., Bobrowski, M., Dehling, D.M., Harris, D.J., Hartig, F., Lischke, H., Moretti, M.D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S.I., Sheppard, C.S., Steinbauer, M.J., Zeuss, D., & Kraan, C. (2018) Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, **27**, 1004–1016.

Dubin, M.J., Zhang, P., Meng, D., Remigereau, M.S., Osborne, E.J., Casale, F.P., Drewe, P., Kahles, A., Jean, G., Vilhjálmsson, B., Jagoda, J., Irez, S., Voronin, V., Song, Q., Long, Q., Rätsch, G., Stegle, O., Clark, R.M., & Nordborg, M. (2015) DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *eLife*, **4**, 1–23.

Dumond, M., Villard, M.A., & Tremblay, É. (2001) Does coyote diet vary seasonally between a protected and an unprotected forest landscape?

Ecoscience, **8**, 301–310.

Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eaton, M.D., Soberón, J., & Peterson, A.T. (2008) Phylogenetic perspective on ecological niche evolution in american blackbirds (Family *Icteridae*). *Biological Journal of the Linnean Society*, **94**, 869–878.

Ellegren, H. & Wolf, J.B.W. (2017) Parallelism in genomic landscapes of differentiation, conserved genomic features and the role of linked selection. *Journal of Evolutionary Biology*, **30**, 1516–1518.

ESA (2017) ESA Climate Change Initiative - Land Cover led by UC Louvain. Online.

F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.

Faurby, S. & Araújo, M.B. (2018) Anthropogenic range contractions bias species climate change forecasts. *Nature Climate Change*, **8**, 252–256.

Feder, J.L., Egan, S.P., & Nosil, P. (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

Feldman, R., Peers, M., Pickles, R., Thornton, D., & Murray, D. (2017) Climate driven niche divergence among host species affects range-wide patterns of parasitism. *Global Ecology and Conservation*, **9**, 1–10.

Feng, S., Jacobsen, S.E., & Reik, W. (2010) Epigenetic reprogramming in plant and animal development. *Science*, **330**, 622–627.

Ferrari, S.L.P. & Cribari-Neto, F. (2004) Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, **31**, 799–815.

Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302–4315.

Flatscher, R., Frajman, B., Schönswetter, P., & Paun, O. (2012) Environmental Heterogeneity and Phenotypic Divergence: Can Heritable Epigenetic Variation Aid Speciation? *Genetics Research International*, **2012**, 1–9.

Foll, M. & Gaggiotti, O. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**, 977–993.

Forster, M.M.R. (2000) Key Concepts in Model Selection: Performance and

Generalizability. *Journal of Mathematical Psychology*, **44**, 205–231.

Foster, J.B. (1964) Evolution of mammals on Islands. *Nature*, **202**, 234–235.

Fraterrigo, J.M., Wagner, S., & Warren, R.J. (2014) Local-scale biotic interactions embedded in macroscale climate drivers suggest Eltonian noise hypothesis distribution patterns for an invasive grass. *Ecology Letters*, **17**, 1447–1454.

Freeman, E.A. & Moisen, G.G. (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, **217**, 48–58.

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., & Nakao, M. (2003) Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *Journal of Biological Chemistry*, **278**, 24132–24138.

Galpern, P., Manseau, M., Hettinga, P., Smith, K., & Wilson, P. (2012) Allelematch: An R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. *Molecular Ecology Resources*, **12**, 771–778.

GBIF.org (2017) GBIF Occurrence Download http://doi.org/10.15468/dl.qi0pur (8th March 2017). Online.

Geffen, E., Adnerson, M.J., & Wayne, R.K. (2004) Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology*, **13**, 2481–2490.

Genner, M.J., Turner, G.F., Barker, S., & Hawkins, S.J. (1999) Niche segregation among Lake Malawi cichlid fishes? Evidence from stable isotope signatures. *Ecology Letters*, **2**, 185–190.

Godsoe, W., Franklin, J., & Blanchet, F.G. (2016) Effects of biotic interactions on modeled species' distribution can be masked by environmental gradients. *Ecology and Evolution*, 654–664.

Griffith, J.S. & Mahler, H.R. (1969) DNA ticketing theory of memory. *Nature*, **223**, 580–582.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Guo, W., Zhu, P., Pellegrini, M., Zhang, M.Q., Wang, X., & Ni, Z. (2017) CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics*, **34**, 381–387.

van Gurp, T.P., Wagemaker, N.C.A.M., Wouters, B., Vergeer, P., Ouborg, J.N.J., & Verhoeven, K.J.F. (2016) epiGBS: reference-free reduced representation bisulfite sequencing. *Nature Methods*, **13**, 322–324.

Han, L., Su, B., Li, W.-H., & Zhao, Z. (2008) CpG island density and its correlations with genomic features in mammalian genomes. *Genome Biology*, **9**, R79.

Hayes, R.D., Farnell, R., Ward, R.M.P., Carey, J., Dehn, M., Kuzyk, G.W., Baer, A.M., Gardner, C.L., & Donoghue, M.O. (2003) Experimental Reduction of Wolves in the Yukon: Ungulate Responses and Management Implications. *Wildlife Monographs*, **152**, 1–35.

Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G., & Körber, J.H. (2007) Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography*, **16**, 754–763.

Ho, M.W. & Saunders, P.T. (1979) Beyond neo-Darwinism-an epigenetic approach to evolution. *Journal of Theoretical Biology*, **78**, 573–591.

Jablonka, E. & Raz, G. (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly review of biology*, **84**, 131–76.

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jombart, T. (2008) Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jones, F.C., Grabherr, M.G., Chan, Y.F., et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Jones, P.A. (2012) Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, **13**, 484–492.

Jones, P.A. & Takai, D. (2001) The role of DNA methylation in mammalian epigenetics. *Science*, **293**, 1068–1070.

Kamvar, Z.N., Tabima, J.F., & Grünwald, N.J. (2014) *Poppr* : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Keller, T.E., Lasky, J.R., & Yi, S. V. (2016) The multivariate association between genomewide DNA methylation and climate across the range of *Arabidopsis thaliana*. *Molecular Ecology*, **25**, 1823–1837.

Kelsey, G., Stegle, O., & Reik, W. (2017) Single-cell epigenomics: Recording the past and predicting the future. *Science*, **358**, 69–75.

Kern, A.D. & Hahn, M.W. (2018) The Neutral Theory in Light of Natural Selection. *Molecular Biology and Evolution*, **35**, 1366–1371.

Khidas, K., Duhaime, J., & Huynh, H.M. (2013) Morphological Divergence of Continental and Island Populations of Canada Lynx. *Northeastern Naturalist*, **20**, 587–608.

Kilgo, J.C., Ray, H.S., Ruth, C., & Miller, K. V (2010) Can Coyotes Affect Deer Populations in Southeastern North America? *Journal of Wildlife Management*, **74**, 929–933.

Kimura, M. (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, **267**, 275–276.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., Mcinerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.C., Zimmermann, N.E., & O'Hara, R.B. (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Kittle, A.M., Anderson, M., Avgar, T., Baker, J.A., Brown, G.S., Hagens, J., Iwachewski, E., Moffatt, S., Mosser, A., Patterson, B.R., Reid, D.E.B., Rodgers, A.R., Shuter, J., Street, G.M., Thompson, I.D., Vander Vennen, L.M., & Fryxell, J.M. (2017) Landscape-level wolf space use is correlated with prey abundance, ease of mobility, and the distribution of prey habitat. *Ecosphere*, **8**, 1–17.

Knouft, J.H., Losos, J.B., Glor, R.E., & Kolbe, J.J. (2006) Phylogenetic analysis of the evolution of the niche in lizard of the *Anolis sagrei* group. *Ecology*, **87**, S29–S38.

Koen, E.L., Bowman, J., & Wilson, P.J. (2015) Isolation of peripheral populations of Canada lynx (*Lynx canadensis*). *Canadian Journal of Zoology*, **93**, 521–530.

Kolenosky, G.B. (1971) Hybridization between wolf and coyote. *Journal of Mammalogy*, **52**, 446–449.

Krueger, F. (2012) Trim Galore! : A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Available at https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. Online.

Krueger, F. & Andrews, S.R. (2011) Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.

Kyle, C.J., Johnson, A.R., Patterson, B.R., Wilson, P.J., Shami, K., Grewal, S.K., & White, B.N. (2006) Genetic nature of eastern wolves : Past, present and future. *Conservation Genetics*, **7**, 273–287.

Laine, V.N., Gossmann, T.I., Schachtschneider, K.M., et al. (2016) Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*, **7**, 1–9.

Lande, R. (1981) Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences*, **78**, 3721–3725.

Lea, A.J., Altmann, J., Alberts, S.C., & Tung, J. (2016) Resource base influences genome-wide DNA methylation levels in wild baboons (*Papio cynocephalus*). *Molecular Ecology*, **25**, 1681–1696.

Lehman, N., Eisenhawer, A., Hansen, K., Mech, L.D., Peterson, R., Gogan, P., & Wayne, R.K. (1991) Introgression of coyote mitochondrial DNA into sympatric North American gray wolf populations. *Evolution*, **45**, 104–119.

Leonard, J.A., Vilà, C., & Wayne, R.K. (2005) Legacy lost: Genetic variability and population size of extirpated US grey wolves (*Canis lupus*). *Molecular Ecology*, **14**, 9–17.

Levene, H. (1953) Genetic Equilibrium When More Than One Ecological Niche is Available. *The American Naturalist*, **87**, 331–333.

Lischer, H.E.L. & Excoffier, L. (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.

Lomolino, M. V. (2005) Body size evolution in insular vertebrates: Generality of the island rule. *Journal of Biogeography*, **32**, 1683–1699.

Lorincz, M.C., Dickerson, D.R., Schmitt, M., & Groudine, M. (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature Structural and Molecular Biology*, **11**, 1068–1075.

Le Luyer, J., Laporte, M., Beacham, T.D., Kaukinen, K.H., Withler, R.E., Leong, J.S., Rondeau, E.B., Koop, B.F., & Bernatchez, L. (2017) Parallel epigenetic modifications induced by hatchery rearing in a Pacific salmon. *Proceedings of the National Academy of Sciences*, **114**, 12964–12969.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2018) cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1.

Malone, E.W., Perkin, J.S., Leckie, B.M., Kulp, M.A., Hurt, C.R., & Walker, D.M. (2018) Which species, how many, and from where: Integrating habitat suitability, population genomics, and abundance estimates into species reintroduction planning. *Global Change Biology*, **24**, 3729–3748.

Maran, T. & Henttonen, H. (1994) Why is the European mink (*Mustela lutreola*) disappearing ? — A review of the process and hypotheses. *Annales Zoologici Fennici, Vol. 32, No. 1, II North European Symposium on the Ecology of Small and Medium-sized Carnivores, Lammi, Finland*, **8**, 47–54.

Markow, T.A. (2015) The secret lives of *Drosophila* flies. *eLife*, **4**, 1–9.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., et al. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, **466**, 253–257.

Mech, L.D. (2011) Non-genetic Data Supporting Genetic Evidence for the Eastern Wolf. *Northeastern Naturalist*, **18**, 521–526.

Mech, L.D., Asa, C.S., Callahan, M., Christensen, B.W., Smith, F., & Young, J.K. (2017) Studies of wolf x coyote hybridization via artificial insemination. *PLoS ONE*, **12**, 1–12.

Mech, L.D., Barber-Meyer, S.M., & Erb, J. (2016) Wolf (*Canis lupus*) generation time and proportion of current breeding females by age. *PLoS ONE*, **11**, 1–13.

Mech, L.D., Christensen, B.W., Asa, C.S., Callahan, M., & Young, J.K. (2014) Production of hybrids between western gray wolves and western coyotes. *PLoS ONE*, **9**, 1–7.

Meiri, S., Dayan, T., & Simberloff, D. (2005) Area, isolation and body size evolution in insular carnivores. *Ecology Letters*, **8**, 1211–1217.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E.H., Ek, M.B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., & Shi, W. (2006) North American regional reanalysis. *Bulletin of the American Meteorological Society*, **87**, 343–360.

Messier, F. (1994) Ungulate Population Models with Predation : A Case Study with the North American Moose. *Ecological Society of America*, **75**, 478–488.

Messier, R. (1995) Trophic interactions in two northern wolf-ungulate systems. *Wildlife Research*, **22**, 131–146.

Murray, D.L., Steury, T.D., & Roth, J.D. (2008) Assessment of Canada Lynx Research and Conservation Needs in the Southern Range: Another Kick at the Cat. *Journal of Wildlife Management*, **72**, 1463–1472.

Naimi, B. & Araújo, M.B. (2016) sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography*, **39**, 368–375.

Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133–142.

Newsome, T.M., Greenville, A.C., Ćirović, D., Dickman, C.R., Johnson, C.N., Krofel, M., Letnic, M., Ripple, W.J., Ritchie, E.G., Stoyanov, S., & Wirsing, A.J. (2017) Top predators constrain mesopredator distributions. *Nature Communications*, **8**, 1–7.

Nosil, P. (2008) Speciation with gene flow could be common. *Molecular Ecology*,

17, 2103–2106.

Nowak, R.M. (2002) The Original Status of Wolves in Eastern North America. *Southeastern Naturalist*, **1**, 95–130.

Otis, J., Thornton, D., Rutledge, L., & Murray, D.L. (2017) Ecological niche differentiation across a wolf-coyote hybrid zone in eastern North America. *Diversity and Distributions*, **23**, 529–539.

Owens, H.L., Campbell, L.P., Dornak, L.L., Saupe, E.E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C.M., Myers, C.E., & Peterson, A.T. (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, **263**, 10–18.

Palacio, F.X. & Girini, J.M. (2018) Biotic interactions in species distribution models enhance model performance and shed light on natural history of rare birds: a case study using the Straight-billed Reedhaunter (*Limnoctites rectirostris*). *Journal of Avian Biology*, **Accepted**, https://doi.org/10.1111/jav.01743.

Paquet, P.C. (1992) Prey Use Strategies of Sympatric Wolves and Coyotes in Riding Mountain National Park, Manitoba. *Journal of Mammalogy*, **73**, 337–343.

Partridge, L., Barrie, B., Fowler, K., & French, V. (1994) Evolution and Development of Body Size and Cell Size in *Drosophila melanogaster* in Response to Temperature. *Evolution*, **48**, 1269.

Peixoto, F.P., Villalobos, F., & Cianciaruso, M. V. (2017) Phylogenetic conservatism of climatic niche in bats. *Global Ecology and Biogeography*, **26**, 1055–1065.

Pembleton, L.W., Cogan, N.O.I., & Forster, J.W. (2013) StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, **13**, 946–952.

Peterson, R., Page, R., & Dodge, K. (1984) Wolves, Moose, and the Allometry of Population Cycles. *Science*, **224**, 1350–1352.

Phillips, S., Anderson, R., & Schapire, R. (2006) Maximum Entropy Modeling of Species Geographic Distributions. *Ecological Modelling*, **190**, 231–259.

Pickles, R.S.A., Thornton, D.H., Feldman, R.E., Marques, A., & Murray, D.L. (2013) Predicting shifts in parasite distribution with climate change: A multitrophic level approach. *Global Change Biology*, **19**, 2645–2654.

Prentice, M.B., Bowman, J., Khidas, K., Koen, E.L., Row, J.R., Murray, D.L., & Wilson, P.J. (2017) Selection and drift influence genetic differentiation of insular Canada lynx (*Lynx canadensis*) on Newfoundland and Cape Breton Island. *Ecology and Evolution*, **7**, 3281–3294.

Pritchard, J.K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Puttock, G.D., Shakotko, P., & Rasaputra, J.G. (1996) An empirical habitat model for moose, *Alces alces*, in Algonquin Park, Ontario. *Forest Ecology and Management*, **81**, 169–178.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. Online.

Radosavljevic, A. & Anderson, R.P. (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, **41**, 629–643.

Razgour, O., Persey, M., Shamir, U., & Korine, C. (2018) The role of climate, water and biotic interactions in shaping biodiversity patterns in arid environments across spatial scales. *Diversity and Distributions*, **24**, 1440–1452.

Robertson, M. & Richards, C. (2015) Opportunities and challenges of next-generation sequencing applications in ecological epigenetics. *Molecular Ecology*, **24**, 3799–3801.

Roemer, G.W., Gompper, M.E., & Van Valkenburgh, B. (2009) The ecological role of the mammalian mesocarnivore. *BioScience*, **59**, 165–173.

Row, J.R., Gomez, C., Koen, E.L., Bowman, J., Murray, D.L., & Wilson, P.J. (2012) Dispersal promotes high gene flow among Canada lynx populations across mainland North America. *Conservation Genetics*, **13**, 1259–1268.

Row, J.R., Wilson, P.J., Gomez, C., Koen, E.L., Bowman, J., Thornton, D.H., & Murray, D.L. (2014) The subtle role of climate change on population genetic structure in Canada lynx. *Global Change Biology*, **20**, 2076–2086.

Roy, M.S., Geffen, E., Smith, D., Ostrander, E. a, & Wayne, R.K. (1994) Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular biology and evolution*, **11**, 553–570.

Rueness, E.K., Stenseth, N.C., O'Donoghue, M., Boutin, S., Ellegren, H., & Jakobsen, K.S. (2003) Ecological and genetic spatial structuring in the Canadian lynx. *Nature*, **425**, 69–72.

Rundle, H.D., Nagel, L., Boughman, J.W., & Schluter, D. (2000) Natural Selection and Parellel Speciation in Sympatric Sticklebacks. *Science*, **287**, 306–308.

Rutledge, L.Y., Devillard, S., Boone, J.Q., Hohenlohe, P.A., & White, B.N. (2015) RAD sequencing and genomic simulations resolve hybrid origins within North American *Canis*. *Biology Letters*, **11**, 20150303. 1-4.

Rutledge, L.Y., Garroway, C.J., Loveless, K.M., & Patterson, B.R. (2010) Genetic differentiation of eastern wolves in Algonquin Park despite bridging gene flow between coyotes and grey wolves. *Heredity*, **105**, 520–531.

Rutledge, L.Y., Wilson, P.J., Klütsch, C.F.C., Patterson, B.R., & White, B.N. (2012) Conservation genomics in perspective: A holistic approach to understanding *Canis* evolution in North America. *Biological Conservation*, **155**, 186–192.

Saykally, J.N., Dogan, S., Cleary, M.P., & Sanders, M.M. (2009) The ZEB1 Transcription Factor Is a Novel Repressor of Adiposity in Female Mice. *PLoS ONE*, **4**, e8460.

Schluter, D. (2009) Evidence for Ecological Speciation and Its Alternative. *Science*, **323**, 737–741.

Schwartz, M.K., Mills, L.S., McKelvey, K.S., Ruggiero, L.F., & Allendorf, F.W. (2002) DNA reveals high dispersal synchronizing the population dynamics of Canada lynx. *Nature*, **415**, 520–522.

Shafer, A.B.A. & Wolf, J.B.W. (2013) Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.

Smith, T.A., Martin, M.D., Nguyen, M., & Mendelson, T.C. (2016) Epigenetic divergence as a potential first step in darter speciation. *Molecular Ecology*, **25**, 1883–1894.

Smolik, M.G., Dullinger, S., Essl, F., Kleinbauer, I., Leitner, M., Peterseil, J., Stadler, L.M., & Vogl, G. (2010) Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *Journal of Biogeography*, **37**, 411–422.

Soberón, J. & Nakamura, M. (2009) Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences, USA*, **106**, 19644–19650.

Stenseth, N.C., Chan, K.-S., Tong, H., Boonstra, R., Boutin, S., Krebs, C.J., Post, E., O'Donoghue, M., Yoccoz, N.G., Forchhammer, M.C., & Hurrell, J.W. (1999) Common Dynamic Structure of Canada Lynx Populations Within Three Climatic Regions. *Science*, **285**, 1071–1073.

Stenseth, N.C., Shabbar, A., Chan, K.-S., Boutin, S., Rueness, E.K., Ehrich, D., Hurrell, J.W., Lingjaerde, O.C., & Jakobsen, K.S. (2004) Snow conditions may create an invisible barrier for lynx. *Proceedings of the National Academy of Sciences*, **101**, 10632–10634.

Stronen, A. V., Tessier, N., Jolicoeur, H., Paquet, P.C., Hénault, M., Villemure, M., Patterson, B.R., Sallows, T., Goulet, G., & Lapointe, F.-J. (2012) Canid hybridization: contemporary evolution in human-modified landscapes. *Ecology and Evolution*, **2**, 2128–2140.

Tabery, J. (2008) R. A. Fisher, Lancelot Hogben, and the Origin(s) of Genotype-Environment Interaction. *Journal of the History of Biology*, **41**, 717–761.

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, **408**, 796–815.

Thornton, D.H. & Murray, D.L. (2014) Influence of hybridization on niche shifts in expanding coyote populations. *Diversity and Distributions*, **20**, 1355–1364.

Turelli, M., Barton, N.H., & Coyne, J.A. (2001) Theory and speciation. *Trends in Ecology & Evolution*, **16**, 330–343.

Uyeda, J.C., Arnold, S.J., Hohenlohe, P.A., & Mead, L.S. (2009) Drift promotes speciation by sexual selection. *Evolution*, **63**, 583–594.

Van Valen, L. (1973) Pattern and the balance of nature. *Evolutionary Theory*, **1**, 31–49.

Via, S. (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 451–460.

vonHoldt, B.M., Brzeski, K.E., Wilcove, D.S., & Rutledge, L.Y. (2018) Redefining the Role of Admixture and Genomics in Species Conservation. *Conservation Letters*, **11**, 1–6.

VonHoldt, B.M., Cahill, J.A., Fan, Z., Gronau, I., Robinson, J., Pollinger, J.P., Shapiro, B., Wall, J., & Wayne, R.K. (2016) Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Science Advances*, 1–14.

vonHoldt, B.M., Kays, R., Pollinger, J.P., & Wayne, R.K. (2016) Admixture mapping identifies introgressed genomic regions in North American canids. *Molecular Ecology*, **25**, 2443–2453.

vonHoldt, B.M., Pollinger, J.P., Earl, D.A., Knowles, J.C., Boyko, A.R., Parker, H., Geffen, E., Pilot, M., Jedrzejewski, W., Jedrzejewska, B., Sidorovich, V., Greco, C., Randi, E., Musiani, M., Kays, R., Bustamante, C.D., Ostrander, E.A., Novembre, J., & Wayne, R.K. (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Research*, **21**, 1294–1305.

Warren, D.L., Glor, R.E., & Turelli, M. (2010) ENMTools : a toolbox for comparative studies of environmental niche models. *Ecography*, **33**, 607–611.

Warren, D.L. & Seifert, S. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335–342.

Wayne, R.K., George, S.B., Gilbert, D., Collins, P.W., Kovech, S.D., Girman Derek, & Lehman, N. (1991) A morphologic and genetic study of the Island

fox, *Urocyon littoralis*. *Evolution*, **45**, 1849–1869.

Weber, K.L., Fischer, R.S., & Fowler, V.M. (2007) Tmod3 regulates polarized epithelial cell morphology. *Journal of Cell Science*, **120**, 3625–3632.

Wheeldon, T.J. & Patterson, B.R. (2012) Genetic and morphological differentiation of wolves (*Canis lupus*) and coyotes (*Canis latrans*) in northeastern Ontario. *Canadian Journal of Zoology*, **90**, 1221–1230.

Wheeldon, T.J. & White, B.N. (2009) Genetic analysis of historic western Great Lakes region wolf samples reveals early *Canis lupus / lycaon* hybridization. *Biology Letters*, **23**, 101–104.

Wilson, P.J., Grewal, S.K., Lawford, I.D., Heal, J.N.M., Granacki, A.G., Pennock, D., Theberge, J.B., Theberge, M.T., Voigt, D.R., Waddell, W., Chambers, R.E., Paquet, P.C., Goulet, G., Cluff, D., & White, B.N. (2000) DNA profiles of the eastern Canadian wolf and the red wolf provide evidence for a common evolutionary history independent of the gray wolf. *Canadian Journal of Zoology*, **78**, 2156–2166.

Wisz, M.S., Pottier, J., Kissling, W.D., et al. (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Wolf, J.B.W. & Ellegren, H. (2017) Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, **18**, 87–100.

Wolf, J.B.W., Lindell, J., & Backström, N. (2010) Speciation genetics: current status and evolving approaches. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 1717–1733.

Wu, H., Caffo, B., Jaffee, H.A., Irizarry, R.A., & Feinberg, A.P. (2010) Redefining CpG islands using hidden Markov models. *Biostatistics*, **11**, 499–514.

Wutz, A. & P. Barlow, D. (1998) Imprinting of the mouse Igf2r gene depends on an intronic CpG island. *Molecular and Cellular Endocrinology*, **140**, 9–14.

Yoder, A.D., Poelstra, J., Tiley, G.P., & Williams, R. (2018) Neutral Theory is the Foundation of Conservation Genetics. *Molecular Biology and Evolution*, **35**, 1–5.

Van Zyll De Jong, C.G. (1975) Differentiation of the Canada lynx, *Felis (Lynx) canadensis subsolana*, in Newfoundland. *Canadian Journal of Zoology*, **53**, 699–705.

APPENDIX A: CHAPTER 2 SUPPORTING INFORMATION

Genotyping and Genetic Analyses

We initially filtered our genotypic dataset for duplicate individuals and those with an excess of missing alleles using *allelematch* (Galpern et al., 2012), which removed individuals that shared ≥ 21 alleles or had ≥ 4 missing alleles. Missing data across individuals was low (0.41% average across all loci) and affected all groups and loci comparably (Supplemental Figure S14).

We identified canid species and hybrids from our genotypic data of individuals with unknown ancestry (n = 1,230) using the Bayesian clustering program *structure v2.3.4* (Pritchard et al., 2000) with parameterized admixture models using a burn-in of 5×10^5 followed by 1×10^6 MCMC iterations, which were run in parallel in UNIX using *strauto* (Chhatre & Emerson, 2017). The simulations were replicated 10 times for K = 1 - 10 and the entire analysis was replicated with and without the `PopFlag` parameter, which predefines individuals of known ancestry. Our individuals of known ancestry included 47 coyotes from Saskatchewan, 41 gray wolves from the Northwest Territories, and 48 of the highest assigned (*Q*-value) eastern wolves from Ontario, openly available on DRYAD (Rutledge et al., 2010). Gray wolves and coyotes from these areas were chosen as they lie outside the putative admixture zone and are more likely to be unadmixed (Mech, 2011).

The optimal value of genetic clusters (K) was determined using *Structure Harvester* (Earl & vonHoldt, 2012) based on the highest ΔK relative to MeanLnP(K). These criteria identified K = 2 as the optimal number of clusters (Supplemental Figure S15), which likely corresponds to the hypothesized new

world (*C. lycaon, C. latrans*), and old world (*C. lupus*) originations of these species. We chose K = 3 for our final analysis, which captured the substructure that exists between eastern wolves and eastern coyotes and correctly assigned 91.9% (n = 125) of our individuals of known ancestry (n = 136). *Structure* results for this value of K were averaged with *CLUMPP* using the `greedy` clustering algorithm (Jakobsson & Rosenberg, 2007) and were plotted using with a PCA in *adegenet* (Jombart, 2008). Some individuals (n = 199; 14.6%) received variable group assignment with fluctuations in *Q*-value threshold (e.g., 0.7, 0.8, 0.9) and were examined qualitatively with a PCA plot (Supplemental Figure S16). Individuals with known ancestry were similarly visualized with a PCA plot to examine the distribution of raw genotypic data (Supplemental Figure S17). Additionally, a traditional structure-style plot was created using a function in R for all genotyped individuals ((R Core Team, 2017); Supplemental Figure S18), and assignment was compared with known individuals for both K = 2 and K = 3 (Supplemental Figure S18). We assessed the sensitivity of our *Q*-value threshold (0.8) by repeating the niche modeling process for *Q*-values at both 0.7 and 0.9 and calculated niche overlap and Pearson's correlations to ensure consistent conclusions (Supplemental Figure S19).

Environmental Data and the Pre-Modeling Checklist

Due to the extensive latitudinal range of this study, we projected all occurrences and environmental data into an equal area projection (North America Albers Equal-Area Projection).

Large mammals have extensive dispersal capabilities, so we created Euclidean distance rasters from extracted land class features in order to avoid categorical designations. We extracted *Urban Areas* (Value 190, (ESA, 2017)*)* values from a land class raster and converted the values to a Euclidean distance matrix across the study extent, which formed our variable '*Distance to Human Areas'*. We repeated this process for the same land class layer, except extracting all *Flooded* raster values, as we hypothesized this may be a habitat feature associated with moose habitats (values 160, 170, 180 (ESA, 2017).

Occurrence locations potentially suffered from sampling bias due to opportunistic sampling, so spatial data was thinned at three distance classes based on environmental autocorrelation. Moose and white-tailed deer spatial data were primarily obtained from online geodatabases and putatively contained a higher degree of sampling bias, so we thinned this dataset at a more conservative scale of 50 – 75 km. Data were thinned at these distance classes based on an environmental heterogeneity raster created within sdmtoolbox (Brown et al., 2017), which determines environmental variation in species-specific covariates using a principal components analysis (PCA) and thins species occurrences based on environmental autocorrelation. A total of 250 spatially autocorrelated occurrences were removed from the moose dataset, while 742 were removed from the white-tailed deer dataset. This left a remaining 142 occurrence locations for moose and 99 for white-tailed deer, which we deemed sufficient for our needs (Supplemental Figure S2). Canid occurrence locations were thinned by species-group at three distance classes (5, 15, 25 km) based on an environmental heterogeneity raster

created using their specific covariates. This process was repeated for all groups across the three *Q*-value thresholds (0.7, 0.8, 0.9) for the sensitivity analysis, giving a variable number of total occurrences per group (Supplemental Table S2).

Prey distribution models

Environmental variables for white-tailed deer distribution modeling included both climatic and biotic variables: 1) mean annual temperature; 2) snow precipitation; 3) percent tree cover; and 4) distance to human-use areas (DeFries et al., 2000; Adaptwest, 2015; ESA, 2017). Moose models were created using the same climatic and tree cover data (1-3) as well as: 5) distance to flooded areas (instead of distance to human-use areas, #4). These variables were created using landcover data (ESA, 2017) and specifically chosen because white-tailed deer distributions are distinguished by their higher tolerance to anthropogenic activity and avoidance of deep snow (Carbaugh et al., 1975), whereas moose distributions are strongly influenced by colder winter conditions and more tree cover (Puttock et al., 1996).

Ensemble models for white-tailed deer suggest a latitudinally-driven gradient in suitable habitat, with most suitable habitat occurring in the southern range (Supplemental Figure S9). All models had reasonably strong fit (mean ± SE: AUC_{test} = 0.72 ± 0.04), with the most important relationship being a negative link with mean annual temperature (permutation importance = 32 ± 1.6%), followed by a positive association with distance to human areas (51 ± 0.8%; Supplemental Table S8, Supplemental Figure S10). Similarly, moose habitats (AUC_{test} = 0.64 ± 0.04; Supplemental Figure S9) were well predicted by mean a negative link to

annual temperature (22 ± 1.7%) but were also associated positively with snow precipitation (12 ± 1.3%). MaxEnt response curves for mean annual temperature revealed a gradual increase in white-tailed deer habitat, with increasing mean annual temperature, while moose have a preferred average annual temperature around 5° C (Supplemental Figure S11). Interestingly, moose and white-tailed deer have inverse responses to precipitation as snow, with moose habitat quality increasing with greater amounts of snow (Supplemental Figure S11).

Niche modelling evaluation

We evaluated ensemble niche models using both AUC_{test} and TSS. The TSS has a lower dependence on prevalence than kappa for model evaluation (Allouche et al., 2006), and is calculated using sum of sensitivity and specificity minus one. Our criteria for optimizing the threshold for evaluation with the TSS was based on maximum sensitivity and specificity, as prevalence was unknown and likely was variable across canid groups (see Freeman and Moisen 2008).

CHAPTER 2 SUPPLEMENTAL TABLES

Supplemental Table S1. Reagent quantities for genotyping canid samples in three multiplexed PCRs, as well as PCR cycling conditions.

Reagent	Concentration	Units	Per Reaction (µl)
Buffer	5	X	3
dNTPs	10	mM	0.3
$MgCl_2$	25	mM	0.9
CXX377-F	40	µM	0.2
CXX377-R	40	µM	0.2
CXX225-F	40	µM	0.1
CXX225-R	40	µM	0.1
CXX123-F	40	µM	0.1
CXX123-R	40	µM	0.1
CXX200-F	40	µM	0.1
CXX200-R	40	µM	0.1
Taq	5	U/ µl	0.15
DNA	2.5	ng/µl	1
ddH_2O		µl	8.65
		Total	15

PCR Cycling Conditions			
94° C			5 min
x 29	[	94° C	30 sec
		56° C	1 min
		72° C	1 min
60° C			45 min

Reagent	Concentration	Units	Per Reaction (µl)
Buffer	5	X	3
dNTPs	10	mM	0.3
$MgCl_2$	25	mM	0.9
CXX109-F	40	µM	0.1
CXX109-R	40	µM	0.1
CXX172-F	40	µM	0.1
CXX172-R	40	µM	0.1
CXX204-F	40	µM	0.1
CXX204-R	40	µM	0.1
CXX250-F	40	µM	0.1
CXX250-R	40	µM	0.1
Taq	5	U/ µl	0.15
DNA	2.5	ng/µl	1
ddH_2O		µl	8.85
		Total	15

Reagent	Concentration	Units	Per Reaction (µl)
Buffer	5	X	3
dNTPs	10	mM	0.3
$MgCl_2$	25	mM	0.9
CXX253-F	40	µM	0.1
CXX253-R	40	µM	0.1
CXX147-F	40	µM	0.1
CXX147-R	40	µM	0.1
CXX410-F	40	µM	0.1
CXX410-R	40	µM	0.1
CXX442-F	40	µM	0.1
CXX442-R	40	µM	0.1
Taq	5	U/ µl	0.15
DNA	2.5	ng/µl	1
ddH_2O		µl	8.85
		Total	15

Supplemental Table S2. Occurrence locations at each *Q*-value threshold (for canids), before (n) and after (n-t) spatial thinning at three rarefication distance classes (5, 15, 25 km), based on environmental autocorrelation. Prey species were thinned at two distance classes (50, 75 km), with raw occurrence points indicated in the first column, and thinned occurrences in the second.

Species Group	n (0.7)	n-t (0.7)	n (0.8)	n-t (0.8)	n (0.9)	n-t (0.9)
Canis latrans	496	150	479	149	442	141
Canis lycaon x Canis latrans	50	29	75	35	107	46
Canis lycaon	189	41	165	33	136	23
Canis lycaon x Canis lupus	34	25	72	50	126	70
Canis lupus	466	144	444	141	406	131
Moose	392	142	-	-	-	-
White-tailed Deer	841	99	-	-	-	-

Supplemental Table S3. Pearson correlation coefficients between canid environmental variables, used to assess multicollinearity between predictor variables.

	Deer Habitat	Distance to Humans	Moose Habitat	Snow Precipitation	Snow Depth	Tree Cover
Deer Habitat	1	-0.7	-0.35	-0.74	-0.72	-0.13
Distance to Humans	NA	1	-0.09	0.39	0.55	0.02
Moose Habitat	NA	NA	1	0.57	0.23	0.36
Snow Precipitation	NA	NA	NA	1	0.48	0.31
Snow Depth	NA	NA	NA	NA	1	0.08
Tree Cover	NA	NA	NA	NA	NA	1

Supplemental Table S4. All combinations of variables and IDs for model permutations. Permutations of ensemble SDMs were done to determine the impacts of incorporating prey variables into predator SDMs. Three covariates were used in each permutation, and each permutation was assigned a value (*Prey Type*) based on which prey covariates were included in the permutation. Each permutation was created using the same six algorithms used in the full models in triplicate, and model fit was evaluated with AUC_{test} and TSS.

PAS: precipitation as snow, SD: snow depth, Moose: moose habitat, DEER: deer habitat, DISHUM: distance to human areas, TCOV: tree cover, MAT: mean annual temperature, TAVE: average winter temperature.

Numerical ID	Permutation ID	1st Covariate	2nd Covariate	3rd Covariate	Percent Prey	Prey Type
1	A	PAS	SD	MOOSE	0.33	MOOSE
2	B	PAS	SD	DEER	0.33	DEER
3	C	PAS	SD	TCOV	0	NONE
4	D	PAS	SD	DISHUM	0	NONE
5	E	PAS	MOOSE	DEER	0.66	BOTH
6	F	PAS	MOOSE	TCOV	0.33	MOOSE
7	G	PAS	MOOSE	DISHUM	0.33	MOOSE
8	H	PAS	DEER	TCOV	0.33	DEER
9	I	PAS	DEER	DISHUM	0.33	DEER
10	J	PAS	TCOV	DISHUM	0	NONE
11	K	SD	MOOSE	DEER	0.66	BOTH
12	L	SD	MOOSE	TCOV	0.33	MOOSE
13	M	SD	MOOSE	DISHUM	0.33	MOOSE
14	N	SD	DEER	TCOV	0.33	DEER
15	O	SD	DEER	DISHUM	0.33	DEER
16	P	SD	TCOV	DISHUM	0	NONE
17	Q	MOOSE	DEER	TCOV	0.66	BOTH
18	R	MOOSE	DEER	DISHUM	0.66	BOTH
19	S	MOOSE	TCOV	DISHUM	0.33	MOOSE
20	T	DEER	TCOV	DISHUM	0.33	DEER
21	AA	PAS	SD	MAT	0	NONE
22	BB	PAS	SD	TAVE	0	NONE
23	CC	PAS	MOOSE	MAT	0.33	MOOSE
24	DD	PAS	MOOSE	TAVE	0.33	MOOSE
25	EE	PAS	DEER	MAT	0.33	DEER
26	FF	PAS	DEER	TAVE	0.33	DEER
27	GG	PAS	TCOV	MAT	0	NONE
28	HH	PAS	TCOV	TAVE	0	NONE
29	II	PAS	DISHUM	MAT	0	NONE

30	JJ	PAS	DISHUM	TAVE	0	NONE
31	KK	PAS	MAT	TAVE	0	NONE
32	LL	SD	MOOSE	MAT	0.33	MOOSE
33	MM	SD	MOOSE	TAVE	0.33	MOOSE
34	NN	SD	DEER	MAT	0.33	DEER
35	OO	SD	DEER	TAVE	0.33	DEER
36	PP	SD	TCOV	MAT	0	NONE
37	QQ	SD	TCOV	TAVE	0	NONE
38	RR	SD	DISHUM	MAT	0	NONE
39	SS	SD	DISHUM	TAVE	0	NONE
40	TT	SD	MAT	TAVE	0	NONE
41	UU	MOOSE	DEER	MAT	0.66	BOTH
42	VV	MOOSE	DEER	TAVE	0.66	BOTH
43	WW	MOOSE	TCOV	MAT	0.33	MOOSE
44	XX	MOOSE	TCOV	TAVE	0.33	MOOSE
45	YY	MOOSE	DISHUM	MAT	0.33	MOOSE
46	ZZ	MOOSE	DISHUM	TAVE	0.33	MOOSE
47	AAA	MOOSE	MAT	TAVE	0.33	MOOSE
48	BBB	DEER	TCOV	MAT	0.33	DEER
49	CCC	DEER	TCOV	TAVE	0.33	DEER
50	DDD	DEER	DISHUM	MAT	0.33	DEER
51	EEE	DEER	DISHUM	TAVE	0.33	DEER
52	FFF	DEER	MAT	TAVE	0.33	DEER
53	GGG	TCOV	DISHUM	MAT	0	NONE
54	HHH	TCOV	DISHUM	TAVE	0	NONE
55	III	TCOV	MAT	TAVE	0	NONE
56	JJJ	DISHUM	MAT	TAVE	0	NONE

Supplemental Table S5. Percent increase in model evaluation metrics (AUC, TSS) per canid depending on prey variables included in model permutations. Permutation models that included no prey were considered baseline, and percent increase was calculated by the average model performance of the indicated group divided by the baseline. (Top) shows the impacts on model evaluation metrics for a categorical variable of prey type, and the bottom table shows the impact when considering the number of prey in models as a discrete percentage. No qualitative difference was seen between the two, and both were necessarily correlated, so we chose to report the categorical prey type variable as that provides the most insight into comparing responses to specific prey.

	Prey Type	Fold Change AUC (TSS)
C. latrans	Both Prey	0.99 (0.96)
	Deer Only	0.99 (0.96)
	Moose Only	0.98 (0.95)
C. lycaon x C. latrans	Both Prey	1.02 (1.02)
	Deer Only	1 (1)
	Moose Only	1 (0.99)
C. lycaon	Both Prey	1.04 (1.07)
	Deer Only	1 (1)
	Moose Only	1.05 (1.09)
C. lycaon x C. lupus	Both Prey	1 (1.02)
	Deer Only	0.99 (0.97)
	Moose Only	1.02 (1.05)
C. lupus	Both Prey	0.99 (0.97)
	Deer Only	1.01 (1.04)
	Moose Only	0.98 (0.94)

	Prey %	Fold Increase AUC (TSS)
C. latrans	0.33	0.98 (0.95)
	0.66	0.99 (0.96)
C. lycaon x C. latrans	0.33	1 (1)
	0.66	1.02 (1.02)
C. lycaon	0.33	1.03 (1.04)
	0.66	1.04 (1.07)
C. lycaon x C. lupus	0.33	1.01 (1.01)
	0.66	1 (1.02)
C. lupus	0.33	1 (0.99)
	0.66	0.99 (0.97)

Supplemental Table S6. Linear mixed-model results per species-group examining the effects of TSS (response variable) against a categorical variable of prey type (i.e. both preys habitats, moose only, or deer only). 5,040 model permutations with all combinations of covariates were used to assess the impacts of incorporating prey into predator SDMs. This is similar to supplemental table S9 except with the TSS instead of AUC_{test}.

Species-Group	Both Prey TSS (*t-statistic*)	Moose TSS (*t-statistic*)	Deer TSS (*t-statistic*)	R2M	R2C
C. latrans	0.37 (-2.66)	0.37 (-4.52)	0.37 (-3.12)	0.02	0.28
C. lycaon x C. latrans	0.53 (1.13)	0.51 (-0.54)	0.52 (0.3)	0	0.17
C. lycaon	0.59 (3.25)	0.6 (5.43)	0.55 (0.21)	0.03	0.17
C. lycaon x C. lupus	0.42 (0.9)	0.43 (2.88)	0.4 (-2.14)	0.02	0.1
C. lupus	0.38 (-1.7)	0.37 (-4.09)	0.41 (2.78)	0.04	0.13

Supplemental Table S7. Pairwise F_{ST} matrices for canids outside the admixture zone (top; *C. lupus* – Northwest Territories, *C. lycaon* – Algonquin Provincial Park, *C. latrans* – Saskatchewan), and the 408 individuals used for this study (bottom) within the admixture zone.

	C. lupus	C. lycaon	C. latrans
C. lupus	0	-	-
C. lycaon	0.136	0	-
C. latrans	0.126	0.096	0

	C. latrans	C. lycaon x C. latrans	C. lycaon	C. lycaon x C. lupus	C. lupus
C. latrans	0	-	-	-	-
C. lycaon x C. latrans	0.009	0	-	-	-
C. lycaon	0.029	0.021	0	-	-
C. lycaon x C. lupus	0.035	0.019	0.014	0	-
C. lupus	0.066	0.026	0.022	0.01	0

Supplemental Table S8. (top) Permutation importance for covariates used for creating prey ensemble models. Means were obtained by averaging permutation importance across all component models that were created to generate ensembles (N = 54) and included 9 replicate models of 6 different modeling algorithms. (bottom) Prey ensemble model fit, evaluated using both AUC_{test} and TSS.

Covariate	Species	Mean	SD	SE
Distance to Human Areas	Deer	0.051	0.062	0.008
Mean Annual Temperature	Deer	0.32	0.118	0.016
Snow Precipitation	Deer	0.028	0.031	0.004
Tree Cover	Deer	0.007	0.016	0.002
Distance to Flood	Moose	0.015	0.019	0.003
Mean Annual Temperature	Moose	0.217	0.122	0.017
Snow Precipitation	Moose	0.118	0.096	0.013
Tree Cover	Moose	0.017	0.021	0.003

	Evaluation Metric	Mean	SD	SEM
White-tailed Deer	AUC	0.72	0.043	0.006
	TSS	0.403	0.052	0.007
Moose	AUC	0.644	0.044	0.006
	TSS	0.274	0.062	0.008

Supplemental Table S9. Model evaluation metrics for final ensembles and each algorithm independently per canid group. Mean AUC and TSS for overall ensembles were calculated by averaging across nine replicates for all six algorithms. BRT: Boosted regression tree, RF: random forest, GAM: generalized additive models, MARS: multivariate adaptive regression splines, MDA: mixture discriminant analysis.

Group	Evaluation Metric	Overall	MaxEnt	BRT	RF	GAM	MARS	MDA
C. latrans	AUC	0.74 ± 0.03	0.76 ± 0.01	0.7 ± 0.03	0.72 ± 0.02	0.76 ± 0.01	0.76 ± 0.01	0.75 ± 0.01
	TSS	0.41 ± 0.05	0.43 ± 0.02	0.35 ± 0.02	0.39 ± 0.04	0.45 ± 0.03	0.43 ± 0.03	0.45 ± 0.03
C. lycaon x C. latrans	AUC	0.79 ± 0.07	0.82 ± 0.04	0.74 ± 0.05	0.74 ± 0.05	0.81 ± 0.08	0.82 ± 0.08	0.83 ± 0.03
	TSS	0.55 ± 0.12	0.6 ± 0.1	0.45 ± 0.08	0.48 ± 0.08	0.59 ± 0.13	0.6 ± 0.1	0.61 ± 0.08
C. lycaon	AUC	0.8 ± 0.04	0.79 ± 0.03	0.8 ± 0.04	0.79 ± 0.03	0.84 ± 0.04	0.81 ± 0.03	0.76 ± 0.03
	TSS	0.6 ± 0.08	0.58 ± 0.07	0.61 ± 0.09	0.58 ± 0.06	0.66 ± 0.08	0.61 ± 0.07	0.58 ± 0.05
C. lycaon x C. lupus	AUC	0.72 ± 0.07	0.73 ± 0.05	0.7 ± 0.04	0.72 ± 0.06	0.74 ± 0.08	0.72 ± 0.08	0.7 ± 0.06
	TSS	0.42 ± 0.09	0.44 ± 0.06	0.39 ± 0.05	0.42 ± 0.09	0.45 ± 0.12	0.44 ± 0.09	0.39 ± 0.08
C. lupus	AUC	0.76 ± 0.04	0.78 ± 0.03	0.72 ± 0.04	0.76 ± 0.04	0.78 ± 0.03	0.77 ± 0.03	0.77 ± 0.02
	TSS	0.47 ± 0.07	0.51 ± 0.04	0.39 ± 0.08	0.47 ± 0.06	0.51 ± 0.04	0.48 ± 0.06	0.48 ± 0.03

Supplemental Table S10. Permutation importance for covariates used to build canid model ensembles. Mean and standard error are indicated. Averages were calculated from component models (N = 54) that were used to build ensembles. Moose and deer habitat variables were species distribution models created in this study.

	Deer Habitat	Moose Habitat	Distance to Human Areas	Snow Precipitation	Snow Depth	Tree Cover
C. latrans	24.4 ± 0.7	3.2 ± 0.3	19.2 ± 0.8	14.4 ± 0.5	3.1 ± 0.3	1.1 ± 0.1
C. lycaon x C. latrans	33.7 ± 1.1	12.7 ± 1	8.4 ± 0.6	20.9 ± 1.1	4.9 ± 0.5	1.5 ± 0.2
C. lycaon	6.8 ± 0.7	35.1 ± 1.2	7.2 ± 0.7	18.9 ± 1.2	1.9 ± 0.2	1 ± 0.2
C. lycaon x C. lupus	10.3 ± 0.7	19.9 ± 0.8	8.4 ± 0.7	16.1 ± 1	5.4 ± 0.5	1.1 ± 0.2
C. lupus	21.2 ± 1	16.7 ± 0.6	16.6 ± 0.7	21.9 ± 0.5	6 ± 0.3	1.8 ± 0.2

Supplemental Table S11. Differential covariate importance across canid groups (Δ permutation importance of AUCtest, t-statistic), compared against C. latrans as baseline, using linear mixed models. Canid group was the explanatory variable and algorithm was a random factor. R^2M refers to variance explained by differences in canid group, while $R^2C – R^2M$ indicates variance attributed to choice of algorithm

	Moose Habitat Δ Imp (t-stat)	Deer Habitat Δ Imp (t-stat)	Distance to Human Areas Δ Imp (t-stat)	Snow Precipitation Δ Imp (t-stat)	Snow Depth Δ Imp (t-stat)	Tree Cover Δ Imp (t-stat)
C. lycaon x C. latrans	0.09 (8.8) ***	0.09 (9.9) ***	-0.11 (-13.9)***	0.06 (6.6)***	0.02 (3.7) ***	0.00 (2.6) **
C. lycaon	0.32 (29.6)***	-0.18 (-18.7)***	-0.12 (-15.4)***	0.04 (4.6)***	-0.01 (-2.7)**	0.00 (-0.5)
C. lycaon x C. lupus	0.17 (15.5)***	-0.14 (-15.0)***	-0.11 (-13.9)***	0.02 (1.8)	0.02 (5.0) ***	0.00 (0.0)
C. lupus	0.14 (12.6)***	-0.03 (-3.4) ***	-0.03 (-3.4) ***	0.07 (7.7)***	0.03 (6.2) ***	0.01 (4.3) ***
R^2M (Species-Group)	0.49	0.43	0.22	0.05	0.08	0.02
$R^2C – R^2M$ (Algorithm)	0.09	0.25	0.33	0.45	0.34	0.44

* *p-value* < 0.05, ** *p-value* < 0.01, *** *p-value* < 0.001

Supplemental Table S12. The effect of incorporating specific covariates into predator SDMs, measured by AUC_{test} with our first baseline approach. AUC_{test} ± SE is indicated. Significance was identified using linear mixed models, with algorithm as a random effect.

	Baseline AUC ± SE	MAT AUC ± SE	Deer Only AUC ± SE	Moose Only AUC ± SE	Moose & MAT AUC ± SE	Both Prey AUC ± SE	R^2M (Treatment)	R^2C-R^2M (Algorithm)
C. latrans	0.72 ± 0.01	0.75 ± 0***	0.74 ± 0**	0.72 ± 0.01	0.74 ± 0*	0.74 ± 0***	0.06	0.46
C. lycaon x C. latrans	0.77 ± 0.01	0.80 ± 0.01*	0.79 ± 0.01*	0.77 ± 0.01	0.80 ± 0.01*	0.79 ± 0.01	0.03	0.25
C. lycaon	0.77 ± 0.01	0.82 ± 0.01***	0.79 ± 0.01**	0.80 ± 0.01***	0.82 ± 0.01***	0.80 ± 0.01**	0.08	0.37
C. lycaon x C. lupus	0.73 ± 0.01	0.77 ± 0.01***	0.73 ± 0.01	0.73 ± 0.01	0.76 ± 0.01**	0.72 ± 0.01	0.09	0.32
C. lupus	0.73 ± 0.01	0.78 ± 0***	0.75 ± 0**	0.75 ± 0***	0.78 ± 0***	0.76 ± 0.01***	0.21	0.28

* *p-value* < 0.05, ** *p-value* < 0.01, *** *p-value* < 0.001

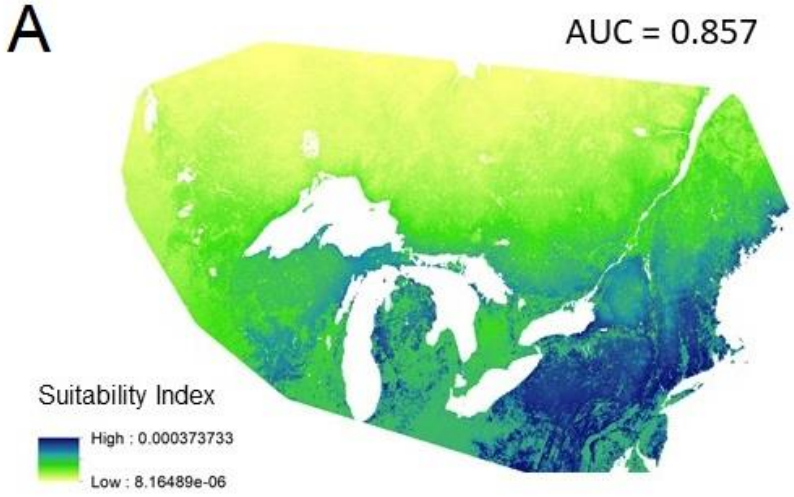
Supplemental Table S13. The effect of incorporating specific covariates into predator SDMs, measured by AUC_{test} with our second permutation approach. 1,008 permutations were evaluated per canid group with all combinations of variables to assess the impacts on model fit. AUC_{test} ± SE is indicated. Significance was identified using linear mixed models, with algorithm as a random effect.

	Base Covariates AUC ± SE	MAT AUC ± SE	Both Prey Habitats AUC ± SE	Deer Habitat AUC ± SE	Moose Habitat AUC ± SE	Moose & MAT AUC ± SE	R^2M (Treatment)	R^2C-R^2M (Algorithm)
C. latrans	0.71 ± 0.00	0.72 ± 0.00	0.71 ± 0.00	0.71 ± 0.00	0.70 ± 0.00***	0.72 ± 0.00*	0.04	0.26
C. lycaon x C. latrans	0.76 ± 0.01	0.78 ± 0.00**	0.78 ± 0.01**	0.76 ± 0.01	0.75 ± 0.01	0.79 ± 0.01***	0.03	0.18
C. lycaon	0.76 ± 0.01	0.77 ± 0.01	0.79 ± 0.01***	0.77 ± 0.01	0.80 ± 0.00***	0.81 ± 0.01***	0.06	0.14
C. lycaon x C. lupus	0.70 ± 0.00	0.72 ± 0.00***	0.71 ± 0.01	0.69 ± 0.00	0.72 ± 0.00**	0.73 ± 0.00***	0.04	0.12
C. lupus	0.72 ± 0.00	0.72 ± 0.00	0.71 ± 0.00*	0.72 ± 0.00	0.71 ± 0**	0.71 ± 0.00	0.01	0.13

* *p-value* < 0.05, ** *p-value* < 0.01, *** *p-value* < 0.001

CHAPTER 2 SUPPLEMENTAL FIGURES

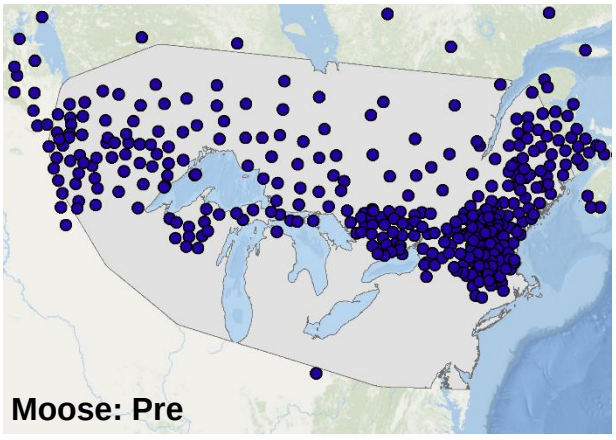
A

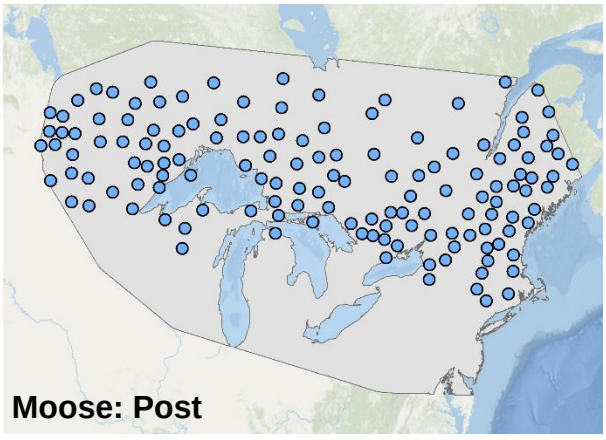


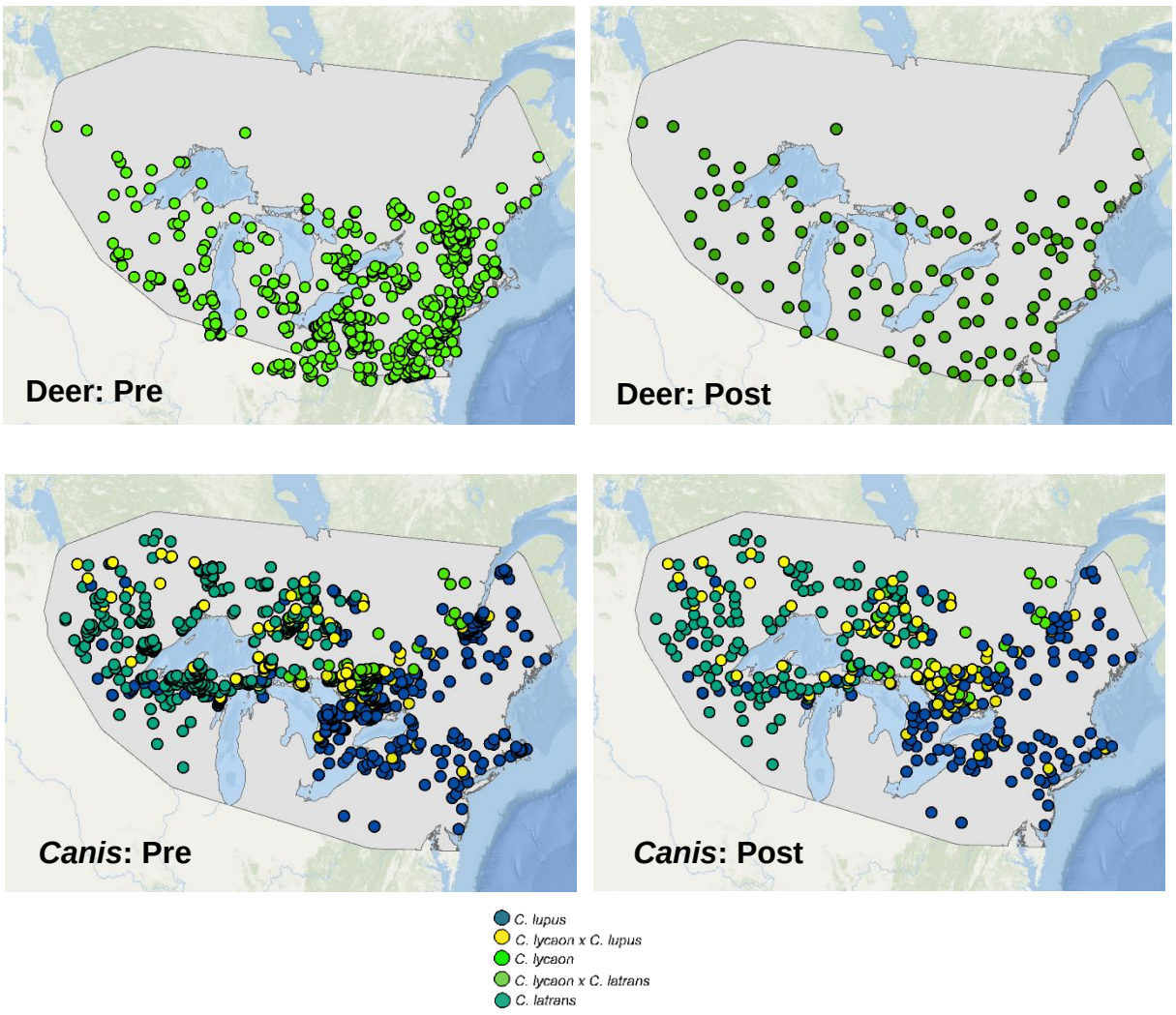
B

Variable	Percent Contribution
Average Winter Temperature	49.40%
Distance to Water Bodies	43%
Tree Cover	7.60%

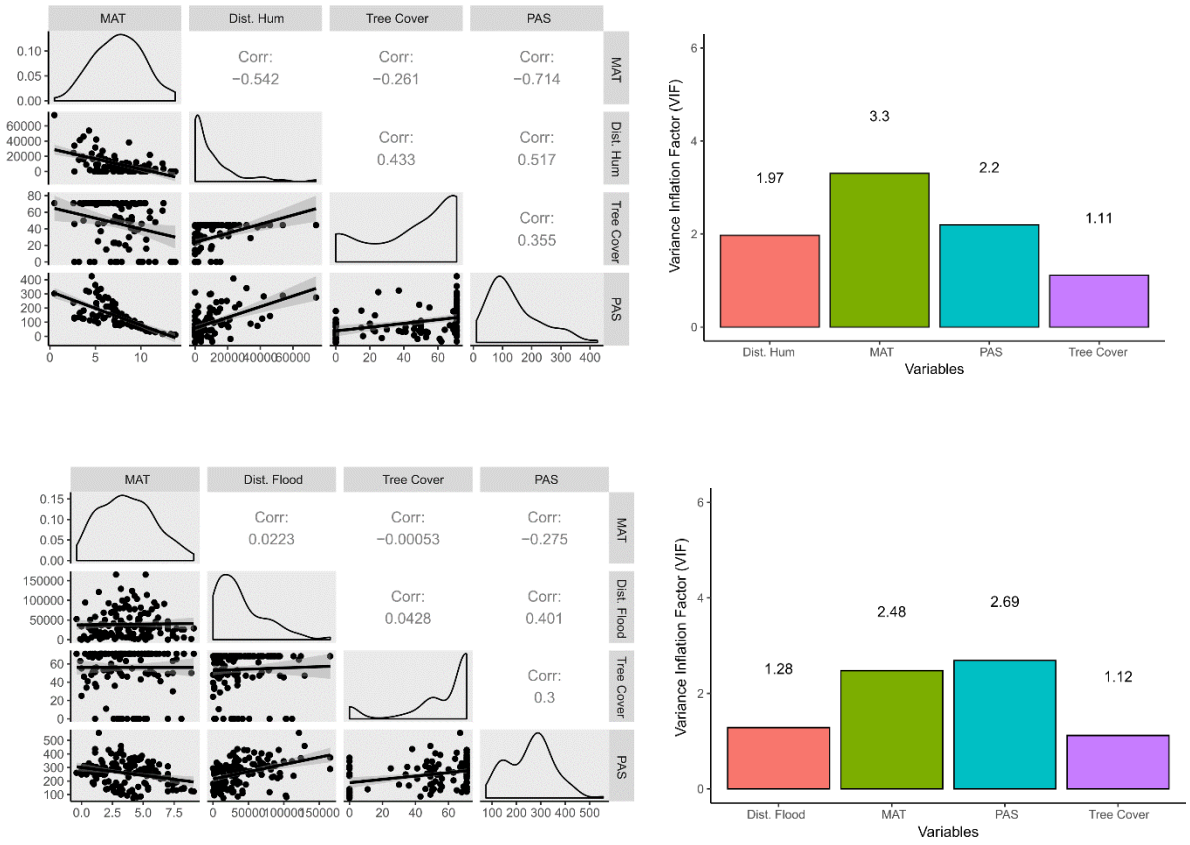
Supplemental Figure S1. Species distribution model for beaver, dropped because of multicollinearity (Pearson's > 0.8) with white-tailed deer model. A) Habitat suitability map for beaver across the study extent, note that this particular map is in raw format, hence the low suitability scores. B) Covariate importance for beaver habitat.



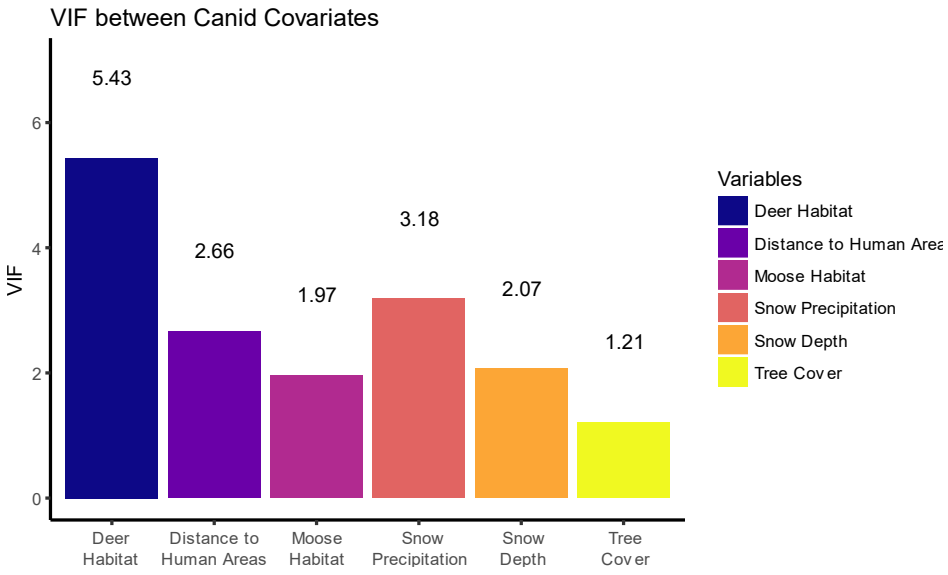




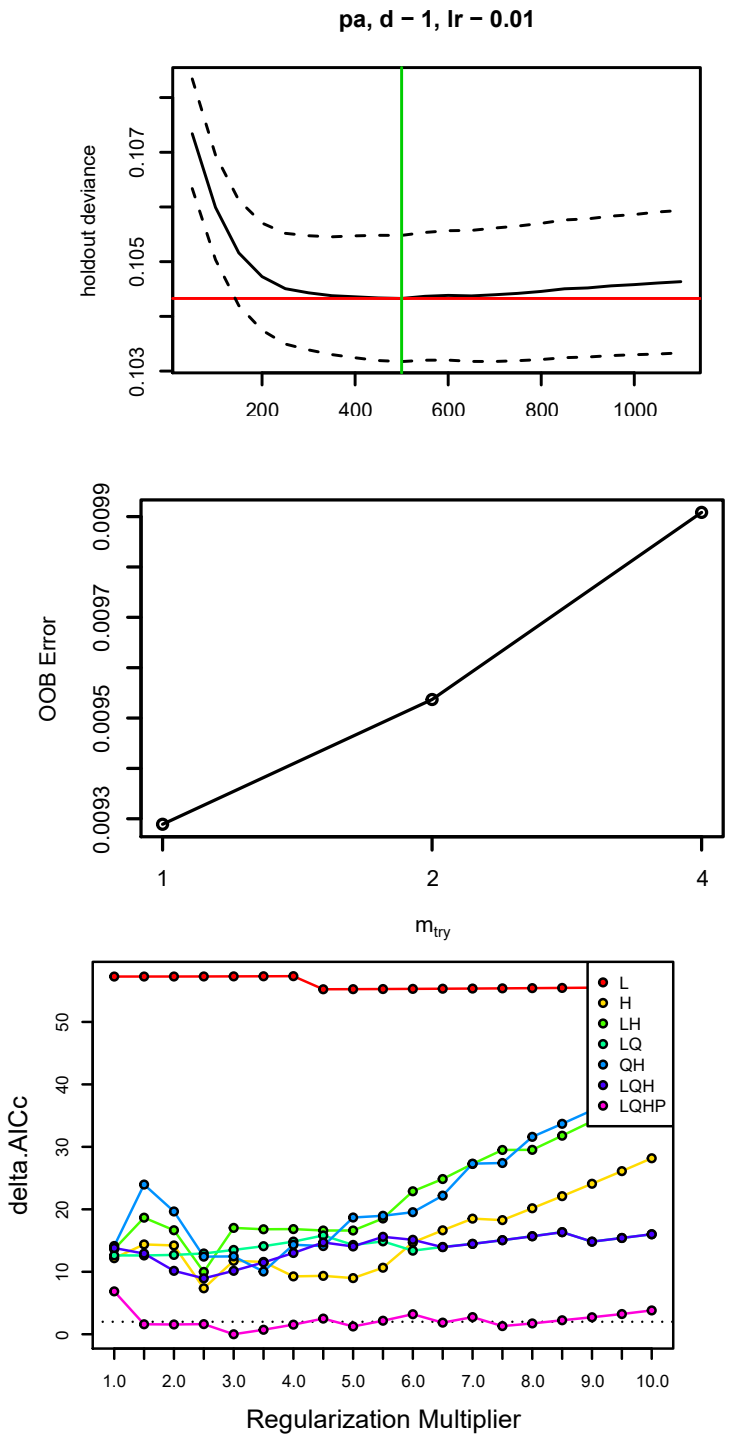
Supplemental Figure S2. Per-species results of spatial thinning, to reduce spatial autocorrelation due to opportunistic sampling bias. Raw spatial data is represented on the left and thinned data on the right. Moose and white-tailed deer occurrence locations were thinned at 50 – 75 km based on environmental heterogeneity using *sdmtoolbox* 2.0, while canid data was thinned at a smaller resolution (5 – 25 km) due to putatively less opportunistic sampling. Prey locations located outside the minimum bounding geometry of the canids were removed from the analysis.



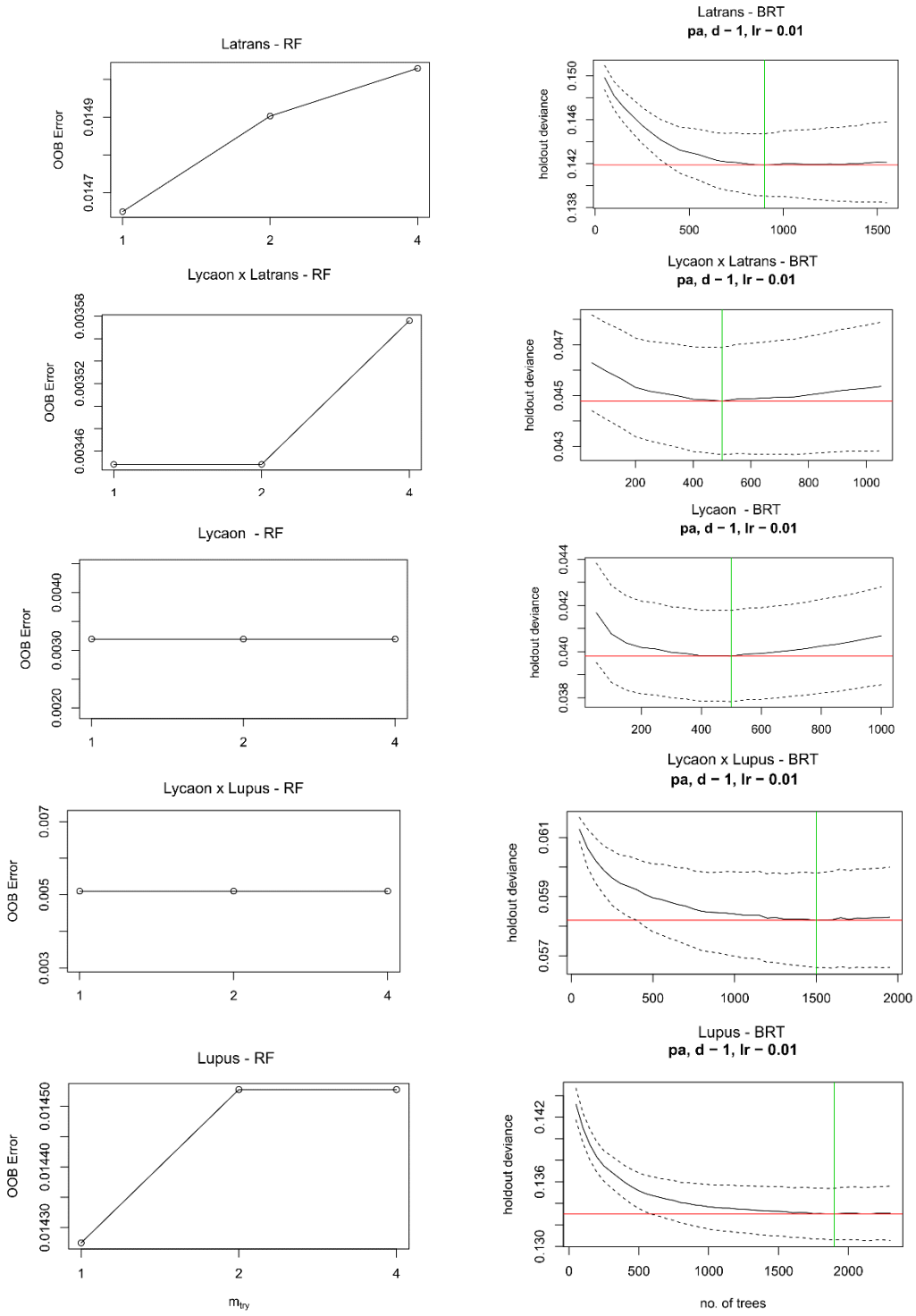
Supplemental Figure S3. Multicollinearity between prey environmental covariates; Pearson's correlations (left), and the variance inflation factor (VIF; right), for white-tailed deer (top) and moose (bottom).



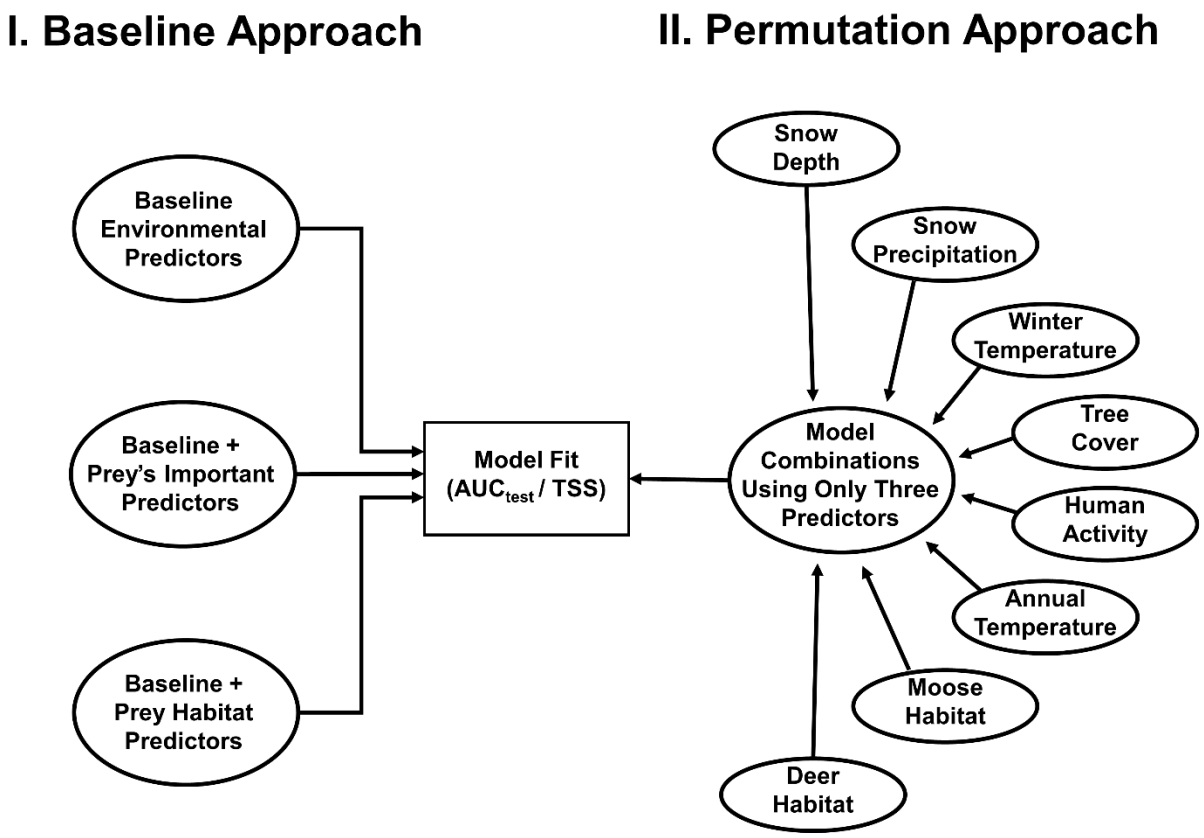
Supplemental Figure S4. Variance inflation factor (VIF) between canid covariates used for ensemble models.



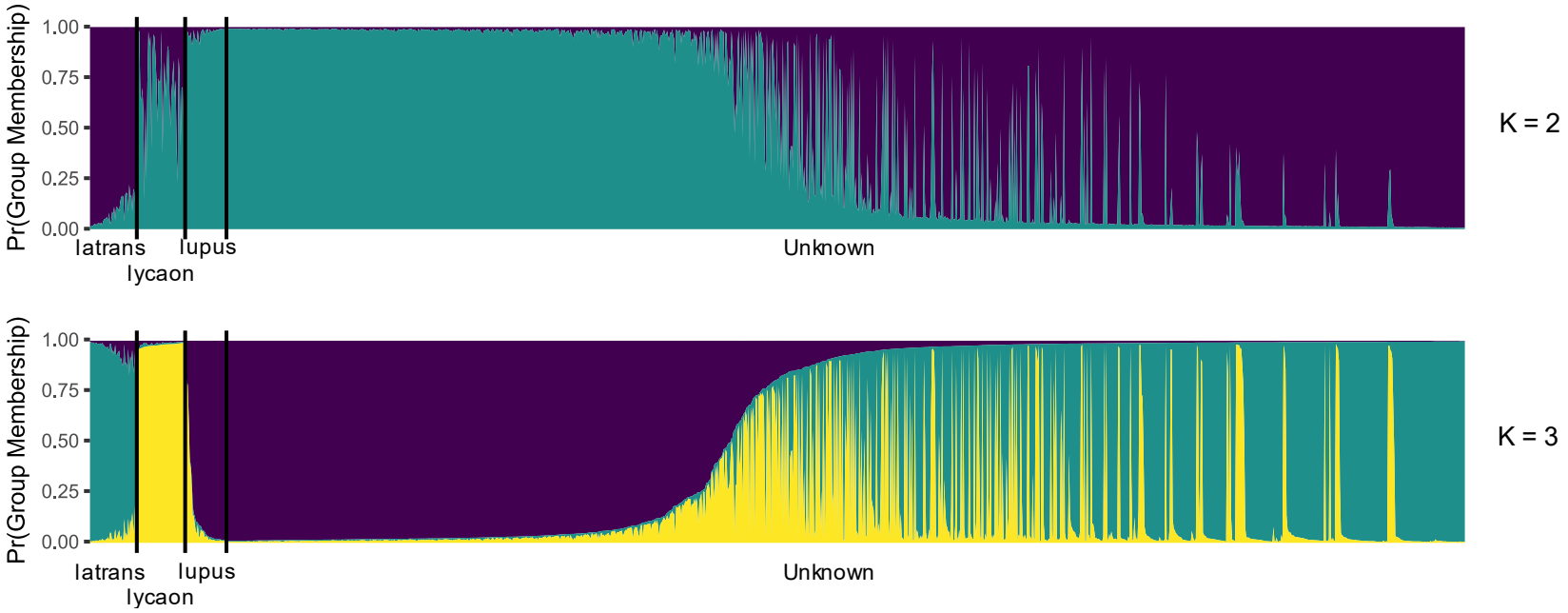
Supplemental Figure S5. Parameter optimization for moose used to identify optimal settings for ensembles. (Top) BRT optimization using GBM.step, (Middle) RF optimization using TuneRF, (Bottom) ENMeval plot for optimizing MaxEnt, showing the regularization and feature settings that minimize ΔAICc. This process was repeated for all species.



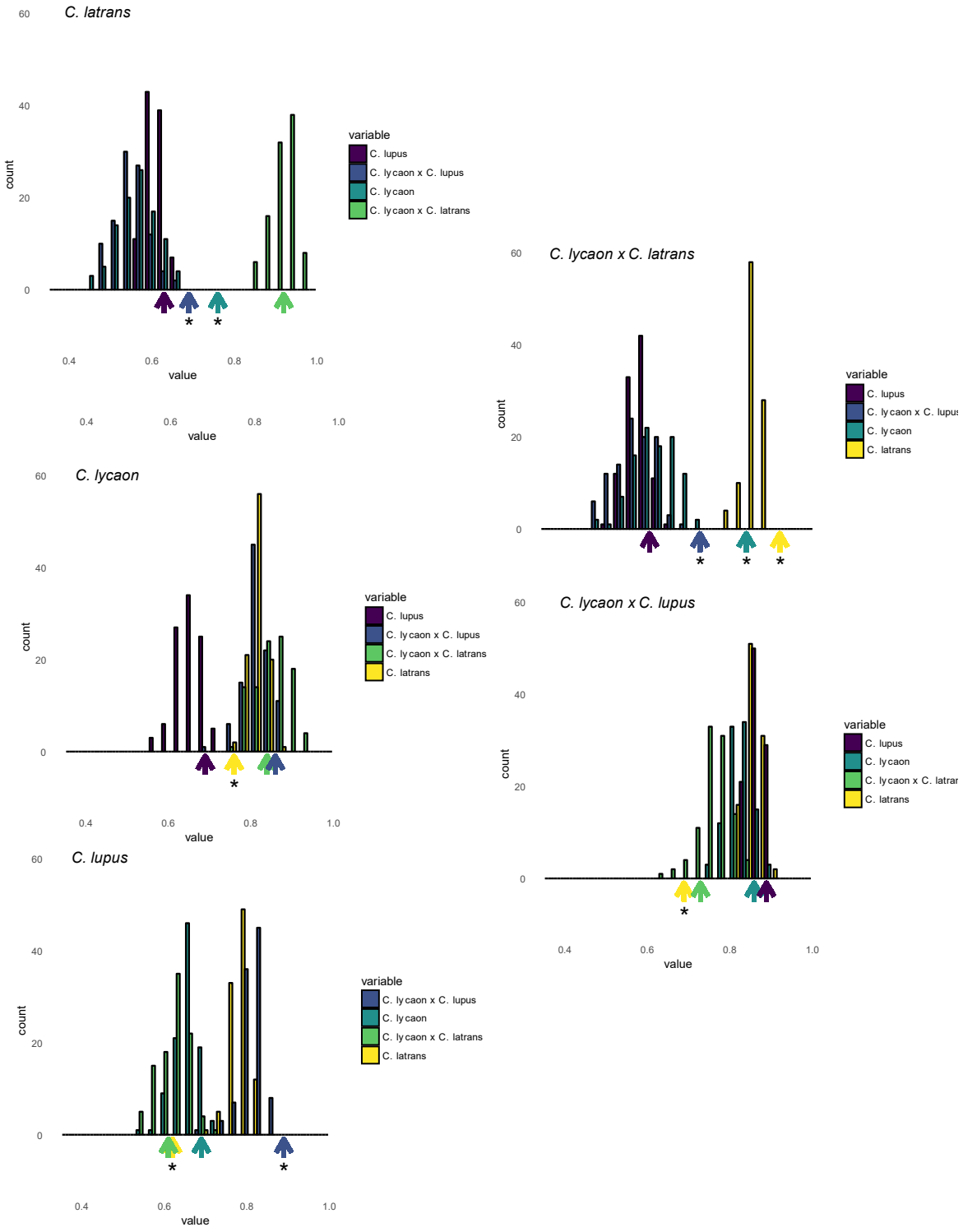
Supplemental Figure S6. Random forest (RF) and boosted regression tree (BRT) optimization plots for all five canid groups. This process was done to optimize the ‘mtry’ and number of trees parameters for final ensemble parameterization.



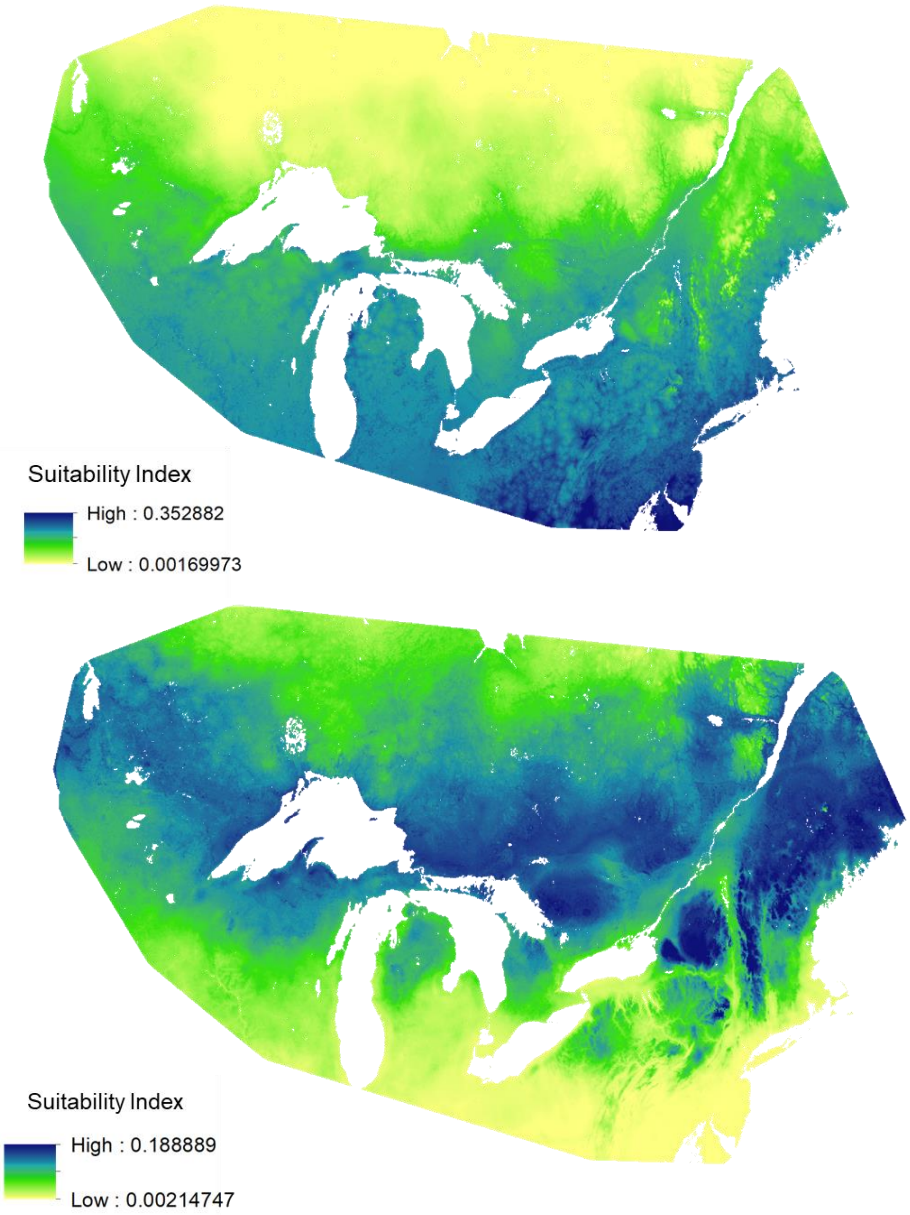
Supplemental Figure S7. Conceptual design of our two-prong approach to assess the impacts of biotic interactions on model fit. The first approach assesses model fit between baseline models that contain only biologically relevant environmental variables, models that include baseline as well as prey habitat variables, and model that include baseline as well as the primary variable predicting prey distributions. Our second approach assesses model fit between model permutations that include specific variables, using 56 different combinations of variables (Supplemental Table S4). Both approaches were analyzed similarly with linear mixed effect models.



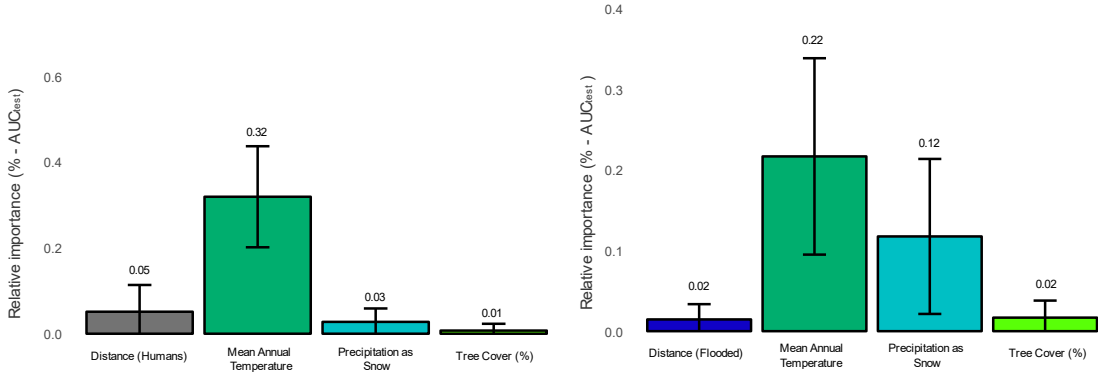
Supplemental Figure S8. Raw structure output for all genotyped canids (n = 1367) at both K = 2 and K = 3. Only previously assigned individuals are labeled on the left side, while unassigned individuals are indicated in the *Unknown* section. Vertical lines indicate individuals, with membership coefficients on the Y-axis. Individuals are placed in the same order across both plots and are directly comparable.



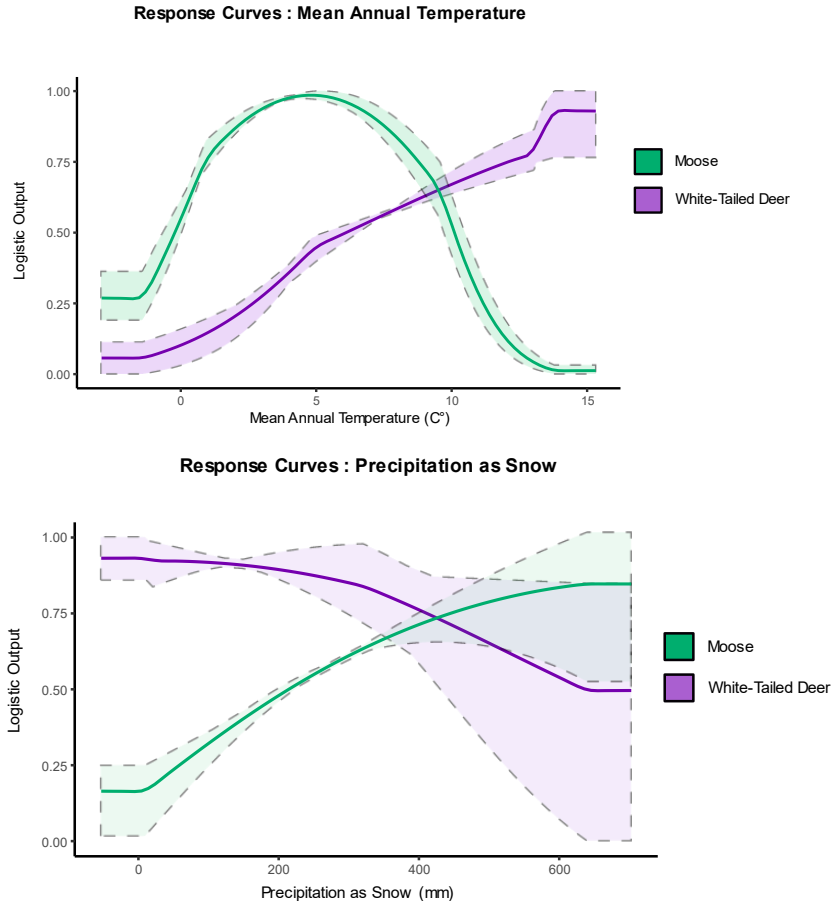
Supplemental Figure S9. Background similarity tests for canid groups, assessed using ENMTools. Histograms represent a null distribution of 100 MaxEnt models generated within the background environment of the focal species, and the arrows indicate actual observed niche overlap. Asterisks denote significance ($p < 0.05$).



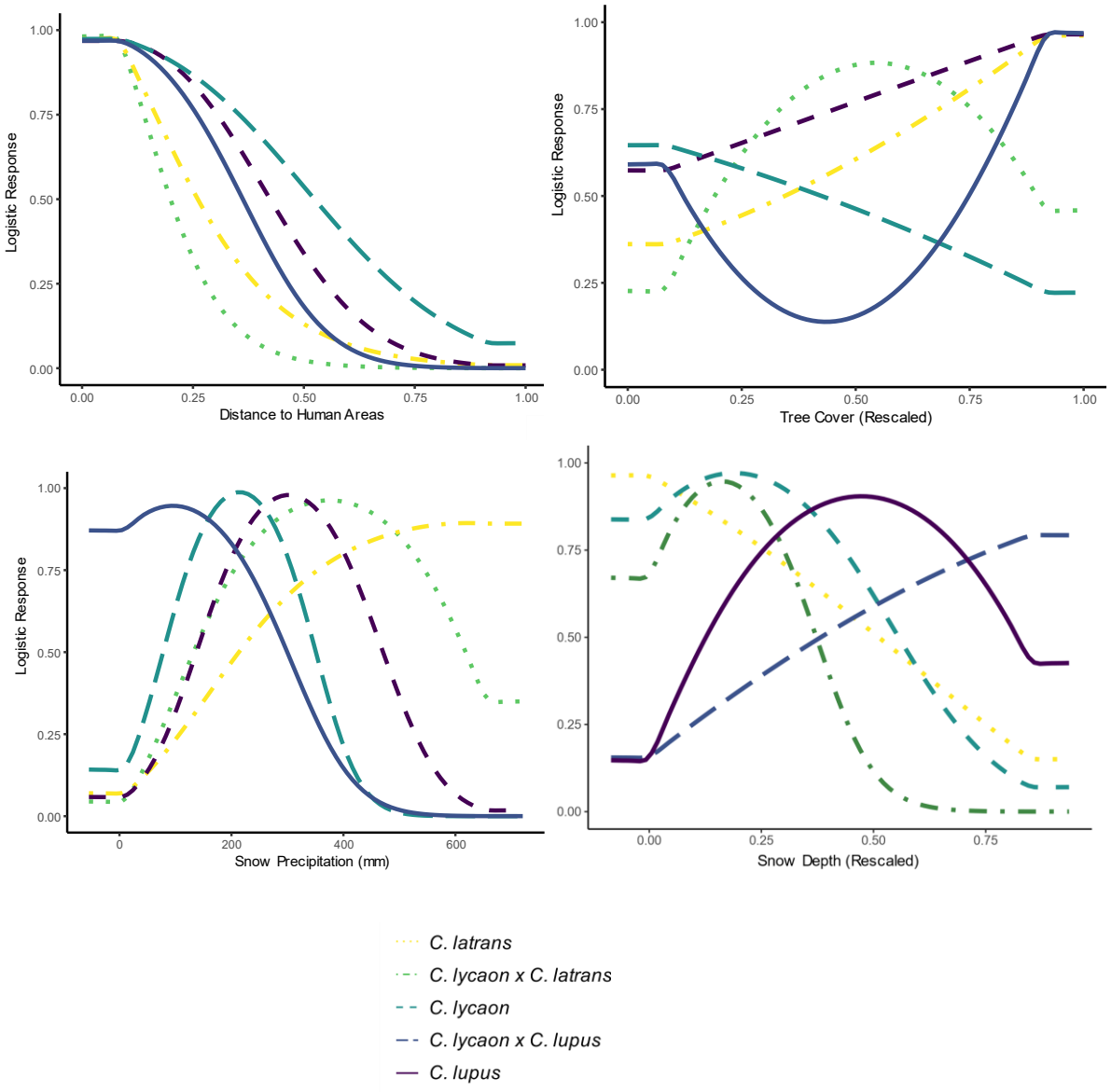
Supplemental Figure S9. Ensemble habitat suitability maps for white-tailed deer (top) and moose (bottom). Ensemble models were created with *sdm* using six different algorithms, optimized using applicable methods. These distribution models were used as covariates in canid distribution models to determine macro-scale associations of predators with their prey.



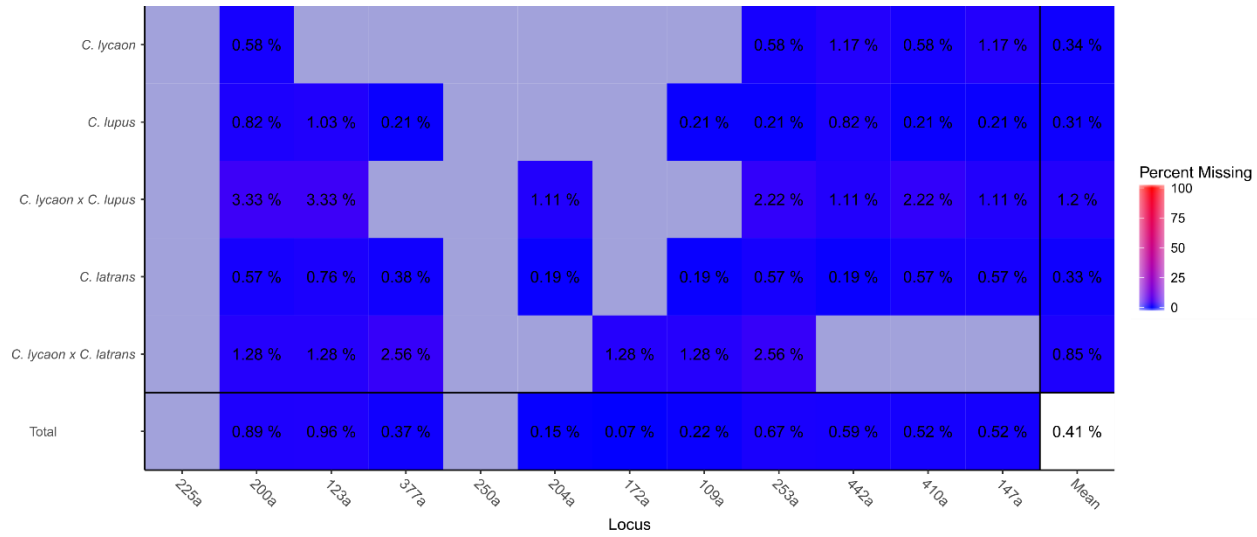
Supplemental Figure S10. Covariate importance for white-tailed deer (left) and moose (right), based on their contribution to AUC_{test}. Averages and standard deviations were calculated across all 54 replicate models.



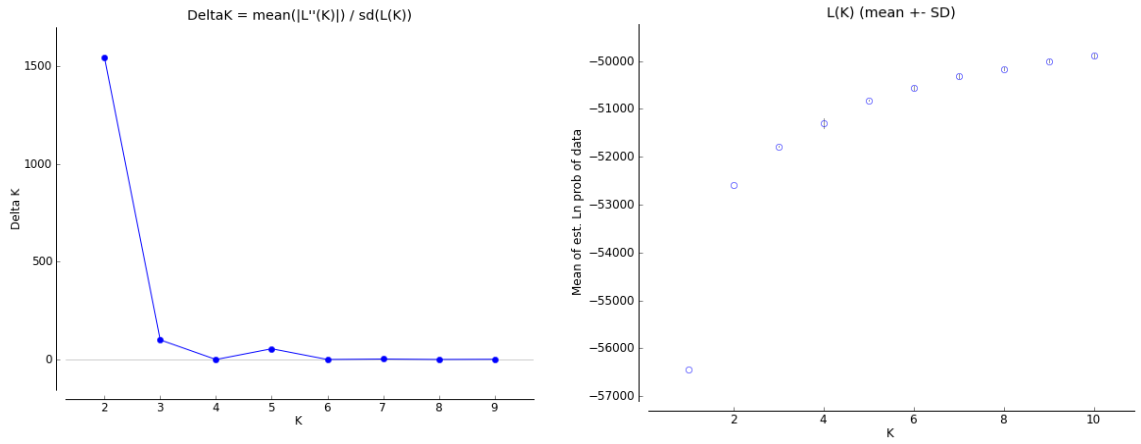
Supplemental Figure S11. Response curves for moose and white-tailed deer for two environmental covariates, depicting the relative importance of each covariate at different levels for predicting suitable habitat for prey. Solid line indicates the mean and the dashed line indicate standard deviation from 10 runs.



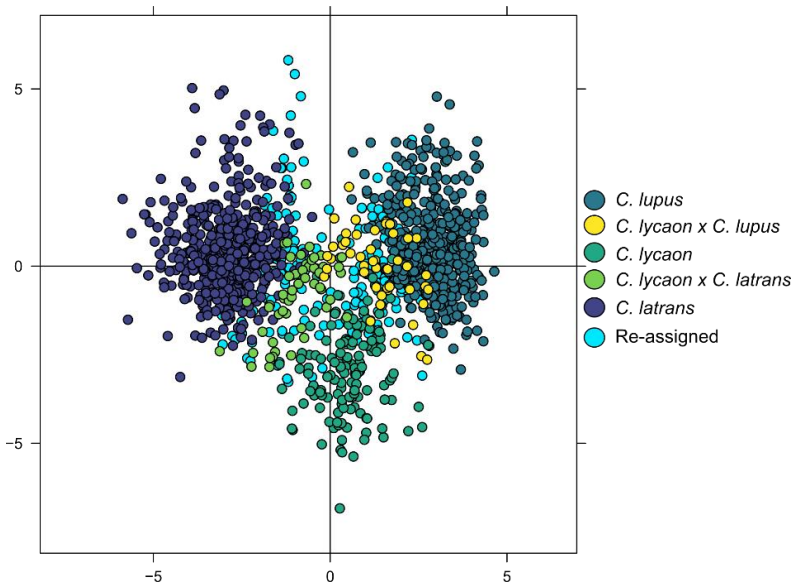
Supplemental Figure S13. Response curves for the four additional environmental variables not shown in the main text. The logistic response of each variable per canid group is shown, with the variable denoted on the X axis. Note that some variables have been rescaled from 0 to 1 for this figure.



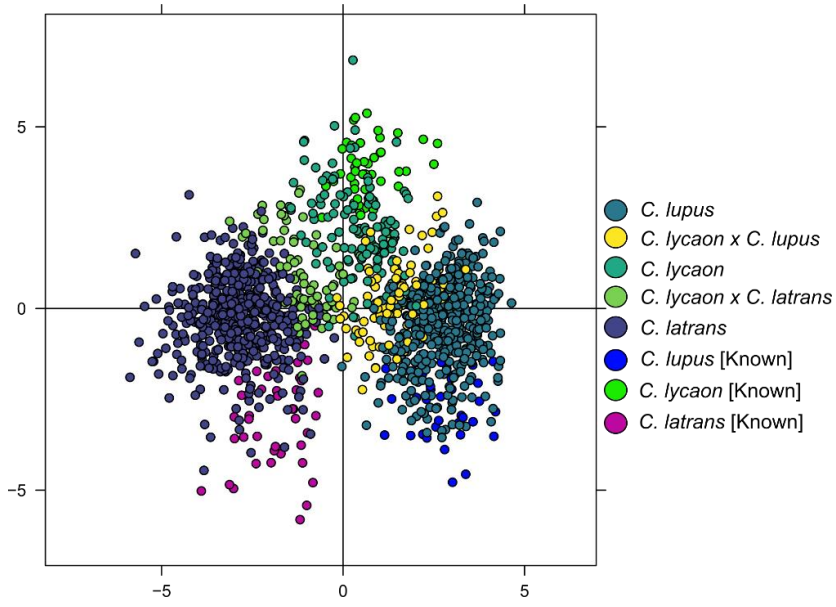
Supplemental Figure S14. Missing data per locus across all canid groups. Lavender coloured boxes with no numerical coefficients indicate no missing data. *Locus* refers to each individual diploid microsatellite.



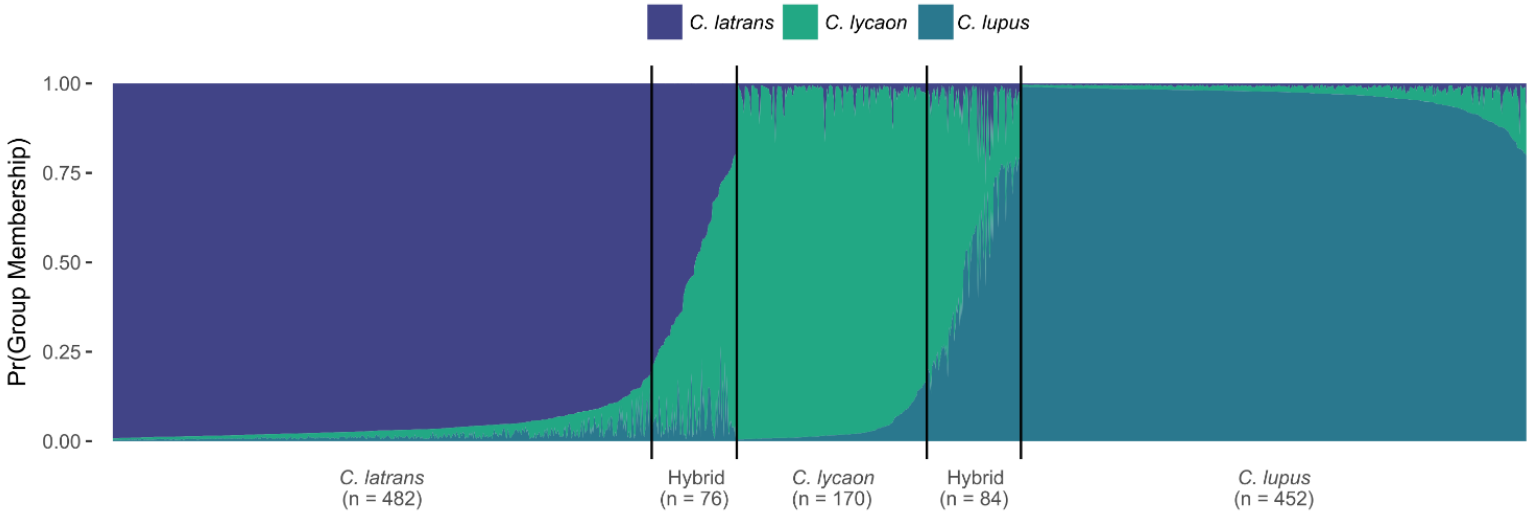
Supplemental Figure S15. Identifying optimal genetic clusters using *structureharvester*. The largest ΔK (left), occurred at K = 2, putatively reflecting the new world (*C. latrans, C. lycaon)* and old world (*C. lupus*) origin of these groups. We designated K = 3 for our final analyses, as this delineation captured the substructuring between *C. lycaon* and *C. latrans* as evidenced through our individuals of known ancestry.



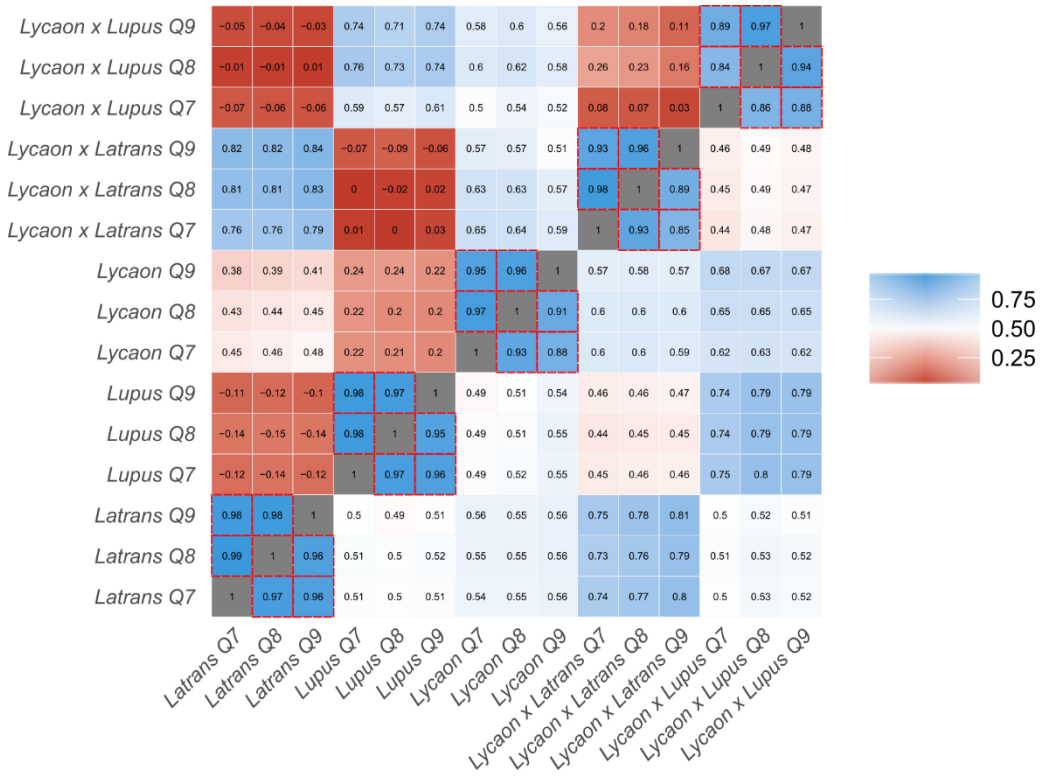
Supplemental Figure S16. PCA plot created using *adegenet* that depicts all 1350 genotyped canid samples. Individuals in the *Re-assigned* category were reassigned to a hybrid group depending on the *Q*-value chosen. The implications of these individuals with variable assignment (n = 199) were assessed with a sensitivity analysis that quantified niche overlap with *Q*-value designations at 0.7, 0.8, and 0.9 with no changes in overall conclusions



Supplemental Figure S17. PCA plot created using *adegenet,* showing the canid group designations at *Q*-value = 0.8, with the individuals of known ancestry (n = 136) demarcated as *Known*.



Supplemental Figure S18. Genetic results from *structure* for all genotyped individuals (n = 1367) at K = 2, visualized using R. Individuals are represented by a single vertical line with their assigned canid group delineated by colour.



Supplemental Figure S19. Sensitivity analysis depicting niche overlap (Schoener's D; bottom/right) and Pearson's correlations (upper/left) calculated from ensemble distribution models for varied *Q*-value thresholds (0.7, 0.8, 0.9). Red dashed boxes highlight comparisons within canid groups between *Q*-values.

APPENDIX B: CHAPTER 3 SUPPORTING INFORMATION

RRBS Library Preparation

Epidermal tissue samples were collected from Canada lynx (*Lynx canadensis*) pelts from either North American Fur Auctions (Toronto, Ontario) or Fur Harvester's Auctions, Inc. (North Bay, Ontario). Pelts were only dried, and not subjected to chemical processing that may impact molecular analyses. Small subsamples were taken consistently from the left side of the back leg on each pelt and were stored in individually barcoded envelopes until DNA extraction. Approximately 20 mg of pelt was removed from this subsample and suspended in 200 µl of lysis buffer and 20 µl of proteinase K to break down proteins and cell walls in preparation for DNA extraction. DNA was then isolated using magnetic beads with a MagneSil Blood Genomic Max Yield System (Promega) on an automated Janus robot (Perkin-Elmer). DNA concentrations were quantified using a Quant-It Picogreen Assay and were subsequently standardized to 20 ng/µl.

Full protocol and additional details can be found in van Gurp *et al.,* (2016) (van Gurp et al., 2016). A total of 95 Canada lynx samples and 1 sample of non-methylated lambda phage DNA (Sigma-Aldrich: D3654) were digested overnight (17 hours) at 37°C in individual 40 µl reactions containing 4 µl NEB 3.1 buffer, 1 µl AseI (10 units/µl; NEB: R0526S), 2 µl NsiI (20 units/µl; NEB: R0127L), 12 µl of UltraPure distilled water (Invitrogen), and 20 µl of DNA (400 ng total). We added individually barcoded in-line methylated adapters (Supplemental Table S1) to each digest product in a reaction containing 6 µl T4 DNA Ligase buffer (10x), 2 µl T4 DNA ligase (2,000,000 units/mL; NEB: M0202M), 4 µl AseI adapter (600 pg/µl), 4

µl NsiI adapter (600 pg/µl), and 8 µl of UltraPure H_2O. This 60 µl reaction was placed in a thermal cycler for three hours at 22°C, and then left on ice overnight at 4°C, with no heat inactivation.

Individually-barcoded samples were then amalgamated into 8 pools with 12 samples in each pool using a QIAQuick PCR Purification Kit (Qiagen: 28106), with 10 µl of 3M sodium acetate (NaOAC) added to each pool to neutralize pH. Pools were eluted in 50 µl and were then size-selected using AMPure XP magnetic beads with a ratio of 0.8x and freshly prepared 80% ethanol (Beckman Coulter: A63881). Residual ethanol was dried by resting samples on a heat block at 56°C until a hairline fracture was seen in the beads. Pools were then eluted in 23 µl UltraPure H_2O. Nicks in the adapters were then repaired in a 25 µl reaction with 19.25 µl of size-selected library, 2.5 µ 5-mC-dNTP mix (10 mM; Zymo: D1030), 2.5 µl NEB buffer 3.0 (10x), and 0.75 µl DNA polymerase I (10 units/µl; NEB: M0209L). This reaction was carried out on a thermal cycler for 60 minutes at 15°C.

Pools were subjected to bisulfite conversion using an EZ DNA Methylation-Lightning Kit (Zymo: D5030T) in a reaction with 23 µl size-selected and repaired library and 149.5 µl conversion reagent in a thermal cycler with an initial step at 98°C for 8 minutes followed by 60 minutes at 54°C. The remaining steps were performed according to manufacturer's instructions with a 20-minute desulphonation time and a final elution in 12 µl. Bisulfite-converted DNA was amplified in triplicate PCRs with 2 µl of bisulfite-converted library, 5 µl KAPA HiFi Uracil+ (Roche: KK2802), 0.3 µl of a forward Illumina primer, 0.3 µl reverse Illumina

primer (Supplemental Table S2), and 2.4 µl UltraPure H_2O. Cycling conditions can be seen in Supplemental Table S3.

Triplicate PCRs were combined using a QIAQuick PCR Purification Kit with 10 µl 3M NaOAC and a final elution in 50 µl. We quantified the concentration of pools with a Qubit 3.0 (ThermoFisher Scientific) and respective quantities were amalgamated into a superpool with equal contribution from each pool. The superpool was cleaned with AMPure XP magnetic beads with a 0.8x ratio and was eluted in 24 µl. Final library fragment distribution was quantified with on an Agilent Bioanalyzer 2100 and successful flowcell ligation was confirmed with qPCR. 125-bp paired-end sequencing was then performed on a single lane on Illumina HiSeq2500 at the Hospital for Sick Children (Toronto, Ontario, Canada).

Genetic Population Structure

SNPs were identified using CGmapTools (Guo et al., 2017), and vcf files were converted using PGDSpider v2.1.1.3(Lischer & Excoffier, 2012) for use in BayeScan v2.0 (Foll & Gaggiotti, 2008). We ran BayeScan on this SNP matrix with default settings (50,000 burn-in, 20 pilot runs with a length of 5,000 chains, and 5,000 outputted iterations) four times, with prior odds for the neutral model modified between runs to 1, 10, 100, and 10,000. We then assessed relative population-level F_{ST} using the .sel output, which is visualized along the outer bar plots in Supplemental Figure S2. Pair-wise FST was calculated using StaMPP v 1.5.1 (Pembleton et al., 2013). Outlier loci were assessed with BayeScan v2.0, but no loci were identified.

Methylation Calling and Bisulfite Conversion Efficiency

Methylated sites in a CpG context were called using the *Bismark* methylation extractor function after removing any reads that were likely candidates for incomplete bisulfite conversion (> 3 methylated sites in a CHH or CHG context) (Krueger & Andrews, 2011). This step was deemed necessary as our non-methylated lambda phage DNA control identified minor incomplete bisulfite conversion. These results reinforce our decision to pool samples prior to bisulfite conversion, so that any reaction inconsistencies will apply to all samples and universalize any biases. Furthermore, we strongly recommend that all bisulfite sequencing experiments employ the use of a non-methylated lambda phage DNA as we demonstrate that incomplete bisulfite conversion does occur despite rigorous adherence to manufacturer's protocols.

We also assessed relative levels of methylation directly around all transcripts using *Seqmonk*. We discovered a reduction in methylation < 5,000-bp upstream of transcripts, with average levels resuming outside of the putative promoter region, consistent with results around the TSS in other research (Laine et al., 2016).

Sensitivity Analyses

Missing Data. First, we assessed the impacts of missing data by calculating Euclidean dissimilarity matrices, performing principal coordinates analyses (PCoAs), and distance-based redundancy analyses (db-RDAs) for all datasets at three different levels of completeness. Filtering was done by removing individuals with the most amount of missing data, while keeping equal representation for all

populations. We evaluated the impacts of completeness at three sample levels (N = 95; N = 40; N = 24). These data were summarized with distance matrices using the function *daisy* (Maechler et al., 2018) and visualized with a PCoA using the function *dudi.pco* (Jombart, 2008) (Fig. S1A).

We then exported the axes of these PCoAs that explained > 30% of the total variation and performed db-RDAs on these axes as a response variable to determine if this variation in missing data would modify our inferences. We found a linear relationship between the amount of the missing data and the total explanatory power of the model (adj. R^2), with coefficients increasing as the level of missing data decreases (Fig. S1B).

Temporal variation. We assessed temporal variation in methylation patterns by analyzing a subset of the data where one population was sampled across two years. Due to sample access, our Canada lynx samples were collected from 2008-2012 (Table S1). We performed a db-RDA using the axes summarizing methylation patterns from the Alaskan population (n = 23), which has individuals from both 2009 and 2010, with year as an independent explanatory variable. db-RDAs for methylation over CpG islands and unannotated regions found that year was not significantly correlated to methylation patterns within this population (Fig. S1C).

Axis retention for db-RDAs. We examined the ramifications of our threshold for axis retention for db-RDA response variables by performing db-RDAs with variable amounts of PCoA axes. We examined the effect on adj. R^2 when using axes that explained 30%, 50%, 75%, and 95% of cumulative variation as response variables

for all three datasets (SNPs, CpG island and gene body methylation, unannotated methylation). We found a linear relationship between adj. R^2 and cumulative variation explained (Fig. S1D), consistent with the notion that including axes which explain less individual variation will decrease overall model fit, likely owing to explaining noise in the data.

Landscape Genomics and Methylomics

We assessed the relationship between biogeographical variables and our molecular data by performing db-RDAs with PCoA axes that explain > 30% of the cumulative variation as response variables. Our explanatory variables included a binary variable with *Newfoundland* and *Mainland* as predictors, to represent the likely impermeable aquatic barrier between the island of Newfoundland and the mainland. Our second variable was geographic distance, which served as a null model as we would expect distribution in allele frequencies to correlate with distance across the landscape. To summarize this variable in a single vector and eliminate collinearity between latitude and longitude, we performed a PCoA on latitude and longitude. The first axis of the PCoA explained overall variation exceptionally [PCo1 = 99.73% of the variation]. To summarize as much variation as possible in climatic patterns into a single variable, we performed a PCA on bioclimatic variables using *sdmtoolbox*(Brown et al., 2017). We chose three variables with suspected biological significance for Canada lynx (*L. canadensis)*, particularly emphasizing variables that correlate with winter conditions. We summarized three bioclimatic variables from WorldClim(Fick & Hijmans, 2017) including the annual temperature range (BIO7), minimum temperature of the

coldest month (BIO6), and precipitation of the coldest quarter (BIO19). The first axis of this PCA described a majority of the variation [PC1 = 85.62% of the variation]. We also included a continuous variable of tree cover (DeFries et al., 2000), and a randomly-generated vector to use as a null variable to mimic noise. Environmental data was extracted from each georeferenced lynx sample using R v3.4.2 (R Core Team, 2017).

We then performed step-wise model selection using the PCoA axes of molecular data as response variables against these biogeographical explanatory variables. Step-wise selection was performed in *vegan* (Dixon, 2003) using the function *ordistep,* performing forward and reverse step selection with 100 steps and 10,000 permutations. Variables were included when their p-value fell below a specified threshold at 0.05 and were removed from the model when its p-value rose above 0.2. All db-RDA results, including partial db-RDAs, can be found within Supplemental Table S4 and Fig. S2. Collinearity between explanatory variables was assessed by calculating the variance inflation factor (VIF; Table S5).

CHAPTER 3 SUPPLEMENTAL TABLES

Supplemental Table S2. Total reads and cytosines identified for the non-methylated lambda phage DNA control. Conversion rate was calculated by dividing the methylated cytosines by the total number of cytosines analyzed.

Total Reads	Methylated Cytosines	Non-Methylated Cytosines	Conversion Rate
4,260,069	16,374,124	173,911,904	0.91395

Supplemental Table S3. Total cytosines retained for analyses after filtering for shared positions between 95 individuals. Unique positions are listed first, with total cytosines shared between individuals in parentheses.

Methylation Dataset	Total Cytosines Analyzed (After Filtering)	Total 5KB Windows (After Filtering)	Average Coverage
Unannotated Regions	5,031 (67,279)	376	43.00 ±16.89%
CpG Islands / Gene Bodies	4,611 (58,305)	329	38.23 ± 23.92%

Supplemental Table S4 db-RDA results, with molecular datasets as response variables and environmental variables as explanatory variables. The full model indicates the model with all covariates, prior to step-selection. Final model is the model identified via simultaneous forward and backward step selection. p-db-RDAs were done to identify the independent explanatory power of each variable, after removing the variation explained by the other variables. *F* is a pseudo-*F* test statistic as described within *vegan* (Dixon, 2003). RDA1 and RDA2 describe the variation explained by each axis of the db-RDA, respectively. The values for the five biogeographical variables indicate p-values.

Dataset	Model	*F*	ANOVA	R^2	adj. R^2	RDA1%	RDA2%	RDA1 (*F*)	RDA2 (*F*)	Insular	Distance	Climate	Tree Cover	Null
SNPs	Full	15.983	0.001	0.473	0.444					0.000	0.306	0.000	0.953	0.328
	p-db-RDA: Climate	13.707	0.001	0.081	0.076						1.177			
	p-db-RDA: Insularity	26.326	0.001	0.155	0.151					26.326		51.125		
	Final Model	38.726	0.001	0.457	0.445	0.380	0.077	64.413	13.038	0.001		0.001		
Unannotated Methylation	Full	10.196	0.001	0.364	0.328					0.000	0.000	0.000	0.131	0.633
	p-db-RDA: Distance	15.722	0.001		0.108									
	p-db-RDA: Climate	6.748	0.001		0.042									
	p-db-RDA: Insularity	10.026	0.001		0.066					10.026	26.284	12.496		
	Final Model	16.269	0.001	0.349	0.328	0.208	0.097	29.136	13.520	0.001	0.001	0.001		
CpG Islands & Gene Body Methylation	Full	9.980	0.001	0.359	0.323					0.000	0.000	0.000	0.734	0.739
	p-db-RDA: Distance	17.465	0.001		0.120									
	p-db-RDA: Climate	6.899	0.001		0.043									
	p-db-RDA: Insularity	13.662	0.001		0.093					13.662	22.245	12.863		
	Final Model	16.257	0.001	0.349	0.327	0.171	0.131	23.878	18.330	0.001	0.001	0.001		

Supplemental Table S5. Collinearity between explanatory variables.

Variance Inflation Factor		
Distance	Insularity	Climate
2.0932	2.747	3.7715

Supplemental Table S6. Oligonucleotides used for massively multiplexed sequencing, where X denotes a methylated cytosines (5mC), and N denotes a degenerated base. Adapters were ligated onto DNA that was digested with AseI and NsiI restriction enzymes. The bottom, colour-coded examples indicate the degenerated bases in red, the barcode sequence in blue, the methylation control nucleotide in orange, and the restriction enzyme cut-site in green. Further details can be found in Van Gurp *et al.,* 2016.

Barcode	AseI (Top)	AseI (Bottom)
AACT	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNaaxtC-3'	5'-TAGagttNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
CCTA	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNxxtaC-3'	5'-TAGtaggNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
TTAC	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNttaxC-3'	5'-TAGgtaaNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
AGGC	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNaggxC-3'	5'-TAGgcctNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
GAAGA	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNgaagaC-3'	5'-TAGtcttcNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
CCTTC	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNxxttxC-3'	5'-TAGgaaggNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
TTCAA	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNttxaaC-3'	5'-TAGttgaaNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
GCGGC	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNgxggxC-3'	5'-TAGgccgcNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
AGATGC	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNagatgxC-3'	5'-TAGgcatctNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
CATAGC	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNxatagxC-3'	5'-TAGgctatgNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
TTCGAC	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNttxgaxC-3'	5'-TAGgtcgaaNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'
ATGCGC	5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNatgxgxC-3'	5'-TAGgcgcatNNNAGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'

	NsiI (Top)	NsiI (Bottom)
AACT	5'-GagttNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	5'-XTXGGXATTXXTGXTGAAXXGXTXTTXXGATXTNNNaaxtCTGXA-3'
CCAG	5'-GctggNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	5'-XTXGGXATTXXTGXTGAAXXGXTXTTXXGATXTNNNxxagCTGXA-3'
TTGA	5'-GtcaaNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	5'-XTXGGXATTXXTGXTGAAXXGXTXTTXXGATXTNNNttgaCTGXA-3'
GGTC	5'-GgaccNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	5'-XTXGGXATTXXTGXTGAAXXGXTXTTXXGATXTNNNggtxCTGXA-3'
ACTA	5'-GtagtNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	5'-XTXGGXATTXXTGXTGAAXXGXTXTTXXGATXTNNNaxtaCTGXA-3'
CAGC	5'-GgctgNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	5'-XTXGGXATTXXTGXTGAAXXGXTXTTXXGATXTNNNxagxCTGXA-3'
TGAT	5'-GatcaNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	5'-XTXGGXATTXXTGXTGAAXXGXTXTTXXGATXTNNNtgatCTGXA-3'
GTCG	5'-GcgacNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'	5'-XTXGGXATTXXTGXTGAAXXGXTXTTXXGATXTNNNgtxgCTGXA-3'

AseI

5'-AXAXTXTTTXXXTAXAXGAXGXTXTTXXGATXTNNNBBBBC-3'

3'-TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGANNNBBBBGAT-5'

NsiI

5'- GBBBBNNNAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3'

3'-AXGTCBBBBNNNTXTAGXXTTXTXGXXAAGTXGTXXTTAXGGXTX-5'

Supplemental Table S7. Illumina primers used during PCR. The forward primer is the Illumina PE PCR Primer 1.0, while the reverse is the PE PCR Primer 2.0.

Illumina Primers	
Forward Primer	5′- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′
Reverse Primer	5′- TCTAGCCTTCTCGCCAAGTCGTCCTTACGGCTCTGGCTAGAGCATACGGCAGAAGACGAAC-3′

Supplemental Table S8. Pair-wise F_{ST}, calculated with StaMPP using 489 putatively neutral SNPs. QC indicates Quebec, MB: Manitoba, NL: Newfoundland, AK: Alaska.

	QC	MB	NL	AK
QC	0.000	-	-	-
MB	0.051	0.000	-	-
NL	0.104	0.122	0.000	-
AK	0.050	0.057	0.128	0.000

Supplemental Table S9. Total cytosines retained for analyses after filtering for shared positions between 95 individuals. Unique positions are listed first, with total cytosines shared between individuals in parentheses.

Methylation Dataset	Total Cytosines Analyzed (After Filtering)	Total 5KB Windows (After Filtering)	Average Coverage
Unannotated Regions	5,031 (67,279)	376	43.00 ±16.89%
CpG Islands / Gene Bodies	4,611 (58,305)	329	38.23 ± 23.92%

Supplemental Table S10. Results of a sensitivity analysis examining the effects of missing data on explanatory power of the models. Effect size was measured with adj. R^2, and missing data is written as a percentage.

Individuals	Dataset	Missingness	adj. R^2
95	SNP	0.420	0.445
40	SNP	0.250	0.631
24	SNP	0.202	0.889
95	CpGI	0.451	0.308
40	CpGI	0.233	0.331
24	CpGI	0.166	0.377
95	Unannotated	0.449	0.328
40	Unannotated	0.238	0.359
24	Unannotated	0.171	0.442

Supplemental Table S11. The relationship between year sampled and methylation patterns, from our subset of the Alaskan population (n = 23). *F* refers to the pseudo-***F*** generated using db-RDAs within *vegan* against environmental variables.

Dataset	P-Value (Year)	*F* (Year)
CpG Islands / Genes	0.641	0.849
Unannotated	0.306	1.174

Supplemental Table 12. Relationship between adj. R^2 and the number of axes retained for db-RDAs. *Axes Retained* refers to the number of PCoA axes retained as response variables for the db-RDA, which explained amount of variation referred to by the *Cumulative Variation Explained.* The *Distance* covariate was excluded from the p-db-RDAs with SNP data because it was non-significant, as identified with step-wise model selection.

Cumulative Variation Explained	Axes Retained	Dataset	adj. R^2 (Overall)	adj. R^2 (Distance)	adj. R^2 (Climate)	adj. R^2 (Insularity)
0.328	3	SNP	0.445	NA	0.076	0.151
0.308	10	Unannotated	0.328	0.108	0.042	0.066
0.321	10	CpGI / Gene Body	0.308	0.112	0.042	0.089
0.525	7	SNP	0.284	NA	0.056	0.102
0.506	21	Unannotated	0.193	0.063	0.025	0.039
0.509	20	CpGI / Gene Body	0.202	0.073	0.029	0.058
0.764	18	SNP	0.195	NA	0.040	0.071
0.753	41	Unannotated	0.124	0.040	0.017	0.025
0.757	40	CpGI / Gene Body	0.129	0.047	0.018	0.037
0.951	48	SNP	0.154	NA	0.031	0.056
0.950	72	Unannotated	0.093	0.030	0.012	0.018
0.952	71	CpGI / Gene Body	0.097	0.035	0.013	0.028

Supplemental Table S13. Metadata on Canada lynx samples

Experiment_ID	Species	R1_Barcode	R2_Barcode	Repository_ID	Population	Nearest_Town	Pelt Size	Year.Trapped	Auction
1	*L. canadensis*	AACT	AACT	a103	QC	MONTREAL	Adult	2009	NAFA
3	*L. canadensis*	CCTA	AACT	a15	QC	ST FELICIEN	Adult	2009	NAFA
4	*L. canadensis*	TTAC	AACT	a18	QC	SACRE COEUR SAGUENAY	Adult	2009	NAFA
7	*L. canadensis*	AGGC	AACT	a234	QC	RIVIERE ST JEAN	Adult	2009	NAFA
8	*L. canadensis*	GAAGA	AACT	a28	QC	Métabetchouan–Lac-à-la-Croix	Adult	2009	NAFA
10	*L. canadensis*	CCTTC	AACT	a33	QC	CHARLESBOURG	Adult	2009	NAFA
11	*L. canadensis*	TTCAA	AACT	a395	QC	LAC A LA TORTUE	Adult	2009	NAFA
14	*L. canadensis*	GCGGC	AACT	a509	QC	BONAVENTURE	Adult	2009	NAFA
15	*L. canadensis*	AGATGC	AACT	a510	QC	PARENT	Adult	2009	NAFA
16	*L. canadensis*	CATAGC	AACT	a597	QC	ROUYN-NORANDA	Adult	2009	NAFA
18	*L. canadensis*	TTCGAC	AACT	a697	QC	ST OMER	Adult	2009	NAFA
24	*L. canadensis*	ATGCGC	AACT	a81	QC	GODBOUT	Adult	2009	NAFA
26	*L. canadensis*	ATGCGC	CCAG	a88	QC	MONTREAL	Adult	2009	NAFA
28	*L. canadensis*	TTCGAC	CCAG	a92	QC	ST MICHEL DES SAINTS	Adult	2009	NAFA
29	*L. canadensis*	CATAGC	CCAG	b120	QC	BAIE COMEAU	Adult	2009	NAFA
30	*L. canadensis*	AGATGC	CCAG	b121	QC	BAIE TRINITE	Adult	2009	NAFA
32	*L. canadensis*	AACT	CCAG	b321	QC	ROUYN-NORANDA	Adult	2009	NAFA
33	*L. canadensis*	CCTA	CCAG	b336	QC	SENNETERRE	Adult	2009	NAFA
34	*L. canadensis*	TTAC	CCAG	b338	QC	Lebel-sur-Quévillon	Adult	2009	NAFA
35	*L. canadensis*	AGGC	CCAG	b387	QC	MARIA	Adult	2009	NAFA
41	*L. canadensis*	GAAGA	CCAG	c21	QC	CHICOUTIMI	Adult	2009	NAFA
42	*L. canadensis*	CCTTC	CCAG	c245	QC	LA TUQUE	Adult	2009	NAFA

43	*L. canadensis*	TTCAA	CCAG	c291	QC	LA SARRE	Adult	2009	NAFA
46	*L. canadensis*	GCGGC	CCAG	c529	QC	ST ADELME DE MATANE	Adult	2009	NAFA
49	*L. canadensis*	AACT	TTGA	A128	MB	THOMPSON	Adult	2008-2009	NAFA
50	*L. canadensis*	CCTA	TTGA	A25	MB	WABOWDEN	Adult	2009	NAFA
51	*L. canadensis*	TTAC	TTGA	A310	MB	NORWAY HOUSE	Adult	2008-2009	NAFA
53	*L. canadensis*	AGGC	TTGA	A345	MB	HOLLOW WATER	Adult	2008-2009	NAFA
54	*L. canadensis*	GAAGA	TTGA	A349	MB	POPLAR RIVER	Adult	2008	NAFA
55	*L. canadensis*	CCTTC	TTGA	A371	MB	WABOWDEN	Adult	2008-2009	NAFA
56	*L. canadensis*	TTCAA	TTGA	A372	MB	WABOWDEN	Adult	2008-2009	NAFA
58	*L. canadensis*	GCGGC	TTGA	A383	MB	WABOWDEN	Adult	2008-2009	NAFA
61	*L. canadensis*	AGATGC	TTGA	A392	MB	WABOWDEN	Adult	2008-2009	NAFA
62	*L. canadensis*	CATAGC	TTGA	A396	MB	BIRCH RIVER	Adult	2008-2009	NAFA
64	*L. canadensis*	TTCGAC	TTGA	A416	MB	WABOWDEN	Adult	2008-2009	NAFA
66	*L. canadensis*	ATGCGC	TTGA	A423	MB	NORWAY HOUSE	Adult	2008-2009	NAFA
67	*L. canadensis*	ATGCGC	GGTC	A427	MB	MANIGOTAGAN	Adult	2008-2009	NAFA
69	*L. canadensis*	TTCGAC	GGTC	A457	MB	CRANBERRY PORTAGE	Adult	2008-2009	NAFA
71	*L. canadensis*	CATAGC	GGTC	A479	MB	SWAN RIVER	Adult	2008-2009	NAFA
73	*L. canadensis*	AGATGC	GGTC	A574	MB	SNOW LAKE	Adult	2008-2009	NAFA
74	*L. canadensis*	AACT	GGTC	A59	MB	WABOWDEN	Adult	2008-2009	NAFA
77	*L. canadensis*	CCTA	GGTC	A73	MB	SNOW LAKE	Adult	2008-2009	NAFA
79	*L. canadensis*	TTAC	GGTC	A839	MB	FLIN FLON	Adult	2008-2009	NAFA
82	*L. canadensis*	AGGC	GGTC	B204	MB	FLIN FLON	Adult	2008-2009	NAFA
83	*L. canadensis*	GAAGA	GGTC	B214	MB	WABOWDEN	Adult	2008-2009	NAFA
85	*L. canadensis*	CCTTC	GGTC	B24	MB	WABOWDEN	Adult	2008-2009	NAFA
87	*L. canadensis*	TTCAA	GGTC	B27	MB	HOLLOW WATER	Adult	2008-2009	NAFA
92	*L. canadensis*	GCGGC	GGTC	C302	MB	NORWAY HOUSE	Adult	2008-2009	NAFA
115	*L. canadensis*	AACT	ACTA	12C013	NL	PASADENA	Adult	2012	NAFA

116	*L. canadensis*	CCTA	ACTA	12C050	NL	ST VERONICS	Adult	2012	NAFA
117	*L. canadensis*	TTAC	ACTA	12C057	NL	ST VERONICS	Adult	2012	NAFA
118	*L. canadensis*	AGGC	ACTA	12C132	NL	GLENWOOD	Adult	2012	NAFA
119	*L. canadensis*	GAAGA	ACTA	12C244	NL	ST VERONICS	Adult	2012	NAFA
120	*L. canadensis*	CCTTC	ACTA	12C345	NL	PASADENA	Adult	2012	NAFA
122	*L. canadensis*	TTCAA	ACTA	12C641	NL	PASADENA	Adult	2012	NAFA
124	*L. canadensis*	GCGGC	ACTA	12D087	NL	PASADENA	Adult	2012	NAFA
125	*L. canadensis*	AGATGC	ACTA	12D151	NL	GLENWOOD	Adult	2012	NAFA
126	*L. canadensis*	CCTA	CAGC	12D325	NL	PASADENA	Adult	2012	NAFA
127	*L. canadensis*	TTCGAC	ACTA	12D592	NL	PASADENA	Adult	2012	NAFA
128	*L. canadensis*	ATGCGC	ACTA	12D600	NL	ST VERONICS	Adult	2012	NAFA
130	*L. canadensis*	ATGCGC	CAGC	12D602	NL	PASADENA	Adult	2012	NAFA
131	*L. canadensis*	TTCGAC	CAGC	12D926	NL	GLOVERTOWN	Adult	2012	NAFA
132	*L. canadensis*	CATAGC	CAGC	12D927	NL	CAMPBELLTON	Adult	2012	NAFA
133	*L. canadensis*	AGATGC	CAGC	12D928	NL	GLENWOOD	Adult	2012	NAFA
134	*L. canadensis*	AACT	CAGC	12D929	NL	GANDER	Adult	2012	NAFA
135	*L. canadensis*	CATAGC	ACTA	12D930	NL	GRAND FALLS / WINDSOR	Adult	2012	NAFA
136	*L. canadensis*	TTAC	CAGC	12D931	NL	ST ANTHONY	Adult	2012	NAFA
138	*L. canadensis*	AGGC	CAGC	12D933	NL	LA SCIE WB	Adult	2012	NAFA
139	*L. canadensis*	GAAGA	CAGC	12D934	NL	GLENWOOD	Adult	2012	NAFA
141	*L. canadensis*	CCTTC	CAGC	12D937	NL	LA SCIE WB	Adult	2012	NAFA
142	*L. canadensis*	TTCAA	CAGC	12D938	NL	BUCHANS	Adult	2012	NAFA
143	*L. canadensis*	GCGGC	CAGC	12D939	NL	LA SCIE WB	Adult	2012	NAFA
153	*L. canadensis*	AACT	TGAT	11H905	AK	NA	Adult	2010	FHA
156	*L. canadensis*	CCTA	TGAT	11H944	AK	NA	Adult	2009	FHA
160	*L. canadensis*	TTAC	TGAT	e370	AK	NA	Adult	2009	NAFA
163	*L. canadensis*	AGGC	TGAT	e442	AK	NA	Adult	2009	NAFA

166	*L. canadensis*	GAAGA	TGAT	E503	AK	NA	Adult	2009	NAFA
167	*L. canadensis*	CCTTC	TGAT	E509	AK	NA	Adult	2010	NAFA
170	*L. canadensis*	TTCAA	TGAT	11H904	AK	NA	Adult	2010	FHA
171	*L. canadensis*	GCGGC	TGAT	E519	AK	NA	Adult		NAFA
172	*L. canadensis*	AGATGC	TGAT	11A859	AK	NA	Adult	2010	NAFA
173	*L. canadensis*	CATAGC	TGAT	11A861	AK	NA	Adult	2010	NAFA
174	*L. canadensis*	TTCGAC	TGAT	11A864	AK	NA	Adult	2010	NAFA
175	*L. canadensis*	ATGCGC	TGAT	S72	AK	NA	Adult		NAFA
176	*L. canadensis*	ATGCGC	GTCG	S78	AK	NA	Adult		NAFA
177	*L. canadensis*	TTCGAC	GTCG	S80	AK	NA	Adult	2010	NAFA
178	*L. canadensis*	CATAGC	GTCG	S85	AK	NA	Adult	2010	NAFA
179	*L. canadensis*	AGATGC	GTCG	S97	AK	NA	Adult	2009	NAFA
180	*L. canadensis*	AACT	GTCG	11B623	AK	NA	Adult	2010	NAFA
181	*L. canadensis*	CCTA	GTCG	11B630	AK	NA	Adult	2010	NAFA
182	*L. canadensis*	TTAC	GTCG	11G403	AK	NA	Adult	2009	FHA
184	*L. canadensis*	AGGC	GTCG	11H382	AK	NA	Adult	2010	NAFA
185	*L. canadensis*	GAAGA	GTCG	11H387	AK	NA	Adult	2010	NAFA
186	*L. canadensis*	CCTTC	GTCG	11H397	AK	NA	Adult	2010	NAFA
187	*L. canadensis*	GCGGC	GTCG	11H416	AK	NA	Adult	2010	NAFA
LAM	*phage lambda*	TTCAA	GTCG	-	-		-	-	-

Supplemental Table S13. Metadata on Canada lynx samples (con't).

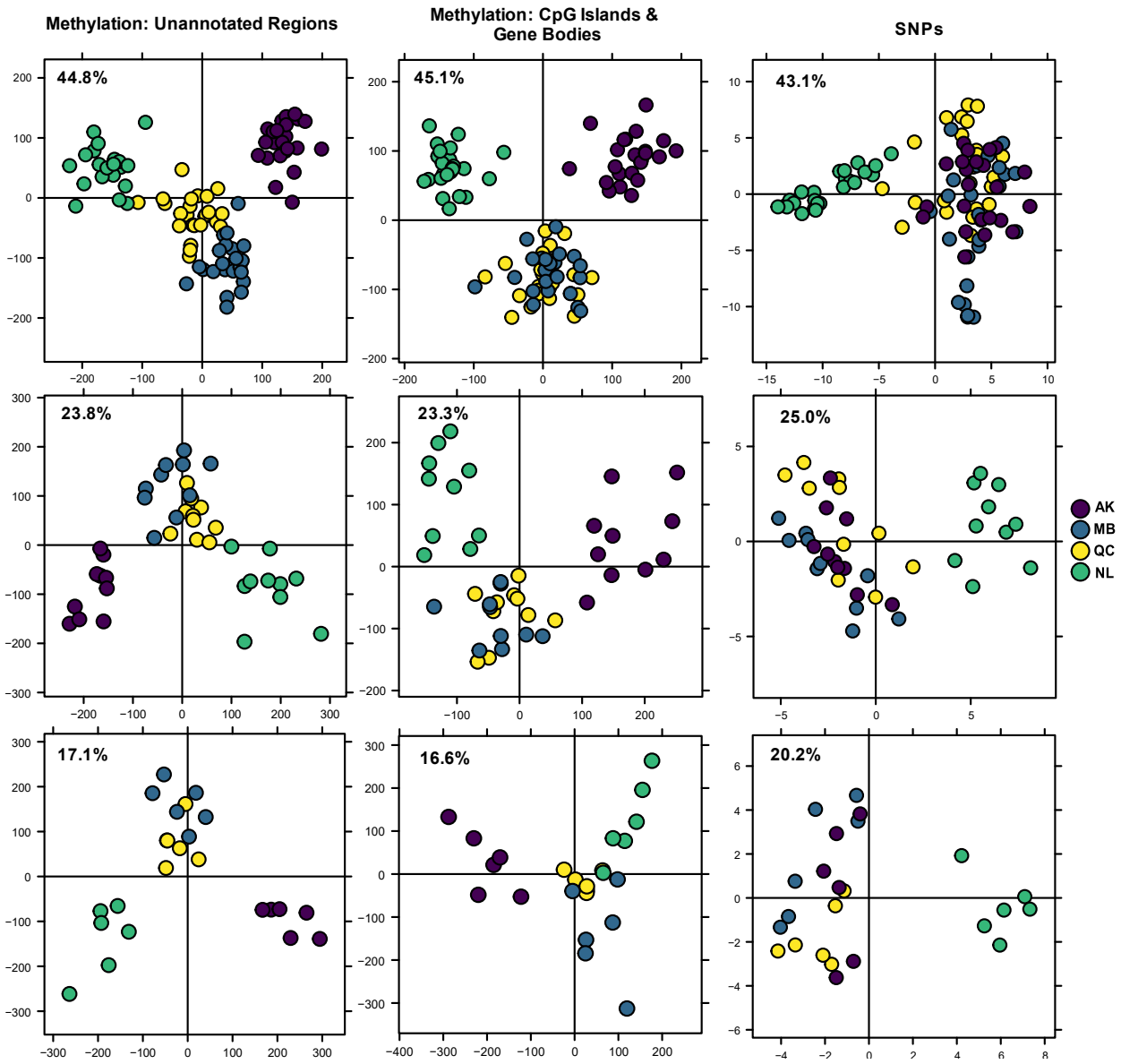
Experiment_ID	Longitude	Latitude	Distance_M	Raw_Read_Count	CAT_Mapping	HUMAN_Mapping	LAMBDA_Mapping
1	-72.715	46.605	4982114.4	1384663	82.80%	3.00%	0.10%
3	-72.683	48.336	4833944	9072401	84.70%	2.80%	0.00%
4	-70.423	48.695	4909052.1	2931915	85.30%	2.80%	0.10%
7	-64.699	50.439	5020042.2	1030075	84.20%	2.90%	0.10%
8	-71.636	48.363	4880992.3	4741702	72.50%	2.50%	0.00%
10	-70.377	46.796	5076123.5	1075804	81.20%	2.80%	0.10%
11	-73.81	46.428	4945337.7	4671944	84.40%	2.70%	0.00%
14	-65.441	48.322	5171105.5	2185040	84.40%	2.60%	0.00%
15	-75.135	47.932	4752214	2419271	72.90%	2.50%	0.00%
16	-78.552	46.836	4681497.2	2203639	82.40%	2.80%	0.00%
18	-66.666	48.171	5128288.3	2519861	82.70%	2.70%	0.00%
24	-66.683	50.178	4991277.4	5238759	85.00%	2.80%	0.00%
26	-72.715	46.605	4982114.4	2104158	82.80%	2.60%	0.00%
28	-74.004	46.601	4921006.1	1297051	82.20%	2.60%	0.00%
29	-68.02	49.329	4991277.4	1236359	81.90%	2.70%	0.00%
30	-67.668	49.748	4945625.9	1784163	83.40%	2.70%	0.00%
32	-79.125	47.845	4567665.5	2081035	85.60%	2.70%	0.00%
33	-76.635	48.361	4643947.9	8569931	84.40%	2.80%	0.00%
34	-76.635	48.361	4643947.9	6582948	84.70%	2.80%	0.00%
35	-65.441	48.322	5171105.5	1168289	86.40%	2.80%	0.00%
41	-71.409	47.698	4949052.9	3291192	78.70%	2.60%	0.00%
42	-73.895	48.172	4790684.4	1163646	83.10%	2.70%	0.00%
43	-78.48	48.629	4532569.8	1406980	82.70%	2.80%	0.10%
46	-67.691	48.622	5041923.8	2298714	83.30%	2.70%	0.00%
49	-99.282	55.928	2953256.6	258845	72.10%	3.00%	0.40%
50	-96.948	53.365	3252155.4	2219425	82.20%	2.80%	0.00%
51	-98.217	54.297	3120901.3	1555243	83.60%	3.00%	0.20%
53	-95.659	51.262	3476771.1	300643	82.30%	2.70%	0.10%
54	-96.424	52.997	3305587.4	3262523	86.50%	3.00%	0.00%
55	-98.425	55.028	3058252.5	350551	81.90%	2.90%	0.10%
56	-98.297	55.17	3054505.3	3811105	85.70%	2.90%	0.00%
58	-98.297	55.17	3054505.3	2094328	81.70%	2.60%	0.00%
61	-98.297	55.17	3054505.3	1910731	82.90%	2.80%	0.00%
62	-101.42	52.642	3084711	1788686	84.60%	3.10%	0.00%
64	-98.297	55.17	3054505.3	3084239	84.50%	2.80%	0.00%
66	-97.805	53.986	3163954.9	4195690	75.90%	2.50%	0.00%
67	-95.659	51.262	3476771.1	536700	82.40%	2.80%	0.00%

69	-101.04	51.759	3171013.9	1598087	85.30%	2.60%	0.00%
71	-101.42	52.642	3084711	1081433	86.20%	3.00%	0.00%
73	-99.781	54.889	3001115.2	3979111	86.90%	2.70%	0.00%
74	-98.297	55.17	3054505.3	1529016	84.70%	2.60%	0.00%
77	-99.781	54.889	3001115.2	4110144	86.40%	2.70%	0.00%
79	-101.04	51.759	3171013.9	4063722	87.20%	2.80%	0.00%
82	-101.04	51.759	3171013.9	555046	87.30%	2.90%	0.00%
83	-98.297	55.17	3054505.3	4044929	87.50%	2.90%	0.00%
85	-98.297	55.17	3054505.3	968604	85.70%	2.70%	0.00%
87	-95.659	51.262	3476771.1	3143592	87.40%	2.80%	0.00%
92	-97.805	53.986	3163954.9	1528319	85.80%	2.50%	0.00%
115	-57.658	49.055	5451354.2	735196	86.20%	3.10%	0.10%
116	-55.894	47.956	5626536.4	3812193	86.20%	3.00%	0.10%
117	-55.828	48.028	5622913.2	1796568	84.40%	2.80%	0.10%
118	-54.926	49.048	5568628.1	442548	87.00%	3.00%	0.10%
119	-55.665	47.974	5634751.6	2054790	86.60%	2.90%	0.10%
120	-57.683	49.021	5453294.5	589481	86.20%	2.60%	0.10%
122	-57.67	48.964	5458976.7	1611028	87.30%	3.00%	0.00%
124	-57.599	48.941	5464144.7	1637354	86.90%	2.80%	0.10%
125	-54.921	49.049	5568807	1425640	86.20%	3.10%	0.00%
126	-57.54	48.971	5463959.2	1864162	87.00%	3.00%	0.00%
127	-57.478	48.99	5464905.5	699902	85.20%	2.80%	0.00%
128	-55.675	47.904	5640679.4	2976735	84.70%	2.80%	0.00%
130	-57.491	49.029	5460811.2	3076186	85.80%	2.40%	0.00%
131	-54.108	48.678	5636805.2	693652	87.10%	2.60%	0.00%
132	-54.879	49.261	5551317.5	403027	87.00%	2.70%	0.00%
133	-54.805	49.022	5576104.9	1130494	84.90%	3.20%	0.00%
134	-54.584	49.055	5582497.7	726879	87.70%	3.00%	0.00%
135	-55.694	49.025	5538116.8	425169	88.10%	3.00%	0.00%
136	-55.681	51.426	5323083.4	1773794	87.60%	2.60%	0.00%
138	-55.7	49.911	5457930.5	310390	86.10%	1.80%	0.10%
139	-54.785	48.975	5581198.7	1226585	87.30%	2.70%	0.00%
141	-55.677	49.884	5461348.9	330885	87.40%	2.10%	0.00%
142	-56.894	48.881	5499798.1	664791	87.90%	2.20%	0.00%
143	-55.641	49.938	5458039.1	602316	86.70%	3.50%	0.00%
153	-145.16	62.054	355625.95	1341774	87.50%	2.50%	0.00%
156	-150.99	63.9	40347.196	6432556	85.30%	2.60%	0.00%
160	-160.04	63.942	432660.44	3392154	88.30%	2.60%	0.00%
163	-146.89	61.56	318670.35	525535	87.30%	2.40%	0.00%
166	-142.43	61.132	533331.76	3319855	88.00%	2.90%	0.00%
167	-144.7	62.894	340567.32	1201544	84.20%	2.40%	0.00%

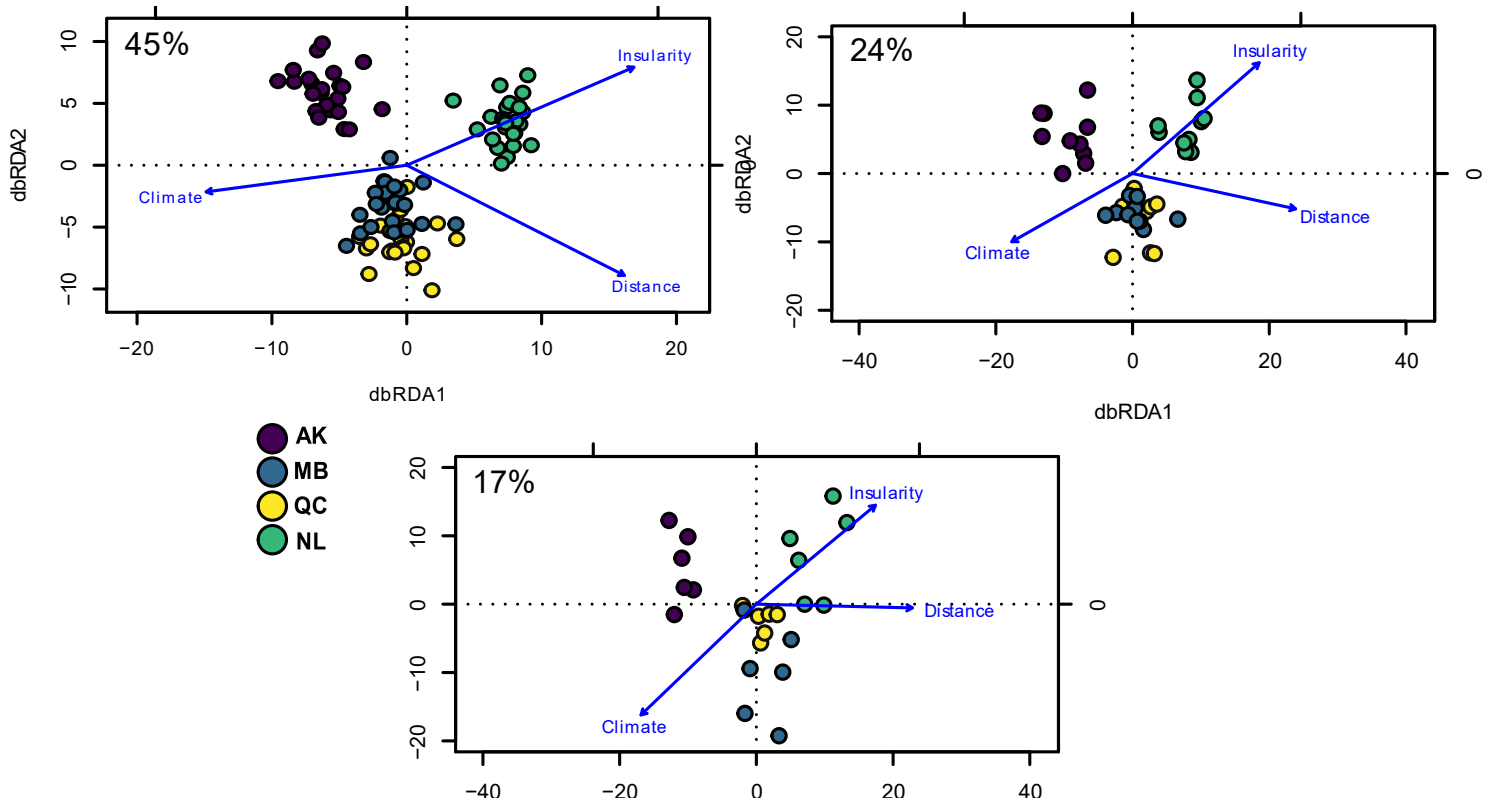
170	-146.55	66.964	438970.11	1985404	87.50%	2.60%	0.00%
171	-145.85	62.088	323348.72	1719589	87.70%	2.30%	0.00%
172	-146.22	62.756	271653.82	2969901	87.60%	2.70%	0.00%
173	-161.34	63.302	501195.29	1082929	87.40%	2.50%	0.00%
174	-161.34	63.302	501195.29	1948388	88.00%	2.60%	0.00%
175	-146.02	64.543	279905.41	2694131	88.10%	2.50%	0.00%
176	-151.81	64.576	115195.05	2956596	86.10%	2.50%	0.20%
177	-161.34	63.302	501195.29	1662249	84.50%	2.50%	0.20%
178	-161.34	63.302	501195.29	1313675	86.40%	2.50%	0.20%
179	-161.21	66.872	590907.26	5613249	88.60%	2.70%	0.10%
180	-145.87	61.793	341397.72	1875769	87.00%	2.90%	0.10%
181	-145.56	61.297	389645.57	5902498	88.00%	2.60%	0.30%
182	-148.86	64.822	183514.67	2063216	85.20%	2.50%	0.70%
184	-161.34	63.302	501195.29	811653	88.20%	2.20%	0.30%
185	-148.97	65.549	247665.09	2913570	88.30%	2.70%	0.40%
186	-150.42	66.032	278050.1	455667	85.70%	2.80%	1.90%
187	-145.81	62.744	291889.78	1151693	85.10%	2.80%	0.50%
LAM				4260069			

CHAPTER 3 SUPPLEMENTAL FIGURES

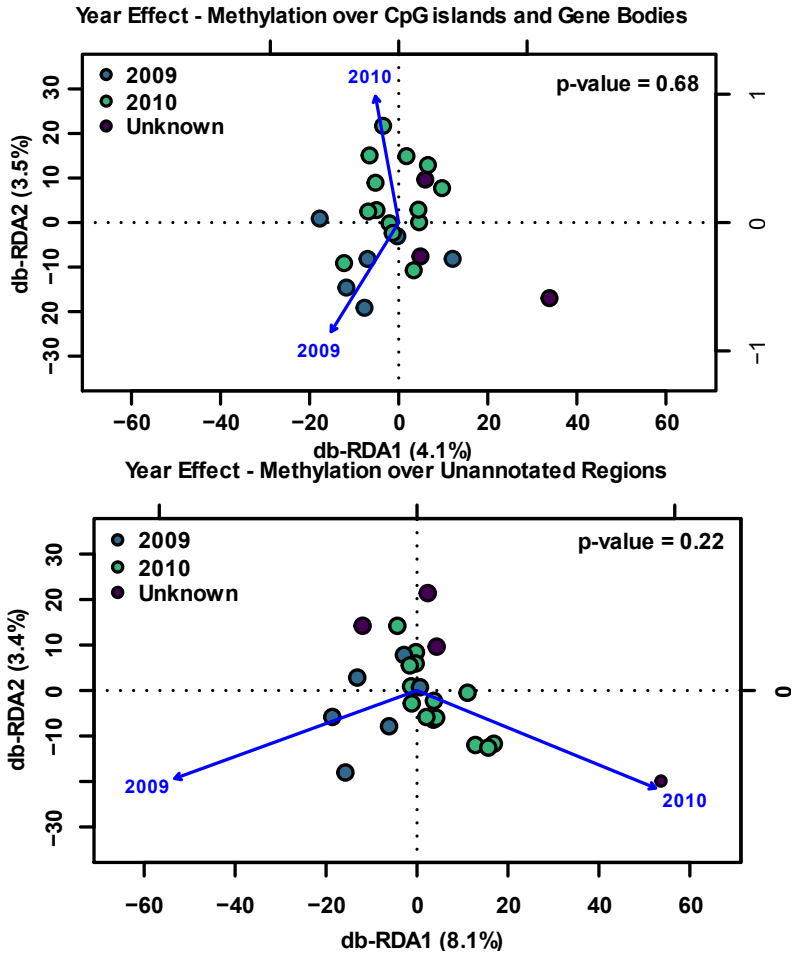
A.



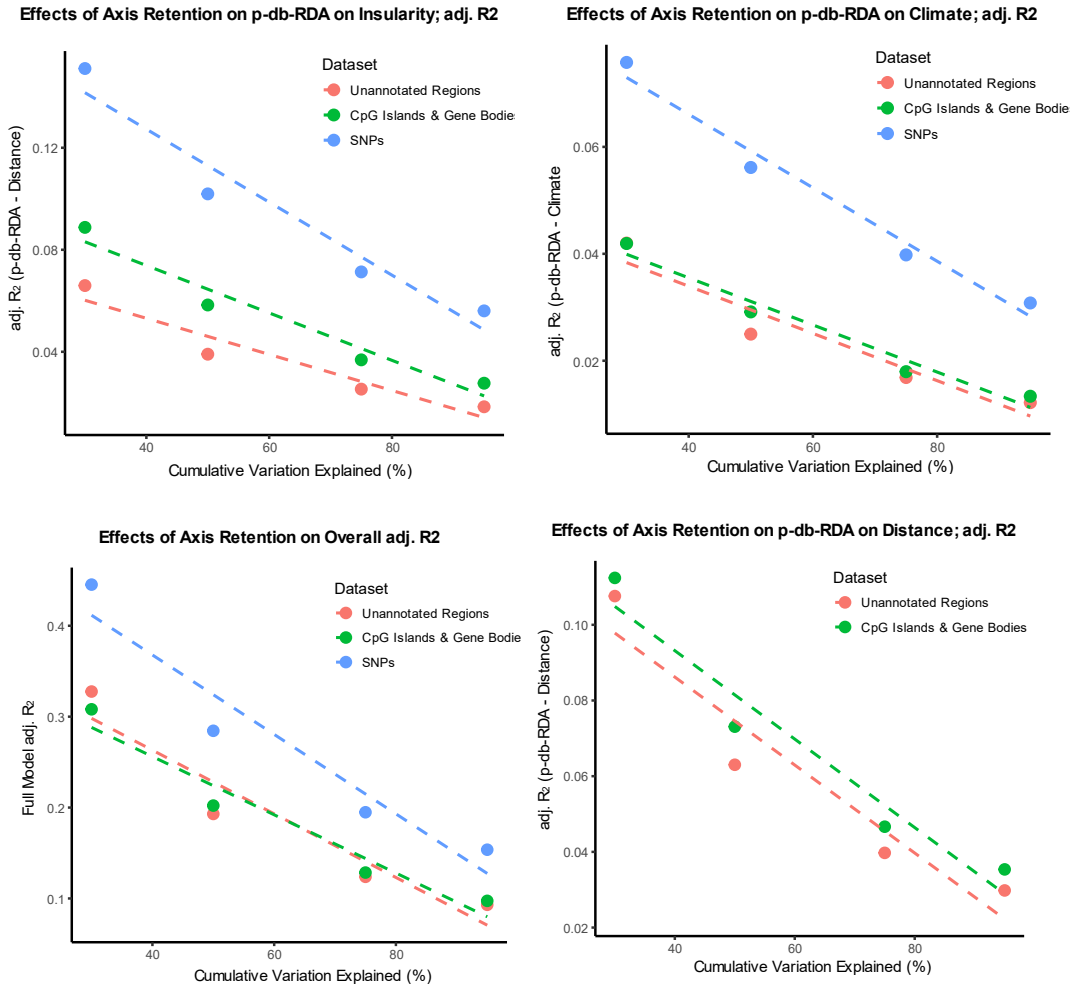
B.


45%
Insularity
Climate
Distance
dbRDA1
dbRDA2
24%
Insularity
Climate
Distance
dbRDA1
dbRDA2
17%
Insularity
Distance
Climate
AK
MB
QC
NL

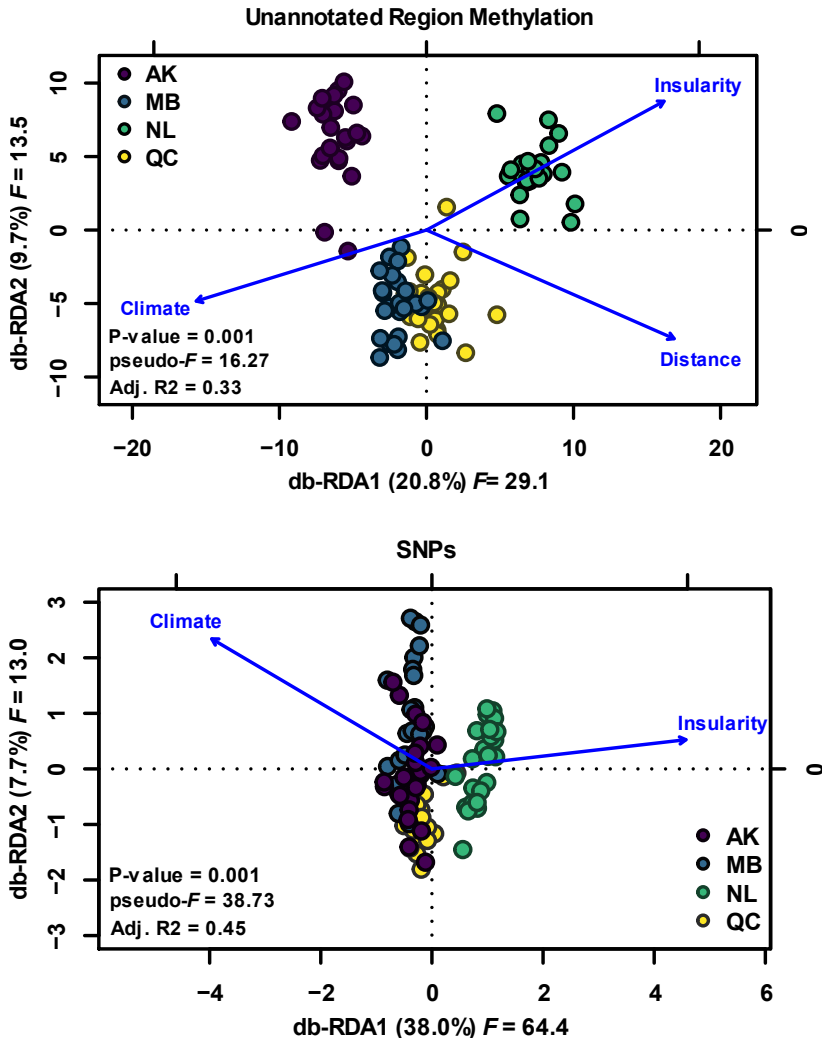
C.


Year Effect - Methylation over CpG islands and Gene Bodies
2009
2010
Unknown
p-value = 0.68
db-RDA2 (3.5%)
db-RDA1 (4.1%)
Year Effect - Methylation over Unannotated Regions
p-value = 0.22
db-RDA2 (3.4%)
db-RDA1 (8.1%)

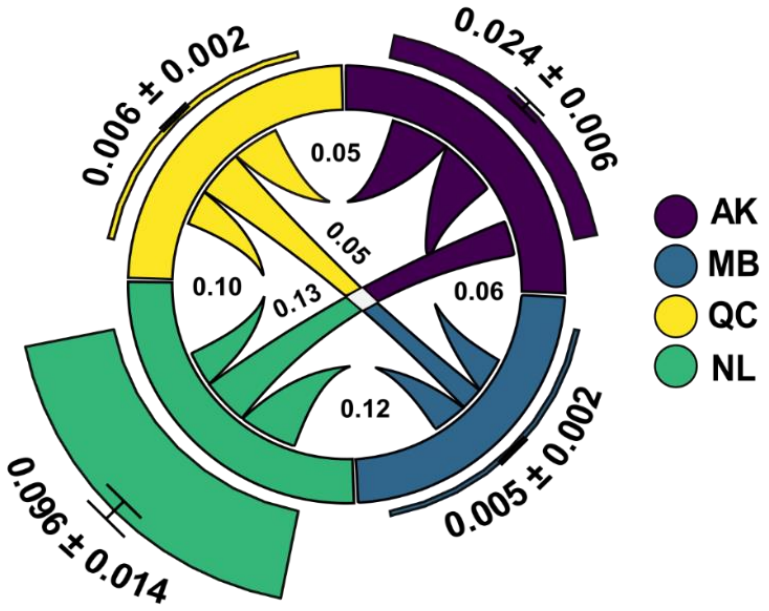
D.



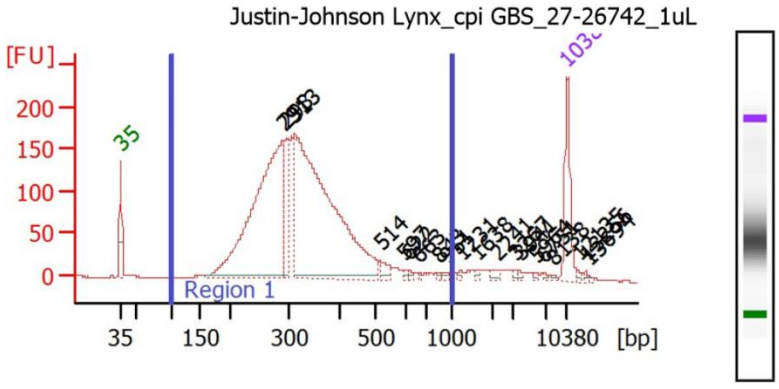
Supplemental Figure S1. Sensitivity analyses on four aspects of our analyses. **A.** Effects of missing data on PCoA results. **B.** Effects of missing data on db-RDA results. **C.** Temporal variation in methylation patterns. db-RDA biplots on methylation patterns over CpG islands / gene bodies and unannotated regions within the Alaska population, with a binary year variable as an explanatory variable. **D.** Plot matrix on the effect of axis retention for db-RDAs. The dotted lines are the predicted linear regression lines, with datasets delineated by colour.



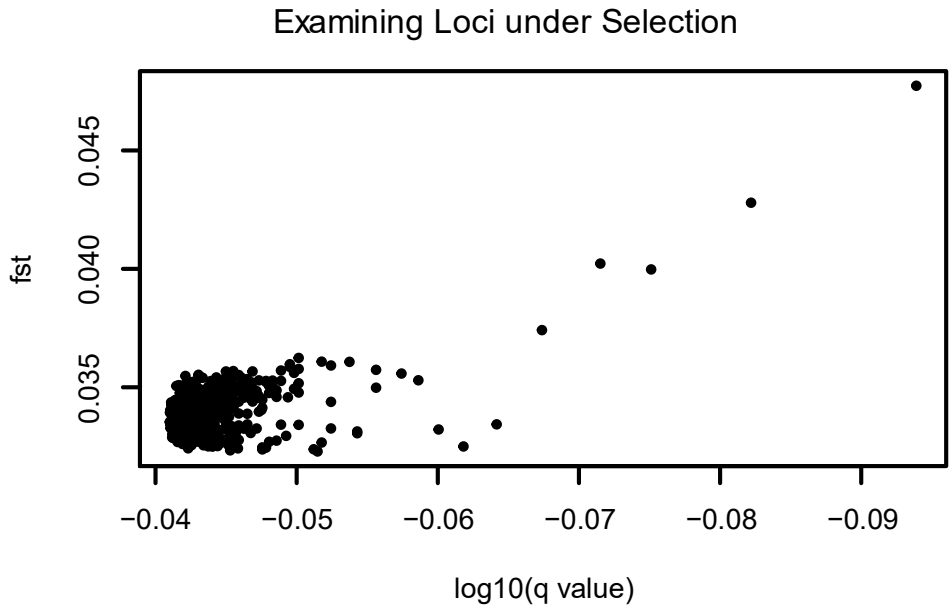
Supplemental Figure S2. Associations between environmental variables and SNP / annotated region methylation. db-RDA biplots showing the relationship between biogeographical variables and molecular data (SNPs, unannotated methylation). Axes explaining > 30% of the cumulative variation were used as response variables. Populations are delineated by colour.



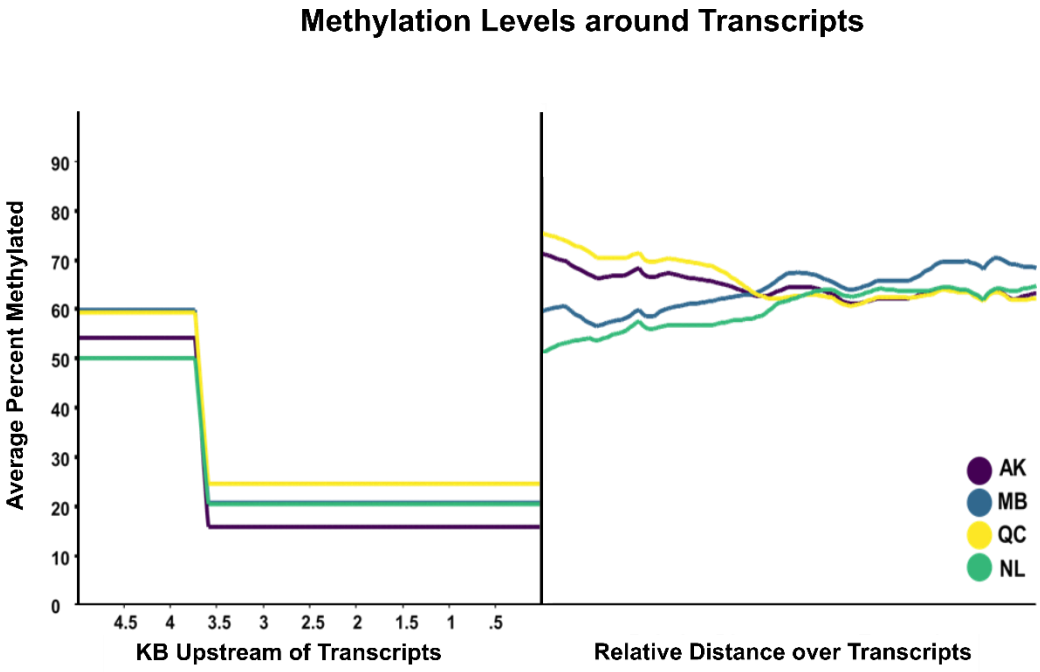
Supplemental Figure S3. Genetic differentiation between populations using 489 SNPs, called from bisulfite converted reads. Outside bars represent relative differentiation ($F_{ST)}$ against a baseline null model, while the inside ribbons indicate pair-wise F_{ST}.



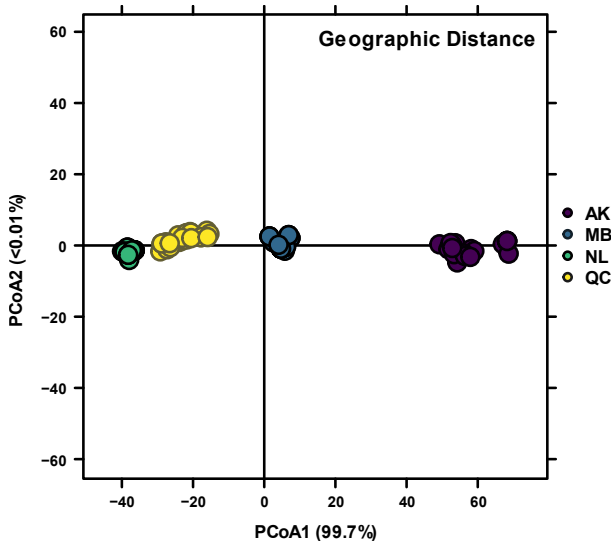
Supplemental Figure S4. Final library quality check for sequencing. Agilent Bioanalyzer 2100 results, showing final library fragment distribution before sequencing.



Supplemental Figure S5. Examining loci under selection. BayeScan Visual outlier plot from to determine if any loci are under selection, none were identified.



Supplemental Figure S6. Methylation levels around transcripts. The left pane shows methylation levels (%) 5,000-bp upstream of transcripts, while the right pane shows average methylation levels across the entire transcript.



Supplemental Figure S7. PCoA summarizing a Euclidean distance matrix on geographic distance (latitude and longitude), used as a variable for db-RDAs to assess biogeographical relationships.

Supplemental Figure S8. PCA summarizing variation in climate across the study area. Variables included annual temperature range, minimum temperature of the coldest month, and precipitation of the coldest quarter, retrieved from WorldClim. The first axis of this PCA explained 85.62% of the variation and was created to summarize winter conditions in a single vector for db-RDAs against molecular data.