

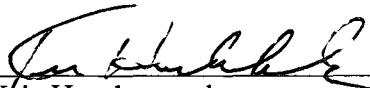
GENETIC ANCESTRY MODELING AND PERFORMANCE ASSOCIATION IN


THE ALASKAN SLED DOG


By

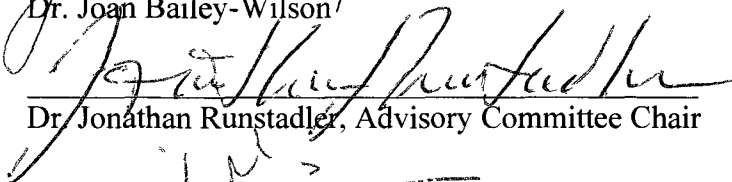
Heather Jay Huson

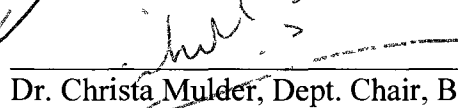
RECOMMENDED:


Dr. Kris Hundertmark

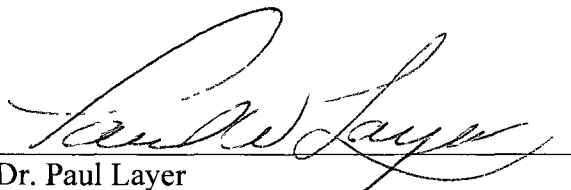

Dr. Elaine Ostrander

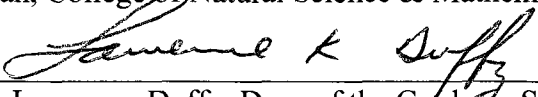

Dr. Joan Bailey-Wilson

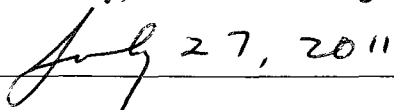

Dr. Jonathan Runstadler, Advisory Committee Chair


Dr. Christa Mulder, Dept. Chair, Biology & Wildlife

APPROVED:


Dr. Paul Layer
Dean, College of Natural Science & Mathematics


Dr. Lawrence Duffy, Dean of the Graduate School


Date

GENETIC ANCESTRY MODELING AND PERFORMANCE ASSOCIATION IN

THE ALASKAN SLED DOG

A

DISSERTATION

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

By

Heather Jay Huson, B.A.

Fairbanks, Alaska

August 2011

UMI Number: 3484670

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

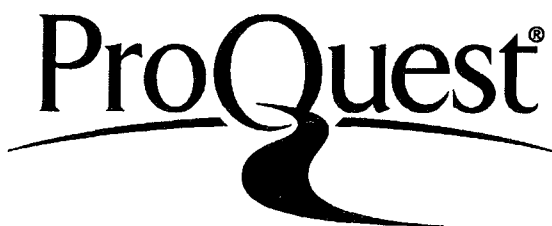
In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3484670

Copyright 2011 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Alaskan sled dogs present us with the unique opportunity to study the development of a population of dogs produced from the selective breeding of high performance athletes. I establish that sled dogs are a genetically distinct population of dogs that segregate into two sub-groups based on their racing style of “sprint” or short distance and “distance” or long distance. The practice of interbreeding Alaskan sled dogs with various purebred dogs over the past century has allowed us to investigate the impact of these domestic breeds on the sled dog genome and their potential contribution to athletic performance. Here, I establish genetic profiles of both the sprint and the distance racing dogs using microsatellite-based markers, single-nucleotide polymorphism (SNP) arrays, and ancestry modeling. Population structure is assessed using clustering and principle component analyses. Inbreeding patterns are examined through population structure, inbreeding statistics, estimations of linkage disequilibrium, and autozygosity. Purebred breed components and their potential role in influencing performance attributes of Alaskan sled dogs were determined through genetic breed identification. Ancestry modeling was used to localize genomic regions of specific breed selection. These breeds were then analyzed for their genetic contribution to regions experiencing selection within the sprint or distance racing dogs. I determined regions of selective sweep and genome-wide association to the sprint or distance racing dogs. A genome-wide association analysis of heat tolerance performance in sprint dogs identified SNPs potentially regulating the *MYH9* gene. This was the first genetic assessment of ancestry, inbreeding, and performance genes attributed to racing Alaskan sled dogs.

Table of Contents

	Page
Signature Page	i
Title Page	ii
Abstract	iii
Table of Contents	iv
List of Figures	ix
List of Tables	xi
Acknowledgements	xiii
Chapter 1: Introduction	1
The Alaskan Sled Dog	3
The Canine Genome	8
Performance Genetics	11
The Alaskan Sled Dog as a Model System for the Study of Performance Genetics	14
References	16
Chapter 2: A Genetic Dissection of Breed Composition and Performance	
Enhancement in the Alaskan Sled Dog	24
Abstract	24
<i>Background</i>	24
<i>Results</i>	25
<i>Conclusion</i>	25

	Page
Background.....	26
Methods.....	28
<i>Sample Collection</i>	28
<i>Performance Ratings</i>	29
<i>Microsatellite Genotyping</i>	31
<i>Statistical Analysis of Population Structures</i>	32
<i>Breed Composition</i>	33
<i>Inbreeding Coefficients</i>	34
Results.....	34
<i>Alaskan Sled Dog Breed Identification</i>	34
<i>Ancestral Groupings</i>	37
<i>Population Structure and Breed Composition</i>	38
<i>Inbreeding</i>	40
<i>Breed Contribution to Performance</i>	42
Discussion	43
Conclusions.....	48
Authors' Contributions	49
Authors' Information	49
Acknowledgements.....	49
References.....	50
Figures.....	52

	Page
Tables.....	60
Appendix	65
 Chapter 3: A SNP within the Angiotensin-Converting Enzyme Distinguishes	
between Sprint and Distance Performing Alaskan Sled Dogs in a Candidate Gene	
Analysis	66
Abstract	66
Introduction.....	67
Methods.....	70
<i>Sample Collection</i>	70
<i>Performance Ratings</i>	71
<i>Sample Selection for ACE Gene Analysis</i>	72
<i>Sample Selection for MSTN Gene Analysis</i>	73
<i>DNA Amplification and Sequencing</i>	73
<i>Statistical Analysis for the ACE Gene</i>	74
<i>Statistical Analysis for the MSTN Gene</i>	76
Results.....	77
<i>ACE Gene</i>	77
<i>MSTN Gene</i>	81
Discussion	82
Acknowledgements.....	87
References.....	88

	Page
Figures	91
Tables	93
Appendix	96
 Chapter 4: Selection for Breed-Specific Ancestry and Genome-Wide Association	
Analysis Target SNPs Associating the <i>MYH9</i> Gene to Heat Tolerance within	
Performing Alaskan Sled Dogs	99
Abstract	99
Introduction.....	100
Methods.....	102
<i>Sample Collection and SNP Array Genotyping</i>	102
<i>Performance Ratings</i>	104
<i>Population Structure</i>	107
<i>Linkage Disequilibrium and Homozygosity Analysis</i>	107
<i>Selective Sweep</i>	108
<i>Genome-wide Association Study (GWAS)</i>	109
<i>Sequencing of the <i>HINT1</i> and <i>MYH9</i> Genes</i>	111
<i>Phasing</i>	113
<i>Modeling Ancestry</i>	113
Results.....	116
<i>Population Structure</i>	116
<i>Linkage Disequilibrium and Homozygosity Analysis</i>	116

	Page
<i>Selective Sweep</i>	118
<i>Genome-wide Association Study (GWAS)</i>	119
<i>Ancestry Modeling</i>	120
<i>Fine-mapping of the HINT1 and MYH9 Genes</i>	125
Discussion	128
References	136
Figures	141
Tables	148
Appendix	154
Chapter 5: Conclusion	157
Population Structure and Inbreeding	157
Relationship of the Alaskan Sled Dog to Purebred Breeds	158
<i>Ancestral Grouping and Breed Components</i>	158
<i>Ancestry Modeling</i>	160
The Genetics of Performance	162
<i>Breed Influence</i>	162
<i>Candidate Gene Investigation</i>	163
<i>Whole Genome Association Studies and Admixture Mapping</i>	165
Project Retrospection	169
References	173

List of Figures

	Page
Figure 2.1: Alaskan sled dogs are a mixed breed dog selected strictly for their racing performance.	52
Figure 2.2: Work ethic was scored on a three-tiered system based on the dog's willingness to run.	53
Figure 2.3: Population structure of purebred dogs and Alaskan sled dogs.	54
Figure 2.4: Population structure of sprint and distance sled dogs during successive increase in assigned population numbers.	55
Figure 2.5: Ancestral breed clustering and composition differences of the sprint and distance Alaskan sled dogs.	56
Figure 2.6: Breed composition of Alaskan sled dogs reflected by three populations based on racing style.	57
Figure 2.7: Breed composition differences between Alaskan sled dog sub-populations.	58
Figure 2.8: The degree of inbreeding F_{IS} within breed populations.	59
Figure 3.1: ACE gene linkage disequilibrium within sprint and distance sled dogs and eight domestic breeds attributed with speed or endurance performance.	91
Figure 4.1: Alaskan sled dogs are a mixed breed dog selected strictly for their performance.	141
Figure 4.2: Principle component analysis plots of Alaskan sled dogs (A & B) and four ancestry reference breeds (B) using a panel of 7K highly ($F_{ST}>0.35$) informative SNPs.	142

Figure 4.3: The estimated decay of linkage disequilibrium and degree of autozygosity among Alaskan sled dogs and their four ancestral component breeds.	143
Figure 4.4: Genome-wide association plot of elite versus poorly performing heat tolerance sprint sled dogs.....	144
Figure 4.5: A visual comparison of the most prevalent diploid state ancestry blocks across the genome of the sprint and distance sled dogs.....	145
Figure 4.6: A comparison of the genome-wide frequency of four ancestral reference breeds within the distance and sprint sled dog populations.....	146
Figure 4.7: The difference in frequency of the four ancestral breeds between the distance and sprint sled dogs on chromosome 16.....	147

List of Tables

	Page
Table 2.1: The percentage breed composition of four Alaskan sled dog sprint sub-populations and four Alaskan sled dog distance sub-populations.	60
Table 2.2: Sled dogs were 5x lower in inbreeding and 15% higher in observed heterozygosity than purebreds.	62
Table 2.3: The percentage change in Alaskan sled dog breed composition between high and low performing individuals.	64
Table 2A.1: STRUCTURE run input information including the population groupings, the number of populations designated, and the result objective.	65
Table 3.1: Associated SNPs across the <i>ACE</i> gene comparing distance and sprint sled dogs.	93
Table 3.2: <i>ACE</i> gene Marker 56 shows P-value scores ≤ 0.001 in separate association tests of Alaskan sled dogs and purebred breeds.	94
Table 3.3: Minor allele frequencies of three <i>MSTN</i> gene polymorphisms found in sprint Alaskan sled dogs, 16 domestic breeds, and a Golden Jackal.	95
Table 3A.1: <i>ACE</i> gene polymorphisms identified by manual sequence screening of Alaskan sled dogs and eight domestic breeds.	96
Table 4.1: The genome-wide frequency (<i>f</i>) of individual ancestral populations and their respective diploid ancestry states within the distance and sprint sled dog populations..	148

Table 4.2: The overall number, median length, and genome-wide frequency (<i>f</i>) of diploid ancestry blocks found private to either the distance (A) or sprint (B) sled dog populations.	149
Table 4.3 Tabulation of the genetic loci demonstrating the highest degree of interest for population differentiation or performance association within Alaskan sled dogs.	150
Table 4.4: SNPs within and surrounding the <i>MYH9</i> gene which are associated with heat tolerance performance in sprint racing Alaskan sled dogs.	152
Table 4.5: <i>MYH9</i> gene SNPs associated to either the Alaskan Malamute or German Shorthaired Pointer breeds.	153
Table 4A.1: Genomic loci demonstrating the greatest degree of selective sweep within distance and sprint sled dogs.....	154
Table 4A.2: Single-nucleotide polymorphism markers showing genome-wide significance in association with either population differentiation between sprint and distance sled dogs or heat tolerance in sprint sled dogs.....	155
Table 4A.3: The top 5 th percentile of SNPs showing the highest degree of ancestral breed differentiation between the distance and sprint sled dog populations.	156

Acknowledgements

The driving force behind this research project has been my passion for sled dogs and a desire to understand what makes them the incredible athletes that they are. Here lies the beginning of my learning and goals on this endeavor to utilize science to explain the athletic prowess of one of my best friends.

I would like to thank my family and husband who have been a life line of support throughout these long years of speed bumps and road blocks with the intermittent highlight. I finally finished! Thank you!

Chapter 1

Introduction

For thousands of years the dog has proven itself an invaluable asset to humans by serving as hunter, protector, transportation, aide, and loyal companion. Contemporary culture has diversified dogs into breeds inherently developed for specific characteristics and trained to fulfill modern needs. The twenty-first century finds specific breeds as well as mixed breeds used in hunting, law enforcement, search and rescue, racing, emotional therapy, and as service dogs to individuals with medical impairments. These breeds provide an excellent genetic resource for investigating the molecular components responsible for the variety of traits for which the dogs are bred. Modern technology allows the dog to serve another role, that of a model biological system. Humans study dogs' physiology and their genetic composition to gain a better understanding of biomedical science and exercise physiology. This research offers insight into the biological mechanisms of dogs as well as increasing our understanding of human biological systems.

The dog's domestication and breed development has paralleled the changing culture of humans. Archaeological evidence has dated the earliest dog remains to between 9,000 to 31,000 years ago, with the suggestion that these first dogs were used in

the tracking, capture, and transport of large “ice age” game as well as an emergency source of food [1-3]. Eventually, pastoral societies gave rise to home and flock guard dogs and herding dogs. As civilizations moved to conquer and defend, dogs became weapons of war [4, 5]. The Victorian Era saw the creation of the majority of breeds known today. Breeds were developed for purposes of companionship, with particular emphasis placed upon appearance and conformity with styles of hunting common to the time period and social culture [6]. Researchers have used molecular methodology to estimate dog domestication as having occurred 15,000-100,000 years ago [7-11]. Recent genetic evidence suggests that multiple lineages and founding events occurred over the course of domestication, with earliest populations originating from the gray wolf in the Middle East [11]. However, the vast majority of the roughly 350-400 domestic breeds recognized throughout the world today are the results of intensive artificial selection for distinctive characteristics. This selection began only 200 years ago, during the Victorian Era. Registries such as the American Kennel Club (AKC), which recognizes 165 breeds within the United States, have formalized and standardized breeds based on conformity, appearance, and behavior [12].

Aside from working and domestic companion roles, dogs have also become a major asset in medical genetic research. Approximately 360 canine diseases are analogous to human diseases, therefore establishing dogs as an excellent model organism for the identification and study of disease loci [13-16]. Canine biomedical research has encompassed diseases ranging from cancer to muscular, behavioral, and neurological disorders [16-22]. Research benefits include commonality between dogs and humans in

disease presentation and treatment response, outcome, and underlying susceptibility, as well as the shared environmental factors affecting cohabitating humans and dogs. While this fact in itself makes dogs an invaluable research tool, dogs also provide a system to investigate the genetics of morphology and athletic performance mechanisms. This includes the identification of genes and mutations regulating overall body size and bone growth, coat variation, and racing enhancement [23-27].

It is in the realm of athletic performance that we examine the Alaskan sled dog. We look at both physiological and behavioral aspects of performance, as well as purebred breed influence since the Alaskan sled dog is a mixed breed dog. This is the first genetic investigation focused on performance, using an admixed population selectively developed for its athletic abilities. The following sections provide an introduction to the history and development of sled dog racing, the characteristics of the canine genome, gene mapping and association studies, and the current knowledge of performance genetics in both humans and animals. This introduction concludes with a discussion of how the Alaskan sled dog can further our understanding of the genetic complexity affecting performance.

The Alaskan Sled Dog

The Alaskan sled dog presents us with the unique opportunity to study the development of a genetically distinct breed of dog produced from the selection of high performance athletes. The early Arctic dog was originally bred for hauling cargo-laden sleds across the frozen tundra, as well as for providing both protection and

companionship to the people of these northern cultures [28-30]. The theory of origin of Alaskan sled dogs varies widely depending upon the evidence one relies upon. Inuit dogs are often regarded as the earliest and purest line of sled dogs, deriving from dogs migrating across the Bering Sea with the Thule people between 500-1100 C.E. The Thule people are ancestors of the Inuit people, and include natives from Canada, Denmark (Greenland), Russia (Siberia) and the U.S. (Alaska), all areas from which sled dog origins are claimed [31]. However, literature often cites the Inuit dog as being specific to the native cultures of Greenland, Canada, and Alaska, while excluding Russia [6, 32]. The Chukchi, an indigenous Siberian (Russian) people, are credited with the development of the Siberian Husky (breed name recognized by the AKC), also known as the Arctic Husky (breed name recognized by the United Kennel Club; UK), approximately 3,000 years ago as a long distance sled dog. The Siberian Husky was brought into Alaska during the early 1900s [6, 12]. Concomitantly, Inuit dogs within Alaska and Canada began to be referred to as Eskimo dogs or Huskies. An Alaskan native tribe, the Mahlemuts, was credited with the development of the Alaskan Malamute from these lineages [6, 12]. From this information, we can speculate that the Alaskan sled dog, Siberian Husky, and Alaskan Malamute all originated from the same stock of early, pre-migratory Inuit dogs and were selected as a northern climate working dog. From origination through today, breeders have continued refining the Alaskan sled dog based primarily upon their performance in northern climates. This refinement has included cross breeding Alaskan sled dogs with early lines of Alaskan Malamutes and Siberian Huskies as well as other breeds in an effort to enhance particular athletic attributes. The

breeding strategy employed throughout the years for Alaskan sled dogs contrasts modern selection in Alaskan Malamutes and Siberian Huskies, both recognized purebred breeds. While the Alaskan sled dog continues to be selected for performance, the purebred Alaskan Malamute and Siberian Husky are selected upon standards of appearance and confined to breeding among registered members within the breed [12].

The “Era of the Sled Dog” extended from the late 1800s to the early 1900s, encompassing the days of the Alaska Gold Rush and early polar exploration [33]. The Royal Canadian Mounted Police were enforcing the law in northern territories with sled dog patrols as early as 1873, while local residents and gold prospectors relied on sled dogs for their individual transportation as well as a valuable freight system delivering mail, supplies, and passengers throughout the north [33, 34]. In 1908, the All-Alaska Sweepstakes, the first formally organized sled dog race, provided a distraction from the long dark winters, an excuse for celebration and gambling, and the opportunity for dog drivers to prove their team’s skill and win prize money [33]. While modern transportation eventually retired the working class sled dog, dog drivers transitioned them into high performance athletes competing in the sport of sled dog racing.

The Alaskan sled dog is not a formally recognized breed and is therefore not restricted in physical appearance (*i.e.*, of size, coat, and color), which is standardized in purebred breeds, or confined to a closed breeding population. However, selective breeding for athletic attributes and behavior as well as ability to perform in northern arctic climates has led to the Alaskan sled dog being informally referred to as a “breed.” Performance selection has given sled dogs commonly shared physical attributes such as a

quick, efficient gait, increased pulling strength, and superior endurance. Weight (averaging 21kg) and density and length of hair coat vary depending upon racing style, geographic location, lineages, and cross breeding to purebred lines.

Sled dog racing has progressed into two distinct styles over the past century based upon the mileage teams travel. Long distance racing covers several hundred miles over multiple days and requires endurance ability in the dog teams. Teams generally consist of 12-14 dogs, averaging speeds between 8-12 miles per hour (12-20 km/h) and carrying approximately 250 pounds (113 kg) of survival gear, including food, cook stove, sleeping bag, and ax. Two of the most well-known races are the Iditarod and Yukon Quest, both of which cover approximately 1,000 miles (1,609 kilometers) with record times taking eight and nine days respectively [35, 36].

The fundamental element of sprint or short distance racing is speed. Top sprint teams average 18-25 miles per hour (29-40km/h) with optimal snow and trail conditions. Sleds can be as light as 14 pounds (30 kg), requiring only minimal gear such as the dog-bag and snow-hook. Where distance racing is reminiscent of a marathon, sprint racing is more analogous to track events, with classes defined by the number of dogs on a team, which in turn dictates the mileage run. There are five common classes held at sprinting events: the four-dog, six-dog, eight-dog, ten-dog, and Open or Unlimited class (10 to 24 dogs). On average, the course distance at which each class competes equals a mile for each dog on the team, with Open class dogs running up to 30 miles per heat. A single heat is run per day, with the fastest combined time over 2-3 days of racing determining the winner [37]. The extreme differences in sprint and distance racing styles have led to a

divergence in the Alaskan sled dog population based on the primary physiological attributes of endurance and speed.

The ability of elite sled dogs to excel in performance while under extreme physical and mental stress has gained them public and scientific fame. Distance dogs in particular have been the focus of numerous physiological studies for their ability to traverse 1,609 kilometers in harsh environmental conditions. For instance, one study found that repetitive endurance exercise resulted in electrocardiographic changes reflecting cardiac hypertrophy [38]. Another associated an increased prevalence of gastric lesions to elite distance dogs, as is common in elite human and equine athletes [39]. Data have also suggested that only modest exercise is required to increase intestinal protein loss, while substantial exercise is required to cause alterations in the proximal gastrointestinal tract [40]. Several studies have evaluated hematologic, hormonal, and enzymatic levels in association with long distance training and racing [41-46]. In particular, enhanced endurance performance has been associated with sled dogs having higher plasma vitamin E concentrations ($>40.7 \mu\text{l}$). Dogs with higher vitamin E proved 1.9 times likelier to finish the Iditarod Sled Dog Race [47]. Numerous studies have investigated the energy expenditure of Alaskan sled dogs with corresponding nutritional work identifying methods of meeting such demands. Specifically, high fat diets have been found to increase stamina and maximize energy production, while high protein diets prevent training-induced anemia [48-50].

While the physiological and nutritional aspects of performance-related traits are well documented in sled dogs, the genetics underlying these traits have yet to be

investigated. Genetics bring us full circle to the starting point of innate performance ability as coded by an individual's DNA. An understanding of performance genetics will help explain the variations in cardiac output, hematologic, hormonal, enzymatic levels, and energy metabolism that have been documented in sled dogs. It can also lead to a better understanding of the differences in metabolic pathways among other athletic species.

The Canine Genome

The use of the Alaskan sled dog to map performance genes requires an understanding of the canine genome. The unique population structure of the canine is primarily attributed to factors of domestication and breed development. Despite the diversity among domestic breeds, they are all of one species, *Canis familiaris*, with the gray wolf as a common ancestor [7, 10, 11, 51, 52]. The relatively short time period in which purebred breeds have evolved due to intense breeding selection has accelerated the pace of both genetic selection and drift within populations. The crossing of lineages during breed development while restricting propagation to members within a breed after breed formalization has had the effect of creating substantial intrabreed homogeneity while retaining considerable interbreed heterogeneity [10]. Within-breed homogeneity results from several factors, including a small number of founders, population bottlenecks, strong selection for specific traits, and the repeated use of popular sires in breeding programs.

In 2005, a draft assembly of the Boxer dog genome was published utilizing a 7.8 fold sequence redundancy, expanding upon the 2x sequence of the standard poodle [10, 53]. With over 2.5 million putative polymorphic nucleotides determined in the canine genome, genome-wide single-nucleotide polymorphism (SNP) arrays are now a common tool used in association and linkage studies of disease, morphology, and behavior [10]. Canine SNP arrays are commercially available, with 50K to 200K SNPs spread genome-wide [54, 55]. However, this is considerably lower than the 900K SNPs found on human arrays [54, 55]. The following description of canine population structure illuminates how a relatively small number of SNPs in the canine genome is capable of successfully mapping genes in comparison to the more concentrated SNP arrays necessary for human mapping projects.

An understanding of population structure and genomic variation is integral to utilizing the canine genome and SNP arrays to their fullest potential. To this end, studies assessed average linkage disequilibrium (LD) within dogs in comparison to other species and LD variance among breeds of dog. Results demonstrated that short-range LD varied locally across the canine genome while long-range LD was approximately 100 times greater in dogs than that observed in humans. The extensive LD in dogs is comparable to that found in species such as pigs, sheep, and cattle. Both domestic farm animal and dog breed populations experience popular sire breeding which may account for the overwhelming difference in long-range LD between humans and domestic animals [56]. The variance of short- and long-range LD within dogs is attributed to the ancient bottleneck during dog domestication shared among breeds and dogs having a relatively

recent breed establishment [10]. It is also noted that the average extent of LD varies considerably between breeds ranging from the Labrador Retriever at 20Kb to the Akita at 4.5Mb. This variation can have a significant impact on the ability to localize causative mutations during genome-wide studies and therefore impacts study design when determining the appropriate breeds to utilize [56]. Along with LD, haplotype structure and homozygosity have been examined across the dog genome, which allows for improved project design. Increased homozygosity and larger blocks of haplotype structure are found among breeds of similar heritage, with a lesser degree of these elements among breeds that have experienced isolation (due to geographical location, for example) [56]. The comparative analysis of LD, haplotype structure, and homozygosity between the human and dog genomes established that only 5K to 30K markers would be needed for genome-wide association in dogs, as opposed to the several hundred thousand markers necessary for a comparable study in humans [56].

A primary goal of gene association mapping is to identify causative mutations of the phenotype in question. While SNPs are commonly found as causative mutations changing protein structure, other sources of genomic variation include copy number variation (CNV), short- and long-interspersed nucleotide elements (SINEs or LINEs), and the identification of areas of DNA slippage during replication. CNVs represent duplications or deletions of entire stretches of genomic DNA. Each of these types of variants, as well as areas of slippage, has been implicated in canine disease and morphology [10, 53, 57-60]. This demonstrates the wide variety of genetic possibilities potentially affecting individual phenotypes within dogs.

The distinctive characteristics of the dog genome, including LD, haplotype structure, homozygosity, and types of genetic variants, create a genetic environment in dogs which is highly conducive to the mapping of a variety of phenotypic and disease loci. The versatility of genomic variation among purebred breeds has been essential to many successful mapping studies [23, 61]. However, it is important to recognize that the use of Alaskan sled dogs, which are an admixed population, has particular effects on association mapping in comparison to commonly used purebred populations. Specifically, the use of an admixed population may increase the difficulty of initially identifying loci demonstrating genome-wide association. On the positive side, an admixed population would potentially narrow the region of association, more quickly focusing the area to fine-map for causative mutations. These two issues are directly related to differences in LD, haplotype structure, and heterogeneity of admixed populations as opposed to purebred populations. Chapter 2 will identify purebred breed components of Alaskan sled dogs and chapter 4 will discuss the utility of ancestry modeling to assist in the localization of performance loci.

Performance Genetics

This brings us to our phenotypic area of focus, athletic performance. Phenotyping, or defining the elite athlete, has been a particular challenge in studies of athletic performance. The majority of performance genetic research has focused on human athletes. It has encompassed an enormous array of sporting events, from weight lifting to cycling, biking, swimming, running, rowing, and mountain climbing. There can

be tremendous variation within a sport, for instance comparing speed cycling versus distance cycling events (such as the Tour de France), as well as variation between sports. A compilation of human research through 2007 cited 214 autosomal genes or quantitative trait loci, as well as seven others on the X chromosome, that demonstrated a positive performance association [62]. These positive associations are generally found when comparing a spectrum of athletes within a common sport. A review of human performance genetics highlighted a few of these genes named in multiple studies demonstrating performance association [63]. The *ACE* gene, which encodes a protein that converts angiotensin I to its active form of angiotensin II, has been repeatedly associated with athletes exhibiting endurance or speed/power [64-69]. Genes such as *ACTN3* and *MSTN*, which affect muscle structure and composition, have been associated with the trait of speed [27, 70-74]. Performance association has also been emphasized in genes important to cardiac and respiratory function as well as blood flow efficiency [63]. Other aspects relevant to the genetics of performance include the effects of race and ancestry and implications of findings towards gene therapy and gene doping [63].

Not all studies derive the same conclusion from their results. Occasionally, a more recent human performance study will contradict a previous association of a specific gene with performance attributes. Researchers speculate that contradictory results pertaining to specific genes may be due to differences in the physiological demands when comparing elite athletes from varying sports and therefore stress the importance of project design when dealing with such complex traits [75].

Sporting animals, predominantly racing horses and dogs, are another asset to the study of performance genetics. In 2009, over a hundred horse performance candidate genes located within Thoroughbred racing horses. Their approach utilized a set of 394 microsatellite-based markers and employed a population-based hitchhiking mapping approach assuming genes under selection would be near the microsatellite-based markers. They targeted regions deviating from expected heterozygosity (D_h/s_d) as well as regions having markers demonstrating high F_{ST} values when comparing four different horse breeds. The strongest areas of selection highlighted the genes *ACSS1*, *ACTA1*, *ACTN2*, *ADHFE1*, *MTFR1*, *PDK4*, and *TNC*, related to muscle composition, fatty-acid oxidation, and insulin sensitivity [76]. Within both horses and dogs, racing performance has been associated with variants of the *MSTN* gene. Thoroughbreds carrying a *C/C* SNP variant were associated with fast, short-distance races, while the *T/T* variant was associated with horses having greater stamina. Horses carrying the heterozygous *C/T* state performed the best at middle-distance races. While a more in-depth study of the *MSTN* gene was undertaken in horses, a functional variant has not yet been correlated to performance [77]. An investigation into the cause of a “double muscling” phenotype seen in whippets identified a two base-pair deletion within the *MSTN* gene. Although the homozygous state of the mutation was deleterious to racing whippets, dogs carrying a heterozygous state proved to have enhanced racing performance along with an intermediate muscle mass in comparison to either homozygous state [27]. In all, genetic studies of performance have been very limited in animals despite their being selectively bred for athletic attributes. Genetic structure due to breed management and performance

selection, as well as recent assemblies of the horse and dog genomes, make these animals prime model species for the continued refinement and understanding of performance associated genetics.

The Alaskan Sled Dog as a Model System for the Study of Performance Genetics

We now look at the suitability of the Alaskan sled dog for the study of performance genetics. Sled dogs embody a population chosen strictly for its athletic abilities. Therefore, genes contributing to performance have historically been, and currently remain, under strong selective pressure. Two primary approaches can be taken to compare the athletic attributes within sled dogs; the first focuses on the variation between sprint and distance racing dogs that have diverged over the past century based on their specialization for either speed or endurance. A second, more refined approach is the comparison of elite and poor performers within each of these distinct racing styles based on particular attributes. The utilization of a dog model allows for a smaller number of SNPs necessary for association mapping, as opposed to a human model [10, 56]. The use of the Alaskan sled dog addresses potential genetic mapping issues related to athletes performing in different sports while maximizing phenotypic variation seen in the extreme difference between sprint and distance sled dog racing. From the standpoint of a dog enthusiast, the exploration of the Alaskan sled dog also allows for a better understanding of their mixed-breed background's contribution to performance. The identification of breeds influencing sled dog performance can then be used in an admixture mapping

approach to assist in the localization of critical genetic loci associated with athletic attributes of the Alaskan sled dog.

To this end, we have sampled 340 Alaskan sled dogs from nine “high performance” racing kennels. Distance dogs were sampled from four kennels, all of which finished in the top 15% of competitors for the Yukon Quest or Iditarod races during the two consecutive years (2007-2008) of sample collection. Sprint dogs were sampled from five kennels, each placing in the top 25% of the International Sled Dog Racing Association (ISDRA) points-ranking medal program during the sampling years (2005-2007). Pedigrees, body measurements, and phenotypic ratings of speed, endurance, work ethic, heat tolerance, and mental stress tolerance were collected (Chapter 2, p 28; Chapter 4, p 103).

We have used genome-wide microsatellite and SNP panels to create genetic profiles detailing population structure, inbreeding, and ancestry modeling within Alaskan sled dogs. We used genetic signatures distinct to individual purebred breeds to identify domestic breed components of modern sled dogs. This information was then used to determine the evolutionary relationship of Alaskan sled dogs to established breeds. Variation in the proportion of these breed components was further explored based upon sled dog racing style and performance ability. We went on to sequence two genes, *MSTN* and *ACE*, both of which were previously associated to performance, in a direct approach to identify potential performance enhancing polymorphisms in Alaskan sled dogs. In a more intensive investigation, we employed genome-wide association studies and admixture mapping to highlight genetic loci associated with athletic attributes and to infer

breed contribution to such attributes. In all, a comprehensive investigation of the genetic make-up of Alaskan sled dogs was performed, illuminating distinctive characteristics of sprint and distance dogs and emphasizing genetic loci associated with performance.

References

1. Germonpre M, Sablin Mikhail V, Stevens R, Hedges REM, Hofreiter M, Stiller M, Despres V: **Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes.** *Journal of Archaeological Science* 2009, **36**:473-490.
2. Tito RY, Belknap SL, 3rd, Sobolik KD, Ingraham RC, Cleeland LM, Lewis CM, Jr.: **Brief communication: DNA from early Holocene American dog.** *American Journal of Physical Anthropology* 2011, **Aug**; **145**(4):653-657.
3. Fiedel S: **Man's best friend - mammoth's worst enemy? A speculative essay on the role of dogs in Paleoindian colonization and megafaunal extinction.** *World Archaeology* 2005, **37**(1):11-25.
4. Thurston ME: **The history of the canine race: Our 15,000-year love affair with dogs**, 1st edn. Kansas City: Andrews and McMeel; 1996.
5. Coren S: **The intelligence of dogs.** New York: Bantam Books; 1994.
6. Wilcox B, Walkowicz: **Atlas of Dog Breeds of the World**, vol. 5th Edition, 5th Edition edn. Neptune City: T.F.H. Publications, Inc; 1995.
7. Vila C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK: **Multiple and ancient origins of the domestic dog.** *Science* 1997, **276**(5319):1687-1689.
8. Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T: **Genetic evidence for an East Asian origin of domestic dogs.** *Science* 2002, **298**(5598):1610-1613.
9. Pang JF, Kluetsch C, Zou XJ, Zhang AB, Luo LY, Angleby H, Ardalan A, Ekstrom C, Skollermo A, Lundeberg J *et al*: **mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves.** *Molecular Biology and Evolution* 2009, **26**(12):2849-2864.

10. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, 3rd, Zody MC *et al*: **Genome sequence, comparative analysis and haplotype structure of the domestic dog.** *Nature* 2005, **438**(7069):803-819.
11. Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A *et al*: **Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication.** *Nature* 2010, **464**(7290):898-902.
12. **American Kennel Club- AKC.org** [<http://www.akc.org/>] Accessed Jan 2011.
13. Parker HG, Ostrander EA: **Canine genomics and genetics: running with the pack.** *PLoS Genetics* 2005, **1**(5):e58.
14. Wayne RK, Ostrander EA: **Lessons learned from the dog genome.** *Trends in Genetics* 2007, **23**(11):557-567.
15. Chase K, Jones P, Martin A, Ostrander EA, Lark KG: **Genetic mapping of fixed phenotypes: disease frequency as a breed characteristic.** *Journal of Heredity* 2009, **100 Suppl 1**:S37-41.
16. Shearin AL, Ostrander EA: **Leading the way: canine models of genomics and disease.** *Disease Models and Mechanisms* 2010, **3**(1-2):27-34.
17. Breen M, Modiano JF: **Evolutionarily conserved cytogenetic changes in hematological malignancies of dogs and humans--man and his best friend share more than companionship.** *Chromosome Research* 2008, **16**(1):145-154.
18. Khanna C, Lindblad-Toh K, Vail D, London C, Bergman P, Barber L, Breen M, Kitchell B, McNeil E, Modiano JF *et al*: **The dog as a cancer model.** *Nature Biotechnology* 2006, **24**(9):1065-1066.
19. Thomas R, Wang HJ, Tsai PC, Langford CF, Fosmire SP, Jubala CM, Getzy DM, Cutter GR, Modiano JF, Breen M: **Influence of genetic background on tumor karyotypes: evidence for breed-associated cytogenetic aberrations in canine appendicular osteosarcoma.** *Chromosome Research* 2009, **17**(3):365-377.
20. Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X, Qiu X, de Jong PJ, Nishino S, Mignot E: **The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene.** *Cell* 1999, **98**(3):365-376.
21. Sharp NJ, Kornegay JN, Van Camp SD, Herbstreith MH, Secore SL, Kettle S, Hung WY, Constantinou CD, Dykstra MJ, Roses AD *et al*: **An error in**

- dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy.** *Genomics* 1992, **13**(1):115-121.
22. Dodman NH, Karlsson EK, Moon-Fanelli A, Galdzicka M, Perloski M, Shuster L, Lindblad-Toh K, Ginns EI: **A canine chromosome 7 locus confers compulsive disorder susceptibility.** *Molecular Psychiatry* 2009, **15**(1):8-10.
 23. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, Vonholdt BM, Rhue A, Boyko A, Byers A *et al*: **Coat variation in the domestic dog is governed by variants in three genes.** *Science* 2009, **326**(5949):150-153.
 24. Chase K, Carrier DR, Adler FR, Jarvik T, Ostrander EA, Lorentzen TD, Lark KG: **Genetic basis for systems of skeletal quantitative traits: principal component analysis of the canid skeleton.** *Proceedings of the National Academy of Sciences* 2002, **99**(15):9930-9935.
 25. Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, Spady TC, Elkahouloun A, Cargill M, Jones PG *et al*: **An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs.** *Science* 2009, **325**(5943):995-998.
 26. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG *et al*: **A single *IGF1* allele is a major determinant of small size in dogs.** *Science* 2007, **316**(5821):112-115.
 27. Mosher DS, Quignon P, Bustamante CD, Sutter NB, Mellersh CS, Parker HG, Ostrander EA: **A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs.** *PLoS Genetics* 2007, **3**(5):e79.
 28. Rennick P (ed.): **Dogs of the North.** Anchorage: Alaska Geographic Society; 1987.
 29. Collins MJC: **Dog Driver: A Guide for the Serious Musher,** First edn. Crawford, CO: Alpine Publications; 1991.
 30. Vaudrin B (ed.): **Racing Alaskan Sled Dogs.** Anchorage, AK: Alaska Northwest Publishing Company; 1977.
 31. Montcombroux G: **The Canadian Inuit Dog: Canada's Heritage,** 2nd edn. Winnipeg: Whippoorwill Printing; 2002.

32. **Inuit Sled Dog International**
[<http://www.inuitsleddoginternational.com/home.html>] Accessed Jan 11, 2011.
33. Wendt R: **Alaska Dog Mushing Guide: Facts and Legends**, Fifth edn.
Fairbanks: Goldstream Publications; 1999.
34. **ISDRA; History of Racing**
[<http://www.webheads.co.uk/sdcom/press/info/003history.html>] Accessed Jan 11, 2011.
35. **The Official Site of the Iditarod** [<http://www.iditarod.com/>] Accessed Aug 2010.
36. **Yukon Quest; Sled Dog Race; Alaska, Yukon** [<http://www.yukonquest.com/>]
Accessed Aug 2010.
37. **ISDRA; International Sled Dog Racing Association** [<http://www.isdra.org/>]
Accessed Jan 2011.
38. Constable PD, Hinchcliff KW, Olson JL, Stepien RL: **Effects of endurance training on standard and signal-averaged electrocardiograms of sled dogs.**
American Journal of Veterinary Research 2000, **61**(5):582-588.
39. Davis MS, Willard MD, Nelson SL, Mandsager RE, McKiernan BS, Mansell JK, Lehenbauer TW: **Prevalence of gastric lesions in racing Alaskan sled dogs.**
Journal of Veterinary Internal Medicine 2003, **17**(3):311-314.
40. Davis M, Willard M, Williamson K, Royer C, Payton M, Steiner JM, Hinchcliff K, McKenzie E, Nelson S, Jr.: **Temporal relationship between gastrointestinal protein loss, gastric ulceration or erosion, and strenuous exercise in racing Alaskan sled dogs.**
Journal of Veterinary Internal Medicine 2006, **20**(4):835-839.
41. Davis MS, Davis WC, Ensign WY, Hinchcliff KW, Holbrook TC, Williamson KK: **Effects of training and strenuous exercise on hematologic values and peripheral blood leukocyte subsets in racing sled dogs.**
Journal American Veterinary Medical Association 2008, **232**(6):873-878.
42. Durocher LL, Hinchcliff KW, Williamson KK, McKenzie EC, Holbrook TC, Willard M, Royer CM, Davis MS: **Effect of strenuous exercise on urine concentrations of homovanillic acid, cortisol, and vanillylmandelic acid in sled dogs.**
American Journal of Veterinary Research 2007, **68**(1):107-111.
43. McKenzie E, Holbrook T, Williamson K, Royer C, Valberg S, Hinchcliff K, Jose-Cunilleras E, Nelson S, Willard M, Davis M: **Recovery of muscle glycogen**

- concentrations in sled dogs during prolonged exercise.** *Medicine & Science in Sports & Exercise* 2005, **37**(8):1307-1312.
44. McKenzie E, Lupfer C, Banse H, Hinchcliff K, Love S, Nelson S, Jr., Davis M, Payton M, Pastey M: **Hypogammaglobulinemia in racing Alaskan sled dogs.** *Journal of Veterinary Internal Medicine* 2009, **24**(1):179-184.
 45. McKenzie EC, Hinchcliff KW, Valberg SJ, Williamson KK, Payton ME, Davis MS: **Assessment of alterations in triglyceride and glycogen concentrations in muscle tissue of Alaskan sled dogs during repetitive prolonged exercise.** *American Journal of Veterinary Research* 2008, **69**(8):1097-1103.
 46. McKenzie EC, Jose-Cunilleras E, Hinchcliff KW, Holbrook TC, Royer C, Payton ME, Williamson K, Nelson S, Willard MD, Davis MS: **Serum chemistry alterations in Alaskan sled dogs during five successive days of prolonged endurance exercise.** *Journal American Veterinary Medical Association* 2007, **230**(10):1486-1492.
 47. Piercy RJ, Hinchcliff KW, Morley PS, Disilvestro RA, Reinhart GA, Nelson SL, Schmidt KE, Morrie Craig A: **Association between vitamin E and enhanced athletic performance in sled dogs.** *Medicine & Science in Sports & Exercise* 2001, **33**(5):826-833.
 48. Hill RC: **The nutritional requirements of exercising dogs.** *Journal of Nutrition* 1998, **128**(12 Suppl):2686S-2690S.
 49. Kronfeld DS: **Diet and the performance of racing sled dogs.** *Journal American Veterinary Medical Association* 1973, **162**(6):470-473.
 50. Kronfeld DS, Hammel EP, Ramberg CF, Jr., Dunlap HL, Jr.: **Hematological and metabolic responses to training in racing sled dogs fed diets containing medium, low, or zero carbohydrate.** *American Journal of Clinical Nutrition* 1977, **30**(3):419-430.
 51. Agnarsson I, Kuntner M, May-Collado LJ: **Dogs, cats, and kin: a molecular species-level phylogeny of Carnivora.** *Molecular Phylogenetics and Evolution* 2009, **54**(3):726-745.
 52. Eizirik E, Murphy WJ, Koepfli KP, Johnson WE, Dragoo JW, Wayne RK, O'Brien SJ: **Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences.** *Molecular Phylogenetics and Evolution* 2010, **56**(1):49-63.

53. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM *et al*: **The dog genome: survey sequencing and comparative analysis**. *Science* 2003, **301**(5641):1898-1903.
54. **Affymetrix** [<http://www.affymetrix.com/>] Accessed April 15, 2011.
55. **Illumina** [<http://www.illumina.com>] Accessed April 15, 2011.
56. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA: **Extensive and breed-specific linkage disequilibrium in *Canis familiaris***. *Genome Research* 2004, **14**(12):2388-2396.
57. Chen WK, Swartz JD, Rush LJ, Alvarez CE: **Mapping DNA structural variation in dogs**. *Genome Research* 2009, **19**(3):500-509.
58. Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM: **The genomic architecture of segmental duplications and associated copy number variants in dogs**. *Genome Research* 2009, **19**(3):491-499.
59. Minnick MF, Stillwell LC, Heineman JM, Stiegler GL: **A highly repetitive DNA sequence possibly unique to canids**. *Gene* 1992, **110**(2):235-238.
60. Fondon JW, 3rd, Garner HR: **Molecular origins of rapid and continuous morphological evolution**. *Proceedings of the National Academy of Sciences* 2004, **101**(52):18058-18063.
61. Parker HG, Kukekova AV, Akey DT, Goldstein O, Kirkness EF, Baysac KC, Mosher DS, Aguirre GD, Acland GM, Ostrander EA: **Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds**. *Genome Research* 2007, **17**(11):1562-1571.
62. Bray MS, Hagberg JM, Perusse L, Rankinen T, Roth SM, Wolfarth B, Bouchard C: **The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update**. *Medicine & Science in Sports & Exercise* 2009, **41**(1):35-73.
63. Ostrander EA, Huson HJ, Ostrander GK: **Genetics of athletic performance**. *Annual Review of Genomics and Human Genetics* 2009, **10**:407-429.
64. Gayagay G, Yu B, Hambly B, Boston T, Hahn A, Celermajer DS, Trent RJ: **Elite endurance athletes and the ACE I allele--the role of genes in athletic performance**. *Human Genetics* 1998, **103**(1):48-50.

65. Luciano M, Hansell NK, Lahti J, Davies G, Medland SE, Raikkonen K, Tenesa A, Widen E, McGhee KA, Palotie A *et al*: **Whole genome association scan for genetic polymorphisms influencing information processing speed**. *Biological Psychology* 2011, **86**(3):193-202.
66. Montgomery H, Clarkson P, Barnard M, Bell J, Brynes A, Dollery C, Hajnal J, Hemingway H, Mercer D, Jarman P *et al*: **Angiotensin-converting-enzyme gene insertion/deletion polymorphism and response to physical training**. *Lancet* 1999, **353**(9152):541-545.
67. Montgomery HE, Marshall R, Hemingway H, Myerson S, Clarkson P, Dollery C, Hayward M, Holliman DE, Jubb M, World M *et al*: **Human gene for physical performance**. *Nature* 1998, **393**(6682):221-222.
68. Tsianos G, Sanders J, Dhamrait S, Humphries S, Grant S, Montgomery H: **The ACE gene insertion/deletion polymorphism and elite endurance swimming**. *European Journal of Applied Physiology* 2004, **92**(3):360-362.
69. Woods DR, Montgomery HE: **Angiotensin-converting enzyme and genetics at high altitude**. *High Altitude Medicine & Biology* 2001, **2**(2):201-210.
70. Yang N, MacArthur DG, Gulbin JP, Hahn AG, Beggs AH, Eastel S, North K: **ACTN3 genotype is associated with human elite athletic performance**. *American Journal of Human Genetics* 2003, **73**(3):627-631.
71. Ahmetov, II, Druzhevskaya AM, Astratenkova IV, Popov DV, Vinogradova OL, Rogozkin VA: **The ACTN3 R577X polymorphism in Russian endurance athletes**. *British Journal of Sports Medicine* 2008, **44**(9):649-652.
72. Druzhevskaya AM, Ahmetov, II, Astratenkova IV, Rogozkin VA: **Association of the ACTN3 R577X polymorphism with power athlete status in Russians**. *European Journal of Applied Physiology* 2008, **103**(6):631-634.
73. Niemi AK, Majamaa K: **Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes**. *European Journal of Human Genetics* 2005, **13**(8):965-969.
74. Schuelke M, Wagner KR, Stolz LE, Hubner C, Riebel T, Komen W, Braun T, Tobin JF, Lee SJ: **Myostatin mutation associated with gross muscle hypertrophy in a child**. *New England Journal of Medicine* 2004, **350**(26):2682-2688.
75. Chakravarti A, Green E (eds.): **Annual Review of Genomics and Human Genetics**, 1 edn. Palo Alto, CA: Annual Reviews; 2009.

76. Gu J, Orr N, Park SD, Katz LM, Sulimova G, MacHugh DE, Hill EW: **A genome scan for positive selection in thoroughbred horses.** *PLoS One* 2009, **4**(6):e5767.
77. Hill EW, Gu J, Eivers SS, Fonseca RG, McGivney BA, Govindarajan P, Orr N, Katz LM, MacHugh DE: **A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses.** *PLoS One* 2010, **5**(1):e8645.

Chapter 2

A Genetic Dissection of Breed Composition and Performance Enhancement in the Alaskan Sled Dog¹

Abstract

Background

The Alaskan sled dog offers a rare opportunity to investigate the development of a dog breed based solely on performance, rather than appearance, thus setting the breed apart from most others. Several established breeds, many of which are recognized by the American Kennel Club (AKC), have been introduced into the sled dog population to enhance racing performance. We have used molecular methods to ascertain the constitutive breeds used to develop successful sled dog lines, and in doing so, determined the breed origins of specific performance-related behaviors.

One hundred and ninety-nine Alaskan sled dogs were genotyped using 96 microsatellite markers that span the canine genome. These data were compared to that from 141 similarly genotyped purebred dog breeds. Sled dogs were evaluated for breed composition based on a variety of performance phenotypes including speed, endurance and work ethic, and the data stratified based on population structure.

¹Huson, H.J., Parker, H.G., Runstadler, J., Ostrander, E.A., (2010). A Genetic Dissection of Breed Composition and Performance Enhancement in the Alaskan Sled Dog. *BMC Genetic* **11**:71.

Results

We observe that the Alaskan sled dog has a unique molecular signature and that the genetic profile is sufficient for identifying dogs bred for sprint versus distance. When evaluating contributions of existing breeds we find that the Alaskan Malamute and Siberian Husky contributions are associated with enhanced endurance; Pointer and Saluki are associated with enhanced speed and the Anatolian Shepherd demonstrates a positive influence on work ethic.

Conclusion

We have established a genetic breed profile for the Alaskan sled dog, identified profile variance between sprint and distance dogs, and established breeds associated with enhanced performance attributes. These data set the stage for mapping studies aimed at finding genes that are associated with athletic attributes integral to the high performing Alaskan sled dog.

Background

“Alaskan sled dogs” are a recognized population of dogs of Northern breed ancestry. They were specifically developed as working dogs to haul cargo-laden sleds across the Arctic terrain[1, 2]. They served as humans’ primary means of transportation, protection, and companionship in northern snow-dominated climates for many years. Indeed, the late 1800’s to early 1900’s was termed the “Era of the Sled Dog” due to the breed’s dominating presence in polar exploration and the boom of the Alaskan gold rush[3]. While the Alaskan sled dog experienced a decline in popularity as more modern modes of transportation became accessible in northern climates, they have recently undergone a rediscovery with the birth of sled dog racing, beginning in the late 1930’s[2, 3]. Concomitant with this rebirth has been a transition from working class dog to high performance athlete. While not recognized by the American Kennel Club as a distinct breed, consistency in behavior has led to them being informally referred to as a “breed”. The long-term goal of this study is to understand the genetic underpinnings associated with both the genetic heritage and the elite athletic performance of Alaskan sled dogs.

The Alaskan sled dog is comprised of several different lineages, optimized for different racing styles (long or short distance)[4-6], and we hypothesize that each will have a unique breed composition. Thus, we sought to identify breed composition profiles associated with expertise at specific tasks. The identification of these breeds would not only set the stage for genome wide association studies (GWAS) aimed at finding the underlying genes, but could theoretically explain why the introduction of certain breeds

enhance performance traits in the Alaskan sled dog, while others have disappeared from the genetic make-up of today's sled dog.

The Alaskan sled dog is unique in that it is not confined to a breed standard of size or appearance, as are most AKC-recognized breeds. Rather, they are a mixed breed dog, with Northern breed ancestry, currently selected for high performance in sled dog racing. This selection for athletic ability has produced dogs of a particular physique. They are known for their quick, efficient gait, pulling strength, and endurance. Weight, averaging 21kg, and density of hair coat, vary depending upon racing style, geographic location, lineages, and cross breeding to purebred lines (Figure 2.1).

Sprint and distance sled dog racing are vastly different in terms of the distance traversed during a race and the speed at which this is accomplished. Long distance racing includes races of several hundred miles lasting multiple days, such as the Yukon Quest and Iditarod of over 1,609 kilometers in the subarctic winter [7, 8]. Sprint racing is more analogous to track and field with multiple competition events defined by the size of the dog team [9]. The extreme differences in these racing styles, ranging from 48 kilometers in one day to 1,609 kilometers in less than ten days has led to a divergence within the Alaskan sled dog population based on the essential physiological athletic attributes of endurance and/or speed as well as “work ethic,” which encapsulates an animal's desire to perform. In this study we define the purebred dog profiles that have given rise to lineages of Alaskan sled dogs differing in their speed, endurance, and work ethic.

Methods

Sample Collection

One hundred and ninety-nine Alaskan sled dogs were sampled from eight “high performance” racing kennels. “High performance” sprint kennels are those whose dogs finished in the top 25% of sprint competitors in the International Sled Dog Racing Association [9] annual points standings. High performance distance kennels are those that had a primary team finish in the top 15% of all competitors for the Yukon Quest or Iditarod races during the two consecutive years that sample collection was undertaken [7, 8]. Ninety percent of Alaskan sled dogs sampled from sprint racing kennels were from open and 8-dog racing classes with the remaining 10% competing in the 6-dog class. This was done to maintain consistency in the dogs sampled in that they were being trained and raced at similar conditions of speed and distance based on their respective racing styles.

Prior to blood collection, all owners signed an informed consent document, consistent with NHGRI Animal Care and Use Committee rules. Whole blood samples were collected from the cephalic vein in 3-5ml EDTA or ACD tubes. Dogs were sampled at their home kennels. Purebred dogs were sampled at AKC-sanctioned events. Samples were stored at 4°C prior to extraction, and genomic DNA was isolated using standard proteinase K/phenol extraction methods by Health Gene (Toronto, Canada) or RX Bioscience (Rockville, MD, USA). DNA samples were stripped of identifiers, numerically coded, and aliquoted for long-term storage at -70°C. Detailed pedigrees were collected for each individual sampled and entered into our database.

In addition to the Alaskan sled dogs, we sought to expand our reference data set of purebred dogs to be used for the comparison study. Towards that end, 44 purebred dogs, representing nine AKC breeds, Japanese Chin, Tibetan Spaniel, Anatolian Shepherd, Briard, Swedish Vallhund, Yorkshire Terrier, German Pinscher, Havanese, and English Springer Spaniel, were sampled and that data added to the existing data set of 132 breeds [10]. As we have done previously, only dogs who shared no common grandparents were selected for analyses [10, 11]. Eight of the new breeds were represented by five unrelated individuals, while the German Pinscher breed was represented by three. Blood draw and sample preparation has been described previously [10, 11].

Performance Ratings

The sled dogs were rated on three aspects of performance: speed, endurance, and work ethic, using previously defined criteria specified for the distinct racing styles of sprint and distance [9]. The performance phenotypes and rating criteria were defined by one of us (H.H.) and reviewed by five professional and independent dog mushers. For both sprint and distance measures, five elite performers were selected for detailed study. In addition, we included five individuals from the other end of the spectrum who were consistently poor performers for each category. The sprint racing category however had only three, two, and four representatives available for low speed, poor endurance, and poor work ethic, respectively. Distance racers had only four individuals sampled for low work ethic. Elite and poorly performing dogs were compared to one another within

racing style. Because of the extreme distance involved, distance dogs were each scored only one time during the calendar year. Sprinting dogs were scored on a weekly basis throughout training and racing season (approximately seven months) to assess consistency. However, the score collected in April, at the end of the season during peak performance, was used for this analysis. These performance scores accounted for the dog's overall performance throughout the entire year. Weekly scores for the sprint dogs were compared to the final score given to monitor consistency.

Speed was defined as an individual dog's ability to successfully maintain the necessary speed of the team. A dog was ranked 1 if it was capable of maintaining the speed of the team, 29-40km/h for sprint dogs and 13-19km/h for distance dogs; or ranked 2 if it was unable to maintain the required speed. Speeds were based on the performance levels of the kennels represented in the study.

Endurance was broken into three ranks: dogs were capable of covering the required mileage in good or poor condition (rank 1 or 2, respectively), or they were unable to finish the required mileage (rank 3). Mileage requirements ranged from 13-48 kilometers for sprint dogs and 1,595-1,851 kilometers for the distance sled dogs and were set according to race length requirements.

The final trait, work ethic, was based on the dog's willingness to run, and was defined by the "effort" the dog displayed to pull the sled throughout the run. Effort was determined by the amount of tension a dog placed on their individual tug-line. The tug-line is the point at which the dog's harness attaches into the main line connecting the team to the sled. A three-tiered system was used in scoring (Figure 2.2). Dogs who

scored a 1 showed the strongest effort, as evidenced by having constant tug-line tension throughout the entire run. Dogs scoring a 2 had occasional tension throughout the run, but maintained the speed of the team. The poorest performers, with a score of 3, showed no tension in the tug line throughout the run, but were capable of the speed and mileage. Dogs were not penalized in their rankings due to the effects of injury.

Microsatellite Genotyping

A panel of 96 previously-described microsatellite-based markers were genotyped using DNA isolated from all dogs [11]. A clustering algorithm from the program STRUCTURE was used to both differentiate dog breeds as well as establish breed composition for the Alaskan sled dogs. Data from 132 previously genotyped breeds (5 dogs/breed) were included in this study [10, 11] as well as the nine new breeds, described in the Methods, for a total of 141 purebred breeds.

One hundred-ninety nine Alaskan sled dogs were genotyped with the same 96 microsatellite-based markers. This included 116 dogs from four sprint racing kennels and 83 dogs from four distance racing kennels. Dogs were chosen for even distribution among all eight kennels, unrelated through the grandparent generation, in order to maximize the number of lineages tested. Dogs were also selected so that approximately equal numbers of high and low performers were included.

PCR amplification of microsatellite markers was done using a protocol similar to that published previously [11], but with the following slight modifications: 1 µl of 1.0 mM dNTP's, 0.1-0.2 µl of 10 µM forward and reverse primers, 0.315-0.42 µl of 50 mM

MgCl, 0.05µl TaqGold, 1µl of 10x TaqGold buffer, 0.1µl of 10pmol/µl of M13 primer covalently linked with either 6FAM, VIC, NED, or PET fluorescent dyes (ABI), and 5ng genomic DNA. Amplification was done at 95°C for 10min, followed by 35 cycles at 94°C for 30sec, 55°C or 58°C for 30sec, and 72°C for 30sec, followed by 10min at 72°C. Samples were denatured in Hi-Di formamide with 15pmol of GeneScan-500LIZ size standard (ABI, Foster City, CA). All samples were run in either 96 or 384 well plates with a positive and negative control on an ABI 3730xl capillary electrophoresis unit. Genotypes were called using GeneMapper 4.0 (ABI). All genotyping calls were checked manually with a positive control to assure consistent allele binning.

Statistical Analysis of Population Structures

Population structure was assessed based on an allele frequency model using the program STRUCTURE at 100,000 iterations after a burn-in of 20,000 iterations[12-14]. K represents the number of populations assigned during each clustering run. Each run was repeated five times with populations being manually determined by breed membership, and then averaged over all runs. Population representatives of the purebred dogs and Alaskan sled dogs were unrelated through the grandparent generation. An optimum of five individual dogs were chosen to represent individual Alaskan sled dog lineages, however, in a few cases, only two dogs were available and both were thus used. These numbers are consistent with the number of individual dogs representing AKC breeds in the clustering analyses established by Parker et al [10, 11]. The population clustering values of each individual representative sled dog were averaged for an overall

breed composition of the specified sled dog population. Cluster analyses were run on 30 datasets of the Alaskan sled dogs and the purebred breeds to determine population representatives, breed composition profiles, and ancestral origins (Table A2.1). These datasets varied by grouping dogs from sprint and distance racing kennels, the performance rankings of the individual dogs, and the number of domestic breeds represented. One particular dataset of 42 sprint dogs who were unrelated at the grandparent generation and 42 similarly unrelated distance dogs were used for the analysis of population structure for all Alaskan sled dogs. Alaskan sled dog population representatives were determined by choosing five population members from the 84 unrelated dogs with a likelihood of ≥ 0.9 of belonging within that population. Performance ability was not a factor when determining the individuals representing each dataset used for population structure.

Breed Composition

Component breeds of the Alaskan sled dog were identified using the previously defined microsatellite-based markers [11]. One hundred forty-one purebred dog breeds were genotyped to ascertain the subset most closely related to the Alaskan sled dog. The analysis was restricted to two populations ($K = 2$), assigned based on similar allelic patterns. The data set also included 84 Alaskan sled dogs. The Alaskan sled dogs and purebred breeds of similar heritage clustered into one population. The second population consisted of purebred breeds with the most divergent allelic patterns to population one. The breeds that had a minimum of two out of five possible individuals clustered with the

Alaskan sled dogs, and had a population score of ≥ 0.3 , were utilized for the subsequent Alaskan sled dog breed composition analyses.

Inbreeding Coefficients

Inbreeding values and heterozygosity were calculated with the Genetic Data Analysis (GDA) software using the microsatellite data[15]. Similar dataset groupings were used for the inbreeding analyses as had been used for the cluster analyses. All 141 breeds were investigated with and without the Alaskan sled dogs. The Alaskan sled dogs were analyzed as a single population of 84 unrelated dogs or as two independent sub-sets of 42 unrelated sprint dogs and 42 unrelated distance dogs, as previously described. Lastly, the sub-populations of sled dogs were compared. GDA established inbreeding coefficients termed f and θ_P , which are the equivalents of F_{IS} and F_{ST} , respectively and referred to as such hence forth. F_{IS} represents the inbreeding of an individual relative to the subpopulation and F_{ST} represents the inbreeding among subpopulations relative to the total population. Sigma-G represents the variance of alleles within individuals. Expected (H_E) and observed (H_O) heterozygosity and the mean number of alleles per locus (A) were also calculated.

Results

Alaskan Sled Dog Breed Identification

Previously, Parker *et al.* showed that with few exceptions, individual purebred dogs are correctly clustered by breed in an unsupervised clustering analysis using

genotype data from just 96 microsatellite-based markers analyzed using the program STRUCTURE [10, 11]. When the allowed number of clusters is restricted, reproducible groups of breeds are formed, typically encompassing breeds of similar appearance and shared heritage.

There was a 1% marker failure rate of the microsatellite-based genotyping calls for the combined 132 breeds previously genotyped by Parker et al. and the additional nine breeds in this study [10, 11]. The Alaskan sled dogs had a slightly higher marker failure rate at 2.42% with less than a third of the total number of dogs in comparison to the domestic breeds. The highest marker failure rate within an individual Alaskan sled dog was 16%. Twenty-one percent of the domestic breeds had an individual marker failure rate greater than 16%. Only 4% of the domestic breeds had a failure rate higher than 40% (peaked at 61% marker failure seen in one individual).

We first compared the Alaskan sled dogs to 141 domestic dog breeds to determine the subset of breeds that had contributed most to the development of the sled dog. All dogs were genotyped using a set of 96 previously described microsatellite-based markers [11]. To determine the subset that were most related to the Alaskan sled dogs, we ran STRUCTURE using the parameter $K=2$, to assign two populations for the cluster analysis. This placed all sled dogs into one population and most of the domestic breeds into a second. A small subset of domestic breeds showed significant clustering with the sled dogs and were considered likely contributors to the population. These were used in subsequent breed composition analyses. Sixteen recognized breeds were identified with a population score of ≥ 0.3 within the sled dog population, which included both sprint and

distance racing dogs. We next analyzed the sprint and distance dogs separately. Five additional domestic breeds were identified when comparing just the 42 unrelated sprint dogs to the 141 domestic breeds. However, analysis of the 42 unrelated distance dogs versus the 141 breeds did not reveal any additional related purebreds. In total, then, 21 domestic breeds were identified with a population score of ≥ 0.3 within the Alaskan sled dog population and are hence forth referred to as the “related breeds” in all future analysis.

The 21 “related breeds” included the Alaskan Malamute and Siberian Husky, which were expected based on historical information, and the Pointer, which has recently and repeatedly been bred into the population[16]. The Samoyed, Chow Chow, and Akita also have historical roots as northern draft dogs[17]. Other breeds included in the “related breeds” group were the Saluki, Afghan Hound, and Borzoi, which are well known for their speed, the Great Pyrenees and the Anatolian Shepherd, both of which are northern climate guard dogs, and the Weimaraner, a hunting breed of shared ancestral heritage to the Pointer [17]. Additional related breeds were the Japanese Chin, Shar-Pei, Shiba Inu, Shih Tzu, Pekingese, Lhasa Apso, Basenji, Tibetan Spaniel, and Tibetan Terrier, most of whom share an Asian heritage with the exception of the Basenji [10].

In order to determine the breed composition of the Alaskan Sled dog, we compared their genotype data in a cluster analysis to that from the “related breeds” using STRUCTURE. Strikingly, when the number of populations allowed ($K = 21$) was equal to the number of domestic breeds in the analysis, sled dogs did not align with any specific breed, but rather defined their own breed group (Figure 2.3). Interestingly, the Alaskan

Malamute and Siberian Husky often grouped as a single breed, as did the Chow Chow and Shar-Pei. These data suggest, therefore, that the breed signature of the Alaskan sled dog is more distinct than a subset of breeds of similar heritage. Individual sled dogs ranged from 40-90% in terms of their Alaskan sled dog signature, while the remainder of each profile was a mixture of the 21 other breeds. These results establish the Alaskan sled dog as a breed, distinguishable by its genetic profile, regardless of the population's diversity in appearance and its mottled history.

Upon further analysis, the Alaskan sled dogs further separated into two clusters based solely on their racing style: sprint versus distance. This can be seen when representatives of both racing styles are analyzed with the domestic breeds (Figure 2.3) as well as when they are analyzed independently (Figure 2.4A). Five dogs, displaying the most distinctly uniform allelic profiles associated with each racing type (sprint versus distance), and referred to as the “extreme” representatives for each style, were then selected for further analysis.

Ancestral Groupings

We investigated the specific relationship between sled dog populations and purebred dog breeds for clustering based on ancestral heritage. In addition, we looked at the composition of sprint and distance dogs with regard to the five major ancestral clusters defined by Parker et al [10, 11]. An Ancient/Asian group, together with a Herding/Sight hound, the Mastiff/Terrier, Hunting and Mountain groups were previously determined as the most probable clusters from an analysis of 132 breeds [10, 11]. In this

more recent analysis of 141 breeds, both the sprint and distance populations consistently clustered within the Ancient/Asian group(Figure 2.5A). The Ancient/Asian group is the first of the populations to distinguish itself at $K=2$ when all 141 purebred breeds are analyzed. The clustering of the Alaskan sled dogs with the Ancient/Asian group may be attributed to a number of factors. However, the fact that the Alaskan Malamute and Siberian Husky are both members of the Ancient/Asian group and are the primary purebred breed components of the Alaskan sled dogs may be influential. The breed membership to this cluster also illuminates why such unlikely breeds as the Lhasa Apso, Pekingese, and Shih Tzu which are also members of the Ancient/Asian group were found within the related breeds to the Alaskan sled dog.

We identified a generalized breed composition for the sprint and distance dogs based on their membership in these ancestral source groupings (Figure 2.5B). The higher percentage of breed composition attributed to Alaskan Malamute and Siberian Husky within distance sled dogs accounts for a higher clustering value of the distance sled dogs within the Ancient/Asian group. The sprint sled dogs owe a higher portion of their group composition to the Hunting group with a small increase in variation due to the Herding/Sight hound group. In contrast, the distance sled dogs have a slight increase in the Mastiff/Terrier group.

Population Structure and Breed Composition

We next examined the population structure within the two subgroups of Alaskan sled dogs. Clustering analysis of each racing style produced four sub-populations (Figure

2.4). The four distance sub-populations were kennel-specific. Three of the four fell out as distinct populations before any of the sprint sub-populations ($K = 5$) (Figure 2.4B-4D). This suggests that the distance-associated populations are genetically more distinct from one another than are any subset of sprint dogs. The last distance population to separate from the large cluster of sprint dogs was, interestingly, the most successful kennel sampled, as defined by number of wins (Figure 2.4G). The individual dog's performance was not based on winning percentage of the kennel.

By comparison, the four sprint dog populations do not align well with kennel of origin. At $K = 6$ (Figure 2.4E) the sprint population divides into a major and minor group. At $K = 7$ and 8, two additional populations are defined. Interestingly, at $K = 8$ the distance population that was revealed last had some representation in the final sprint group. This suggests that the most successful distance dogs retain some genetic features of sprinters. We next determined which AKC recognized breeds accounted for the majority of the Alaskan sled dog signature (Figure 2.5). To do this, we evaluated the three groups as they clustered at $K = 3$ in Figure 4B; the extreme sprint dogs, extreme distance dogs, and the remaining overlapping sprint and distance sled dogs. At $K = 22$, we found that the sprint dogs had the largest signature for Alaskan Sled Dog (58%) (Figure 2.6, Column 1). They also had the largest signature for Pointer (5.9%), and the smallest signature for Alaskan Malamute (5.9%) and Siberian Husky (13.3%). By comparison, the extreme distance dogs (Figure 2.6, Column 3) had the weakest signature for Alaskan Sled dog (43.9%) and the largest signature for Alaskan Malamute (25%) and Siberian Husky (19%). They also had the smallest signature for Pointer (0.5%). As

expected, the group that overlapped sprint and distance (Figure 2.6, Column 2) had a composite profile. However, it did have the largest component of Saluki (2.9%) in comparison to that observed in the extreme sprint (2.6%) and extreme distance (1.3%) dogs.

We further refined our analysis, by analyzing the purebred breed composition of each of the eight populations defined in Figure 2.4. At $K = 22$, we observe that breed composition differences among the sub-populations highlighted specific trends (Figure 2.7&Table 2.1). The four distance sub-populations showed the greatest variation in terms of Alaskan sled dog, Alaskan Malamute, and Siberian Husky contribution. Distance sub-population one did show a slightly greater contribution from the Weimaraner than did the other distance populations, and all showed some variation in terms of Anatolian Shepherd contribution. By comparison, the sprint sub-populations all showed smaller degrees of variation across a wider range of purebred breeds. Individual sprint sub-populations displayed a particular influence of specific breeds, such as the Pointer in sub-population 1 and 3, Saluki in sub-population 2, and Borzoi in sub-population 4. The refinement of breed composition down to the sub-population level allows for a more focused interrogation of sled dog attributes acquired from specific breeds.

Inbreeding

We next examined the degree of inbreeding in the sled dog populations compared to purebred dogs. Inbreeding and heterozygosity statistics were calculated using the same 96 microsatellite-based markers described previously and the software program

GDA[15]. We observed, first, that inbreeding coefficients of F_{IS} and F_{ST} both demonstrate at least a five-fold higher inbreeding value for purebred breeds than the Alaskan sled dog sprint and distance populations combined (Table 2.2). This is further exemplified by the combined sprint and distance populations of Alaskan sled dogs having a 15% increase in allele variance (σ^2_G) and twice the number of alleles per locus (A) than the average purebred breed. Observed heterozygosity (H_o) calculations of the sprint and distance populations analyzed together or separately had a 15% increase compared to the average purebred breed. These data support the idea that Alaskan sled dogs are generally less inbred than purebred breeds (Table 2.2).

When we compared the level of inbreeding amongst sprint and distance populations we observed that the inbreeding coefficient F_{IS} was 8-fold higher for the distance population than the sprint population (Table 2.2). In Figure 2.8, we compare the individual F_{IS} values for 141 purebred breeds with the individual values for sprint and distance populations. An excess of heterozygosity within the population is represented by a more negative value of F_{IS} , whereas an excess of homozygosity is represented as a more positive value. Zero represents Hardy-Weinberg equilibrium. We observed that the sprint dog population ($-0.20197 F_{IS}$) demonstrated an extreme excess of heterozygosity in comparison to the distance population ($0.65265 F_{IS}$) and purebred breeds.

Breed Contribution to Performance

In the final analysis we sought to determine how the introduction of various purebred breeds could enhance performance attributes. To accomplish this, each sled dog was rated on their individual skills with regard to speed, endurance, and work ethic using a set of defined criteria specified for the different racing styles (See Methods).

Approximately five elite and poor performance representatives for each athletic attribute were analyzed in a cluster analysis with the “related breeds” to identify breed components associated with extremes of performance. The breed composition of elite and poor performance sled dogs was then compared to identify specific breeds that may enhance a performance phenotype (Table 2.3).

The most striking observation was that strong performers in all categories and of both racing types showed a comparative increase in the Alaskan sled dog genetic signature. This was particularly illustrated by a 25% increase in the Alaskan sled dog signature seen in high performers of speed for distance dogs and a 26% and 38% increase in endurance and work ethic respectively for sprint dogs. In addition, endurance, which is obviously important in distance dogs, showed the highest increase of any AKC breed signature with an 11% increase in both Alaskan Malamute and Siberian Husky. This implies a significant role for sustaining the genetic integrity of the Alaskan Malamute and Siberian Husky in the distance sled dogs superior endurance performance.

A small number of breeds contributed disproportionately to elite performance of both racing types. Specifically, Saluki and Anatolian Shepherd had a 3% increase in their

breed contribution for dogs exhibiting better speed in sprint and distance racing styles, respectively, compared to the 0% or negative effect observed otherwise. Also, the Pointer, which has been bred into sprint sled dogs in recent years with the idea of increasing speed, did not correlate with any athletic attributes. Finally, Anatolian Shepherd displayed a 6% increase for distance dogs with a high level of work ethic.

Discussion

The Alaskan sled dog presents a case in which a genetically distinct breed of dog has been developed through the selection and breeding of individuals based solely on their athletic prowess. The creation of the Alaskan sled dog breed happened without the implementation of breed standards of size and appearance, or the closing of the breeding population to only those individuals deemed representative of the breed, as is the norm for AKC-recognized purebred breeds. We observed that inbreeding values were five-fold lower in the Alaskan sled dog population than the average purebred domestic breed. In addition, they demonstrated an excess of heterozygosity. These observations likely reflect the continual out-crossing for athletic enhancement that is common among Alaskan sled dogs. Interestingly, the process still led to the Alaskan sled dog repeatedly producing its own unique genetic signature. Indeed, the Alaskan sled dog breed proved to be more genetically distinct than breeds of similar heritage such as the Alaskan Malamute and Siberian Husky. Thus, we conclude that the breeding practices used to produce dogs of optimal performance have created a distinct breed of dog that developed

using different criteria then are commonly used in the development and propagation of purebred dogs.

Cluster analysis demonstrated that when the Alaskan sled dog population was compared to a large data set of purebred breeds they separated into two groups that align with their racing style, sprint versus distance. The same results were evident when the sled dog population was analyzed independently, without the purebred breeds. These two racing styles have diverged over the past 100 years [3] with athletic emphasis on either speed or endurance, as appropriate to the extreme differences in race length (sprint-48 kilometers, distance-1,609 kilometers). Unsupervised clustering analysis of the Alaskan sled dog population showed that 21% of a subset of 42 unrelated dogs from competitive sprint kennels grouped within the distance sled dog population. However, not a single individual from a distance kennel grouped within the sprint sled dog population. We speculate that a fraction of the dogs competitive in sprint racing are genetically capable of performing as distance dogs, but the reverse is not true.

In a clustering analysis with 141 purebred domestic breeds, both the sprint and distance populations consistently clustered within what has been termed the Ancient/Asian group of the five major ancestral clusters [10]. This supports the theory that at least a subset of early domesticated dogs originated in Asia and migrated with human nomadic tribes across the Bering Strait into North America [11, 18-21]. Identification of compositional differences between the sprint and distance dogs with regard to the five ancestral clusters showed major differences. For instance, the Ancient/Asian group makes up 56% of the sprint and 66% of the distance group

respectively. By comparison, the Hunting group comprises 23% of the sprint and 11% of the distance group. Finally the Mastiff/Terrier group also shows a two-fold difference with 6% associated with sprint while 12% comprise the distance population. We hypothesize that breeds from the Hunting group may enhance speed in sprint dogs and that breeds from the Mastiff group may contribute to the body stature and musculature necessary for the successful distance dogs.

Eight sub-populations, four sprint and four distance, were discovered within the Alaskan sled dogs. The distance populations were more genetically distinct, kennel specific, and had a higher average inbreeding value than the sprint populations. Interestingly, the final distance population to separate itself from the sprint populations is the most successful kennel in terms of race wins during the study period. The average F_{IS} score of the sprint populations demonstrated the highest degree of heterozygosity when compared with the distance populations and 141 purebred breeds. We speculate that breeding programs are more confined within distance kennels while sprint kennel breeding programs cross between kennels. The inbreeding values support a higher level of gene flow among the sprint populations as well as a higher level of gene flow within the entire Alaskan sled dog population compared to the average domestic breed.

The dominant genetic signature found in all Alaskan sled dogs was that of the Alaskan sled dog breed. The extreme sprint population had a 14% higher average Alaskan sled dog signature and 5.5% higher Pointer breed signature than those of the extreme distance population. This suggests that the Pointer and the Alaskan sled dog

signature itself play a major role in the enhancement of speed, which is the primary athletic attribute of sprint dogs. The Alaskan Malamute and Siberian Husky breed signatures were 19% and 6% higher respectively in the extreme distance dogs compared to the sprint dogs. This in turn, suggests that the Alaskan Malamute and Siberian Husky enhance endurance. Given their heavier build and stamina for cold temperatures this is not surprising.

A more detailed investigation of breed composition among the eight Alaskan sled dog populations revealed the particular influence of specific breeds. The four distance subpopulations differed primarily in terms of the contribution of the Alaskan sled dog, Alaskan Malamute, and Siberian Husky. In addition, Weimaraner and Anatolian Shepherd contributed to a few select sub-populations. Sprint sub-populations followed a different pattern with their primary breed component being Alaskan sled dog and their remaining composition consisting of smaller degrees of contributions from a wider range of purebred breeds. Individual sprint populations displayed a strong contribution from the Pointer, Saluki, Borzoi, and Weimaraner breeds.

The most compelling discoveries in the investigation of how various breeds may enhance performance attributes were seen with a 25%, 26%, and 38% increase in Alaskan sled dog breed for speed in distance dogs, and endurance and work ethic in sprint dogs, respectively. The highest contribution of a purebred was an 11% increase in both the Alaskan Malamute and Siberian Husky breed signatures for elite endurance performance in distance dogs as opposed to distance dogs who are known to exhibit poorer endurance abilities. This implies that strong genetic selection for Alaskan

Malamute and Siberian Husky contributions has contributed to the elite performing distance dogs in terms of their endurance. The Saluki and Anatolian Shepherd both had minor positive influences for exceptional speed performance in sprint and distance populations, respectively. Unexpectedly, the Anatolian Shepherd, whose heritage describes a large, powerful, and independent northern livestock guardian dog, demonstrated a 6% positive influence in distance sled dogs of high work ethic. The Anatolian Shepherd's breed description of loyalty, independence, and hardiness portray a breed of similar character and strength to that of the distance sled dogs [17].

Interestingly, the Pointer, which has been repeatedly bred into the sprint dogs in recent years with the idea of enhancing speed, was not found to positively affect speed performance. One hypothesis is that the integration of Pointers with the sled dog population has not successfully enhanced speed performance. However, a comparison of four high profile sprint races' finish times in 1998 versus 2007 all showed faster completion times in 2007 [22]. There was a much lower degree of Pointer ancestry in the sled dog population in 1998, as compared to pedigree analysis of dogs in 2007 [23].

Another hypothesis that could explain the lack of significant Pointer contribution to speed may reflect a lack of significant difference in speed measures among the most and least successful sprint dogs. It may be that the contribution of the Pointer for speed is only evident when comparing sprint versus distance dogs, or if more extreme representatives of elite versus less successful sprint dogs were assayed.

Conclusions

We have identified component breeds of Alaskan sled dogs and defined specific breeds that have influenced the athletics of the dogs in terms of speed, endurance, and work ethic. It should be noted, however, that the crossing of purebred dogs into the Alaskan sled dog population does not guarantee performance enhancement in one generation. Rather, several generations of repeated selection and breeding of elite performers may be needed to permit optimal athletic performance to be obtained, as alleles of multiple genes are likely to play a role.

These experiments set the stage for genomic exploration into the phenotypes associated with successful sprint and distance performing Alaskan sled dogs. Whole genome association studies, currently underway, are likely to reveal genetic loci introduced by the various breeds that enhance each attribute of athletic performance. Such studies are optimal for situations where there are a small number genes of major effect. We know little about the starting stock of Alaskan sled dogs and the number of major and or minor genes that contribute to these phenotypes. Current genomic tools, however, permit us to ask such questions, and our long term studies aim to paint a complete picture of the genetics of canine athletic performance, providing a complete palate of genes and alleles that contribute to these complex phenotypes.

Authors' Contributions

HJH generated the hypothesis and led the study design phase together with JR and EAO. Sample collection was conducted by HJH and JR. EAO and JR served as primary graduate mentors for HJH. Genotyping was completed by HJH for wet-lab portions. Statistical analysis was directed by HGP in collaboration with HJH. All authors were involved in scientific discussion of the project. HJH and EAO drafted the manuscript with all authors providing comments and final approval.

Authors' Information

HJH has 25 years of professional sled dog sprint racing and breeding experience.

Acknowledgements

We acknowledge the INBRE grant 5P20RR016466 from NCRR and we thank the Intramural program of the National Human Genome Research Institute for their support. We also thank Kenneth and Lori Chezik, Dr. Dawn Brown, Greg Sellentin, and Deborah McGrath for sharing their expertise and review of the performance phenotype rating criteria. We acknowledge Greg Sellentin for his contribution of dog photos. We appreciate collection assistance from Danielle Dillon, Keiko Herrick, Lori Gildehaus, Ian Herriott, and the kennel owners and handlers and we gratefully acknowledge the sled dog

owners who have generously contributed performance ratings and canine DNA samples to our study.

References

1. Collins MJC: **Dog Driver: A Guide for the Serious Musher**, First edn. Crawford, CO: Alpine Publications; 1991.
2. Rennick P (ed.): **Dogs of the North**. Anchorage: Alaska Geographic Society; 1987.
3. Wendt R: **Alaska Dog Mushing Guide: Facts and Legends**, Fifth edn. Fairbanks: Goldstream Publications; 1999.
4. Welch J: **The Speed Mushing Manual: How To Train Racing Sled Dogs**, First edn. Eagle River, AK: Sirius Publishing 1989.
5. Attla GBL: **Everything I Know About Training and Racing Sled Dogs**, 1st Edition edn. Hayward: Typart Publishing Service; 1972.
6. Swenson R: **The Secrets of Long Distance Training and Racing**, 1 edn. Wasilla: L&B Color Printing; 1987.
7. **Yukon Quest; Sled Dog Race; Alaska, Yukon** [<http://www.yukonquest.com/>] Accessed Aug 2010.
8. **The Official Site of the Iditarod** [<http://www.iditarod.com/>] Accessed Aug 2010.
9. **ISDRA; International Sled Dog Racing Association** [<http://www.isdra.org/>] Accessed Jan 2011.
10. Parker HG, Kukekova AV, Akey DT, Goldstein O, Kirkness EF, Baysac KC, Mosher DS, Aguirre GD, Acland GM, Ostrander EA: **Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds**. *Genome Res* 2007, **17**(11):1562-1571.
11. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L: **Genetic structure of the purebred domestic dog**. *Science* 2004, **304**(5674):1160-1164.

12. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**(2):945-959.
13. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations.** *American Journal of Human Genetics* 2000, **67**(1):170-181.
14. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies.** *Genetics* 2003, **164**(4):1567-1587.
15. **GDA; Genetic Data Analysis Software**
[<http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php>] Accessed Mar 2010.
16. **Dogtec.com Web Kennel System** [<http://www.dogtec.com/>] Accessed Jan 2010.
17. **American Kennel Club- AKC.org** [<http://www.akc.org/>] Accessed Jan 2011.
18. Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillen S, Vila C: **Ancient DNA evidence for Old World origin of New World dogs.** *Science* 2002, **298**(5598):1613-1616.
19. Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T: **Genetic evidence for an East Asian origin of domestic dogs.** *Science* 2002, **298**(5598):1610-1613.
20. Vila C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK: **Multiple and ancient origins of the domestic dog.** *Science* 1997, **276**(5319):1687-1689.
21. Sablin Mikhail V, Khlopachev Gennady A: **The Earliest Ice Age Dogs: Evidence from Eliseevichi 1.** *Current Anthropology* 2002, **43**(5):795-799.
22. **Sled Dog Central** [<http://www.sleddogcentral.com/>] Accessed Aug 2010.
23. **Sled Dog Studio: Pedigree Books**
[<http://www.sleddogstudio.com/pedigree/index.html>] Accessed May 2010.

Figures

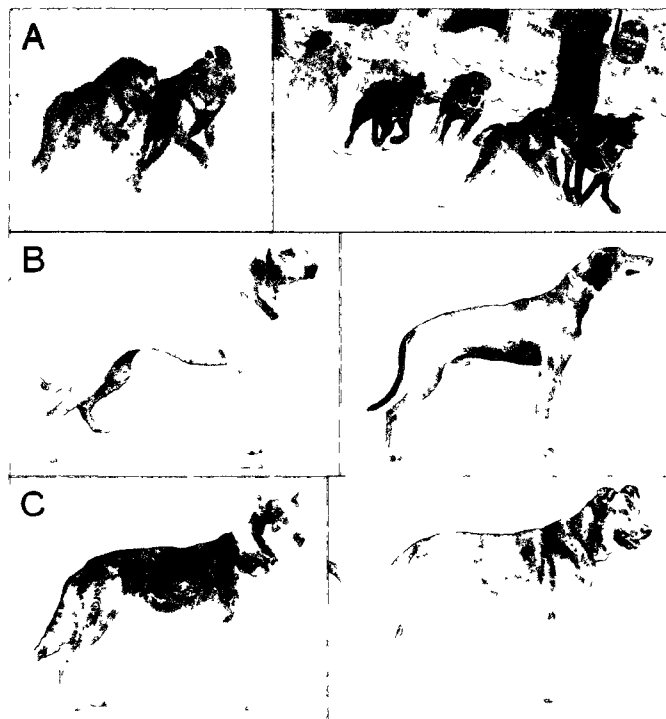


Figure 2.1: Alaskan sled dogs are a mixed breed dog selected strictly for their racing performance.

A) Top row: Sprint racing teams of “traditional” Alaskan sled dogs (no purebred crossings in the last 5 generations) and Pointer crossed Alaskan sled dogs. Spandex dog coats (in blue) are commonly used on shorter-haired Pointer x Alaskan sled dogs when temperatures are $\leq 10^{\circ}\text{F}$. B) Middle row: Sprint sled dogs of 25% or greater Pointer ancestry according to their written pedigree records. C) Bottom row: “Traditional” Alaskan sled dogs from distance racing teams. All photos were taken between 2006-2009 of dogs competitively racing in high performance kennels.



Figure 2.2: Work ethic was scored on a three-tiered system based on the dog's willingness to run.

The effort a dog put forth during a run was determined by the amount of tension a dog placed on their individual tug-line. The tug-line, indicated with yellow, is the line attaching the dog's harness into the main line connecting the dog team to the sled. Dogs demonstrating the strongest effort, defined by having a constant tug-line tension throughout the run, were designated as rank 1 (top line). Rank 2 (middle line) defined dogs that had intermittent tug-line tension throughout the run, but maintained the speed of the team. The poorest performers, rank 3 (bottom line), showed no tug-line tension during the run but were capable of the speed and mileage. Dogs were not penalized due to the effects of injury.

Figure 2.3: Population structure of purebred dogs and Alaskan sled dogs.

The cluster analysis of 21 purebred breeds (Parker *et al*, 2004; Parker *et al*, 2007) and two Alaskan sled dog populations grouped by the racing style (sprint or distance). For each breed we utilized DNA samples from five individuals who were unrelated at the grandparent level. Individuals grouped into breed-specific clusters, denoted as differing colors on Figure 2, based on the percentage of their allelic pattern belonging to the specific cluster. The two Alaskan sled dog groups created their own populations based on their unique genetic signature of microsatellite-based markers.

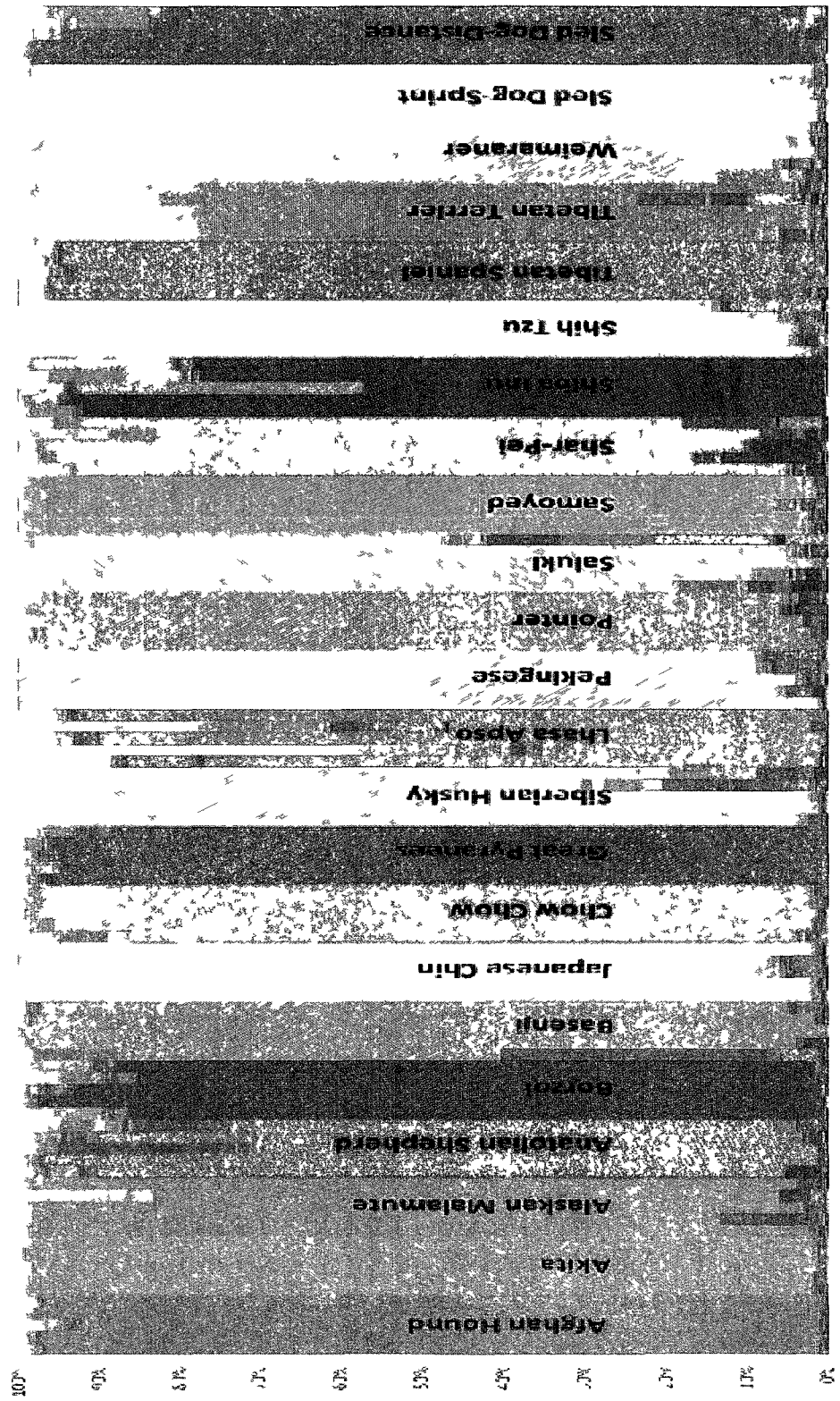


Figure 2.4: Population structure of sprint and distance sled dogs during successive increase in assigned population numbers.

The population structure of 84 unrelated Alaskan sled dogs of even distribution between four sprint and four distance kennels. The 42 Alaskan sled dogs from the sprint kennels are on the left side of the figure and the 42 Alaskan sled dogs from the distance kennels are on the right side of the figure. Each population is designated by a different color in the chart. Individuals are categorized based on the percentage of their allelic pattern belonging to each of the populations. Figures 4A-G show a successive increase in the assigned number of populations from $K=2$ through $K=8$. In total, eight sub-populations, four in sprint dogs and four in distance dogs, were documented from the sampled Alaskan sled dogs.

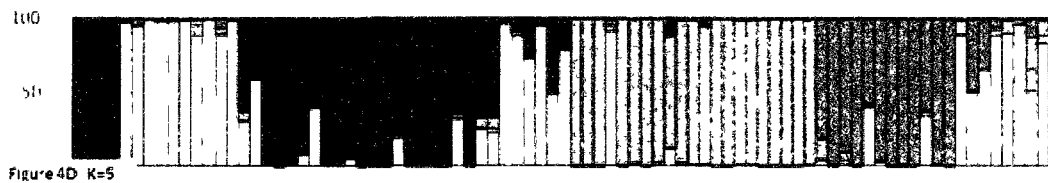
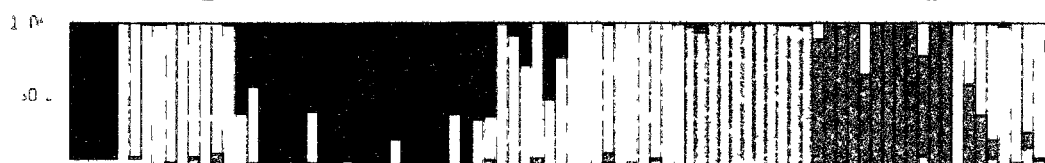
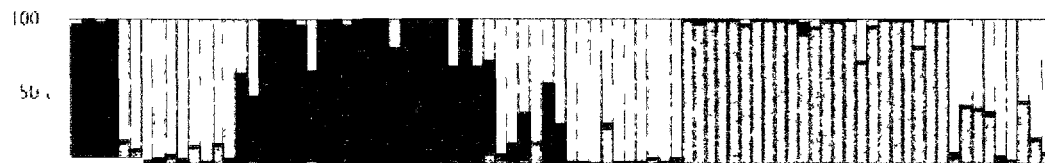
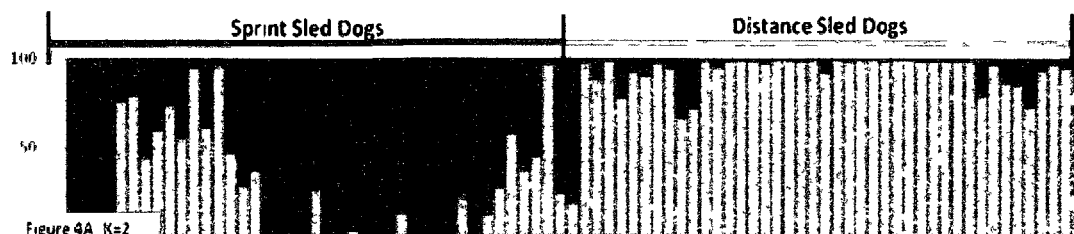
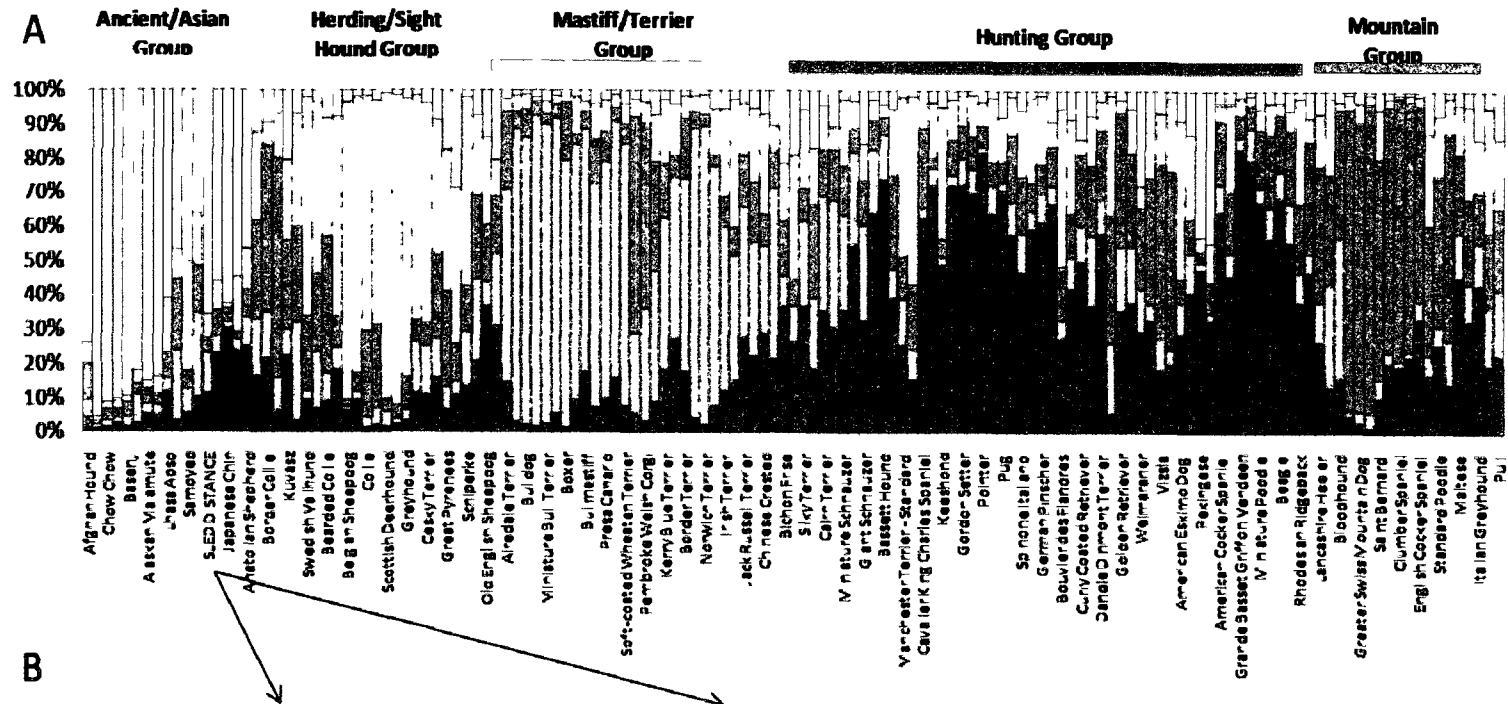
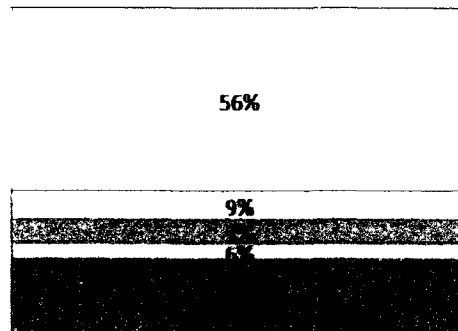


Figure 2.5: Ancestral breed clustering and composition differences of the sprint and distance Alaskan sled dogs.

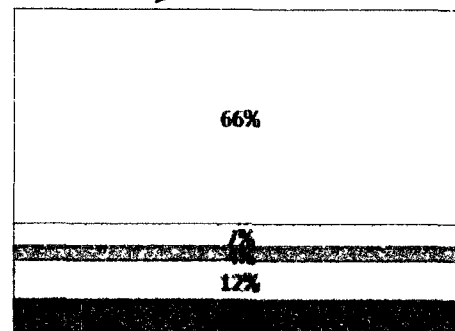
A). The five ancestral breed clusters signified by distinct colors as determined by Parker et al in 2004, 2007. One hundred forty-one purebred breeds (132 breeds from Parker et al with 9 new breeds) and sprint and distance populations of Alaskan sled dogs are represented in the clusters. Both Alaskan sled dog populations repeatedly cluster among the Ancient/Asian Group (yellow). B). The sprint and distance sled dog populations were investigated for ancestral group composition differences. We observe an increase in the Hunting Group (red) contribution among the sprint dogs and an increase in the Mastiff/Terrier Group (blue) contribution among distance dogs.



B



SPRINT SLED DOG



DISTANCE SLED DOG

- ☐ Ancient/Asian Group
- ☐ Herding/Sighthound Group
- ☒ Mountain Group
- ☐ Mastiff/Terrier Group
- ☒ Hunting Group

Figure 2.6: Breed composition of Alaskan sled dogs reflected by three populations based on racing style.

The three populations represented extreme sprint, extreme distance, and a 3rd overlapping population of sprint and distance sled dogs. The Alaskan sled dogs are assigned to three populations based on clustering analysis of microsatellite-based markers that are used to establish breed composition of each group. The percentage of each breed is denoted by a different color. The left most group is comprised of ten individuals representative of the “extreme” sprint sled dogs; the right most group is comprised of ten individuals representative of the “extreme” distance sled dogs; and the middle group is comprised of the remaining ten sprint and ten distance sled dogs which cluster together. There is an overall trend for increased Alaskan sled dog, Pointer, and Saluki signature in sprint sled dogs and an increase in Alaskan Malamute and Siberian Husky signature seen in distance sled dogs.

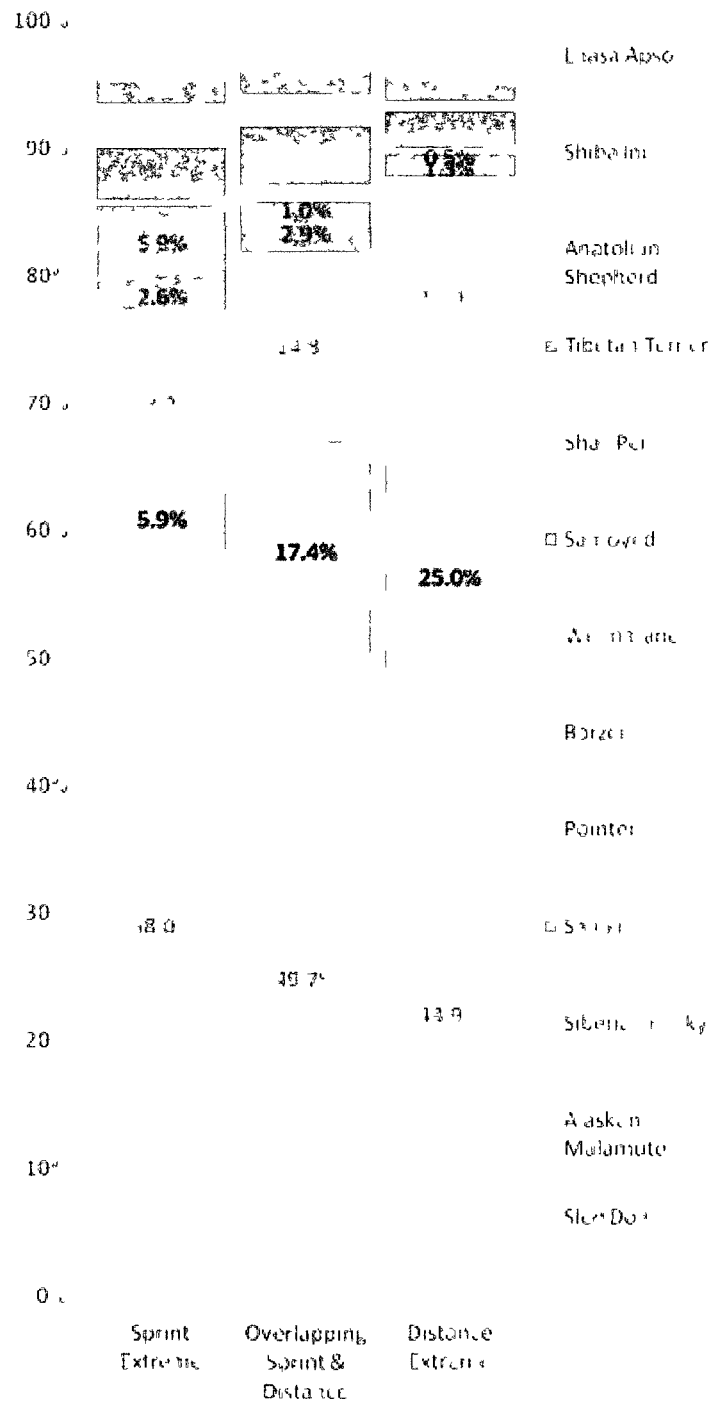
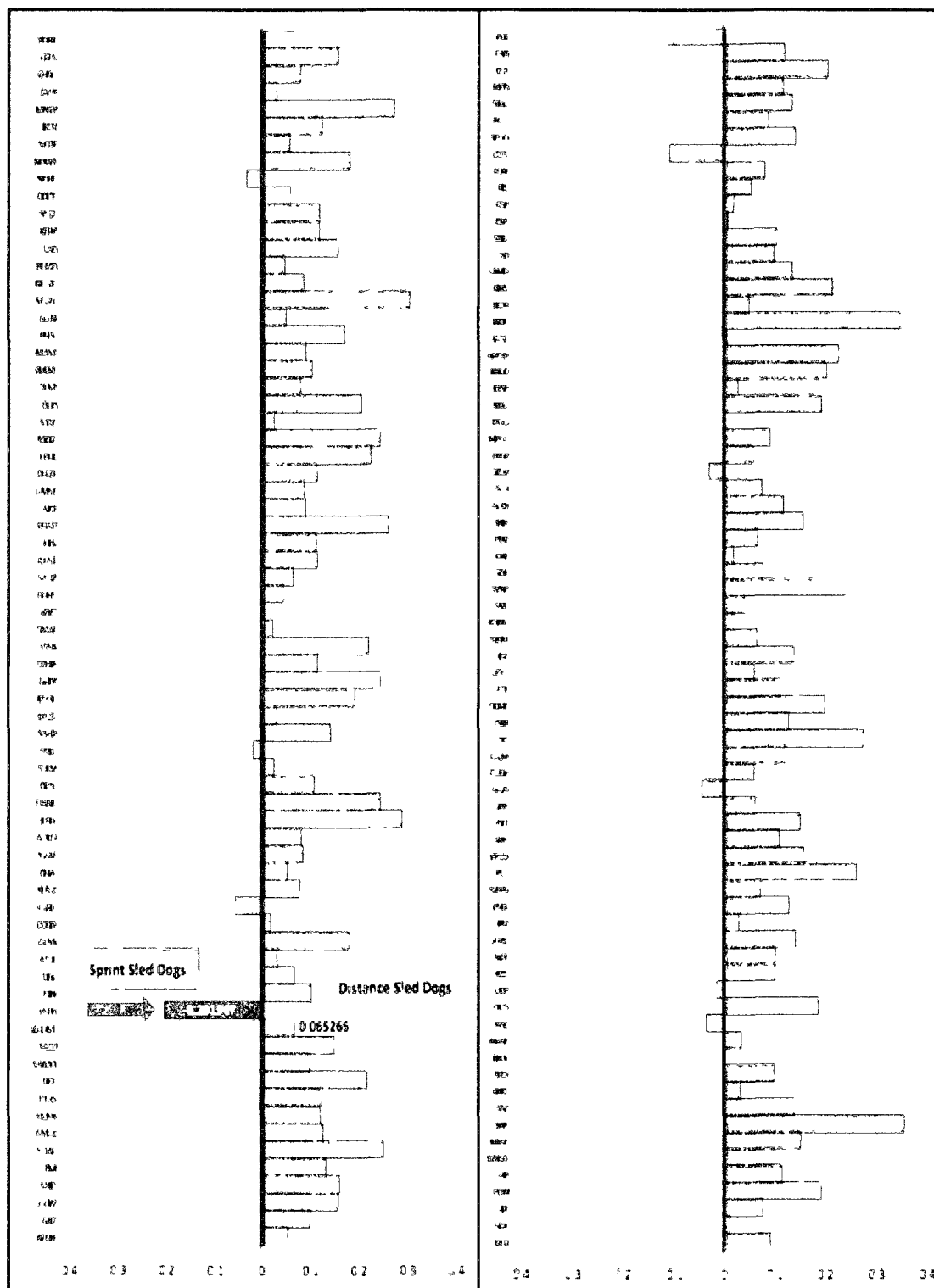


Figure 2.7: Breed composition differences between Alaskan sled dog sub-populations.

Four sprint sub-populations and four distance sub-populations were identified through cluster analysis (Figure 4G). These sub-populations were analyzed for breed composition differences as depicted. Each component breed is represented by a unique color (which corresponds to colors used in Figure 6). The sprint sub-populations illustrate the greatest differences in the Saluki, Pointer, and Weimaraner.

Figure 2.8: The degree of inbreeding F_{IS} within breed populations.

The degree of inbreeding, F_{IS} , as determined for 141 purebred breeds (681 individuals, ~5 individuals per breed) and two Alaskan sled dog sprint and distance racing populations (five sprint and five distance individuals). The 141 breeds are listed in alphabetical order from top to bottom and are viewed in two identically scaled panels side by side. The vertical red line indicates the midpoint in Hardy Weinberg equilibrium, with inbreeding values ranging from -1 (below the red line) to +1 (above the red line). The more negative values indicate an excess of heterozygosity while the more positive values represent an excess of homozygosity. Each blue bar represents the degree of inbreeding for a single breed. The purple bar signifies the sprint dogs, and the green bar indicates the distance dogs. We observe an excess of heterozygosity within the sprint sled dogs.



Tables

Table 2.1: The percentage breed composition of four Alaskan sled dog sprint sub-populations and four Alaskan sled dog distance sub-populations.

	Sled Dog	AlaskanMalamute	SiberianHusky	Saluki	Pointer	Borzoi	Weimaraner	Samoyed	AfghanHound	Shar-Pei	TibetanTerrier	AnatolianShepherd
Sprint Sub-Populations ^A												
Pop 1	59%	3%	11%	3%	7% ^B	0%	2%	4%	1%	3%	1%	2%
Pop 2	55%	15%	4%	6% ^C	1%	0%	4% ^D	1%	1%	3%	2%	1%
Pop 3	55%	8%	15%	2%	4% ^B	1%	1%	1%	0%	4%	2%	2%
Pop 4	50%	11%	15%	2%	2%	5% ^E	3%	1%	1%	4%	2%	0%
Distance Sub-Populations ^A												
Pop 1	53%	18%	15%	1%	0%	0%	4% ^D	1%	0%	1%	1%	3% ^F
Pop 2	55%	20%	15%	1%	0%	0%	1%	2%	0%	1%	1%	2%
Pop 3	35%	30%	23%	1%	0%	0%	2%	1%	0%	1%	2%	4% ^F
Pop 4	41%	24%	23%	3%	0%	0%	2%	2%	0%	2%	2%	0%

- A. The breed composition of each sub-population is the average breed composition of the five most representative members of the sub-population based on haplotype pattern.
- B. Sprint sub-populations 1 and 3 have the highest percentage of Pointer breed composition.

- C. Sprint sub-population 2 has the highest percentage of Saluki breed composition.
- D. Sprint sub-population 2 and Distance sub-population 1 have the highest Weimaraner breed composition.
- E. Sprint sub-population 4 has the highest percentage of Borzoi breed composition.
- F. Distance sub-populations 1 and 4 have the highest percentage of Anatolian Shepherd breed composition.

Table 2.2: Sled dogs were 5x lower in inbreeding and 15% higher in observed heterozygosity than purebreds.

Groupings	F_{IS}^A	F_{ST}^B	Sigma-G ^C	A ^D	He ^E	Ho ^F
Purebred Breeds Only^G	0.1085	0.2538	41.8949	2.6064	0.4729	0.4299
Purebred/Sprint/Distance^H	0.1055	0.2534	42.0894	2.6100	0.4738	0.4318
All Sprint & Distance^I	0.0170	0.0514	57.5737	5.6927	0.6090	0.5962
All Sprint	-0.0034					
All Distance	0.0456					
Unrelated Sprint & Distance^J	0.0292	0.0424	57.5158	5.3646	0.6162	0.5985
Unrelated Sprint	0.0061					
Unrelated Distance	0.0526					
8 Alaskan Sub-Populations^K	-0.0554	0.1349	56.2292	3.0156	0.5581	0.5849
4 Sprint Sub-Populations^L	-0.0729	0.1082	58.7893	3.1016	0.5749	0.6111
4 Distance Sub-Populations^M	-0.0361	0.1422	53.7001	2.9297	0.5413	0.5586

A. F_{IS} is the degree of inbreeding within populations (correlation of alleles within individuals within one sub-population).

B. F_{ST} is the overall inbreeding coefficient (correlation of alleles of different individuals in the same population).

C. Sigma-G is the variance of alleles within individuals.

D. A is the mean number of alleles per locus.

E. He is the expected heterozygosity of a population.

- F. H_o is the observed heterozygosity of a population.
- G. Data set included 141 purebred breed populations (681 individual dogs).
- H. Data set included 141 purebred breed populations and 2 Alaskan sled dog populations using the 5 most representative individuals of the sprint racing and distance racing styles (10 total Alaskan sled dogs).
- I. Data set included all 199 Alaskan sled dogs with microsatellite data. Relatedness not accounted for.
- J. Data set included 42 unrelated to the grand-parent generation Alaskan sled dog representatives of each racing style (sprint versus distance) (84 total Alaskan sled dogs).
- K. Data set included the 5 most representative individuals of each sub-population found within the sprint and distance racing style (40 Alaskan sled dogs).
- L. Data set included the 5 most representative individuals of each of the 4 sprint sub-populations (20 sprint Alaskan sled dogs).
- M. Data set included the 5 most representative individuals of each of the 4 distance sub-populations (20 distance Alaskan sled dogs).

Table 2.3: The percentage change in Alaskan sled dog breed composition between high and low performing individuals.

Performance Phenotype	Racing Style	Sled Dog	Alaskan Malamute	Siberian Husky	Saluki	Pointer	Weimaraner	Samoyed	Anatolian Shepherd
Speed	Sprint ^A	5% ^B	-6%	-3%	3% ^C	-3%	1%	0%	0%
Speed	Distance ^A	25% ^B	-15%	-10%	-6%	-2%	1%	0%	3% ^C
Endurance	Sprint ^A	26% ^B	-10%	-7%	0%	-9%	-2%	0%	0%
Endurance	Distance ^A	-15%	11% ^D	11% ^D	2%	0%	0%	2%	0%
Work Ethic	Sprint ^A	38% ^B	-23%	-17%	-6%	-6%	-6%	2%	0%
Work Ethic	Distance ^A	11% ^B	-13%	-13%	0%	0%	0%	0%	6% ^E

- A. The average breed composition of the five most representative dogs within the given race style for each athletic attribute.
- B. There is an overall trend of increased Alaskan sled dog signature in higher performing dogs of all athletic phenotypes.
- C. Saluki and Anatolian Shepherd show slight elevation for the speed phenotype.
- D. Alaskan Malamute and Siberian Husky show an increase in representation within distance sled dogs for high endurance performance.
- E. The Anatolian Shepherd is increased for the enhancement of the behavioral trait of work ethic in distance sled dogs.

Appendix

Table 2A.1:STRUCTURE run input information including the population groupings, the number of populations designated, and the result objective.

Description: Thirty datasets were investigated using the software program STRUCTURE in unsupervised cluster analyses. The datasets were categorized based on the purebred breeds and Alaskan sled dog populations utilized for exploring the ancestral origins, breed composition, and population structure of the Alaskan sled dogs.

STRUCTURE Software Settings			
Sled Dog Groupings	Purebred Groupings	# of Populations, K value	Outcome
42 Unrelated Sprint dogs	None	2-6	4 Sprint Sub-populations, 5 individual dog representatives of sub-populations
42 Unrelated Distance dogs	None	2-6	4 Distance Sub-populations, 5 individual dog representatives of sub-populations
84 Unrelated Sprint & Distance dogs	None	2-10	8 Sub-populations, 5 individual dog representatives of the extreme sprint and distance racing styles
42 Unrelated Sprint dogs	141 Purebred Breeds	2	AKC breeds with 30% marker similarity to Alaskan sled dogs
42 Unrelated Distance dogs	141 Purebred Breeds	2	AKC breeds with 30% marker similarity to Alaskan sled dogs
84 Unrelated Sprint & Distance dogs	141 Purebred Breeds	2	AKC breeds with 30% marker similarity to Alaskan sled dogs
5 Extreme Sprint and 5 Extreme Distance dogs	141 AKC Breeds	5-7	Evolutionary clustering of breeds
5 Extreme Sprint and 5 Extreme Distance dogs	Related Breeds	22-23	Breed composition differences between extreme sprint & distance dogs
First ranked representative dogs from all 4 sprint sub-populations (4 individuals)	Related Breeds	22	Averaged with other sub-population representatives to define breed composition of each sprint sub-population
Second ranked representative dogs from all 4 sprint sub-populations (4 individuals)	Related Breeds	22	
Third ranked representative dogs from all 4 sprint sub-populations (4 individuals)	Related Breeds	22	
Fourth ranked representative dogs from all 4 sprint sub-populations (4 individuals)	Related Breeds	22	
Fifth ranked representative dogs from all 4 sprint sub-populations (4 individuals)	Related Breeds	22	
First ranked representative dogs from all 4 distance sub-populations (4 individuals)	Related Breeds	22	Averaged with other sub-population representatives to define breed composition of each distance sub-population
Second ranked representative dogs from all 4 distance sub-populations (4 individuals)	Related Breeds	22	
Third ranked representative dogs from all 4 distance sub-populations (4 individuals)	Related Breeds	22	
Fourth ranked representative dogs from all 4 distance sub-populations (4 individuals)	Related Breeds	22	
Fifth ranked representative dogs from all 4 distance sub-populations (4 individuals)	Related Breeds	22	
5 Best speed performers in sprint racing	Related Breeds	22	Breed composition of top speed performers for sprint racing
5 Worst speed performers in sprint racing	Related Breeds	22	Breed composition of worst speed performers for distance racing
5 Best speed performers in distance racing	Related Breeds	22	Breed composition of top speed performers for sprint racing
5 Worst speed performers in distance racing	Related Breeds	22	Breed composition of worst speed performers for distance racing
5 Best endurance performers in sprint racing	Related Breeds	22	Breed composition of top endurance performers for sprint racing
5 Worst endurance performers in sprint racing	Related Breeds	22	Breed composition of worst endurance performers for distance racing
5 Best endurance performers in distance racing	Related Breeds	22	Breed composition of top endurance performers for sprint racing
5 Worst endurance performers in distance racing	Related Breeds	22	Breed composition of worst endurance performers for distance racing
5 Best work ethic performers in sprint racing	Related Breeds	22	Breed composition of top work ethic performers for sprint racing
5 Worst work ethic performers in sprint racing	Related Breeds	22	Breed composition of worst work ethic performers for distance racing
5 Best work ethic performers in distance racing	Related Breeds	22	Breed composition of top work ethic performers for sprint racing
5 Worst work ethic performers in distance racing	Related Breeds	22	Breed composition of worst work ethic performers for distance racing

Chapter 3

A SNP within the Angiotensin-Converting Enzyme Distinguishes between Sprint and Distance Performing Alaskan Sled Dogs in a Candidate Gene Analysis¹

Abstract

The Alaskan sled dog offers a unique mechanism for studying the genetics of elite athletic performance. They are a group of mixed breed dogs, comprised of multiple common breeds, and a unique breed entity seen only as a part of the sled dog mix. Alaskan sled dogs are divided into two primary groups as determined by their racing skills. Distance dogs are capable of running over 1,609 kilometers in ten days, while sprint dogs run much shorter distances, approximately 48 kilometers, but in faster times, i.e. 29-40 km/h. Finding the genes that distinguish these two types of performers is likely to illuminate genetic contributors to human athletic performance. In this study we tested for association between polymorphisms in two candidate genes: *angiotensin-converting enzyme (ACE)* and *myostatin (MSTN)* and enhanced speed and endurance performance in one hundred seventy-four Alaskan sled dogs. We observed eighty-one novel genetic variants within the *ACE* gene and four within the *MSTN* gene, including a polymorphism

¹Huson, H.J., Byers, A.M., Runstadler, J., Ostrander, E.A., (2011). A SNP within the Angiotensin-Converting Enzyme Distinguishes between Sprint and Distance Performing Alaskan Sled Dogs in a Candidate Gene Analysis. *Journal of Heredity in press*

within the *ACE* gene that significantly ($P\text{-value} = 2.38 \times 10^{-5}$) distinguished the sprint versus distance populations.

Introduction

Alaskan sled dogs are a population of dogs with Northern breed ancestry originally developed as a primary means of human transportation over snow-covered terrain [1, 2]. While they were an integral part of human habitation in northern climates in the late 1800's to early 1900's, sled dogs eventually experienced a decline in popularity as modern modes of transportation came into common use. However, in the 1930's there was a resurgence in Alaskan sled dog popularity, with the launching of competitive sled dog racing and the transition of the Alaskan sled dog into a high performance athlete [2, 3].

The sport has diverged over the past century into two distinct racing styles: sprint, or short distance racing, and the popular long distance events [4-7] which cover several hundred miles over multiple days such as the Iditarod and Yukon Quest [8, 9]. Sprint racing is more analogous to short distance track events with multiple competition events defined, in this case, by the size of the dog team. The status of elite endurance and sprinting athlete has established the Alaskan sled dog as a new system for exploring the genetics of performance enhancing polymorphisms (PEP) [10].

There are no limits regarding the size or appearance of Alaskan sled dogs as they are not an established breed recognized by any registering body such as the American

Kennel Club (AKC) in the United States. Dogs are selected and bred solely on their athletic ability and have a unique physique that is capable of a quick, efficient gait, pulling strength, and endurance. Their weight, averaging 21 kg, and the density of their coat varies, depending on racing style, location, and lineage.

We have shown previously, that as a population, Alaskan sled dogs are a mix of Northern breeds including the Alaskan Malamute and Siberian Husky, together with other AKC-recognized breeds which were introduced to enhance different racing aspects of performance such speed or endurance [11]. Our previous study showed, further, that the sprint versus distance populations could be genetically distinguished following analysis of a panel of 96 microsatellite based-markers. Although there are likely multiple genetic differences between the sprint and distance populations, in this study we sought to look for allelic differences in the two populations with regard to just two candidate genes: *ACE* and *MSTN*.

Variants within the *ACE* gene were among the first PEPs found in humans [12] and *ACE* variants have been widely studied in the context of elite athletes, particularly high altitude mountaineers [13]. *ACE* is part of the rennin-angiotensin system and is responsible for degradation of the vasodilator bradykinin, regulation of inflammatory reactions in the lung, respiratory drive, erythropoiesis, tissue oxygenation, and the regulation of skeletal muscle efficiency [14, 15]. The most common PEP associated with the human *ACE* gene is the I/D polymorphism, a 287bp intronicindel. The I allele is associated with lower serum and tissue *ACE* activity and improved performance in sports requiring high levels of endurance, such as marathon running [16, 17]. The I allele is

believed to facilitate the maximization of oxidative fuel for metabolism [18].

Conversely, the D allele is associated with higher serum and tissue activity and superior performance in sports requiring short bursts of power [14]. The D allele is also associated with greater increase in left ventricular mass, higher VO_2max , and greater strength gain in response to training [19].

We also investigated the role of the *MSTN* gene, for which we have previously demonstrated the presence of deletion mutations that are, in turn, associated with increased racing speed in whippet dogs [20]. Dogs heterozygous for the mutation exhibited a more muscular phenotype and consistently excelled in competition with faster race times than dogs homozygous for the wild genotype. Dogs carrying two copies of the deleterious mutation are heavily muscled [20], and the resultant phenotype is colloquially termed “double muscling”, [21]. Individuals homozygous for various *MSTN* mutations have been reported in mice [22], cattle [23, 24], sheep [25], and humans [26], all of whom share similar phenotypes.

The *ACE* and *MSTN* genes were chosen for investigation in Alaskan sled dogs due to their previous association with endurance or speed enhancement respectively. The sprint and distance populations of Alaskan sled dogs have diverged over the past decades due to selection of dogs’ speed or endurance capabilities respective to the different performance requirements of the two racing styles. Therefore we investigated whether any corresponding relationships existed between the candidate genes and population differentiation or population performance enhancement.

Methods

Sample Collection

A total of one hundred and seventy-four Alaskan sled dogs were sampled from eight “high performance” racing kennels. Four sprint kennels were deemed high performers by their points ranking which placed them within the top 25% of sprint or short distance sled dog as recorded by the International Sled Dog Racing Association during sampling years (2005-2007). Ninety percent of the sprint racing Alaskan sled dogs were from open (10 or more dogs in a team) and 8-dog racing classes, with the remaining 10% competing in the 6-dog class. All dogs were conditioned at similar increasing mileage and speed throughout the training and racing season, which extends for approximately seven months. This allowed for consistency in sample collection for the relative speeds and distance. The other four kennels were deemed “high performance” distance kennels because they finished in the top 15% of competitors for the Yukon Quest or Iditarod races during the two consecutive years (2007-2008) that sample collection was undertaken[8, 9]. In addition to the Alaskan sled dogs, eighty purebred dogs from eight breeds including the Alaskan Malamute, Siberian Husky, Greyhound, Whippet, Mastiff, Staffordshire Bull Terrier, German Shorthaired Pointer, and English Pointer were included in the *ACE* gene study only. These breeds were selected based on either our previous study which demonstrated that they contributed to the makeup of the modern Alaskan sled dog (Alaskan Malamute, Siberian Husky, English and German Shorthaired Pointer) or because their athletic attributes made them

reasonable candidates to consider for contribution. To maximize diversity, we selected dogs from the same breed that were unrelated at the grandparent level or further removed.

Prior to blood collection, all owners signed an informed consent document, consistent with NHGRI Animal Care and Use Committee rules. Whole blood samples were collected from the cephalic vein in 3-5ml EDTA or ACD tubes. Sled dogs were sampled at their home kennels. Purebred dogs were sampled at AKC-sanctioned events. Samples were stored at 4°C prior to extraction, and genomic DNA was isolated using standard proteinase K/phenol extraction methods by Health Gene (Toronto, Canada) or RX Bioscience (Rockville, MD). DNA samples were stripped of identifiers, numerically coded, and aliquoted for long-term storage at -70°C. Detailed pedigrees were collected for each individual sampled and entered into an anonymous database.

Performance Ratings

Sled dogs were rated in terms of both speed and endurance with respect to the distinct styles of sprint and distance[7-9]. Sprint dogs were reviewed for these criteria by competing at 29-40 km/h for sixteen – forty-eight kilometers while distance dogs were rated at standards of 13-19 km/h over 1,609 kilometers. The performance phenotypes and rating criteria were defined by one of us (H.H.) and reviewed by five professional and independent dog mushers.

Speed was defined as an individual dog's ability to run at a specific rate of kilometers per hour that the team is traveling. A dog was ranked one if it was capable of maintaining the speed of the team during each run; 29-40 km/h for sprint dogs and 13-19

km/h for distance dogs; or ranked two if it was unable to maintain the required speed. Speed requirements were based on the performance levels of the kennels represented in the study.

Endurance was measured by assigning dogs to one of three ranks. Dogs who achieved Rank 1 covered the required mileage in good condition, while Rank 2 dogs completed the required mileage but struggled to do so. Rank 3 dogs were unable to finish the required mileage. Mileage requirements ranged from 13-48 kilometers for sprint dogs and 1,595-1,850 kilometers for the distance dogs, and were set according to race length requirements.

Sample Selection for ACE Gene Analysis

DNA from 20 Alaskan sled dogs (10 sprint and 10 distance) and eighty purebred dogs was initially sequenced using Sanger methodology for 99% of the *ACE* gene. Ten individuals unrelated within three generations and belonging to each group of sprint, distance, and the eight domestic breeds were used to represent the distinct populations. The Alaskan sled dogs ranked elite for their speed and endurance within their respective racing populations. The region sequenced spanned approximately 20 Kb on canine chromosome 9 (CFA9) and included 48 overlapping amplicons, averaging 700bp in length. Amplicons covered all 28 exons, the associated introns, and putative flanking regions. Sequence data were analyzed for polymorphisms as described below and a total of 81 polymorphisms were found. The 10 distance and 10 sprint dogs were then compared at all polymorphisms to test for population-associated differences in the *ACE*

gene sequence. In addition, each polymorphism was tested for a putative association with endurance and speed. In separate analysis, we compared allele distribution and frequency for each marker in the set of 80 purebred dogs versus the sprint and distance populations. Sixty-three additional Alaskan sled dogs, 24 elite distance and 39 elite sprint, were selected for subsequent genotyping of markers showing statistically significant differences in allele frequency between sprint and distance dogs.

Sample Selection for MSTN Gene Analysis

The *MSTN* gene was sequenced using Sanger methodology using DNA from 91 sprint Alaskan sled dogs and two whippets, the latter of which served as controls who carried a previously reported two base pair deletion in exon 3 [20]. The canine *MSTN* gene spans approximately five Kb on CFA37. All exons and non-coding regions were sequenced, as well as the flanking regions of the gene except for a 1,039 bp GC-rich region in intron one. Sequencing was done using 12 overlapping amplicons averaging 700bp in length. Dogs sequenced included 46 sprint dogs ranked as elite performers and 37 sprint dogs ranked as poor performers. An additional eight sled dogs, without performance rating measurements, were also sequenced.

DNA Amplification and Sequencing

PCR amplification for both genes was performed in a 10 μ l volume containing 20ng genomic DNA, 1.8 μ l GC melt, 1 μ l of 10X TaqGold buffer, 0.1 μ l of TaqGold (Applied Biosystems, www.appliedbiosystems.com), 1 μ l of 1mM dNTP's, 0.3 μ l of

50mM MgCl₂, 1μl of both forward and reverse 3μM primers, and 1.8μl water.

Touchdown PCR was carried out as follows: 95°C for 7 min, followed by 20 cycles of 94°C for 30 sec, then decreasing by 0.5°C per cycle starting at 65°C down to 55°C for annealing for 30 sec, followed by 20 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, and a final extension of 72°C for 7 min.

PCR products were sequenced using Big Dye version 3.1 on an ABI 3730x1 capillary electrophoresis unit (Applied Biosystems). Sequence reads were aligned and analyzed using Sequencher 4.8 software (Gene Codes, <http://www.genecodes.com/>). Single nucleotide polymorphisms (SNP) and indels were identified by manual comparison to the available canine reference sequences using UCSC Gene Browser (<http://ucsc.genome.edu>). Polymorphisms were numerically labeled as individual markers.

Statistical Analysis for the ACEG Gene

Association tests using SNPs, indels, or haplotypes, and permutation testing for association significance along with linkage disequilibrium (LD) plots were completed using Haploview 4.1 software (<http://www.broadinstitute.org/mpg/haploview>). All analyses used a case/control format and conducted pairwise comparisons between all markers with the exclusion of individuals with greater than 50% missing genotypes. Permutation tests were run for 10,000 cycles and performed separately for the *MSTN* and *ACE* genotypes with cases and controls set as described below.

All 81 *ACE* gene polymorphisms found by comparing DNA sequence from the sprint, distance and purebred dogs were used in association testing in an attempt to identify markers that distinguished sprint versus distance dogs. Sixty-three additional Alaskan sled dogs (39 elite sprint, 24 elite distance) were genotyped for three markers which showed a statistically significant difference ($p \leq 0.005$; permutation $p \leq 0.05$) between the 10 sprint and 10 distance dogs. All of the Alaskan sled dog genotype data was combined for a total of 49 elite sprint dogs and 34 elite distance dogs which were reanalyzed for population association at the statistically significant SNPs.

To investigate whether the three *ACE* markers found to differentiate between the sprint and distance dog populations were also associated with endurance or sprinting/power, the same markers were also genotyped in the 80 domestic dogs. It was critical to first develop a scheme that would ensure that we were testing each marker for association to performance attributes rather than breed differences. Thus, we first compared allele frequency and distribution between the 49 sprint and 34 distance dogs and obtained p-values. We then independently assigned the eight pure breeds to four sets of two pairs, with each pair representing an athletic attribute. For speed we paired Greyhound/Whippet, and German Shorthaired Pointer/English Pointer. For endurance/strength we paired Alaskan Malamute/Siberian Husky and Mastiff/Staffordshire Bull Terrier. We then analyzed each of the three markers for each group of 20 dogs assigned to a pairing versus the other 60, and obtained p-values. For example, allele frequencies and distribution were compared between the 20 Greyhound/Whippet (10 each breed), set as cases, in comparison to the 60 controls from

the Alaskan Malamute, Siberian Husky, German Shorthaired Pointer, English Pointer, Mastiff, and Staffordshire Bull Terrier. Markers with raw p-values of ≤ 0.005 in **both** sled dog only analysis and the purebred dog analysis were deemed to be potentially associated with an athletic attribute.

Linkage disequilibrium was investigated using the entire panel of polymorphic markers found in the *ACE* gene during the initial sequence analysis of the 20 Alaskan sled dogs and the 80 purebred dogs. LD plots were analyzed under the four gamete rule for the Alaskan sled dogs and the domestic breed pairs used in the association tests. This allowed us to test whether Alaskan sled dogs had LD patterns similar to any of the pure breeds.

Statistical Analysis for the MSTN Gene

To investigate *MSTN* marker association with speed performance, we were able to use a simpler scheme because we had samples from both elite and poor sprint performers. We thus compared genotypes at four markers from 46 elite performing sprint sled dogs (cases) versus 37 poor performers (controls) using Haploview 4.1 software. In addition, we genotyped sequence reads from 77 purebred dogs and a wild canid (*Canis aureus*) sequenced previously by Mosher et al. [20] for the four *MSTN* SNPs. This allowed us to determine if any variation seen between elite and poor performing sled dogs was also observed in domestic dog breeds. The 16 domestic dog breeds sequenced by Mosher et al. included an average of 4 dogs per breed (Afghan Hound, Akita, Australian Shepherd, Basenji, Beagle, Boxer, Bull Terrier, Flat-Coated Retriever, German Wirehaired Pointer,

Greyhound, Ibizan Hound, Italian Greyhound, Mastiff, Pharaoh Hound, Saluki, Siberian Husky, and Whippet) and a Golden Jackal [20].

Results

ACE Gene

We hypothesized that the canine *ACE* gene might play a role in differentiating between the sprint and distance populations of Alaskan sled dogs. We also wished to determine if any SNPs in the *ACE* gene were associated with speed versus endurance performance. Thus, the coding sequences, introns, and putative regulatory flanking regions of the *ACE* gene were initially sequenced in a set of 20 Alaskan sled dogs of which 10 were sprint and 10 were distance dogs. In addition, a set of 80 purebred dogs representing 10 dogs of each of eight breeds (Alaskan Malamute, Siberian Husky, Staffordshire Bull Terrier, Mastiff, Whippet, Greyhound, German Shorthaired Pointer, and English Pointer) were also fully sequenced. Purebred dogs were selected because they had either been shown to contribute to the genetic makeup of Alaskan sled dogs, or because of their performance attributes. Analysis of the sequence reads from the combined dataset of 100 dogs identified eighty-one polymorphisms. Two alleles, a G at Marker 60 and a C at Marker 75 were unique to Alaskan sled dogs. Seventeen polymorphisms were present only in the domestic breeds of which a T at Marker 28 was specific to the German Shorthaired Pointer breed (Supplementary Table 3.1).

Of the 81 variants, only four were located in exons. One was in an untranslated region of exon 28 (marker 80), two resulted in synonymous changes (markers 23 & 82) in

exons 8 and 7 respectively, and a fourth, marker 46, caused a non-synonymous change in exon 17. Both alleles of marker 80 were found in sprint and distance dogs and all of the domestic breeds, with both homozygotes and heterozygotes present. Marker 23 caused an A to G change from the canonical sequence, which did not change the encoded amino acid glycine. Both alleles were found in the sled dog and purebred dog population. Marker 46 caused a C to T alteration that changed a threonine to methionine at position 1,4632,203bp. It was observed in both sprint and distance dogs, as well as purebreds. Finally, marker 82 was homozygous for the T allele in all Alaskan sled dogs and all purebred dogs tested, as opposed to the reported C allele in the boxer reference sequence (NW876331.1) (Table A3.1).

Although a number of variants were found in non-coding regions, two (markers 24 and 56) were distinguished by the fact that they contained an allele distinct from that reported in the boxer reference sequence (NW876331.1). In the case of marker 24, we observed dogs that had homozygous and heterozygous deletions of the C allele at nucleotide 14627332 bp in the sprint, distance, and purebred populations. Frequencies were 0.400, 0.444, and 0.500 respectively (Table A3.1). In the case of marker 56, the 20 sled dogs showed the G allele only present in the sprint dogs at a frequency of 0.35 (Table 3.1). Sequencing of an additional 63 sled dogs, described below for performance association, found the G allele in both the sprint and distance dogs with a frequency of 0.305 and 0.031, respectively (Table 3.2). Interestingly, of the eight domestic breeds tested, only the Alaskan Malamute and Siberian Husky breeds carried the G allele with a frequency of 0.265 with the 20 dogs combined.

We next compared linkage disequilibrium (LD) between the Alaskan sled dogs and the purebred dogs. Sprint and distance dogs were considered separately and LD plots produced (Figure 3.1A and 3.1B). We used the same groupings of the purebred dogs that we had developed for analysis of the *ACE* gene. Thus, closely related breeds were paired: Alaskan Malamute/Siberian Husky, Whippet/Greyhound, Mastiff/Staffordshire Bull Terrier, and German Shorthaired Pointer/English Pointer and LD plots were produced for those four combinations (Figure 3.1C, 3.1D, 3.1E and 3.1F). The LD plots demonstrate, first, that the sprint and distance dogs contain a substantial amount of LD in the *ACE* gene, but that it differs between the two populations. There is considerably less LD in each of the four purebred pairs. None of the four purebred pairs is reminiscent of the patterns observed in the Alaskan sled dogs, highlighting, again, the uniqueness of the breed.

We next wanted to determine if any of the markers could be used to distinguish between sprint and distance populations of the Alaskan sled dogs. Using Haploview(<http://www.broadinstitute.org/mpg/haploview>), both SNP and association tests were used to evaluate all 81 markers for association with either the sprint or distance population. In the initial analysis ten sprint and ten distance dogs were compared at all markers. We observed that three; 42, 56, and 74, located at 14631037 bp, 14635693 bp, 14637900 bp, respectively, demonstrated p-values lower than ≤ 0.005 (permutation p-values ≤ 0.05) with variances in allele frequencies separating elite sprint and distance sled dogs (Table 3.1).

Also, we were interested in how the sprint and distance populations related to the purebred dog populations with regard to the three markers mentioned above. We hypothesize that since all the markers found were in the *ACE* gene, and they easily distinguished the sprint and distance populations, they might highlight specific performance patterns associated with the purebred dogs. We first compared all sets of two purebred breeds against all other purebreds in order to determine if there were markers within the *ACE* gene that distinguished, specifically, purebred breeds associated with speed (greyhound/whippet) versus endurance (Malamute/Husky). Marker 56 which generates a single-base-pair change from A/G at 14635693 bp within intron 19 (A allele in reference sequence NW876331.1), was the only marker to have a significant p-value (7.57×10^{-8}), for any purebred pair, in this case when we compared the Alaskan Malamute/Siberian Husky pair to all other purebreds (Table 3.2).

This was a particularly interesting marker as the previous analysis demonstrated that marker 56 was one of three which was useful for distinguishing sprint versus distance dogs ($p = 0.0036$, Table 3.1). That analysis, however only involved 10 dogs of each type. We expanded the analysis to include 63 additional for a total of 83 Alaskan sled dogs (49 sprint and 34 distance) and obtained a p-value of 2.38×10^{-5} (Table 3.2).

We did however observe significant difference in minor allele frequencies (MAF) between the Alaskan Malamute/Siberian Husky group compared to the Alaskan sled dog distance population. Specifically, we observed a MAF (G allele) of 0.265 in the Malamute/Husky group, while all other purebred dogs were homozygotes for the A allele. In addition, the G allele was found at a frequency of 0.305 in the sprint sled dogs

and 0.031 for the distance dogs. Alleles were in Hardy Weinberg equilibrium with respect to the sprint and distance populations. Had the marker been functional with respect to the performance aspects of the distance versus sprint populations, we would have expected that the distance dogs would carry the G allele much more frequently, more analogous to the Malamute/Husky group, from whom they presumably “inherit” a significant portion of their endurance, and a much lower frequency of the G allele, with respect to the sprint dogs. The fact that we observe the opposite suggests that while marker 56 is useful for distinguishing populations with the Alaskan sled dog sprint and distance groups as a population level, the marker is not a hallmark of any putative contribution the *ACE* gene may be making to performance.

MSTN Gene

We sequenced the coding region, introns, and putative regulatory regions flanking the *MSTN* gene in 91 sprint dogs. No obvious deleterious mutations were found including the two base pair deletion at nucleotide 939-940, which we have previously reported in racing whippets [20]. Four polymorphisms were found in non-coding regions of the *MSTN* gene during the analysis of the sled dog sequence reads. One polymorphism in intron 2, a four bp indel at nucleotide 3731257 (marker 01), was observed in 24 dogs. Also, a T/A SNP, downstream of the last coding exon at nucleotide 3720985 bp (marker 02), was found in 23 dogs. Two other polymorphisms were found upstream of the 5' end of the gene, an A to G change at 3736327 bp (marker 03) which was observed in 11 dogs

and an A insertion at 3739468 bp (marker 04) which was observed in 73 dogs (Table 3.3).

The four SNPs were then analyzed in a panel of 77 purebred dogs and a wild canid (*Canis aureus*) previously genotyped by Mosher et al. to determine if the allele was truly unique to the sprint dog population. There was an average of four dogs representing each of 16 breeds (Afghan Hound, Akita, Australian Shepherd, Basenji, Beagle, Boxer, Bull Terrier, Flat-Coated Retriever, German Wirehaired Pointer, Greyhound, Ibizan Hound, Italian Greyhound, Mastiff, Pharaoh Hound, Saluki, Siberian Husky, and Whippet) and a Golden Jackal [20]. We observed all polymorphisms in the domestic breeds, thus demonstrating that they are not unique to the sprint sled dogs. Marker 03 was deemed poor quality due to less than 50% genotype call rate in both the sled dogs and the purebred dogs. MAF showed a difference of ≤ 0.095 between the sled dogs and the purebred panel for the three remaining markers (Table 3.3).

SNP association tests were performed to identify whether any of the four markers were in association to sprint sled dog performance. Haploview 4.1 software found that neither the *MSTN* gene nor surrounding markers had a significant p-value (raw $p \leq 0.005$) that would have suggested an association with either sled dogs performing poorly in speed or being ranked elite in their speed performance.

Discussion

The Alaskan sled dog provides researchers with a unique system in which to study the genetics of athletic performance. In this study we focus on understanding

specific genes that are candidates for distinguishing the sprint and distance populations of Alaskan sled dogs along with being potentially influential in athletic performance. We targeted two genes; *angiotensin-converting enzyme (ACE)* and *myostatin (MSTN)* and tested for association between gene polymorphisms and both sled dog population differentiation and athletic attributes such as endurance, speed and power. Novel genetic variants were found within both genes. Four *MSTN* gene polymorphisms were found in the screening of 91 sprint dogs and confirmed in a panel of 77 purebred dogs from 16 breeds: Afghan Hound, Akita, Australian Shepherd, Basenji, Beagle, Boxer, Bull Terrier, Flat-Coated Retriever, German Wirehaired Pointer, Greyhound, Ibizan Hound, Italian Greyhound, Mastiff, Pharaoh Hound, Saluki, Siberian Husky, and Whippet and a Golden Jackal. However none of the *MSTN* variants showed association with any traits of interest.

Eighty-one polymorphic markers were identified in the *ACE* gene through sequence analysis of 100 dogs, 10 each from the sprint sled dogs, distance sled dogs, Alaskan Malamute, Siberian Husky, Whippet, Greyhound, German Shorthaired Pointer, English Pointer, Mastiff, and Staffordshire Bull Terrier. Variation in *ACE* gene markers 60 and 75 was only found within the Alaskan sled dogs.

We previously demonstrated that a panel of 96 genome wide microsatellite-based markers successfully differentiates Alaskan sled dogs into two populations based on their racing style of sprint (30 miles at 18-25mph) or distance (1,000 miles at 8-12mph) [11]. Here, we successfully identified three individual *ACE* gene markers (markers 42, 56, and

74) with a raw p-value of ≤ 0.005 (permutation p-value ≤ 0.05) that differentiated between the sprint and distance sled dog populations.

We hypothesized that these three *ACE* gene markers had the potential to be associated with athletic attributes such as endurance or speed/power exhibited respectively by the distance and sprint populations. Separate analysis of allele frequency and distribution within the pure breed pairings established marker 56 as having significant G allele association (p-value of 7.57×10^{-8}) between the Alaskan Malamute/Siberian Husky pairing and the other six domestic breeds. An expanded analysis of the Alaskan sled dogs to include 49 sprint and 34 distance dogs improved the p-value from 0.0036 (10 sprint and 10 distance) to 2.38×10^{-5} for marker 56 (Table 3.2). However, we observed a significant difference in the MAF (G allele) between the Alaskan Malamute/Siberian Husky group at 0.265 compared to the Alaskan sled dog distance population at 0.031. In contrast, the MAF (G allele) of the sprint dogs at 0.305 was more analogous to the Alaskan Malamute/Siberian Husky breed pairing. In a previous study by our group, the distance dogs showed a 25% higher degree of Alaskan Malamute and Siberian Husky in their total breed composition than the sprint dogs. We also found an 11% increase in these two breeds when comparing high and low endurance performance distance sled dogs. We therefore expected the distance sled dogs to be similar in allele frequency to the Alaskan Malamute and Siberian Husky based upon their common athletic attribute of endurance and previous identification of these two purebred breeds being higher component breeds within the distance sled dog population [11]. The fact that we observe the opposite suggests that while marker 56 is useful for

distinguishing between sprint and distance populations of Alaskan sled dogs, the marker is not a hallmark of any putative contribution the *ACE* gene may be making to sprint versus distance performance.

One explanation for these results may relate to the founder populations of the Northern breeds that created the sled dogs over a century ago are genotypically different, especially for performance genes, from the registered AKC Alaskan Malamutes and Siberian Huskies we sampled at conformation events. Dogs shown in conformation events are selected based on AKC standards for body structure, not performance abilities [28]. Another explanation is that the G allele may have arisen separately in the sprint dog population and hence is not in the domestic dogs, or it may have come from a lineage that was not investigated.

The A allele of marker 56 appears to be near fixation in the distance sled dogs, while the G allele has been selected for in the sprint dogs (Table 3.2). This suggests that the G allele may be under selection in the sprint dogs for a trait other than endurance that was not in evidence when we compared domestic breeds such as the greyhounds and whippets, who share the attribute of speed. The association of marker 56 in the sled dogs may also reflect that this SNP is in LD with another other variants that are more biologically relevant. It would be interesting to sample and genotype working Alaskan Malamutes and Siberian Huskies to determine if they demonstrate selection for marker 56 in the working dogs, as opposed to the dogs bred for conformation.

The inclusion of purebred dogs in the analysis was important for several reasons. First it was necessary for carrying out tests of performance association. Second, and more

importantly, our knowledgebase regarding the composition of sprint and distance dogs is built upon our previous clustering analyses, which identified distinct contributions of multiple pure breeds, uniquely, to both sprint and distance populations[11]. There are limitations to the use of domestic breeds as a control for determining whether markers were associated with performance attributes selected for in sprint and distance sled dogs as opposed to being a population identifier. While performance is a genetically complex trait, this approach required a specific marker to demonstrate association to both the sled dog population attributed with speed (sprint) or endurance (distance) along with the respective domestic breeds displaying the same athletic attribute, therefore assuming the same genetic mechanism effecting performance in the different populations. However, the genetic components integral to the speed exhibited by greyhounds and whippets may be significantly different from those selected for in sprint sled dogs. While we hypothesized that the genetic basis for endurance is more likely to be similar between distance sled dogs and their component breeds of Alaskan Malamute and Siberian Husky those similarities may not lie within the *ACE* gene. Nevertheless, understanding allele frequencies of critical SNPs in the context of both the sled dog and the contributing purebred dog populations was thus important. Comparison of LD across the *ACE* gene in the Alaskan sled dogs and the four pairs of domestic breeds analyzed corroborated our earlier findings, highlighting the uniqueness of the Alaskan sled dog breed. A substantial amount of LD in the *ACE* gene was found in both the sprint and distance dogs, but with noticeable differences in pattern between the two populations. By comparison, purebred

pairs showed considerably less LD and no pattern similarity to either Alaskan sled dog population (Figure 3.1).

Identifying genes and their subsequent markers that distinguish between elite endurance performing distance sled dogs and elite sprinting sled dogs has the potential to illuminate contributors in the complex genetic arena of human athletic performance. Although none of the variants identified in the *ACE* or *MSTN* genes were significantly associated with any behavioral traits, the finding of markers within the *ACE* gene which distinguishes these two populations of Alaskan sled dogs, and the developed understanding of how the populations relate to one another as well as various purebred breeds, sets the stage for genome-wide association studies aimed at finding performance-associated genes.

Acknowledgements

We acknowledge INBRE grant 5P20RR016466 from NCRR (J.R.) and the Intramural program of the National Human Genome Research Institute for their support (H.H., H.P. E.A.O). We also thank Kenneth and Lori Chezik, Dr. Dawn Brown, Greg Sellentin, and Deborah McGrath for sharing their expertise and review of the performance phenotype rating criteria. We appreciate collection assistance from Danielle Dillon, Keiko Herrick, Lori Gildehaus, Ian Herriott, and the kennel owners and handlers and we gratefully acknowledge the sled dog owners who have generously contributed information and DNA to our studies.

References

1. Collins MJC: **Dog Driver: A Guide for the Serious Musher**, First edn. Crawford, CO: Alpine Publications; 1991.
2. Rennick P (ed.): **Dogs of the North**. Anchorage: Alaska Geographic Society; 1987.
3. Wendt R: **Alaska Dog Mushing Guide: Facts and Legends**, Fifth edn. Fairbanks: Goldstream Publications; 1999.
4. Atlla GBL: **Everything I Know About Training and Racing Sled Dogs**, 1st Edition edn. Hayward: Typart Publishing Service; 1972.
5. Welch J: **The Speed Mushing Manual: How To Train Racing Sled Dogs**, First edn. Eagle River, AK: Sirius Publishing 1989.
6. Swenson R: **The Secrets of Long Distance Training and Racing**, 1 edn. Wasilla: L&B Color Printing; 1987.
7. **ISDRA; International Sled Dog Racing Association** [<http://www.isdra.org/>] Accessed Jan 2011.
8. **Yukon Quest; Sled Dog Race; Alaska, Yukon** [<http://www.yukonquest.com/>] Accessed Aug 2010.
9. **The Official Site of the Iditarod** [<http://www.iditarod.com/>] Accessed Aug 2010.
10. Ostrander EA, Huson HJ, Ostrander GK: **Genetics of athletic performance**. *Annual Review of Genomics and Human Genetics* 2009, **10**:407-429.
11. Huson HJ, Parker HG, Runstadler J, Ostrander EA: **A genetic dissection of breed composition and performance enhancement in the Alaskan sled dog**. *BMC Genetics* 2010, **11**:71.

12. Gayagay G, Yu B, Hambly B, Boston T, Hahn A, Celermajer DS, Trent RJ: **Elite endurance athletes and the ACE I allele—the role of genes in athletic performance.** *Human Genetics* 1998, **103**(1):48-50.
13. Montgomery HE, Marshall R, Hemingway H, Myerson S, Clarkson P, Dollery C, Hayward M, Holliman DE, Jubb M, World M *et al*: **Human gene for physical performance.** *Nature* 1998, **393**(6682):221-222.
14. Thompson WR, Binder-Macleod SA: **Association of genetic factors with selected measures of physical performance.** *Physical Therapy* 2006, **86**(4):585-591.
15. Zhang X, Wang C, Dai H, Lin Y, Zhang J: **Association between angiotensin-converting enzyme gene polymorphisms and exercise performance in patients with COPD.** *Respirology* 2008, **13**(5):683-688.
16. Danser AH, Schalekamp MA, Bax WA, van den Brink AM, Saxena PR, Riegger GA, Schunkert H: **Angiotensin-converting enzyme in the human heart. Effect of the deletion/insertion polymorphism.** *Circulation* 1995, **92**(6):1387-1388.
17. Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, Soubrier F: **An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels.** *Journal of Clinical Investigation* 1990, **86**(4):1343-1346.
18. Montgomery H, Clarkson P, Barnard M, Bell J, Brynes A, Dollery C, Hajnal J, Hemingway H, Mercer D, Jarman P *et al*: **Angiotensin-converting-enzyme gene insertion/deletion polymorphism and response to physical training.** *Lancet* 1999, **353**(9152):541-545.
19. Woods DR, Montgomery HE: **Angiotensin-converting enzyme and genetics at high altitude.** *High Altitude Medicine & Biology* 2001, **2**(2):201-210.
20. Mosher DS, Quignon P, Bustamante CD, Sutter NB, Mellersh CS, Parker HG, Ostrander EA: **A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs.** *PLoS Genetics* 2007, **3**(5):e79.

21. Girgenrath S, Song K, Whittemore LA: **Loss of myostatin expression alters fiber-type distribution and expression of myosin heavy chain isoforms in slow- and fast-type skeletal muscle.** *Muscle Nerve* 2005, **31**(1):34-40.
22. Szabo G, Dallmann G, Muller G, Patthy L, Soller M, Varga L: **A deletion in the myostatin gene causes the compact (Cmpt) hypermuscular mutation in mice.** *Mammalian Genome* 1998, **9**(8):671-672.
23. McPherron AC, Lee SJ: **Double muscling in cattle due to mutations in the myostatin gene.** *Proceedings of the National Academy of Sciences* 1997, **94**(23):12457-12461.
24. Grobert L, Martin LJ, Poncelet D, Pirottin D, Brouwers B, Riquet J, Schoeberlein A, Dunner S, Menissier F, Massabanda J *et al*: **A deletion in the bovine myostatin gene causes the double-musled phenotype in cattle** *Nature Genetics* 1997, **17**:71-74.
25. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen JM, Eychenne F *et al*: **A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep.** *Nature Genetics* 2006, **38**(7):813-818.
26. Schuelke M, Wagner KR, Stolz LE, Hubner C, Riebel T, Komen W, Braun T, Tobin JF, Lee SJ: **Myostatin mutation associated with gross muscle hypertrophy in a child.** *New England Journal of Medicine* 2004, **350**(26):2682-2688.
27. **American Kennel Club; Conformation Events**
[<http://www.akc.org/events/conformation/>] Accessed 1 Aug 2010.
28. Club AK: **The Complete Dog Book**, vol. 19, 19 edn. New York: Howell Book House; 1997.

Figures

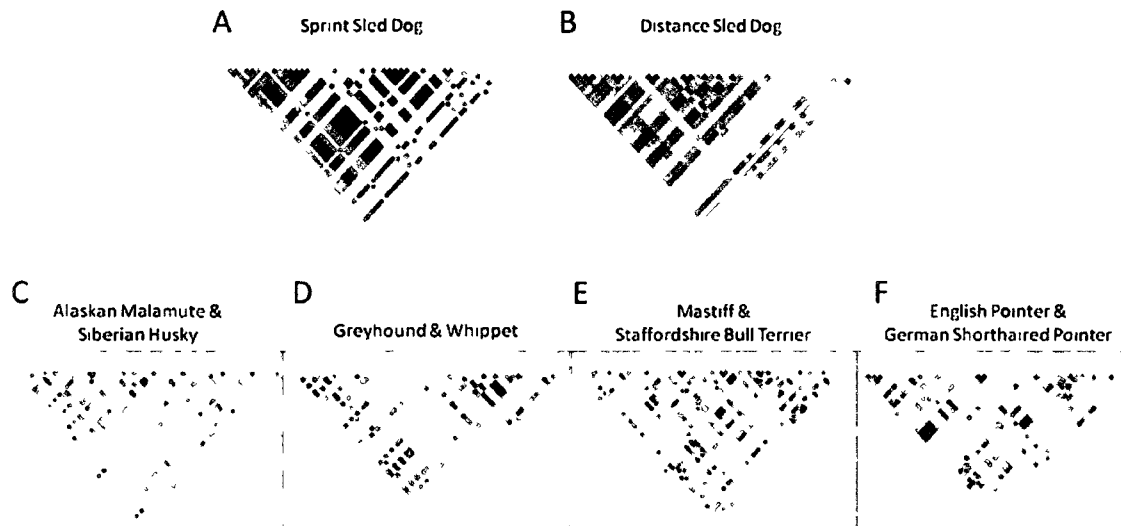


Figure 3.1: ACE gene linkage disequilibrium within sprint and distance sled dogs and eight domestic breeds attributed with speed or endurance performance.

ACE gene linkage disequilibrium (LD) plots were analyzed separately for sprint and distance dogs and closely related breed pairs to determine if sled dogs have similar LD patterns to the domestic breeds. LD was determined for each group based on the 81 polymorphisms found from manual sequence screening of the 20 Alaskan sled dogs (10 sprint, 10 distance) and 80 purebred dogs. The LD plots were analyzed with the four gamete rule and depicted using the alternate D'/LOD color scheme in Haploview 4.1. Low LOD and high LOD scores with low D' scores are in white, low LOD with high D' scores are in shades of pink, and high LOD with high D' scores are in black. 1A) Ten sprint dogs 1B) Ten distance dogs 1C) 10 Alaskan Malamutes and 10 Siberian Huskies 1D) 10 Greyhounds and 10 Whippets 1E) 10 Mastiffs and 10 Staffordshire Bull Terriers 1F) 10 English Pointers and 10 German shorthaired Pointers. The Alaskan sled dogs demonstrate a substantial amount of LD in the *ACE* gene which differs between the sprint

and distance populations. The purebred pairs show less LD in the *ACE* gene and no similarity in pattern to the Alaskan sled dog populations.

Tables

Table 3.1: Associated SNPs across the *ACE* gene comparing distance and sprint sled dogs.

Marker	Chr.	Location (bp)	Gene Position	Allele	Distance ¹ (Case) MAF	Sprint ² (Control) MAF	p-value	Permutation (10000 cycles) p-value
42	9	14631037	Intron 15	A:G	0	0.35	0.0036	0.0156
56	9	14635693	Intron 19	A:G	0	0.35	0.0036	0.0156
74	9	14637900	Intron 21	G:A	0	0.389	0.0051	0.0399

1. Ten elite distance dogs were set as cases
2. Ten elite sprint dogs were set as controls.

Table 3.2: *ACE* gene Marker 56 shows P-value scores ≤ 0.001 in separate association tests of Alaskan sled dogs and purebred breeds.

Group	Marker	Chr.	Location(bp)	Gene Position	Allele	Case MAF	Control MAF	p-value
Sled dog ¹	56	9	14635693	Intron19	A:G	0.031	0.305	2.38E-05
Purebreed ²	56	9	14635693	Intron19	A:G	0.265	0	7.57E-08

1. SNP association test comparing 34 elite distance dogs (cases) and 49 elite sprint dogs (controls).
2. SNP association test comparing 10 Alaskan Malamutes and 10 Siberian Huskies (20 cases) and 60 purebred dogs (10 dogs from each of the following breeds; Whippet, Greyhound, Mastiff, Staffordshire Bull Terrier, English Pointer, and German Shorthaired Pointer)

Table 3.3: Minor allele frequencies of three *MSTN* gene polymorphisms found in sprint Alaskan sled dogs, 16 domestic breeds, and a Golden Jackal.

Gene	Group	Marker	Chr.	Location (bp)	Gene Position	Allele	MAF
MSTN	Sprint sled dog ¹	Marker01	37	3731257	Intron 2	C:A	0.281
MSTN	Sprint sled dog ¹	Marker02	37	3720985	Downstream 3'	T:A	0.125
MSTN	Sprint sled dog ¹	Marker04	37	3739468	Upstream 5'	A:C	0.151
MSTN	Pure breeds ²	Marker01	37	3731257	Intron 2	C:A	0.297
MSTN	Pure breeds ²	Marker02	37	3720985	Downstream 3'	T:A	0.22
MSTN	Pure breeds ²	Marker04	37	3739468	Upstream 5'	A:C	0.161

1. 91 sprint Alaskan sled dogs.
2. 77 total purebred dogs averaging 4 dogs per 16 breeds (Afghan Hound, Akita, Australian Shepherd, Basenji, Beagle, Boxer, Bull Terrier, Flat-Coated Retriever, German Wirehaired Pointer, Greyhound, Ibizan Hound, Italian Greyhound, Mastiff, Pharaoh Hound, Saluki, Siberian Husky, and Whippet) and a Golden Jackal.

Appendix

Table A3.1: *ACE* gene polymorphisms identified by manual sequence screening of Alaskan sled dogs and eight domestic breeds.

Marker	Location (bp)	Gene Position	Sprint ¹ Alleles	Sprint ¹ MAF	Distance ² Alleles	Distance ² MAF	Pure- bred ³ Allele	Pure- bred ³ MAF
01	14621929	Intron 1	C:C	0.000	C:T	0.071	C:T	0.075
02	14623981	Intron 4	C:A	0.062	C:A	0.214	A:C	0.396
03	14623991	Intron 4	A:A	0.500	A:C	0.429	C:A	0.306
04	14624070	Intron 4	A:A	0.500	A:G	0.286	G:A	0.224
05	14624087	Intron 4	C:C	0.000	C:C	0.000	C:T	0.045
06	14624154	Intron 4	A:A	0.500	A:T	0.143	A:T	0.470
07	14624190	Intron 4	A:A	0.500	A:C	0.286	C:A	0.273
08	14624293	Intron 4	A:A	0.000	A:A	0.000	G:A	0.118
09	14624317	Intron 4	C:A	0.400	C:A	0.286	A:C	0.399
10	14624417	Intron 4	A:A	0.000	A:G	0.143	A:G	0.171
11	14624652	Intron 5	G:T	0.350	T:G	0.357	G:T	0.468
12	14624734	Intron 5	G:C	0.400	C:G	0.357	G:C	0.475
13	14624765	Intron 5	A:G	0.350	G:A	0.357	G:A	0.466
14	14625167	Intron 5	C:T	0.350	C:C	0.500	C:T	0.231
15	14625202	Intron 5	C:C	0.000	C:C	0.000	C:G	0.172
16	14625216	Intron 5	C:T	0.350	C:C	0.500	C:T	0.239
17	14625245	Intron 5	T:C	0.350	T:C	0.429	T:C	0.276
18	14625426	Intron 5	C:C	0.000	C:C	0.000	C:A	0.038
19	14625993	Intron 6	T:C	0.400	C:T	0.450	C:T	0.493
20	14626033	Intron 6	G:G	0.000	G:G	0.000	G:A	0.250
21	14626264	Intron 6	C:C	0.000	C:C	0.000	C:A	0.257
22	14626995	Intron 7	G:A	0.300	G:G	0.000	G:A	0.069
23	14627056	Exon 8	A:G	0.400	G:A	0.444	A:G	0.400
24	14627332	Intron 8	C:A	0.400	A:C	0.444	A:A	0.500
25	14627803	Intron 9	A:G	0.389	G:A	0.417	A:G	0.365
26	14627877	Intron 9	G:A	0.350	A:G	0.312	G:A	0.462
27	14628368	Intron 10	C:G	0.389	G:C	0.286	C:G	0.342
28	14628534	Intron 10	C:C	0.000	C:C	0.000	C:T	0.013
29	14629180	Intron 12	C:C	0.000	C:C	0.000	C:T	0.260
30	14629222	Intron 12	C:C	0.000	C:C	0.000	C:T	0.093
31	14629265	Intron 12	G:A	0.350	A:G	0.389	G:A	0.226
32	14629303	Intron 12	G:A	0.350	A:G	0.389	G:A	0.199

Table A3.1 Continued

33	14629333	Intron 12	G:A	0.350	A:G	0.389	G:A	0.264
34	14629348	Intron 12	C:C	0.000	C:C	0.000	C:T	0.372
35	14630053	Intron 14	G:T	0.312	G:G	0.000	G:T	0.041
36	14630097	Intron 14	A:C	0.125	A:C	0.350	A:C	0.243
37	14630238	Intron 14	C:T	0.222	C:T	0.111	C:T	0.007
38	14630239	Intron 14	A:C	0.389	C:A	0.450	A:C	0.220
39	14630504	Intron 14	G:G	0.000	G:G	0.000	G:A	0.264
40	14630624	Intron 14	A:C	0.333	C:A	0.450	C:A	0.427
41	14630834	Intron 15	A:C	0.333	A:C	0.450	A:C	0.240
42	14631037	Intron 15	A:G	0.350	A:A	0.000	A:G	0.164
43	14631054	Intron 15	G:A	0.050	G:G	0.000	G:A	0.014
44	14631459	Intron 15	G:G	0.000	G:G	0.000	G:A	0.107
45	14632055	Intron 16	T:C	0.350	C:T	0.278	C:T	0.404
46	14632203	Exon 17	C:T	0.350	T:C	0.389	C:T	0.175
47	14633154	Intron 17	C:T	0.300	C:T	0.350	C:T	0.371
48	14633155	Intron 17	G:A	0.300	G:A	0.350	G:A	0.339
49	14633495	Intron 17	G:A	0.333	A:G	0.350	A:G	0.451
50	14633513	Intron 17	G:G	0.000	G:G	0.000	G:A	0.160
51	14633658	Intron 17	G:A	0.350	A:G	0.450	G:A	0.269
52	14633984	Intron 17	G:A	0.350	A:G	0.450	G:A	0.253
53	14634114	Intron 17	T:C	0.350	C:T	0.350	C:T	0.455
54	14635408	Intron 19	A:G	0.350	G:A	0.300	G:A	0.471
55	14635574	Intron 19	G:A	0.350	A:G	0.450	G:A	0.275
56	14635693	Intron 19	A:G	0.350	A:A	0.000	A:G	0.066
57	14635740	Intron 19	C:T	0.350	T:C	0.450	C:T	0.196
58	14636361	Intron 21	T:G	0.350	G:T	0.350	G:T	0.431
59	14636408	Intron 21	C:T	0.111	C:T	0.250	C:T	0.133
60	14636470	Intron 21	T:G	0.222	T:G	0.100	T:T	0.000
61	14636558	Intron 21	A:G	0.333	G:A	0.400	G:A	0.483
62	14636595	Intron 21	G:A	0.333	A:G	0.450	G:A	0.277
63	14636608	Intron 21	C:T	0.333	T:C	0.300	C:T	0.308
64	14636779	Intron 21	T:T	0.000	T:C	0.050	T:C	0.200
65	14636895	Intron 21	G:T	0.150	G:T	0.111	G:T	0.038
66	14636932	Intron 21	C:T	0.100	C:T	0.111	C:T	0.220
68	14637072	Intron 21	C:C	0.000	C:C	0.000	C:T	0.243
69	14637209	Intron 21	C:C	0.000	C:A	0.111	C:A	0.240
70	14637373	Intron 21	G:A	0.200	G:A	0.056	G:A	0.013
71	14637457	Intron 21	G:G	0.000	G:A	0.111	G:A	0.240
72	14637574	Intron 21	T:C	0.350	C:T	0.389	T:C	0.325

Table A3.1 Continued

73	14637884	Intron 21	C:C	0.000	C:C	0.000	C:T	0.178
74	14637900	Intron 21	G:A	0.389	G:G	0.000	G:A	0.059
75	14639495	Intron 25	A:C	0.188	A:C	0.125	A:A	0.000
76	14639511	Intron 25	C:G	0.444	C:G	0.312	G:C	0.439
77	14639574	Intron 25	A:C	0.111	A:C	0.188	A:C	0.243
78	14640591	Intron 25	T:C	0.389	C:T	0.333	T:C	0.295
79	14640702	Intron 25	C:C	0.000	C:C	0.000	C:T	0.172
80	14642149	Exon 28	C:T	0.111	C:T	0.167	C:T	0.377
		Down- stream 3'						
81	14642295	end	T:T	0.000	T:C	0.167	T:C	0.305
82	14626398	Exon 7	T:T	0.000	T:T	0.000	T:T	0.000

1. The alleles and minor allele frequencies (MAF) found in 10 elite sprint dogs.
2. The alleles and minor allele frequencies (MAF) found in 10 elite distance dogs.
3. The alleles and minor allele frequencies (MAF) found in 10 individual from each of the following domestic breeds; Alaskan Malamute, Siberian Husky, German Shorthaired Pointer, English Pointer, Greyhound, Whippet, Staffordshire Bull Terrier, Mastiff (80 individuals).

Chapter 4

Selection for Breed-Specific Ancestry and Genome-Wide Association Analysis Target SNPs Associating the *MYH9* Gene to Heat Tolerance within Performing Alaskan Sled Dogs¹

Abstract

Alaskan sled dogs present us with a unique opportunity to study the development of a genetically distinct breed of dog produced from the selective breeding of high performance athletes. The practice of interbreeding Alaskan sled dogs with various purebred dogs over the past century has allowed us to investigate the impact of these domestic breeds on the sled dog genome and their potential contribution to athletic attributes. Here, genetic profiles of both the “sprint” or short distance dogs and the “distance” or long distance racing dogs were created using genome-wide single-nucleotide polymorphism (SNP) arrays. The Alaskan Malamute, Siberian Husky, German Shorthaired Pointer, and Borzoi, which we previously established as genetic breed components of Alaskan sled dogs, were used for modeling ancestry in the sprint

¹Huson, H.J., vonHoldt, B.M., Rimbault, M., Byers, A.M., Runstadler, J., Ostrander, E.A., (to be submitted 2011). Selection for Breed-Specific Ancestry and Genome-Wide Association Analysis Target SNPs Associating the *MYH9* Gene to Heat Tolerance within Performing Alaskan Sled Dogs.

and distance dogs. We assessed population structure using principle component analyses and compared linkage disequilibrium and autozygosity to examine inbreeding patterns. We also determined regions of selective sweep and loci demonstrating genome-wide association to population differentiation, endurance, or heat tolerance in both the sprint and distance dogs. Genes potentially influential in differentiating between the sled dog populations and performance were identified by these methods. Coinciding regions of selective sweep or genome-wide association and ancestry selection allowed us to infer breed contribution of these ancestors to particular traits through genetic analysis. Most notably, we identified seven *MYH9* gene single nucleotide polymorphisms (SNPs) significantly associated with heat tolerance performance in sprint sled dogs with two SNPs corresponding to conserved analogous promoter and enhancer regions in human skeletal muscle myoblasts.

Introduction

The Alaskan sled dog has evolved over the past century from a working dog integral to northern arctic culture to become an elite twenty-first century athlete. The early arctic dog was originally developed to haul cargo-laden sleds over snow-covered terrain [1-3]. Their dominating presence in polar exploration and the boom of the Alaskan Gold Rush gave rise to the “Era of the Sled Dog” which reigned from the late 1800s to early 1900s [4]. While modern transportation eventually retired sled dogs from the role of necessary working dog, with the birth of sled dog racing dog drivers transitioned them into a sporting dog. Alaskan sled dogs, though not recognized by the

American Kennel Club (AKC) and not bred towards a physical standard, are selectively bred for climate-specific athletic attributes, which has resulted in a level of genetic distinctiveness comparable to that of AKC-recognized breeds [5]. Performance selection has given sled dogs a common athletic phenotype that includes a quick and efficient gait, superior pulling strength, and increased endurance. However, overall body weight and coat type can vary depending upon racing style, geographic location, lineage, and cross breeding to purebred lines.

Sled dog racing has progressed into two distinct styles based upon the mileage teams' travel. Long distance racing covers approximately 1,000 miles over multiple days with average speeds ranging between 13-19 km/h (*e.g.*, Iditarod and Yukon Quest) [6, 7]. Sprint racing is comprised of multiple events or classes defined by the number of dogs in the team (4-20 dogs), which in turn dictates the mileage run (~6-38 kilometers, dependent upon class). Sprint teams average 29-40 km/h with a single heat run per day, with the fastest combined time over two to three racing days producing the winner. The extreme differences in these racing styles has led to divergent selection on the Alaskan sled dog population, resulting in two genetic subgroups having different functional phenotypes focusing on either endurance or speed (Figure 4.1) [5].

The Alaskan sled dog affords us the rare opportunity to document the history of selection for athletic performance in an admixed breed. They not only have a historical mosaic of both purebred and northern "village" dog ancestry, but a continuation of the practice of crossing purebred dogs into the population for the attainment of desirable athletic attributes. Written pedigrees as well as genetic investigation have shown breeds

such as the Alaskan Malamute, Siberian Husky, Pointer (English and German Shorthaired), Saluki, Borzoi, Irish Setter, Weimaraner, German Shepherd, and Anatolian Shepherd to have been directly crossed into the Alaskan sled dog population or to be a genetic component of modern sled dogs [5, 8]. This complex admixed ancestry model has given us the opportunity to investigate the genetic profile of Alaskan sled dogs, thereby acquiring an understanding of the molecular mechanisms associated with the athletic performance of sled dogs. Here, we have utilized genome-wide panels of single-nucleotide polymorphism (SNP) markers to assess population structure (115,425 SNPs and 27,416 SNPs), perform mapping through ancestry analysis (7,644 SNPs), and to conduct genome-wide association studies (115,425 SNPs) highlighting sprint/distance differentiation as well as the athletic traits of endurance and heat tolerance in Alaskan sled dogs.

Methods

Sample Collection and SNP Array Genotyping

DNA was extracted on 150 Alaskan sled dogs and 45 dogs of four domestic breeds, for analysis on genome-wide SNP chip arrays. Prior to blood collection, all owners signed an informed consent document, consistent with NHGRI Animal Care and Use Committee rules. Whole blood samples were collected from the cephalic vein in 3-5ml EDTA or ACD tubes. Sled dogs were sampled at their home kennels. Purebred dogs were sampled at AKC-sanctioned events. Samples were stored at 4°C prior to extraction,

and genomic DNA was isolated using standard proteinase K/phenol extraction methods by Health Gene (Toronto, Canada) or RX Bioscience (Rockville, MD). DNA samples were stripped of identifiers, numerically coded, and aliquoted for long-term storage at -70°C. Detailed pedigrees were collected for each individual sampled, anonymized, and entered into a database.

A total of 150 Alaskan sled dogs, 65 from distance racing kennels and 85 from sprint racing kennels, were genotyped using the Illumina HD Canine SNP array [9]. Smaller groups of sled dogs belonging to either the sprint or distance populations and having elite or poor performance rankings for specific athletic attributes were used in the subsequent analyses. In addition, 45 AKC-registered dogs from four breeds previously established as purebred components of Alaskan sled dogs were sampled to represent ancestral populations to the sled dogs [5]. To this effect, 10 Alaskan Malamutes, 12 Siberian Huskies, 11 German Shorthaired Pointers, and 12 Borzois were genotyped using the Affymetrix v2.0 Canine SNP array [10]. Genome Studio and PLINK software were used for SNP quality control filtering [9, 11, 12]. A total of 115,425 SNPs were produced on the sled dogs using the Illumina HD Canine SNP array. All dogs had $\geq 93\%$ SNP call rate and less than 10% missing genotypes. Rare SNPs with less than 10% frequency were removed. Purebred dogs had a total of 48,716 SNPs produced from the Affymetrix Canine SNP array after the same quality control thresholds were applied. An analysis of overlapping SNPs from the Illumina and Affymetrix panels left a total of 27,416 SNPs common to all dogs. The 115,425 and 27,416 thousand SNP panels are referred to in the following analyses.

Performance Ratings

Sled dogs were individually scored for their abilities in speed, endurance, work ethic, mental stress tolerance, and heat tolerance. The attributes of endurance and heat tolerance are investigated independent of one another in this study with scoring criteria detailed below. Distance dogs were sampled from four kennels, all of which finished in the top 15% of competitors for the Yukon Quest or Iditarod races during the two consecutive years (2007-2008) of sample collection. Sprint dogs were sampled from four kennels, each placing in the top 25% of the International Sled Dog Racing Association points-ranking medal program during the sampling years (2005-2007). Due to the sled dogs being sampled from competitive racing kennels throughout the United States and Canada, there was limited control over environmental factors. Therefore, the following measures were taken to obtain comparable and reliable phenotypic evaluations. All distance kennels maintained similar training regimes concerning mileage (increasing up to ~322 kilometers) and speed (13-19km/h) traveled over the course of fall training through winter racing season (September- March). Sprint kennels were likewise similar to each other concerning mileage (increasing up to ~48 kilometers) and speed (24-40km/h) traveled during this same period. The study did not control for personal driver training styles. The kennels sampled were located throughout the northern continental United States and Alaska, and northern Canada with slight variations in weather and terrain. However, the dogs sampled from these different kennels competed in many of the same races throughout the winter, primarily located in Alaska. Dry dog food varied in brand among kennels but was comparable in total protein (~26-34%) and fat (~14-

20%) content. Each kennel also supplemented the dry diets with either raw meat or meat supplements, particularly in the winter racing months.

Criteria for each athletic attribute were defined by the author (HH) and reviewed by five professional sled dog drivers participating in a scoring test with the author (HH). Drivers and author (HH) independently rated a minimum of the same eight sled dogs after a single training run to review phenotypic scoring reliability between drivers and author (HH). All sled dogs were scored by their respective driver (eight sled dog drivers, not all drivers participated in rater reliability test). Individual distance dogs were scored a single time for their overall performance regarding each phenotype (speed, endurance, work ethic, mental stress tolerance, and heat tolerance) during the peak racing season (~March). Individual sprint dogs were scored on a weekly basis for each phenotype beginning at fall training (~September/October) and continuing through the end of the peak racing season (~March/April). Approximately 80% of the sprint racing dogs were phenotypically scored over consecutive years (2005-2007). To achieve a single score for sprint racing dogs that was comparable to distance dog scoring, the last weekly rating for individual sprint dogs during peak racing season was regarded as their overall year performance score. Consecutive year ratings were compared for individual dogs. If an individual dog's ranking for each attribute (speed, endurance, work ethic, heat tolerance, mental stress tolerance) did not change over consecutive years, that score became the dog's overall performance score. For this study, each athletic attribute was looked at independently of the other four attributes. A dog that had different yearly scores for the athletic attribute being investigated was not included for analysis of that attribute. Due to

the low number of dogs available for analysis of individual athletic attributes, sled dogs were not restricted by age and ranged from one to six years old actively racing dogs. A disparity in male versus female dogs was seen when comparing sprint versus distance racing dogs. Sprint kennels had a higher percentage of female dogs (60% female) while distance kennels had a higher percentage of male dogs (72%). Performance was investigated within sprint and distance dogs separately, therefore, this disparity in sex was the same in both elite versus poorly performing dogs. Lastly, we note that dogs were not penalized in performance scores due to injuries.

The endurance criteria were determined based on the average mileage traversed in a race and is calculated in kilometers. Dogs were ranked in three categories based on their endurance performance. Mileage requirements ranged from 13-48 kilometers for sprint dogs and 1,595-1,850 kilometers for distance dogs. A ranking of 1 was given to dogs completing the required mileage in good condition. Dogs that completed the mileage but struggled to do so were ranked 2, and dogs unable to complete the mileage were ranked 3.

Heat tolerance is a measure of how well a dog functions physically while running in warm temperatures (approximately -7 to 10 degrees Celcius). More specifically, it is a rating of whether a dog reaches or nears a state of heat exhaustion while running. Heat exhaustion is when the body is unable to keep itself cool. Therefore, the body temperature rises, causing increased heart rate, muscle weakness, dizziness or confusion, rapid breathing, nausea, and vomiting. Dogs showing no change in their ability to perform were ranked as 1. A ranking of 2 was given to dogs demonstrating a lower than

normal performance only found when running in warm temperatures and mild signs of heat exhaustion (two or more of the above mentioned signs). Dogs unable to complete the mileage and demonstrating considerable signs of heat exhaustion (collapse or near collapse) were ranked as 3. Physiological measurements were not directly taken upon the dogs displaying signs of heat exhaustion, as instrumentation and adequate human assistance were unavailable due to the unpredictable nature of the occurrence, location, and training schedules.

Population Structure

Population structure was assessed with unrelated individuals using principle component analysis (PCA) between distance (n=19) and sprint (n=27) sled dogs and then between all populations: Alaskan Malamute (n=10), Borzoi (n=12), German Shorthaired Pointer (n=11), Siberian Husky (n=12), distance (n=19), and sprint sled dogs (n=27). Principle component values were calculated using a panel of 7,644 ancestry informative marker SNPs (see AIMS; Modeling Ancestry). This panel was used specifically to analyze the ability of the AIM SNPs selected for ancestry modeling to differentiate between the individual populations. The PCA was run using the software EIGENSTRAT [13].

Linkage Disequilibrium and Homozygosity Analysis

Genome-wide pairwise genotypic associations (r^2) were produced from a panel of 27,416 SNPs common to the sled dogs and purebred dogs (Alaskan Malamute n=10,

Borzoi n=12, German Shorthaired Pointer n=11, Siberian Husky n=12, distance n=19, sprint n=19). R^2 was generated using PLINK software, as an estimate of linkage disequilibrium (LD) [12]. R^2 scores were averaged for a set of inter-SNP distances (Kb) binned into the following classes: 1.25, 2.5, 3.75, 5, 7.5, 10, 15, 20, 30, 40, 60, 80, 115, 150, 212.5, 275, 387.5, 500, 737.5, 975, and 1000Kb [14]. An estimate of LD decay was generated by obtaining the physical distance (Kb) at which the r^2 score dropped below the threshold of 0.5 for each population [15]. Population distances were also calculated using an r^2 threshold of 0.2 for a more direct comparison to a previous study which also used r^2 as opposed to D' as a measure of LD decay [16]. Additionally, we determined the level of autozygosity within each population by surveying runs of homozygous genotypes (ROH) using the 27,416 SNP panel in PLINK. Homozygous tracks were required to be a minimum of 100Kb in length and to contain at least 25 SNPs [14].

Selective Sweep

We next wanted to identify areas of selective sweep as another means of genetically profiling the sprint and distance populations and detecting potentially important genes differentiating these two groups. Four criteria were used to independently distinguish the major areas of selective sweep within the sprint (n=27) and distance (n=19) populations using the full panel of 115,425 SNPs. First, the lower fifth percentile (distance =0 H_O ; sprint <0.0833 H_O) of SNPs demonstrating a loss of heterozygosity were chosen (9,362 SNPs). Loss of heterozygosity was defined as the observed heterozygosity (H_O) being below one standard deviation (H_O value of 1 SD

distance=0.16; sprint=0.22) of the genome average (H_O values determined using PLINK). Regions were then identified as having a minimum of two consecutive loss of heterozygosity SNPs less than 300Kb apart (2,145 regions). These regions were narrowed down further by requiring at least one of the region's SNPs to be in the top fifth percentile of the greatest H_O difference between the sprint and distance populations (5,158 SNPs) and the top fifth percentile of F_{ST} scores (5,621 SNPs) [17]. While we recognize the complexity of this series of restrictions as well as alternative methods and cutoff values, we sought to define a smaller number of sites to interrogate using common scores (H_O and F_{ST}) and percentiles (95th percentile).

Genome-wide Association Study (GWAS)

Genome-wide association analyses were run using EMMAX software which accounts for population relatedness and stratification on the full panel of 115,425 SNPs within the Alaskan sled dogs [18]. To identify SNPs associated with sled dog population differentiation, 19 distance dogs were compared to 27 sprint dogs, all unrelated through the grandparent generation, in a case/control (sprint/distance) analysis. This study investigated the performance attributes of endurance and heat tolerance. The attributes of speed, work ethic, and mental stress tolerance, having sub-minimal numbers of unrelated poorly performing dogs, were not analyzed at this time. Age and sex of performance rated dogs were discussed in the “*Performance Ratings*” (Methods) section and not considered as covariates in the statistical analyses. The attribute of endurance was analyzed in a case/control format of poor (case) versus elite (control) performers within

the sprint or distance populations independently due to the considerable difference in performance requirements between the two groups (sprint dogs, 48 kilometers; distance dogs, 1,609 kilometers). Endurance was tested in sprint dogs, comparing 20 poor performers to 21 elite performers, and in distance dogs comparing 14 and 19 individuals respectively. Heat tolerance was also tested independently within each sled dog population. Due to the fact that environmental temperature conditions were comparable between the two groups even though performance requirements of mileage and rate of speed run were not, an additional GWAS was run which combined the two populations, with all elite heat tolerance performers compared to all poor performers. Heat tolerance compared 17 poor performers to 21 elite performers in the sprint dogs, and in distance dogs comparing 10 and 19 individuals respectively. The combined sled dog population heat tolerance analysis compared 27 poor performers to 40 elite performers. Poorly performing dogs were allowed relatedness after two generations due to the low numbers of poorly performing individuals, while elite individuals were unrelated within three generations. Within the dogs sampled for this study, less than 10% were scored as a 3 for either athletic attribute (endurance or heat tolerance). Therefore, in both the endurance and heat tolerance association analyses, the dogs ranked as 2 and 3 were grouped together as the poorly performing case group. Dogs ranked as 2 demonstrated signs of poor performance (endurance- difficulty completing mileage; heat tolerance- at least 2 signs of heat exhaustion) but were not the extreme representatives of either phenotype which is more ideal for genome-wide association. Significance levels were generated using basic

(adaptive) permutation testing in PLINK for SNPs demonstrating genome-wide association in EMMAX.

Sequencing of the HINT1 and MYH9 Genes

Two candidate genes were selected for sequencing from regions identified by genome-wide association studies (described in Methods & Results: Genome-wide Association Study). The *histidine triad nucleotide binding protein 1 (HINT1)* gene was chosen as a candidate gene potentially associated with population variation between the sprint and distance dogs. The *myosin heavy chain 9 non-muscle type II class A (MYH9)* gene was chosen from the GWAS region associated with elite versus poor sprint performers with respect to heat tolerance. The exons of the *HINT1* and the *MYH9* genes were sequenced in an attempt to identify segregating alleles. The *HINT1* gene is located on canine chromosome 11:22400779-22560252, consisting of four exons totaling 560bp in length. Five amplicons, averaging 550bp, were necessary to cover the *HINT1* gene. Nineteen distance dogs and 27 sprint dogs were sequenced. These were the same sled dogs used in the genome-wide population association analysis. Eight German Shorthaired Pointers and eight Siberian Huskies were also sequenced through the *HINT1* exons due to the sled dog populations having an excess of these two ancestry groups within this same region (Results: Ancestry Modeling).

The *MYH9* gene is considerably larger, consisting of 40 exons and totaling 7,318bp in length. It is located on canine chromosome 10:31135177-31194500. Forty-three amplicons, averaging 620bp in length, were used to cover the *MYH9* exons.

Another 11 amplicons were used to sequence through highly conserved regions both up- and downstream of the gene. Six elite heat tolerance sprint sled dogs and six poorly performing heat tolerance sprint dogs were sequenced through all 54 amplicons to identify genetic variants. An additional 26 unrelated within two generations of poor sprint heat tolerance performers and 15 unrelated (three generations) elite sprint heat tolerance performers were sequenced across 16 *MYH9* SNPs demonstrating association in the preliminary 12 sled dogs analyzed. Therefore, a total of 32 poor heat tolerance dogs were compared to 21 elite heat tolerance dogs to determine significance and stability of association of performance to these SNPs. Eight German Shorthaired Pointers and six Alaskan Malamutes were also sequenced over these 16 SNPs due to the excess of these two ancestries in sprint sled dogs within this region for comparison to the poor and elite heat tolerance sprint sled dogs.

PCR amplification for the *HINT1* and the *MYH9* genes was performed in a 10 μ l volume containing 10ng genomic DNA, 1 μ l of 10x Taqgold Buffer, 0.05 μ l of TagGold[19], 1 μ l of 1mM dNTPs, 0.3 μ l of 50mM MgCl₂, 1 μ l of both forward and reverse 2 μ M primers, and 4.65 μ l water. Touchdown PCR was carried out as follows: 94°C for 10 min, followed by 20 cycles of 94°C for 30 s, then decreasing by 0.5°C/cycle starting at 65°C down to 55°C during annealing for 30 s, and 72°C for 45s, followed by another 20 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 45s, with a final extension phase of 72°C for 10min. A small subset of amplicons within each gene required the following PCR protocol for successful amplification: 10 μ l total volume containing 10ng genomic DNA, 5 μ l of KOD Buffer, 0.2 μ l of KOD [20], 1.6 μ l of

2.5mM dNTPs, and 1.2 μ l of both forward and reverse 2 μ M primers. The annealing temperature was also adjusted in the touchdown PCR decreasing by 0.5°C/cycle for the first 20 cycles from 67°C to 57°C and remaining at 57°C for the second 20 cycles.

PCR products were sequenced using Big Dye version 3.1 on an ABI 3730x1 capillary electrophoresis unit [19]. Sequence reads were aligned and analyzed using Phred, Phrap, and Consed software. PolyPhred software was used to identify heterozygous positions [21-23]. All genetic variants, including SNPs identified by PolyPhred and indels identified by manual sequence analysis, were then analyzed with Haploview 4.2 for allele segregation significance, linkage disequilibrium, and haplotype identification [24].

Phasing

We inferred haplotype phase using the program fastPHASE version 1.4.0 [25] across the 27,416 SNP panel for use in the ancestry modeling analysis. All individual Alaskan sled dogs and purebred dogs were phased together with a 0.05 masking rate applied. We specified the number of haplotype clusters (K) to range from 2 to 9 with an interval of 1.

Modeling Ancestry

The program SABER was utilized for modeling ancestry within the sprint and distance populations. An admixture mapping approach using this information was taken to localize genomic regions of particular ancestry selection within the two sled dog

populations [26, 27]. We used the Alaskan Malamute, Siberian Husky, Pointer, and Borzoi as our reference ancestry breeds due to their previous identification as four main domestic-breed components of Alaskan sled dogs [5]. We note that the (English) Pointer breed identified in our previous microsatellite work [5] was substituted with the German Shorthaired Pointer breed in this analysis due to the availability of SNP array data on the German Shorthaired Pointer. Both of these Pointer breeds have been documented within written pedigree records as being interbred with Alaskan sled dogs. SABER software delineates ancestry blocks in the admixed sled dog populations from the reference domestic breeds by implementing an extended Markov-Hidden Markov Model (MHMM) for inferring ancestry switches across the genome and accounts for background LD. From the original 27,416 SNPs common to the Alaskan sled dogs and purebred array panels we identified a subset of 7,644 ancestry informative markers (AIMs) that are diagnostic of the four reference (ancestry) breeds: Alaskan Malamute (n=10), Siberian Husky (n=12), German Shorthaired Pointer (n=11), and Borzoi (n=12) [27-30]. The AIM SNPs were selected based on Wright's population differentiation statistic, F_{ST} [14], using the program SCATTER { [17]}. The resulting AIM SNPs had an $F_{ST} > 0.35$ with an average spacing of ~300Kb. We specified a 1.0 cM/Mb recombination rate [14] and used a prior of 10 generations from the initial admixture event ($\tau = 10$) for ancestry block assignments across all 38 autosomes using the four reference breeds.

SABER generates diploid ancestry block assignments for individual sled dogs. Using four ancestor populations, ten diploid ancestry states are produced; four states are homozygous for the individual ancestor breeds (Alaskan Malamute, Borzoi, German

Shorthaired Pointer, and Siberian Husky) and six are heterozygous combinations of the breeds (Alaskan Malamute/Borzoi, Alaskan Malamute/German Shorthaired Pointer, Alaskan Malamute/Siberian Husky, Borzoi/German Shorthaired Pointer, Borzoi/Siberian Husky, and German Shorthaired Pointer/Siberian Husky). The sled dogs were grouped with respect to their racing style (distance $n=19$; sprint $n=27$) to identify the most frequent ancestry per SNP for each sled dog population. In order to estimate ancestry block frequency within each sled dog group, we randomly sampled 19 dogs from the unrelated sprint dogs to equal our 19 unrelated distance dogs. We included blocks that had a minimum of three contiguous SNPs therefore excluding potentially false ancestry blocks (due to random chance or lack of information). Longer ancestry blocks that have multiple SNPs of the same ancestry assignment are likely to be true blocks. Ancestry blocks were deemed private to a single sled dog population if they had $>20\%$ frequency in the population they were private to and $<5\%$ frequency in the opposing population. Regions showing excess or deficient selection (>1 standard deviation from each ancestral frequency mean) towards a particular ancestor were identified within the distance and sprint sled dogs based upon the highest degree of ancestral frequency difference between the two populations at consecutive SNPs. The top 5% of AIM SNPs (382 SNPs) showing the highest degree of ancestry frequency difference between the sprint and distance populations were used to determine the genomic regions demonstrating the strongest ancestry selection in each sled dog population. These regions were above two standard deviations from the mean ancestor frequency difference [26].

Results

Population Structure

We conducted a principal component analysis (PCA) to explore the degree of substructure in Alaskan sled dogs. PCA of the Alaskan sled dogs identified two separate but closely related groups (sprint and distance) with PC1 accounting for 6% of the variation and PC2 through PC4 each accounting for 4% of the variation (Figure 4.2A). In a comparison of the domestic breeds as well as the Alaskan sled dogs (Figure 4.2B), the first component (PC1, 16%) separates the Northern breeds (Alaskan Malamute and Siberian Husky) from the Borzoi and German Shorthaired Pointer, with both sled dog populations falling between the breed extremities. PC2 (9.8%) separates the Borzoi from the German Shorthaired Pointer, while PC3 (6.1%) distinguishes the Alaskan Malamute from the Siberian Husky. PC4 (3.6%) separates all Alaskan sled dogs from all of the domestic breeds, while PC5 (1.9%) separates the sprint from the distance sled dogs.

Linkage Disequilibrium and Homozygosity Analysis

To assess inbreeding patterns of the domestic breeds and the Alaskan sled dogs we estimated linkage disequilibrium (LD) using pairwise genotypic association (r^2) and their degree of autozygosity. LD should be greater in inbred than admixed populations [14, 16, 27, 31, 32]; therefore, in comparison to the Alaskan sled dogs which are freely mixed with purebred dogs for performance enhancement, we expected the four parental breeds to show a higher degree of LD due to their closed population breeding patterns [33]. The physical distance at which r^2 decays below the threshold of 0.5 gives an

estimate of LD [12, 15]. This was exemplified by both sled dog populations having a shorter distance to $r^2_{0.5}$ decay (Sprint, 2.5-3.75Kb; Distance, 7.5-10Kb) than any of the purebred groups (German Shorthaired Pointer, 10-15Kb; Siberian Husky, 15-20Kb; Alaskan Malamute and Borzoi, 20-30Kb) (Figure 4.3A). However, due to the fact that previous studies estimating LD decay at a threshold of 0.5 used D' , we also calculated the distance to LD decay at $r^2_{0.2}$. At the threshold of $r^2_{0.2}$, the Alaskan Malamute, Siberian Husky, Borzoi, and distance sled dog populations have greater than 1Mb of LD, ranging from 0.21-0.25 at 1MB (maximum distance calculated). LD at $r^2_{0.2}$ decays at approximately 700Kb in German Shorthaired Pointers and 80Kb in sprint sled dogs. These values are comparable with the range of $r^2_{0.2}$ values reported in 23 purebred breeds examined by Gray *et al* in 2009 [16]. Overall, shorter distances to decay are indicative of admixture or a larger effective population size. We also analyzed the genome-wide degree of autozygosity or identity by descent surveyed as the size distribution of homozygous tracts (runs of homozygosity, ROH) (Figure 4.3B) [14]. Trends were similar to that of LD decay, with the domestic breeds having more homozygous blocks of longer length (>2Mb) than the sled dogs, indicating a higher degree of inbreeding in the domestic breeds. However, the distance dogs had a slight inflation of homozygous blocks of large size (~12Mb) compared to the sprint dogs, concordant with the previous inbreeding assessments within Alaskan sled dogs [5].

Selective Sweep

We next identified areas of selective sweep between the sprint and distance populations to detect potentially important genes differentiating these two groups. The strongest regions demonstrating selective sweep were generated by employing a series of restricting factors based upon observed heterozygosity, the number of informative SNPs, distance between informative SNPs within the region, and SNP F_{ST} scores (see Methods Selective Sweep) (Supplemental Table 4.1, 60 regions total). Eighty-seven percent (52) of the regions showed selective sweep within the distance dogs, while only eight regions demonstrated selection in the sprint dogs. The region of greatest H_O difference (0.833) was on chromosome 3 from 83,775,932 to 83,798,854bp with selection in the distance dogs. One of only two genes (human genome) within 1Mb upstream and 1Mb downstream of this region is the *ADP-ribosylation factor-like 2 binding protein (ARL2BP)*, which is linked to mitochondrial activity, particularly in cardiac and skeletal muscle tissues [34]. Therefore, a gene such as the *ARL2BP* may be influential in the energy output, specifically of the cardiac and skeletal muscles within distance sled dogs. The highest region of H_O (0.75) difference within sprint dogs was on chromosome 17 (8158751-8170123bp) and had no obvious genes based on annotated gene or protein function [35, 36] within 1Mb upstream or 1Mb downstream of the region in the vicinity potentially differentiating sprint from distance dogs.

Genome-wide Association Study (GWAS)

Genome-wide association analyses were performed to identify genetic loci associated with either population differentiation or performance attributes of endurance or heat tolerance. Six loci associated with sprint and distance population variation had p-values less than 4.68×10^{-6} (permuted p-values $< 3 \times 10^{-6}$) (Supplemental Table 4.2). SNP *Chr3.82650187* had the most significant population association (cases equal distance dogs) with a p-value of 1.03×10^{-7} . This SNP is approximately 1Mb from the selective sweep region showing the highest degree of H_O difference between sprint and distance dogs. The gene *ARL2BP* (mitochondrial activity in cardiac and skeletal muscle tissue) discussed previously in selective sweep is one of only two genes within 2MB surrounding the associated SNP and selective sweep region. The next region demonstrating an association to distance sled dogs set as the cases in the analysis was found on chromosome 11, consisting of two SNPs with p-values of 1.00×10^{-6} . Human genome annotation identified 25 genes within a range of 1Mb upstream to 1Mb downstream of the GWAS SNPs. Of the 25 genes annotated within this region, only the *histidine triad nucleotide binding protein 1 (HINT1)* gene, residing approximately 70Kb and 600Kb from these SNPs, was recognized as a potential candidate gene associated to sled dog population variation based upon gene and protein function [35, 36] and literature review. The *HINT1* gene has been associated with anxiety and stress coping behaviors in knockout mice [37, 38].

Elite versus poorly performing distance dogs were compared for the athletic attributes of endurance and heat tolerance. Likewise, elite versus poor performing sprint

dogs were compared for these same attributes. While endurance performance in sprint sled dogs was associated with 15 loci containing SNPs with a p-value less than 1×10^{-6} , permutation testing proved all sites statistically unstable (p-values above 1×10^{-4}). Heat tolerance performance in sprint sled dogs showed stronger association stability by producing a region on chromosome 10 (31089847-31188654bp) with four clustered SNPs having p-values ranging from 4.53×10^{-6} to 5.57×10^{-7} (permuted p-values from 1.20×10^{-5} to 5×10^{-6}) (Figure 4.4; Supplemental Table 4.2). The SNPs highlighting this region are either within or directly upstream of the *myosin heavy chain 9 non-muscle type II class A* (*MYH9*) gene. However, an additional 33 genes are annotated in the human genome from 1MB upstream to 1MB downstream of the outlying GWAS SNPs but do not have gene or protein function information or previous research suggesting an association to heat tolerance. The *MYH9* gene makes for an intriguing performance candidate gene as it has been associated to exercise muscle efficiency and has demonstrated activity level differences due to muscle temperature variation [39-41].

Ancestry Modeling

A genome-wide ancestry profile was generated for the sprint and distance sled dogs to determine genomic regions of ancestry selection based on the four ancestry reference breeds of Alaskan Malamute, Siberian Husky, German Shorthaired Pointer, and Borzoi [5]. The average genome-wide proportion of each ancestor breed was calculated for the sprint and distance dogs (Table 4.1). Despite the genome having an overall mosaic structure in each of the sled dog populations (Figure 4.5), the proportion of

Alaskan Malamute (Distance = 0.32; Sprint = 0.25) and Siberian Husky (Distance = 0.27; Sprint = 0.22) were greater in the distance than in the sprint dogs (Figure 4.6A; Table 4.1). The German Shorthaired Pointer ancestry showed the highest frequency difference with a 10% increase in the sprint dogs (Distance = 0.23, Sprint = 0.33) (Figure 4.6A; Table 4.1). An analysis of genome-wide frequency differences of each diploid ancestry state between sprint and distance showed the top three differences to be in homozygous states of the Alaskan Malamute (6.18% increase in Distance), German Shorthaired Pointer (5.21% increase in Sprint), and Siberian Husky (4.57% increase in Distance) (Figure 4.6B; Table 4.1). The Borzoi/German Shorthaired Pointer showed the largest degree of difference (4.26% increase in Sprint) for a heterozygous ancestry state.

We identified a total of 186 private ancestry blocks in the Alaskan sled dogs (distance: 97 blocks with 447 copies; sprint: 89 blocks with 392 copies) based on size (minimum of 3 consecutive SNPs) and ancestry state (Table 4.2). The median block length in the distance population was 1,337Kb, with a shorter median block length of 1,137Kb in the sprint population. The longest ancestry block in distance dogs was a homozygous state of Alaskan Malamute, while the most frequent private ancestry state was that of the Alaskan Malamute/German Shorthaired Pointer. The respective blocks in sprint dogs were a homozygous state of Siberian Husky and a heterozygous state of Borzoi/German Shorthaired Pointer (Table 4.2).

After creating a genome-wide ancestry profile of the sprint and distance populations we sought to localize genomic regions showing the strongest ancestral selection in either population. We established 48 regions showing the highest degree of

ancestry frequency difference between the sprint and distance populations (Supplemental Table 4.3). The minimum ancestral frequency difference in these regions was 0.33, which was beyond 2 standard deviations from the mean (Mean = 0.095; 2 SD = 0.26). The highest ancestral frequency difference was seen on chromosome 11 (18482294-23584745bp) with a positive selection of 0.510 for Siberian Husky in distance sled dogs. This 5Mb region contains genes such as *fibrillin 1 and 2* (*FBN1*, *FBN2*), which produce proteins integral to the structure and function of connective tissue as well as *acyl-CoA synthetase long-chain family member 6* (*ACSL6*) and *solute carrier family 27, member 6* (*SLC27A6*) genes which are important in fatty acid metabolism and transport respectively [35, 36]. Another gene featured in this region was the *HINT1* gene (stress coping behavior) previously found near two SNPs segregating with the sled dog populations in our GWAS [37, 38]. Overall, nineteen regions demonstrated positive ancestry selection in sprint dogs; two regions selected Borzoi and 17 selected German Shorthaired Pointer ancestry. The remaining regions (29) demonstrated positive selection within distance dogs and included 15 Alaskan Malamute, two Borzoi, one German Shorthaired Pointer, and 11 Siberian Husky ancestries.

We next wanted to identify areas where ancestry selection correlated with selective sweep or GWAS in either the distance or sprint populations. This allowed us to differentiate between an excess in ancestry attributed to random chromosomal inheritance or a product of selection beneficial to the sled dogs. Five areas of selective sweep coincided with four regions of ancestry selection on chromosomes 3, 10, 16, and 28 (Table 4.3). Chromosome 3 contains two areas of selective sweep in distance dogs,

which is overlapped by positive selection for Siberian Husky ancestry in distance dogs. This region includes the gene *solute carrier family 2, member 9 (SLC2A9)*, which is integral to glucose homeostasis as a glucose transporter [35, 36]. Chromosome 10 had coinciding selective sweep and German Shorthaired Pointer ancestry selection, but in opposing distance and sprint populations respectively. Despite the fact that these two methods pointed us towards different sled dog populations, the nearest gene, *methionine sulfoxidereductase B3 (MSRB3)*, produces proteins which perform crucial functions for cell protection against oxidative stress which may be vital for sled dogs performing under extreme physiological and environmental conditions [42]. Three distinct ancestry patterns occur in the region of selective sweep on chromosome 16. The area was highlighted for overlapping with a long region of positive selection of the German Shorthaired Pointer ancestry (0.398 frequency difference, Supplemental Table 4.3) in sprint dogs (Figure 4.7). In addition to the Pointer selection, distance dogs in which the selective sweep occurs show a 0.25 decrease in Alaskan Malamute ancestry frequency coinciding with an increase of 0.25 for Siberian Husky ancestry (Figure 4.7). The gene *protein tyrosine phosphatase, receptor type, N polypeptide 2 (PTPRN2)* falls within this region and is involved in insulin binding and beta cell growth regulation within the insulin granule [43]. Chromosome 28 possessed selective sweep (29046328-29143901bp) and an excess of Alaskan Malamute (0.312 frequency difference) ancestry in distance dogs. There was an even greater selection of German Shorthaired Pointer ancestry in sprint dogs (0.421 frequency difference) in this area. The gene *attractin-like 1 (ATRNL1)*, linked to both physical and behavioral attributes including dysmorphic

facial attributes, toe syndactyly, and cognitive impairment, is a candidate gene within this region [44]. *ATRNL1* was also linked with human information processing speed and IQ in a recent GWAS [45].

We also determined where areas of ancestry selection correlated with regions of interest found during genome-wide association analyses. Two regions differentiating between the sprint and distance populations, located on chromosomes 11 and 32, demonstrated p-values less than 1×10^{-6} , and overlapped with ancestry selection (Table 4.3). The chromosome 11 locus was highlighted by two SNPs and drew our attention back to the gene *HINT1* related to anxiety and stress coping behavior [37, 38]. This gene was previously mentioned as lying in the region of highest ancestry selection, specifically Siberian Husky within distance dogs, but also overlapping with another region of substantial ancestry selection for German Shorthaired Pointer within sprint dogs (Supplemental Table 4.3: Index #18 & 19). The GWAS loci located on chromosome 32 drew our attention to the *hemoglobin zeta (HBZ)* gene. *HBZ* has been tied to the disease Human T-cell Lymphotropic Virus (HTLV-1), which includes symptoms such as progressive weakness, stiff muscles, and muscle spasms [46, 47]. The *MYH9* gene, investigated for its role in heat tolerance, also correlated with positive selection of the German Shorthaired Pointer ancestry within sprint dogs (frequency difference 0.313). This region was not initially pinpointed as overlapping with ancestry selection because the frequency of ancestry fell below the 95th percentile threshold (frequency difference ≥ 0.333) used to determine the major regions of ancestry selection (Supplemental Table

4.3). Overall, eight loci, identified by either GWAS or selective sweep, corresponded with an excess of one of the ancestral reference populations.

Fine-mapping of the HINT1 and MYH9 Genes

The *HINT1* gene was determined a candidate gene potentially associated with sled dog population variation and their respective ancestry selections. As a result, we hypothesized that this gene may account for differences in stress coping abilities between the distance and sprint dogs and proceeded with sequencing to identify segregating haplotypes or causative variants. An excess of Siberian Husky ancestry in distance dogs and German Shorthaired Pointer within sprint dogs, overlapping the *HINT1* gene, supported the theory that these breeds play a role in the two sled dog populations' stress coping behaviors. Direct sequencing of the four *HINT1* exons and their surrounding area produced seven genetic variants found in sprint and distance sled dogs. Six of the variants were found in German Shorthaired Pointers and four were found within Siberian Huskies. One SNP variant, an *A/T* base-pair change, was located 200bp upstream of the gene. All German Shorthaired Pointers were homozygous *A* (major allele) at this region. The Siberian Husky genotype contrasted this with a selection for the minor *T* allele showing a frequency (MAF) of 0.75. Four more SNPs and two indels were located in intronic regions of the *HINT1* gene. Siberian Huskies were homozygous *T* and *C*, both major alleles, for SNPs 57bp and 68bp after the first exon respectively. The German Shorthaired Pointers had a frequency of 0.40 for both of these alleles. None of the variants were found to be associated with the sprint or distance sled dogs.

Our GWAS comparing elite and poor performing heat tolerance sprint sled dogs produced a promising candidate gene, the *MYH9* gene. The GWAS included four significantly associated SNPs, two of which were located within the *MYH9* gene itself, with the other two located 45Kb and 20Kb upstream of the 5' end of the *MYH9* gene. Direct sequencing of six elite sprint heat tolerance dogs and six poorly performing sprint heat tolerance dogs through the 40 *MYH9* exons and conserved regions up and downstream of the gene produced 51 novel variants. Forty-three novel variants were within the *MYH9* gene and included five SNPs within exons, 43 SNPs within introns, and two indels within introns. An additional eight new SNPs and two indels were found upstream of the 5' end of the gene. Synonymous amino acid changes were found in exons 4 (31155024bp, aspartic acid), 9 (31161766bp, leucine), 18 (31175229bp, glutamic acid), 24 (31181751bp, arginine), and 29 (31184517bp, arginine).

A preliminary single marker association analysis comparing sprint elite (6) heat tolerance dogs to poorly performing (6) dogs utilized a total of 72 markers, including novel variants and SNP array markers. This produced 16 SNPs with raw p-values below 0.05 associated to poor heat tolerance performers (cases in the analysis). An additional 26 poor sprint heat tolerance dogs and 15 elite sprint heat tolerance dogs were sequenced for these 16 SNPs. Single marker association analysis of these 16 SNPs comparing a total of 32 poor sprint heat tolerance dogs and 21 elite sprint heat tolerance dogs produced 14 SNPs with raw p-values less than 0.05 (Table 4.4). Seven of these SNPs retained permuted p-values less than 0.05 (Table 4.4) maintaining their association with heat tolerance performance in sprint sled dogs. A pairwise comparison of D' among

these six SNPs proved greater than 0.90 for 65% of the pairwise comparisons, between 0.80 to 0.90 for an additional 21% of the comparisons, with the remaining 14% being between 0.60 to 0.70, defining the degree of linkage disequilibrium among the associated SNPs.

In addition to sprint Alaskan sled dogs, Alaskan Malamutes and German Shorthaired Pointers were similarly sequenced and analyzed for SNP frequencies and association to breed for 16 SNPs within and surrounding the *MYH9* gene. The 16 SNPs investigated in these two breeds were originally associated with heat tolerance performance in sprint sled dogs. These two ancestry breeds were shown in excess within sprint sled dogs spanning this genomic region. Therefore, we sought to identify similarities in allele frequencies between the ancestry breeds and either poor or elite performing heat tolerance sprint dogs. To this end, three of the 16 SNPs were associated to breed (Alaskan Malamutes as cases) with both raw and permuted p-values being below 0.05 (Table 4.5). SNPs *Chr10.31115476* and *Chr10.31123184* had similar allele frequencies between poorly performing heat tolerance sprint dogs (p-value 2.02×10^{-6} , *G* allele 0.840; p-value 0.0024, *T* allele 0.553) and Alaskan Malamutes (p-value 4.92×10^{-5} , *G* allele 0.700; 6.00×10^{-4} , *T* allele 0.700). SNP *Chr10.31156486*, which associated the *A* allele with the Alaskan Malamute breed, did not show significant association to heat tolerance performance after permutation testing. The SNP did have a slightly higher *A* allele frequency in poor heat tolerance sprint dogs (0.487) as opposed to elite heat tolerance sprint dogs (0.310).

Discussion

The Alaskan sled dog is the embodiment of a unique, genetically distinct breed developed solely by the selection and breeding of elite athletes [5]. They possess a dynamic population structure crafted from interbreeding purebred dogs possessing desirable performance traits and choosing specific attributes geared towards either short (up to 48 kilometers) or long (~1,609 kilometers) distance racing events. We demonstrated that a select group of 7,644 SNPs showing a high degree of differentiation are capable of clustering Alaskan sled dogs into sprint versus distance racing groups, corroborating our previous findings in which microsatellite data clustered dogs based on their racing style (Figure 4.2A).

We used the 7,644 SNP panel to model ancestry in the two admixed sled dog populations of distance and sprint dogs using the parental reference breeds of Alaskan Malamute, Siberian Husky, German Shorthaired Pointer, and Borzoi. The average genome-wide proportion of each ancestor showed that the Alaskan Malamute and Siberian Husky had a frequency increase of 7% and 5% respectively within the distance dogs, while the German Shorthaired Pointer increased by 10% within the sprint dogs (Table 4.1). Ten diploid ancestry states were generated using the four parental breeds, with homozygous states of the Alaskan Malamute, German Shorthaired Pointer, and Siberian Husky showing the greatest degree of frequency difference between the two sled dog groups (Table 4.1). This distinct difference in homozygous ancestry states between sprint and distance dogs in comparison to heterozygous states may be due to what breeders refer to as backcrossing (crossing a hybrid with one of its parents or an

individual genetically similar to the parent) or line-breeding (form of inbreeding selecting for a desirable trait seen in a common ancestor). This could be achieved by crossing Alaskan sled dogs back with purebreds of the parental breeds. Backcrossing and line-breeding have been fairly common over the past two decades, with sprint dogs being crossed to German Shorthaired and English Pointers [8] or by interbreeding recent hybrids. Since the frequency of these diploid states of Alaskan Malamute and Siberian Husky are more heavily selected for in distance dogs and German Shorthaired Pointer is selected for in sprint dogs, it would also suggest that the backcrossing or line-breeding occurs separately within the two sled dog populations.

Ancestry blocks were identified across the genome based on size and ancestry state, with the most common ancestry blocks within the sprint and distance dogs depicted in Figure 4.5. Strikingly, chromosomes such as 6, 25, and 30 are nearly identical within both sled dog groups in regard to the most common ancestral state, while other chromosomes such as 16, 24, and 29 show almost entire chromosomal differences. The sprint population had slightly fewer total and private ancestry blocks than the distance (6% and 4% less respectively) and a shorter average block length (200Kb less) (Table 4.2). These results can be explained by a higher degree of inbreeding within the distance dogs. Both previous microsatellite data and our current SNP data supported elevated levels of inbreeding in distance dogs in comparison to sprint dogs [5]. Distance dogs demonstrated a longer distance (Kb) to LD decay (r^2) than sprint dogs regardless of the r^2 threshold analyzed. The sprint dogs had a considerably shorter distance to LD decay ($r^2_{0.5} \approx 3\text{Kb}$; $r^2_{0.2} \approx 80\text{Kb}$), indicative of an out-crossed population, than any of the other

populations examined. Distance dogs also had a greater number of long (~12Mb) homozygous blocks, which was comparable to those found in the Siberian Husky.

Ancestry modeling highlighted 48 loci demonstrating positive ancestry selection within either the sprint or distance populations (Supplemental Table 4.3). To determine whether the selection of a particular ancestry held significance for differentiating between the sprint and distance dogs or was due to random retention of ancestral chromosomes, we identified coinciding loci of selective sweep and regions of interest produced through genome-wide association analyses.

Investigation of H_0 produced sixty regions of selective sweep with 87% of those occurring in distance sled dogs (Supplemental Table 4.1). This may be an indication of complex genetic interactions with smaller individual gene effect in distance sled dogs. We postulate that there are also more attributes under selective pressure within distance dogs, therefore allowing less leeway for variation. In theory, some of these regions may affect characteristics such as fur length and density or hardness of the pads of the feet which are strictly maintained in distance dogs due to the extreme nature of their racing conditions while they traverse over 1,609 kilometers. Modern sprint dogs are commonly found with shorter, less dense coats, which can be tolerated or compensated for during short racing distances (Figure 4.1). In total, five areas of selective sweep overlapped with four regions of ancestry selection, with potentially interesting candidate genes located in proximity (Table 4.3). Of these, chromosome 3 appeared to be a straightforward association of selective sweep and ancestry selection occurring in distance dogs. Therefore, data would suggest that this area of selection may be attributed to Siberian

Husky origins in which the ancestry selection occurred over the same region of selective sweep. The other three regions show a more complex ancestry pattern with selections occurring in both sprint and distance dogs encompassing multiple ancestry breeds. To follow up these findings, it would be beneficial to run a denser set of SNP markers in these regions on both sled dogs and the purebred breeds in order to gain a better understanding of breed influence and to locate a causative variant.

Genome-wide association analyses were used to identify genetic loci associated with either population variation between sprint and distance dogs or the performance attributes of endurance or heat tolerance. While sled dogs are an out-crossed population of dogs, breeding strategies and sampling groups of dogs from within kennels makes it impossible to avoid population structure and relatedness completely. Sampling sled dogs from high performing racing kennels compounded GWAS issues with a considerable lack of rank 3 dogs for either endurance or heat tolerance. It was necessary therefore, to group dogs ranked as 2 and 3 together as poorly performing which decreases phenotypic variation. The low numbers of poor performers also necessitated a higher degree of relationship among poorly performing individuals (two generations as opposed to three generations in elite dogs) which may introduce bias to the analysis. Measures taken to minimize these issues included choosing dogs unrelated to within a minimum of two generations and having approximately the same number of dogs from each kennel represented in both the cases and control groups as well as using software such as EMMAX, which corrects for both population relatedness and structure. Due to intense artificial selection for performance attributes in Alaskan sled dogs it was possible to

utilize relatively small sample sizes of both cases and controls in comparison to human GWAS studies as exemplified by previous GWAS of humans and dogs [48, 49]. This concentrated selection for performance, which has no deleterious effects such as a disease does and therefore no selective pressure against a specific allele, makes it possible for an allele having a major impact for performance phenotypes to be common.

GWAS results corresponded with three loci having an excess of ancestry selection (Table 4.3). Two of these regions, one on chromosome 11 and another on 32, were related to sled dog population differentiation. The *HINT1* gene on chromosome 11 allowed for the intriguing question about the effects of differing ancestries, which are under selection in the sprint and distance dogs, on their respective stress tolerance behaviors. Specifically, the Siberian Husky and German Shorthaired Pointer demonstrated positive selection within distance and sprint dogs respectively. While the Siberian Husky has been characterized as “stubborn and easily bored” despite its hardy working dog nature, the German Shorthaired Pointer is noted for its “ease of training and adaptability” along with its commitment to performing [50]. Anecdotally, the “mental toughness” (ability to deal with stress) of Alaskan sled dogs crossed with German Shorthaired Pointers is a topic of debate among sled dog drivers. The sequencing of exons of the *HINT1* gene within sled dogs, Siberian Huskies, and German Shorthaired Pointers produced seven non-coding variants. While these variants were not associated with either sprint or distance sled dogs, three of the SNPs segregated alleles in the two purebred breeds. Allele association with breed has been previously documented and used for phylogeny studies [17]. Therefore, the association of *HINT1* alleles with Siberian

Husky or German Shorthaired Pointer is not indicative in itself for associating the *HINT1* gene to behavioral differences between the breeds. Actual behavioral phenotypes would be necessary for future association of this gene to the breeds' behavioral attributes distinguishing these allelic trends from the random inheritance patterns unique to purebreds.

Our investigation into the ability of Alaskan sprint sled dogs to perform at warmer temperatures with little or no effect led us to the fine mapping of the *MYH9* gene. A clustering of four GWAS SNPs on chromosome 10, two of which were within the *MYH9* gene, demonstrated genome-wide significance (two highest SNPs with a p-value of 5.57×10^{-7}) to heat tolerance (Figure 4.4; Table 4.3). Previous research has associated an increase in myosin heavy chain with increased cardiac power output [41]. Another study found a decrease in myosin heavy chain and actin within injured mouse extensor muscles to account for approximately 58% reduction in isometric titanic force output [40, 41]. Most notably, the percentage of increase of the myosin heavy chain type II class A within muscle tissue experiencing elevated temperature (ET=37.5°C; Normal,N=34.2°C) correlated with the magnitude of increase for power output. The slight temperature elevation improved muscle fiber power output through an increased rate of anaerobic ATP turnover and muscle fiber conduction velocity (MFCV) [39]. However, the efficiency of muscle contraction actually decreased as temperature rose. Overall, the study concluded that fibers with a high proportion of myosin heavy chain type II class A were the most sensitive to temperature fluctuations [39]. We found that the *MYH9* gene overlapped with a selection of German Shorthaired Pointer ancestry in sprint sled dogs in

which the heat tolerance association was identified. Fine-mapping of this region identified seven SNPs associated with heat tolerance performance in sprint sled dogs (Table 4.4). The two most significantly associated SNPs (*Chr10.31105851*, 7.83×10^{-06} ; *Chr10.31115476*, 2.02×10^{-06}) were found 19Kb and 29 Kb upstream from the 5' end of the *MYH9* gene and in considerable linkage disequilibrium with each other ($D' = 1$) as well as with the other four SNPs ($D' \geq 0.871$). SNP *Chr10.31105851* lies within a conserved region of the dog, human, and mouse genomes. In a comparison of the analogous region of this SNP within the human genome, chromatin profiling of human skeletal muscle myoblasts show this region to be an active promoter site. Chromatin marks (chromatin sites identified as being regulated through histone-binding and DNA methylation) in seven other human cell types within this region showed signs of being strong enhancers. Likewise, a second SNP *Chr10.31121778*, which was in another conserved region upstream of the 5' end, was found in a strong enhancer region in an analogous comparison of the human genome [51, 52]. This allows us to postulate that promoter and enhancer regulatory sites may be the means by which the canine *MYH9* gene potentially affects heat tolerance performance in sprint sled dogs.

The ancestral breeds of Alaskan Malamute and German Shorthaired Pointer were both found in excess within sprint sled dogs within the *MYH9* region. SNPs *Chr10.31115476* and *Chr10.31123184* significantly differentiated between Alaskan Malamutes and German Shorthaired Pointers (Table 4.5). Specifically, sprint dogs with poor heat tolerance (*G*, 0.840; *T*, 0.553) had similar allele frequencies to Alaskan Malamutes (*G*, 0.700; *T*, 0.700). Elite heat tolerance sprint dogs had a decreased allele

frequency (G 0.357; T 0.262) at these two SNPs with German Shorthaired Pointers having the lowest allele frequencies (G 0.091; T 0.062) by comparison. Previous research associating muscle temperature elevation and power output, combined with our GWAS and fine-mapping results, make the *MYH9* gene an intriguing candidate potentially affecting the performance of heat tolerance in sprint sled dogs.

Corresponding *MYH9* gene allele frequencies between sprint dogs with poor heat tolerance and Alaskan Malamutes supports ancestral breed influence for performance attributes. Replication of our results in a larger cohort of Alaskan sled dogs would be an appropriate next step for this research with gene expression analysis as a follow-up.

With this study, we corroborated our previous findings demonstrating that Alaskan sled dogs do indeed divide into two separate, yet closely related, populations based on their racing style. A number of candidate genes potentially affecting performance were highlighted through GWAS and regions of selective sweep within the sled dogs. Component breeds were used for modeling ancestry within the Alaskan sled dogs and to pinpoint genomic regions of particular breed selection. We note that the purebred individuals utilized in this study were from conformational breeding stock as opposed to working dog stock which would be more appropriate, but more difficult to collect, when investigating performance attributes. Measures of selective sweep and genome-wide association were used to identify whether breed selection was related to random chromosomal inheritance or athletic attributes chosen for within the two sled dog populations. The ancestral breed linked to specific performance loci may be responsible for the origination of the genomic mechanism responsible for the desired performance

traits in Alaskan sled dogs. The association of SNPs upstream and within the *MYH9* gene to heat tolerance performance in sprint sled dogs demonstrated the value of researching performance mechanisms within Alaskan sled dogs. By combining admixture mapping to our GWAS analyses we are able to identify possible breed influence regarding the genes that are hypothetically under selection. This research provides a foundation of sled dog performance candidate genes, as well as their potential breed origin, for exploration as more statistically powerful genome-wide association and linkage studies are performed. Our suggestive evidence that the *MYH9* gene plays a role in sled dog heat tolerance performance sets the stage for future questions regarding this gene's influence on mammalian physiology as well as performance.

References

1. Collins MJC: **Dog Driver: A Guide for the Serious Musher**, First edn. Crawford, CO: Alpine Publications; 1991.
2. Rennick P (ed.): **Dogs of the North**. Anchorage: Alaska Geographic Society; 1987.
3. Vaudrin B (ed.): **Racing Alaskan Sled Dogs**. Anchorage, AK: Alaska Northwest Publishing Company; 1977.
4. Wendt R: **Alaska Dog Mushing Guide: Facts and Legends**, Fifth edn. Fairbanks: Goldstream Publications; 1999.
5. Huson HJ, Parker HG, Runstadler J, Ostrander EA: **A genetic dissection of breed composition and performance enhancement in the Alaskan sled dog**. *BMC Genetics* 2010, **11**:71.
6. **The Official Site of the Iditarod** [<http://www.iditarod.com/>] Accessed Aug 2010.

7. **Yukon Quest; Sled Dog Race; Alaska, Yukon** [<http://www.yukonquest.com/>] Accessed Aug 2010.
8. **Alaska Dog Musher's Association** [<http://www.sleddog.org/2010/03/21/open-north-american-results-day-3-final/>] Accessed Jan 2011.
9. **Illumina** [<http://www.illumina.com>] Accessed April 15, 2011.
10. **Affymetrix** [<http://www.affymetrix.com/>] Accessed April 15, 2011.
11. **PLINK_v1.07** [<http://pngu.mgh.harvard.edu/purcell/plink/>] Accessed March 1, 2011.
12. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ *et al*: **PLINK: a tool set for whole-genome association and population-based linkage analyses**. *American Journal of Human Genetics* 2007, **81**(3):559-575.
13. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies**. *Nature Genetics* 2006, **38**(8):904-909.
14. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K, Brisbin A, Parker HG, vonHoldt BM *et al*: **A simple genetic architecture underlies morphological variation in dogs**. *PLoS Biology* 2010, **8**(8):e1000451.
15. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA: **Extensive and breed-specific linkage disequilibrium in *Canis familiaris***. *Genome Research* 2004, **14**(12):2388-2396.
16. Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne RK: **Linkage disequilibrium and demographic history of wild and domestic canids**. *Genetics* 2009, **181**(4):1493-1505.
17. Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A *et al*: **Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication**. *Nature* 2010, **464**(7290):898-902.
18. **Efficient Mixed_model Association eXpediated (EMMAX)** [<http://genetics.cs.ucla.edu/emmax/install.html>] Accessed March 1, 2011.

19. **Applied Biosystems**
[<http://www.appliedbiosystems.com/absite/us/en/home.html>] Accessed May 2011.
20. **EMD Chemicals** [<http://www.emdchemicals.com/>] Accessed May 2011.
21. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Research* 1998, **8**(3):175-185.
22. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Research* 1998, **8**(3):195-202.
23. Bhangale TR, Stephens M, Nickerson DA: **Automating resequencing-based detection of insertion-deletion polymorphisms.** *Nature Genetics* 2006, **38**(12):1457-1462.
24. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21**(2):263-265.
25. Scheet P, Stephens M: **A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase.** *American Journal of Human Genetics* 2006, **78**(4):629-644.
26. Tang H, Choudhry S, Mei R, Morgan M, Rodriguez-Cintron W, Burchard EG, Risch NJ: **Recent genetic selection in the ancestral admixture of Puerto Ricans.** *American Journal of Human Genetics* 2007, **81**(3):626-633.
27. Tang H, Coram M, Wang P, Zhu X, Risch N: **Reconstructing genetic ancestry blocks in admixed individuals.** *American Journal of Human Genetics* 2006, **79**(1):1-12.
28. Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF: **A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping.** *American Journal of Human Genetics* 2006, **79**(4):640-649.
29. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M: **Genome-wide association studies in diverse populations.** *Nature Review Genetics* 2010, **11**(5):356-366.
30. Cheng CY, Reich D, Wong TY, Klein R, Klein BE, Patterson N, Tandon A, Li M, Boerwinkle E, Sharrett AR *et al*: **Admixture mapping scans identify a locus affecting retinal vascular caliber in hypertensive African Americans: the**

- Atherosclerosis Risk in Communities (ARIC) study.** *PLoS Genetics* 2010, **6**(4):e1000908.
31. Pritchard JK, Przeworski M: **Linkage disequilibrium in humans: models and data.** *American Journal of Human Genetics* 2001, **69**(1):1-14.
 32. Gaut BS, Long AD: **The lowdown on linkage disequilibrium.** *Plant Cell* 2003, **15**(7):1502-1506.
 33. **American Kennel Club- AKC.org** [<http://www.akc.org/>] Accessed Jan 2011.
 34. Sharer JD, Shern JF, Van Valkenburgh H, Wallace DC, Kahn RA: **ARL2 and BART enter mitochondria and bind the adenine nucleotide transporter.** *Molecular Biology of the Cell* 2002, **13**(1):71-83.
 35. **University of California, Santa Cruz Genome Browser** [<http://genome.ucsc.edu/>] Accessed April 2011.
 36. **National Center for Biotechnology Information** [<http://www.ncbi.nlm.nih.gov/>] Accessed April 2011.
 37. Varadarajulu J, Lebar M, Krishnamoorthy G, Habelt S, Lu J, Bernard Weinstein I, Li H, Holsboer F, Turck CW, Touma C: **Increased anxiety-related behaviour in Hint1 knockout mice.** *Behavioral Brain Research* 2011, **220**(2):305-311.
 38. Barbier E, Wang JB: **Anti-depressant and anxiolytic like behaviors in PKCI/HINT1 knockout mice associated with elevated plasma corticosterone level.** *BMC Neuroscience* 2009, **10**:132.
 39. Gray SR, De Vito G, Nimmo MA, Farina D, Ferguson RA: **Skeletal muscle ATP turnover and muscle fiber conduction velocity are elevated at higher muscle temperatures during maximal power output development in humans.** *American Journal of Physiology- Regulatory, Integrative and Comparative Physiology* 2006, **290**(2):R376-382.
 40. Ingalls CP, Warren GL, Armstrong RB: **Dissociation of force production from MHC and actin contents in muscles injured by eccentric contractions.** *Journal of Muscle Research and Cell Motility* 1998, **19**(3):215-224.
 41. Burniston JG: **Adaptation of the rat cardiac proteome in response to intensity-controlled endurance exercise.** *Proteomics* 2009, **9**(1):106-115.

42. Kwak GH, Kim JR, Kim HY: **Expression, subcellular localization, and antioxidant role of mammalian methionine sulfoxide reductases in *Saccharomyces cerevisiae*.** *BMB Reports* 2009, **42**(2):113-118.
43. Suckale J, Solimena M: **The insulin secretory granule as a signaling hub.** *Trends Endocrinology & Metabolism* 2010, **21**(10):599-609.
44. Stark Z, Bruno DL, Mountford H, Lockhart PJ, Amor DJ: **De novo 325 kb microdeletion in chromosome band 10q25.3 including ATRNL1 in a boy with cognitive impairment, autism and dysmorphic features.** *European Journal of Human Genetics* 2010, **53**(5):337-339.
45. Luciano M, Hansell NK, Lahti J, Davies G, Medland SE, Raikonen K, Tenesa A, Widen E, McGhee KA, Palotie A *et al*: **Whole genome association scan for genetic polymorphisms influencing information processing speed.** *Biological Psychology* 2011, **86**(3):193-202.
46. Belrose G, Gross A, Olindo S, Lezin A, Dueymes M, Komla-Soukha I, Smadja D, Tanaka Y, Willems L, Mesnard JM *et al*: **Opposite effects of valproate on Tax and HBZ expressions in T-lymphocytes from HTLV-1 asymptomatic carriers and HAM/TSP patients.** *Blood* 2011.
47. **HTLV-1 associated myelopathy/tropical spastic paraparesis**
[<http://rarediseases.info.nih.gov/GARD/Disease.aspx?PageID=4&diseaseID=8208>] Accessed April 2011.
48. Hakonarson H, Grant SF: **Planning a genome-wide association study: Points to consider.** *Annals of medicine* 2011, **43**(6):451-460.
49. Parker HG, Shearin AL, Ostrander EA: **Man's best friend becomes biology's best in show: genome analyses in the domestic dog.** *Annual review of genetics* 2010, **44**:309-336.
50. Wilcox B, Walkowicz: **Atlas of Dog Breeds of the World**, vol. 5th Edition, 5th Edition edn. Neptune City: T.F.H. Publications, Inc; 1995.
51. Ernst J, Kellis M: **Discovery and characterization of chromatin states for systematic annotation of the human genome.** *Nature Biotechnology* 2010, **28**(8):817-825.
52. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al*: **Mapping and analysis of chromatin state dynamics in nine human cell types.** *Nature* 2011, **473**(7345):43-49.

Figures

Figure 4.1: Alaskan sled dogs are a mixed breed dog selected strictly for their performance. A) Left column: Distance racing dogs. B) Right column: Sprint racing dogs.

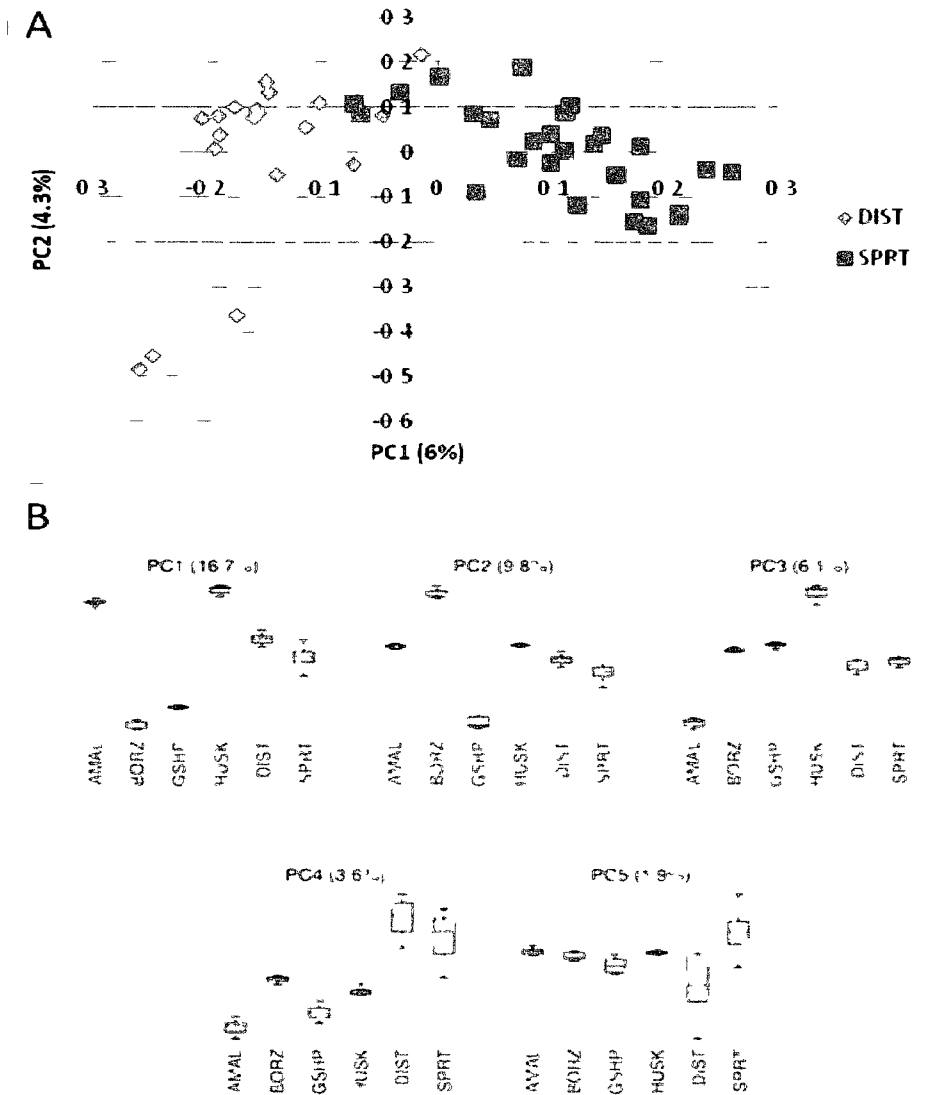


Figure 4.2: Principle component analysis plots of Alaskan sled dogs (A & B) and four ancestry reference breeds (B) using a panel of 7K highly ($F_{ST} > 0.35$) informative SNPs.

A) Alaskan sled dogs from either distance (DIST-blue) or sprint (SPRT-red) racing kennels. B) Four ancestry reference breeds including the Alaskan Malamute (AMAL), Borzoi (BORZ), German Shorthaired Pointer (GSHP), and Siberian Husky (HUSK) as well as Alaskan sled dogs divided into their two populations of distance (DIST) and sprint (SPRT).

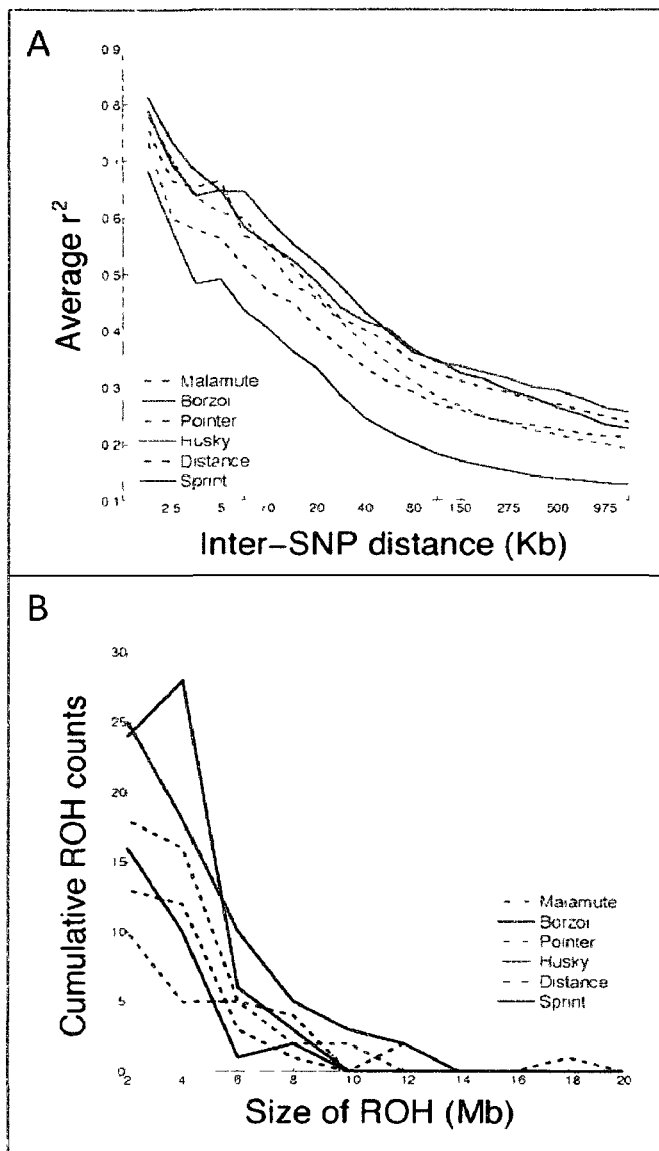


Figure 4.3: The estimated decay of linkage disequilibrium and degree of autozygosity among Alaskan sled dogs and their four ancestral component breeds. Alaskan sled dogs are divided into their two respective racing style populations of “distance” (dashed black line) and “sprint” (solid black line) and compared with their four ancestral reference populations of Alaskan Malamute (dashed blue line), Borzoi (solid blue line), German Shorthaired Pointer (dashed red line), and Siberian Husky (solid red line). A) The decay of linkage disequilibrium (LD) is estimated from the distance at which the genotypic association, r^2 , reaches a threshold of 0.5. B) The degree of autozygosity is determined through the cumulative number of runs of homozygosity (ROH) of various length (Mb).

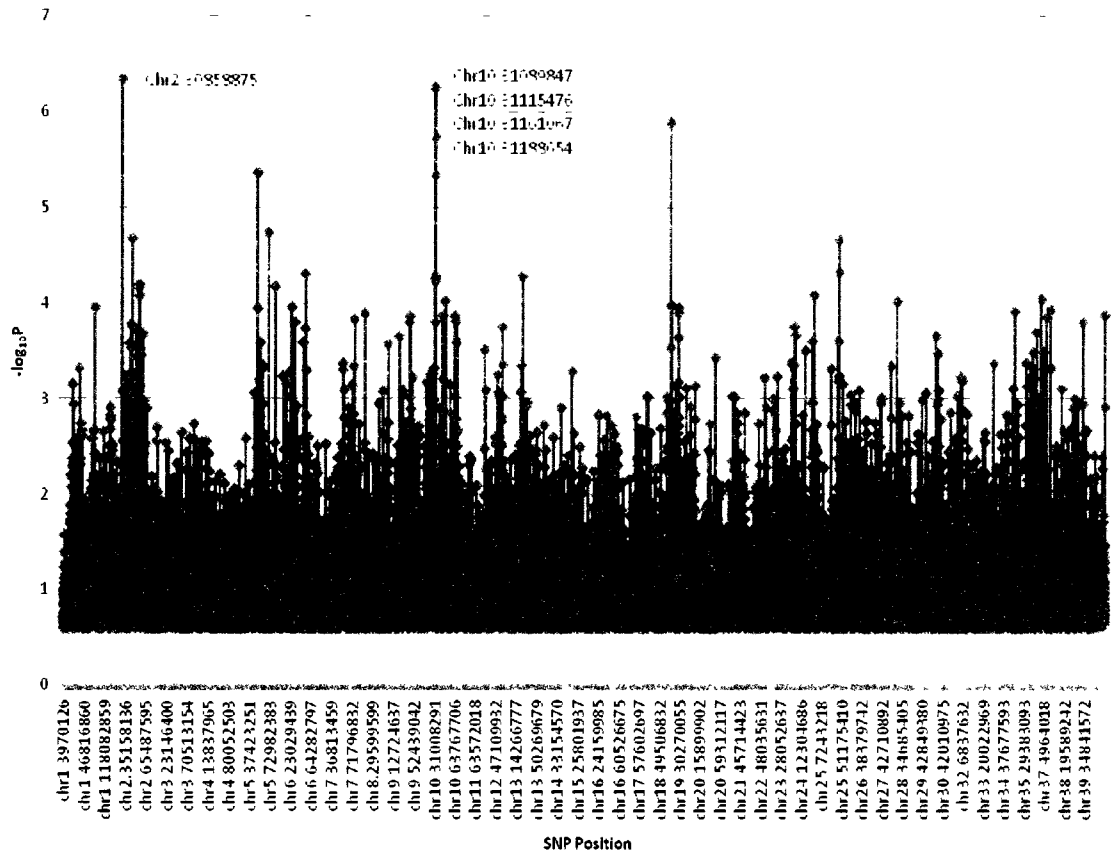


Figure 4.4: Genome-wide association plot of elite versus poorly performing heat tolerance sprint sled dogs.

Two genomic loci located on chromosomes 2 and 10 were identified with SNPs having genome-wide significance in a comparison of 21 elite heat tolerance sprint dogs to 17 poor heat tolerance sprint dogs. A panel of 115,425 SNPs spanning all canine autosomes and the X chromosome was used in the association analysis. The x-axis denotes SNP positions in increasing genomic order from chromosome 1 through 38 and the X chromosome. The y-axis plots the $-\log_{10} p$ -value as determined in an association analysis using the program EMMAX.

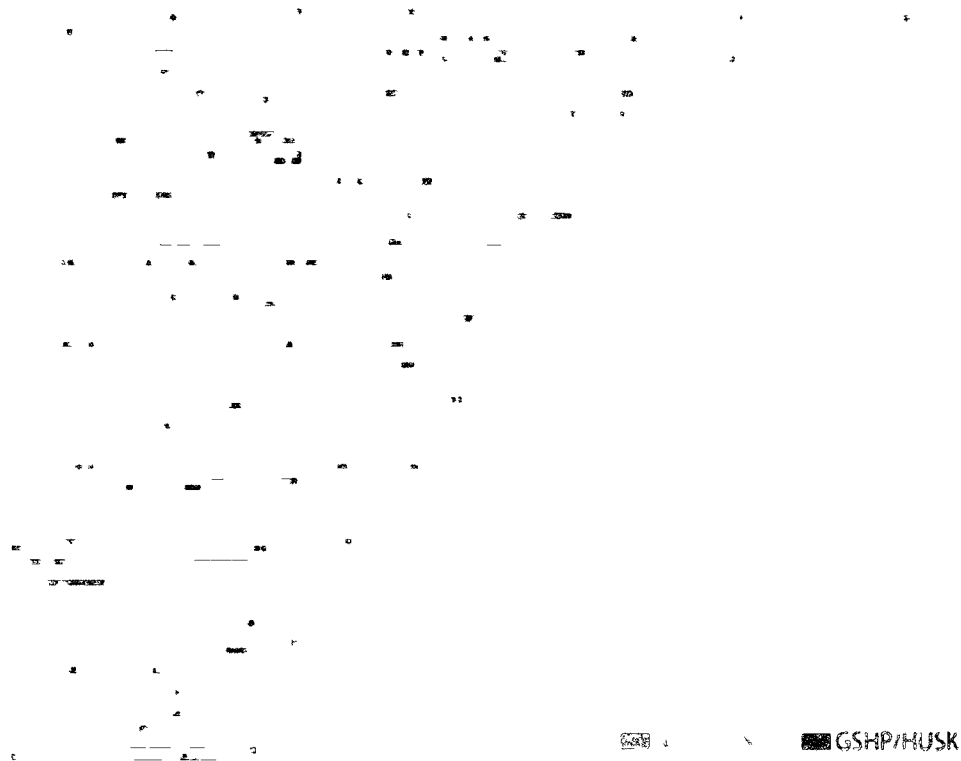


Figure 4.5: A visual comparison of the most prevalent diploid state ancestry blocks across the genome of the sprint and distance sled dogs.

Individual chromosomes are plotted horizontally with the x-axis denoting genomic position (Mb). The most common diploid ancestry blocks across the genome are visualized using the following color scheme: Homozygous state Alaskan Malamute (AMAL)- teal blue; Homozygous state German Shorthaired Pointer (GSHP)- green; Alaskan Malamute/German Shorthaired Pointer (AMAL/GSHP)- pink; German Shorthaired Pointer/Siberian Husky (GSHP/HUSK)- blue; Alaskan Malamute/Siberian Husky (AMAL/HUSK)- yellow; Homozygous state Siberian Husky (HUSK)- orange.

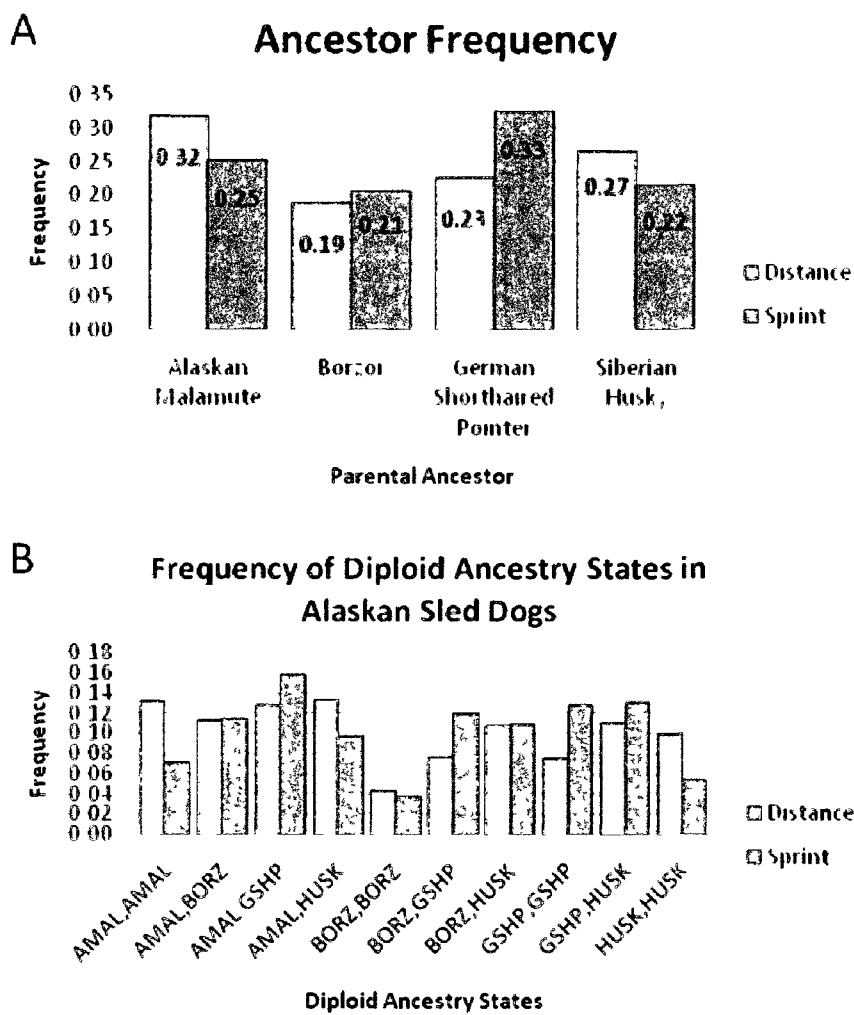


Figure 4.6: A comparison of the genome-wide frequency of four ancestral reference breeds within the distance and sprint sled dog populations.

A) The genome-wide proportion of the individual ancestral reference breeds of Alaskan Malamute, Borzoi, German Shorthaired Pointer, and Siberian Husky within the distance (blue) and sprint (red) populations. B) The genome-wide proportion of diploid ancestry states within the distance (blue) and sprint (red) populations.

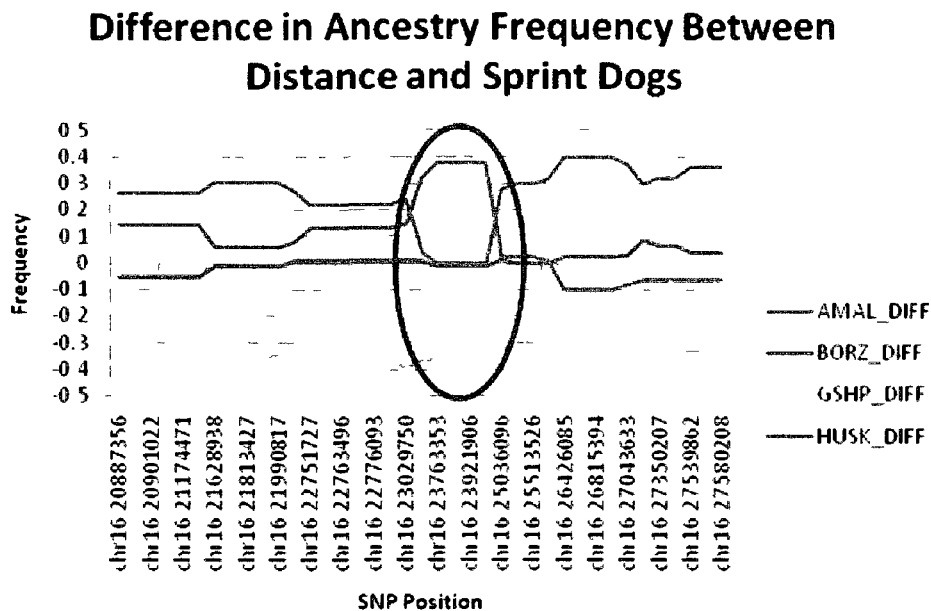


Figure 4.7: The difference in frequency of the four ancestral breeds between the distance and sprint sled dogs on chromosome 16.

The difference in frequency scores (y-axis) between distance and sprint dogs of each ancestral breed is plotted across SNPs (x-axis) evaluated on chromosome 16. A more positive frequency difference corresponds to a higher selection of the ancestral breed within the distance population while the more negative frequency difference corresponds to a greater selection of the ancestral breed within the sprint population. The region within the red circle denotes an area highlighted as being in the top 5% of genomic regions demonstrating the greatest degree of ancestry selection between the sprint and distance sled dogs as well as corresponding to a region of selective sweep within distance dogs. Abbreviations: AMAL- Alaskan Malamute (blue), BORZ- Borzoi (red), GSHP- German Shorthaired Pointer (green), HUSK- Siberian Husky (purple).

Tables

Table 4.1: The genome-wide frequency (f) of individual ancestral populations and their respective diploid ancestry states within the distance and sprint sled dog populations.

Diploid Ancestry State frequency matrix ²	Distance Sled Dogs ¹			
	Alaskan Malamute	Borzoi	German Shorthaired Pointer	Siberian Husky
Alaskan Malamute	0.1328	0.1136	0.1279	0.1337
Borzoi		0.043	0.0768	0.1074
German Shorthaired Pointer			0.0754	0.1096
Siberian Husky				0.0992
Total $f(\text{Distance})^3$	0.3181	0.1884	0.2271	0.2664
Diploid Ancestry State frequency matrix ²	Sprint Sled Dogs ¹			
	Alaskan Malamute	Borzoi	German Shorthaired Pointer	Siberian Husky
Alaskan Malamute	0.071	0.1153	0.1582	0.0966
Borzoi		0.0373	0.1194	0.1093
German Shorthaired Pointer			0.1275	0.1313
Siberian Husky				0.0535
Total $f(\text{Sprint})^3$	0.2533	0.2059	0.3251	0.2158

1. A total of 19 distance dogs and 27 sprint dogs, all unrelated at the grandparent generation, were used to generate population frequencies.
2. A matrix of the diploid ancestry states with their respective genome-wide frequencies (f).
3. The genome-wide frequency (f) of individual ancestral populations shown at the head of each column.

Table 4.2: The overall number, median length, and genome-wide frequency (f) of diploid ancestry blocks found exclusive to either the distance (A) or sprint (B) sled dog populations.

A. Distance	Alaskan Malamute	Borzoi	German Shorthaired Pointer	Siberian Husky
Number of ancestry blocks (447)				
Alaskan Malamute	45	35	80	72
Borzoi		12	35	44
German Shorthaired Pointer			40	52
Siberian Husky				32
Median length (Kb) of ancestry block				
Alaskan Malamute	2354	1660	1464	1428
Borzoi		1046	1062	1151
German Shorthaired Pointer			1740	1083
Siberian Husky				1581
f (ancestry block)				
Alaskan Malamute	0.101	0.078	0.179	0.161
Borzoi		0.027	0.078	0.098
German Shorthaired Pointer			0.09	0.116
Siberian Husky				0.071
B. Sprint	Alaskan Malamute	Borzoi	German Shorthaired Pointer	Siberian Husky
Number of ancestry blocks (392)				
Alaskan Malamute	17	37	54	58
Borzoi		9	65	46
German Shorthaired Pointer			39	44
Siberian Husky				23
Median length (Kb) of ancestry block				
Alaskan Malamute	1671	1054	939	967
Borzoi		1112	926	1112
German Shorthaired Pointer			996	1333
Siberian Husky				1891
f (ancestry block)				
Alaskan Malamute	0.043	0.094	0.138	0.148
Borzoi		0.023	0.166	0.117
German Shorthaired Pointer			0.01	0.112
Siberian Husky				0.059

Method of Identification ¹	Chromosome	Start Position (bp)	End Position (bp)	Block_length (bp)	Sled Dog Population ²
Selective Sweep, SABER	3	71896408	71898732	2,324	Distance
Selective Sweep, SABER	3	72727082	72784438	57,356	Distance
GWAS	3	82650187			
Selective Sweep	3	83775932	83798854	22,922	Distance
Selective Sweep, SABER	10	11081762	11121003	39,241	Distance
GWAS, SABER	10	31089847	31188654	98,807	
SABER	11	18482294	23584745	5,102,451	Distance
GWAS, SABER	11	22331950	23117401	785,451	
Selective Sweep, SABER	16	23391731	23391985	254	Distance
Selective Sweep, SABER	28	29046328	29143901	97,573	Distance
GWAS, SABER	32	8774288			

Table 4.3: Tabulation of the genetic loci demonstrating the highest degree of interest for population differentiation or performance association within Alaskan sled dogs.

Ancestry Population ³	GWAS Association ⁴	Performance Candidate Genes ⁵	Number of Genes within Region ⁶
Siberian Husky		<i>SLC2A9</i>	11
Siberian Husky		<i>SLC2A9</i>	14
	Population	<i>ARL2BP</i>	2
		<i>ARL2BP</i>	2
German Shorthaired Pointer		<i>MSRB3</i>	15
German Shorthaired Pointer	Heat Tolerance	<i>MYH9</i>	34
		<i>FBN1, FBN2, ACSL6.</i>	
Siberian Husky		<i>SLC27A6, HINT1</i>	51
German Shorthaired Pointer/Siberian Husky	Population	<i>HINT1</i>	25
German Shorthaired Pointer		<i>PTPRN2</i>	13
German Shorthaired Pointer		<i>ATRNL1</i>	12
German Shorthaired Pointer	Population	<i>HBZ</i>	11

1. Genomic regions of interest were determined by demonstrating an excess of breed ancestry (SABER), selective sweep, or genome-wide association.
2. Identifies either the sprint or distance sled dog population as being associated with the genomic region of interest.
3. Identifies the ancestral reference breed population found in excess within the designated sled dog population.
4. Identifies whether the genome-wide association study associated the genomic region to either a sled dog population or a specific athletic trait.
5. Candidate genes identified at genomic regions of interest potentially affecting athletic performance.
6. The number of human genes annotated within the genomic region of interest as well as 1MB upstream and 1MB downstream of said region.

Table 4.4: SNPs within and surrounding the *MYH9* gene which are associated with heat tolerance performance in sprint racing Alaskan sled dogs.

Position	Alleles ¹	Poor HT ² MAF	Elite HT ³ MAF	Poor HT	Permutation	
				Assoc Allele	p-value	p-value
31089847	A:C	0.240	0.700	A	1.28E-05	0.0008
31105851	A:G	0.222	0.643	A	7.83E-06	0.0004
31115476	G:A	0.160	0.643	G	2.02E-06	0.0001
31121778	A:G	0.338	0.700	A	2.00E-04	0.0082
31123184	C:T	0.553	0.262	T	0.0024	0.0645
31128725	G:A	0.320	0.643	G	0.002	0.0612
31134023	C:A	0.320	0.643	C	0.002	0.0612
31145292	G:A	0.320	0.650	G	0.0018	0.054
31156535	C:A	0.132	0.350	C	0.0058	0.2197
31161067	C:T	0.263	0.643	C	5.48E-05	0.0024
31172587	T:C	0.385	0.690	T	0.0014	0.0425
31176097	C:T	0.270	0.619	C	2.00E-04	0.0086
31188654	G:A	0.365	0.619	G	0.0083	0.3093
31234860	G:A	0.840	1.000	A	0.0067	0.2402

1. Major allele: Minor allele.
2. Frequency of the minor allele in 32 poorly performing heat tolerance sprint dogs (cases).
3. Frequency of the minor allele in 21 elite performing heat tolerance sprint dogs (controls).

Table 4.5: *MYH9* gene SNPs associated to either the Alaskan Malamute or German Shorthaired Pointer breeds.

Position	Alleles ¹	AMAL ² MAF	GSHP ³ MAF	AMAL Assoc		Permutation
				Allele	p-value	p-value
31115476	A:G	0.700	0.091	G	4.92E-05	0.0001
31123184	C:T	0.700	0.062	T	6.00E-04	0.0047
31156486	G:A	0.700	0.071	A	0.0013	0.0093

1. Major allele: Minor allele
2. Frequency of the minor allele in six Alaskan Malamutes (cases).
3. Frequency of the minor allele in eight German Shorthaired Pointers (controls).

Appendix

Table 4A.1: Genomic loci demonstrating the greatest degree of selective sweep within distance and sprint sled dogs.

Index	Chromosome	Start_bp	End_bp	Block_length (bp)	# SNPs in block	Population	Observed Heterozygosity	
							Distance	Sprint
1	1	12096874	12099093	2,219	2	Distance	0	0.5
2	1	25490435	25495415	4,980	2	Distance	0	0.5833
3	3	8978705	9004559	25,854	4	Distance	0	0.5833
4	3	71896408	71898732	2,324	2	Distance	0	0.5833
5	3	72727082	72784438	57,356	5	Distance	0	0.75
6	3	75921529	75938480	16,951	2	Sprint	0.625	0
7	3	83775932	83798854	22,922	2	Distance	0	0.8333
8	3	83880428	83902791	22,363	4	Distance	0	0.6667
9	5	43612855	43624678	11,823	2	Distance	0	0.5
10	5	44057317	44131651	74,334	3	Distance	0	0.5833
11	7	29601768	29645316	43,548	3	Distance	0	0.5
12	8	28609050	28614015	4,965	2	Sprint	0.5	0
13	10	11081762	11121003	39,241	4	Distance	0	0.5
14	10	27807343	27866141	58,798	2	Distance	0	0.6667
15	11	4022660	4170651	147,991	2	Distance	0	0.5833
16	11	75920799	75930341	9,542	2	Distance	0	0.5
17	12	4115578	4136210	20,632	2	Distance	0	0.5833
18	13	20301899	20309614	7,715	2	Sprint	0.625	0
19	13	24363484	24369167	5,683	2	Distance	0	0.5
20	13	27026973	27094574	67,601	4	Sprint	0.5	0
21	13	28474850	28491108	16,258	2	Distance	0	0.5
22	14	25734879	25739306	4,427	2	Sprint	0.625	0
23	14	42007812	42011513	3,701	2	Sprint	0.625	0
24	15	20230770	20372456	141,686	4	Distance	0	0.5
25	15	20466195	20502372	36,177	3	Distance	0	0.5
26	15	20519390	20532352	12,962	2	Distance	0	0.75
27	15	21889570	21896944	7,374	2	Distance	0	0.6667
28	15	25173358	25207024	33,666	3	Distance	0	0.5833
29	15	54947175	54970726	23,551	3	Distance	0	0.5
30	15	59792355	59810546	18,191	2	Distance	0	0.6667
31	16	12662362	12704048	41,686	4	Distance	0	0.5
32	16	23391731	23391985	254	2	Distance	0	0.5
33	17	8158751	8170123	11,372	2	Sprint	0.75	0
34	18	6675182	6786305	111,123	2	Distance	0	0.6364
35	18	8258000	8325293	67,293	4	Distance	0	0.6667
36	18	10743899	10777232	33,333	3	Distance	0	0.5
37	18	26066128	26072978	6,850	2	Sprint	0.5	0
38	18	31273089	31280656	7,567	2	Distance	0	0.5
39	18	45203778	45209080	5,302	2	Distance	0	0.5833
40	19	47654894	47657540	2,646	2	Distance	0	0.5833
41	20	18064338	18099556	35,218	2	Distance	0	0.5
42	21	9868044	9879444	11,400	3	Distance	0	0.75
43	21	18843588	18864435	20,847	2	Distance	0	0.5
44	21	33148596	33176168	27,572	2	Distance	0	0.75
45	22	31461195	31465512	4,317	2	Distance	0	0.5833
46	22	31506285	31533138	26,853	3	Distance	0	0.5833
47	25	6068291	6115389	47,098	3	Distance	0	0.5
48	25	20144057	20315060	171,003	6	Distance	0	0.5
49	25	23615778	23622576	6,798	2	Distance	0	0.6667
50	25	26649533	26745543	96,010	5	Distance	0	0.75
51	25	31078190	31085035	6,845	3	Distance	0	0.75
52	27	6143935	6156585	12,650	2	Distance	0	0.5
53	28	29046328	29143901	97,573	4	Distance	0	0.6667
54	29	8105695	8142652	36,957	5	Distance	0	0.6667
55	30	25104616	25224575	119,959	6	Distance	0	0.6364
56	30	25401287	25406897	5,610	2	Distance	0	0.5833
57	34	5280024	5306200	26,176	3	Distance	0	0.6667
58	34	26189017	26208328	19,311	4	Distance	0	0.5
59	35	12180438	12238951	58,513	4	Distance	0	0.75
60	37	13284071	13294062	9,991	2	Distance	0	0.5833

Table 4A.2: Single-nucleotide polymorphism markers showing genome-wide significance in association with either population differentiation between sprint and distance sled dogs or heat tolerance in sprint sled dogs.

Phenotype	Racing Style	Chr	Position (bp)	SNP ID	EMMAX_P ¹	EMMAX_ '-logP	PLINK	
							PLINK_P ²	Permutated P ³
Population		3	82650187	chr3.82650187	1.03E-07	6.988887349	1.14E-11	1.00E-06
Population		7	30793374	chr7.30793374	3.27E-06	5.485456735	3.76E-07	1.00E-06
Heat_Tolerance	Sprint	10	31089847	chr10.31089847	1.8078E-06	5.742856832	3.32E-06	2.00E-06
Heat_Tolerance	Sprint	10	31115476	chr10.31115476	5.5671E-07	6.254372012	4.53E-06	5.00E-06
Heat_Tolerance	Sprint	10	31161067	chr10.31161067	5.5671E-07	6.254372012	4.53E-06	5.00E-06
Heat_Tolerance	Sprint	10	31188654	chr10.31188654	4.5326E-06	5.343652913	4.48E-05	1.20E-05
Population		11	22331950	chr11.22331950	1.00E-06	5.998538853	1.44E-06	2.00E-06
Population		11	23117401	chr11.23117401	1.02E-06	5.989601166	1.76E-08	1.00E-06
Population		15	54567557	chr15.54567557	2.36E-06	5.626330451	1.33E-08	1.00E-06
Population		17	48764612	chr17.48764612	4.68E-06	5.330200038	1.41E-06	3.00E-06
Population		32	8774288	chr32.8774288	1.88E-06	5.72566013	7.52E-08	1.00E-06

1. P-value obtained through association analysis using EMMAX software which accounts for population stratification and relatedness.
2. P-value obtained through association analysis using PLINK software.
3. P-value obtained through association analysis using PLINK software and run through a series of adaptive permutation testing (1000000 permutations) to determine significance of stability.

Table 4A.3: The top 5th percentile of SNPs showing the highest degree of ancestral breed differentiation between the distance and sprint sled dog populations.

Index	Chromosome	Start_bp	End_bp	Block_Lengt h (bp)	Ancestry	Positive Selection Within	Frequency Difference ¹
1	2	13356758	13976158	619,400	Alaskan Malamute	Distance	0.341
2	3	53798499	53840474	41,975	Siberian Husky	Distance	0.348
3	3	71155871	72784438	1,628,567	Siberian Husky	Distance	0.382
4	3	89892953	91314456	1,421,503	German Shorthaired Pointer	Sprint	0.390
5	4	61053588	61158405	104,817	Borzoi	Distance	0.372
6	5	28032072	32781895	4,749,823	Alaskan Malamute	Distance	0.404
7	5	32367525	32781895	414,370	Borzoi	Sprint	0.339
8	5	46041552	47375859	1,334,307	Borzoi	Sprint	0.405
9	5	48151351	48637045	485,694	Alaskan Malamute	Distance	0.373
10	5	65327351	65354461	27,110	German Shorthaired Pointer	Distance	0.336
11	5	66488558	66496108	7,550	Siberian Husky	Distance	0.355
12	5	75113818	Single SNP ¹		German Shorthaired Pointer	Sprint	0.373
13	7	77539655	Single SNP ²		Alaskan Malamute	Distance	0.333
14	9	37734122	38673279	939,157	Alaskan Malamute	Distance	0.352
15	9	62664146	64350642	1,686,496	Siberian Husky	Distance	0.391
16	10	10302292	11405132	1,102,840	German Shorthaired Pointer	Sprint	0.347
17	10	17256141	18645753	1,389,612	Alaskan Malamute	Distance	0.428
18	11	18482294	23584745	5,102,451	Siberian Husky	Distance	0.510
19	11	21562728	25706808	4,144,080	German Shorthaired Pointer	Sprint	0.424
20	11	66332944	66718636	385,692	German Shorthaired Pointer	Sprint	0.336
21	13	56695005	59845783	3,150,778	Alaskan Malamute	Distance	0.457
22	14	3701855	Single SNP ²		German Shorthaired Pointer	Sprint	0.344
23	15	23142619	24225236	1,082,617	German Shorthaired Pointer	Sprint	0.350
24	16	20609137	23926737	3,317,600	German Shorthaired Pointer	Sprint	0.398
25	16	23763353	23926737	163,384	Siberian Husky	Distance	0.378
26	16	26426085	27592633	1,166,548	Alaskan Malamute	Distance	0.397
27	16	27539862	27969735	429,873	German Shorthaired Pointer	Sprint	0.337
28	16	32342221	40340989	7,998,768	German Shorthaired Pointer	Sprint	0.427
29	16	50952939	51153803	200,864	Alaskan Malamute	Distance	0.349
30	17	42126493	42347263	220,770	Siberian Husky	Distance	0.344
31	17	43289243	Single SNP		Alaskan Malamute	Distance	0.342
32	17	47995890	48079454	83,564	Siberian Husky	Distance	0.360
33	21	9606104	9664208	58,104	German Shorthaired Pointer	Sprint	0.334
34	22	42507517	42649162	141,645	Borzoi	Distance	0.391
35	23	14275103	15078147	803,044	Alaskan Malamute	Distance	0.404
36	23	18679285	18728624	49,339	German Shorthaired Pointer	Sprint	0.385
37	23	27490920	27543574	52,654	Alaskan Malamute	Distance	0.354
38	24	3289924	5179711	1,889,787	Siberian Husky	Distance	0.338
39	28	26646832	29631108	2,984,276	German Shorthaired Pointer	Sprint	0.421
40	28	33176700	34837841	1,661,141	Alaskan Malamute	Distance	0.404
41	28	34743651	37406408	2,662,757	German Shorthaired Pointer	Sprint	0.364
42	28	38689587	41105630	2,416,043	Alaskan Malamute	Distance	0.352
43	28	39797816	39935167	137,351	German Shorthaired Pointer	Sprint	0.337
44	29	18870564	24947366	6,076,802	German Shorthaired Pointer	Sprint	0.376
45	32	7978135	8980689	1,002,554	German Shorthaired Pointer	Sprint	0.355
46	32	20673802	21772514	1,098,712	Siberian Husky	Distance	0.344
47	34	13477200	14137133	659,933	Siberian Husky	Distance	0.360
48	34	18641563	19272322	630,759	Alaskan Malamute	Distance	0.482

1. The difference in frequency at specified location between the distance and sprint populations of the ancestral reference breed demonstrating the greatest positive selection within a region.

Chapter 5

Conclusion

Population Structure and Inbreeding

The Alaskan sled dog is a genetically distinct breed of dog that has developed through the selection and breeding of individuals based solely on their athletic prowess (Chapter 2). Its creation happened without the implementation of breed standards for size or appearance, or the closing of the breeding population, both of which are customary for AKC-recognized purebred breeds. Nonetheless, they have reached a level of genetic distinctiveness comparable to that of a purebred dog. Overall, they possess a dynamic population structure crafted from interbreeding purebred dogs possessing desirable performance traits and choosing specific attributes geared towards either short (up to 48 kilometers) or long (~1,609 kilometers) distance racing events.

Both cluster analysis using 96 microsatellite-based markers (Figure 2.4A) and principle component analysis using a 115K SNP panel (Figure 4.2A) demonstrated a segregation of Alaskan sled dogs into two populations reflective of their racing style, sprint or distance. Cluster analysis broke these two groups down further into eight sub-populations: four sprint and four distance (Figure 2.4G). Distance groups were determined to be more genetically distinct and kennel-specific, therefore suggesting a breeding practice generally confined to individual kennels.

Overall, Alaskan sled dogs had a fivefold lower inbreeding value than the average domestic breed, with an excess of heterozygosity displayed within the population. In a comparison of inbreeding between sprint and distance populations, all data suggested that the distance population is comparatively more inbred, while the sprint population is the driving force behind the excess heterozygosity (Figure 2.8; Table 2.2). This finding fits with the well-documented continual out-crossing of the sprint population with purebred breeds. Specifically, F_{IS} values were eightfold higher (Table 2.2), the inter-SNP distance estimating LD decay (r^2) was three times longer (Figure 4.3), and there were more large size (~12Mb) homozygous blocks (Figure 4.3) within distance dogs supporting greater inbreeding. Thus, I establish that two major populations exist within Alaskan sled dogs correlating to the racing style for which the dogs were bred. I also postulate a more closed breeding practice among distance kennels in contrast to cross-breeding among sprint kennels and out-breeding to purebred dogs. The genetic data supported what dog mushers have come to see as two different groups of sled dogs specifically chosen for sprint or distance racing. Likewise, these groups are rarely interbred due to the flow of undesirable traits such as decreased speed in sprint dogs, or less endurance and shorter hair coats in distance dogs.

Relationship of the Alaskan Sled Dog to Purebred Breeds

Ancestral Grouping and Breed Components

In Chapter 2, a panel of 96 microsatellite-based markers producing genetic signatures capable of distinguishing between 141 domestic breeds [1, 2] was utilized to

establish the relationship of Alaskan sled dogs to domestic breeds and identify purebred breed components within modern sled dogs. Both the sprint and distance populations consistently and predominantly clustered within the previously established Ancient/Asian group to which we note that both the Alaskan Malamute and Siberian Husky belong [1, 2]. However, both populations displayed a variation in the percentage allele similarity within the Ancient/Asian group (10% increase in distance), as well as within the Hunting group (12% increase in sprint) and Mastiff/Terrier group (6% increase in distance) (Figure 2.5). This was an early indication that the two sled dog populations had different ancestral lineages driving their athletic differentiation.

Variation of breed composition was compared within the sprint and distance populations. This analysis unequivocally demonstrated that Alaskan sled dogs produce their own unique genetic signature, which accounts for the largest portion of their breed composition and identified the Alaskan Malamute and Siberian Husky as the major purebred breed components. Sprint and distance dogs showed variation in the percentage and makeup of breed composition, with their respective sub-populations demonstrating particular breed selection (Figure 2.6 and 2.7). Overall, all distance sub-populations varied primarily in the percentage of Alaskan sled dog, Alaskan Malamute, and Siberian Husky genetic breed signatures, while sprint sub-populations had a lower percentage of the Alaskan Malamute and Siberian Husky and higher percentages of breed signatures such as Pointer, Saluki, and Borzoi (Table 2.1). In Chapter 4, we utilized breed composition and relationship information gained through the microsatellite analyses to embark on an intensive ancestry modeling study of the sprint and distance populations.

Ancestry Modeling

Four ancestral reference breeds (Alaskan Malamute, Siberian Husky, Borzoi, and German Shorthaired Pointer) were chosen for ancestry modeling and admixture mapping. This methodology, using a subset of over seven thousand select ancestry informative SNPs, allowed for the localization of ancestry across the genome and the identification of areas demonstrating specific ancestral breed selection (Chapter 4). The distance dogs showed the same ranking of purebred breeds with Alaskan Malamute being the most prevalent, then Siberian Husky, German Shorthaired Pointer, and Borzoi respectively (Table 4.1; Figure 2.6). In contrast to the microsatellite data, ancestry modeling showed a higher proportion of German Shorthaired Pointer ancestry in sprint sled dogs as opposed to Alaskan Malamute and Siberian Husky ancestry (Table 4.1; Figure 2.6). Two particular factors may contribute to this finding. First, the microsatellite data utilized 13 breeds including the Alaskan sled dog “breed,” while the ancestry modeling factored in only the four reference breeds. Therefore, the percentage of breed composition accounted for by the extra nine breeds in the microsatellite data would be absorbed by the four breeds utilized in the ancestry modeling. Second, and most likely of major impact, was the exponentially greater degree of genome-wide coverage offered through 7K SNPs used in the ancestry modeling as opposed to the 96 microsatellite-based markers used previously. Therefore, we conclude that while ancestry modeling only accounted for the four reference breeds used, it produced a more accurate estimate of the proportion of these primary ancestors represented in the two sled dog populations.

As mentioned, ancestry modeling localizes areas of ancestry selection across the genome and also generates diploid ancestry states utilizing the four reference breeds. A comparison of the diploid ancestry states between the sprint and distance populations showed the greatest degree of variation within the three homozygous states of the German Shorthaired Pointer, Alaskan Malamute, and Siberian Husky, from most to least difference, respectively (Table 4.1). This distinct difference in homozygous ancestry states versus heterozygous states may be due to what breeders refer to as backcrossing (crossing a hybrid with one of its parents or an individual genetically similar to the parent) or line-breeding (form of inbreeding selecting for a desirable trait seen in a common ancestor). This could be achieved by crossing Alaskan sled dogs to purebreds of the parental breeds, hence backcrossing. For example, both German Shorthaired and English Pointers have been commonly crossed with sprint dogs over the past two decades [3]. Another possible reason for the increased difference in the homozygous diploid ancestry states is the interbreeding of recent hybrids or line-breeding. Because the frequency of these diploid states of Alaskan Malamute and Siberian Husky are more heavily selected for in distance dogs, and German Shorthaired Pointer is selected for in sprint dogs, it would also suggest that the backcrossing or line-breeding occurs separately within the two sled dog populations. This supported my previous findings of population structure and breeding strategies using the microsatellite-based markers. Blocks of diploid ancestry were identified across the genome with a comparison of the most common ancestry states between sprint and distance dogs depicted in Figure 4.4. We note the striking disparity among chromosomes, in which some are nearly identical

between both sled dog populations while others are almost entirely different. Overall, the distance population had a 6.0% increase in total and a 4.3% increase in private ancestry blocks. Ancestry blocks were on average 200Kb longer in the distance population as compared to the sprint population (Table 4.2). These minor differences can be explained by the higher degree of inbreeding within the distance population.

The Genetics of Performance

Breed Influence

So far, all of the work described highlights the genetic segregation of sprint and distance racing dogs. Recalling that these two populations have been selectively bred for either speed or endurance, we sought to elucidate genetic components contributing to performance as opposed to random signals of population variation. A general examination identified particular breed components of increased frequency in elite dogs as opposed to poor performing dogs with regard to speed, endurance, or work ethic (Chapter 2). Most influential to overall performance was the Alaskan sled dog “breed”, which increased in frequency for all attributes in elite dogs. The highest contribution of purebred breeds was an 11% increase in the breed signatures of both Alaskan Malamute and Siberian Husky for elite endurance performance in distance dogs. The Saluki and Anatolian Shepherd, respectively, had minor positive influences for exceptional speed in sprint and work ethic in distance dogs (Table 2.3). The inclusion of the Anatolian Shepherd was unexpected as a breed component for positive influence on work ethic. However, the breed is described as a large, powerful, and independent northern livestock

guardian dog known for its loyalty, independence, and hardiness [4] and these characteristics are very similar to descriptions of the historically classic distance sled dog [5-7]. Interestingly, the Pointer, which has been repeatedly bred into sprint dogs over the past 20 years with the intent of enhancing speed, was not found to positively affect speed performance. One hypothesis is that the integration of Pointers with Alaskan sled dogs has not successfully increased their speed. However, a comparison of high profile sprint races shows faster finish times in 2007 versus 1998. These data corresponds to pedigree analyses showing a higher degree of Pointer ancestry in dogs in 2007, and therefore disputes the theory that Pointers have not had a positive influence on speed [8]. A more likely explanation is that there is a lack of significant difference in speed measurements of elite versus poorly performing sprint dogs. Instead, it may be that the contribution of the Pointer to speed enhancement is only evident when comparing sprint (averaging 18-25mph) versus distance (averaging 8-12mph), or if more extreme representatives of elite and poor performing sprint dogs are assayed. These data give insight as to the genetic influence of specific purebred breeds with respect to athletic performance in an admixed population of dogs.

Candidate Gene Investigation

To look at actual gene effect on performance we started with a straightforward candidate gene approach by interrogating two genes previously associated with performance, the myostatin (*MSTN*) and angiotensin-converting enzyme (*ACE*) genes. The *MSTN* gene, which regulates muscle fiber growth and composition, had already been

linked to racing performance enhancement in whippets. Whippets carrying the heterozygous mutation state possessed increased musculature as well as improved racing success [9]. We hypothesized that sprint dogs in particular might maintain a similar polymorphism due to their selection for speed and their more muscular appearance, which are primarily attributed to recent out-crossing to purebred lines. We also chose the *ACE* gene in which an I/D polymorphism has been repeatedly correlated with improved human endurance or power performance respectively [10-12]. Novel variants were found within both genes and confirmed in panels of purebred dogs (Supplementary Table 3.1 and Table 3.3). Three individual *ACE* gene markers (markers 42, 56, and 74) successfully distinguished between the distance and sprint sled dogs analyzed with raw p-values ≤ 0.005 (Table 3.1). We utilized pairs of purebred breeds, which exemplify athletic attributes of endurance or power/speed in an attempt to decipher whether these variants were contributing to athletic differences between sprint and distance dogs versus a more random population difference. We hypothesized that polymorphisms would be similar in state and frequency between purebred breeds attributed with endurance and distance dogs and purebred breeds attributed with speed and sprint dogs. This would support an association of the polymorphism to performance. *ACE* gene marker 56, an intronic allele change from *A* to *G* at base-pair 14635693, did indeed show a significant p-value of 7.57×10^{-8} associating the *G* allele with Alaskan Malamutes and Siberian Huskies. Because these two breeds are credited with improved endurance we postulated that the *G* allele was associated with endurance. However, upon closer examination, the calculated minor allele frequency (MAF) of the *G* allele in the Alaskan

Malamute/Siberian Husky pairing (0.265) was more analogous to the MAF in the sprint dogs (0.305; Distance 0.031) (Table 3.2). This therefore countered our expectation that the distance population would be similar in allele frequency to the Alaskan Malamute/Siberian Husky pairing. Not only do these three groups of dogs share endurance attributes, but our previous analyses showed a 25% higher degree of Alaskan Malamute and Siberian Husky genetic signature in distance dogs than in sprint dogs, as well as an 11% increase of each breed in elite endurance distance dogs (Chapter 2). Thus, we could only positively conclude that *ACE* marker 56 was a variant capable of distinguishing between the sprint and distance populations.

Whole Genome Association Studies and Admixture Mapping

Whereas the candidate gene investigation was a viable starting point, the 221 genes or quantitative trait loci currently associated with performance made the continual use of this approach impractical to achieving our goal of identifying performance genes within Alaskan sled dogs [13]. We therefore utilized a combination of association studies, observed heterozygosity measurements, and admixture mapping to conduct a more thorough genome-wide investigation of performance genetics characteristic to Alaskan sled dogs (Chapter 4). Specifically, these analyses highlighted genomic loci demonstrating either population variation and/or performance association (Supplemental Tables 4.1-3). Independently, the highest degree of genome-wide population association and selective sweep were found within the same region on chromosome 3. The *ARL2BP* gene, which is linked to mitochondrial activity within cardiac and skeletal muscle tissue,

is one of the most intriguing candidate genes for this region [14]. The greatest selection for ancestry, specifically Siberian Husky, encompassed a 5Mb region which included five potentially interesting candidate genes. Two of these genes, *FBN1* and *FBN2*, produce proteins integral to the structure and function of connective tissue. The other two genes, *ACSL6* and *SLC27A6*, are important for fatty acid metabolism and transport which are related, respectively, to energy production [15, 16]. Lastly, the *HINT1* gene also falls within this region and has been associated with anxiety and stress coping behavior in mice [17, 18]. The use of admixture mapping permitted speculation as to the contribution and influence of particular ancestry breeds. Combining admixture mapping with regions demonstrating either selective sweep or genome-wide association allowed us to postulate whether areas with excess ancestry were due to performance selection or to the random inheritance of chromosomal segments. With respect to this, an excess of ancestry corresponding to a region of selective sweep or genome-wide association supports the segregation of said area due to performance selection. Consequently, seven loci had overlapping regions of interest and ancestry selection (Table 4.3). Regions on chromosomes 3, 10, 16, 28, and 32 produced potential candidate genes of *SLC2A9*(glucose homeostasis), *MSRB3*(oxidative stress), *PTPRN2*(insulin binding), *ATRNL1*(information processing), and *HBZ*(muscle weakness) respectively. The most intriguing associations that instigated fine-mapping analyses were of the *HINT1* gene on chromosome 11 related to stress coping behaviors and the *MYH9* gene on chromosome 10 related to heat tolerance.

The *histidine triad nucleotide binding protein 1 (HINT1)* gene has been shown to affect the ability of knockout mice to cope with anxiety and stress. In one study, data suggested that the *HINT1* gene was important to mood regulation [18]. In a second study, *HINT1* knockout mice exhibited increased anxiety throughout a variety of behavioral analyses as well as displaying more “emotional arousal” when confronted with physically adverse situations [17]. Within Alaskan sled dogs, the *HINT1* gene is overlapped by Siberian Husky selection among distance racing dogs and German Shorthaired Pointer selection among sprint racing dogs. Breed descriptions have characterized the Siberian Husky as “stubborn and easily bored” despite its hardy working dog nature, as opposed to the German Shorthaired Pointer, which is noted for its “ease of training and adaptability” along with its commitment to performing [19]. Anecdotally, the “mental toughness” (ability to deal with stress) as well as the eagerness to please of Alaskan sled dogs with a high percentage of German Shorthaired Pointer ancestry is a topic of debate among sled dog drivers. Direct sequencing of sprint dogs, distance dogs, German Shorthaired Pointers, and Siberian Huskies through the four exons of the *HINT1* gene produced seven novel polymorphisms. Lacking allelic association with either sprint or distance populations, we could not confirm an association of the *HINT1* gene to behavioral attributes of either population.

The *myosin heavy chain 9 non-muscle type II class A (MYH9)* gene plays a more obvious role with regard to performance enhancement. Previous research associated an increase in myosin heavy chain with increased cardiac power output [20]. Specifically, an increase in myosin heavy chain type II class A within muscle tissue experiencing

elevated temperature (ET=37.5°C; Normal,N=34.2°C) correlated with an increase of power output. The slight temperature elevation improved muscle fiber power output through an increased rate of anaerobic ATP turnover and muscle fiber conduction velocity (MFCV). However, the efficiency of power output was found to decrease in the temperature-elevated muscles [21]. I identified seven *MYH9* gene SNPs significantly associated (permuted p-values < 0.05) with heat tolerance performance in sprint sled dogs, all of which were within significant LD with one another ($LD \geq 0.871$). Two of the strongest associated SNPs are in conserved regions upstream of the 5' end of the gene. Chromatin segment analysis of human skeletal muscle myoblasts in the analogous human region identified these SNPs to be in active promoter and enhancer sites suggesting potential mechanisms affecting the *MYH9* gene in sled dogs [22, 23]. Ancestry modeling identified the German Shorthaired Pointer and Alaskan Malamute breeds as being within excess in the *MYH9* region for sprint sled dogs. The German Shorthaired Pointer ancestry showed a particular spike in ancestry selection over the *MYH9* gene while the Alaskan Malamute demonstrated a gradual increase over approximately 20Mb throughout this region. Analysis of the sled dog associated *MYH9* gene SNPs within these two breeds showed that sprint dogs with poor heat tolerance had similar allele frequencies to Alaskan Malamutes with elite heat tolerance sprint dogs approaching allele frequencies similar to German Shorthaired Pointers. Ancestry modeling and allele frequency comparisons supports ancestral breed influence for performance attributes such as heat tolerance. Previous research combined with my results pinpoint the *MYH9* gene as an intriguing candidate gene potentially affecting the performance of heat tolerance in

sprint sled dogs as well as documenting the prospective role of ancestral breeds in regards to performance.

By combining admixture mapping, selective sweep, and genome-wide association analysis I was able to determine potential candidate genes, as well as possible breed influences on these genes. This work creates a foundation of both gene and breed contribution from which to explore the genetic mechanisms responsible for performance characteristics of sprint and distance racing sled dogs. The utility of exploring performance in Alaskan sled dogs and combining various methods of analysis was exemplified with the identification of the *MYH9* gene as an intriguing candidate affecting sled dog heat tolerance performance. Future questions can now be posed as to the *MYH9* gene's influence on human physiology as well as human performance.

Project Retrospection

The genetic profiling of a select group of high performance Alaskan sled dogs supported historical and modern views regarding their relationship to purebred breeds, inbreeding, and population structure. The research has illuminated novel genetic characteristics distinct to this admixed population, most notably the identification of a unique breed signature specific to Alaskan sled dogs. However, the primary goal of identifying genes influencing elite sled dog performance has proven rather elusive throughout the project. A weakness of this dataset is the lack of phenotypic variation between high and low performers, as well as very low numbers of poor performers overall. This is due to dogs being sampled from successful racing kennels in order to

minimize false rankings of poor performance due to insufficient training. Unfortunately the selection of successful kennels has had the adverse effect of decreasing the phenotypic variation within the data set, because our poor performers are not necessarily the most extreme representatives. The initial approach of capitalizing on performance variation between sprint and distance dogs for attributes of speed and endurance was confounded due to the fact that dogs divided into two genetically differentiated populations respective of their racing style. Therefore, variants specific to a racing population had to be further distinguished as being associated to either an athletic attribute or a random population variation.

As is the norm with science, projects that start with sound design and methodology do not always look as logical in hindsight. An initial foray into sled dog performance genetics using candidate genes previously associated with endurance and speed was a logical starting point. However, I recognized the futility of continuing this approach due to the overwhelming number of genes previously associated at varying degrees with human performance. This turned me toward the use of genome-wide association, taking advantage of the dog genome assembly and the commercial availability of genome-wide SNP arrays. The necessity of using unrelated individuals limits genome-wide association for this dataset; even though the sled dogs were sampled from nine kennels, there remained substantial relatedness among dogs within each kennel. The sprint dogs further confounded this with relatedness extending across kennel boundaries. Overall, the use of unrelated individuals in genome-wide association became another factor reducing the number of dogs available for analysis. I addressed the low

individual numbers by using multiple methods to identify loci of interest. In this regard, we also looked at selective sweep, F_{ST} scores, and ancestry modeling. Combining these analyses increased my confidence in the association of loci to either performance or population phenotypes. Throughout the highs and lows of this research project I learned fundamental elements necessary for designing a scientific project with integrity and improved probability of success. I understand the value of datasets in terms of phenotyping, relatedness, and number of individuals for analysis. I also have an appreciation of the myriad methods available to answer questions and to increase confidence in the answers. Although the identification of performance genes in Alaskan sled dogs proved challenging, I finish this chapter of my work with a wide array of ideas on how to improve what I currently have and where to go from here.

In this regard, the most obvious improvements would be to obtain DNA and phenotypes from more dogs belonging to different kennels and specifically focus on obtaining poor performing dogs that have received good nutrition and training. I believe the current dataset would also benefit from using a linkage analysis approach. Linkage analysis utilizes family groups of siblings or parents and offspring, therefore more appropriately accommodating the relatedness found among my sample dogs. I also question the idea of a generalized sampling of dogs from the least successful kennels, therefore disregarding the covariate of training received and broadly labeling these dogs as poor performers. To counter potential false phenotyping, I believe that a substantially greater number of elite and poor performers would need to be analyzed rather than the minimal 20 elite and 20 poor performers that I have been currently using. Another option

would be to collaborate with researchers who have done metabolic studies on racing Alaskan sled dogs. The enzyme, hormone, and blood flow parameters they have found to be correlated to performance could then be utilized as quantitative phenotypic markers to perform genome-wide association analyses. Each of these individual approaches has potential for success, and combined they may increase the rate of success and confidence in the results.

The current research focused a great deal on identifying purebred breed components within modern Alaskan sled dogs and localizing areas of the genome demonstrating an excess of ancestry selection. The purebred dogs used in my research came from conformation lines (selective breeding based upon physical conformation of a dog to meet breed standards), as opposed to working dog lines. It would be intriguing to compare purebred dogs selectively bred for their working abilities with the Alaskan sled dogs and the purebred conformation dogs used in my previous analyses. I would also like to obtain samples from Inuit dogs found in northern Canada and Greenland and “village dogs” from rural Alaska. Adding these groups of dogs, even though they are not purebred breeds, would give insight as to the historical relationship of these groups of dogs to modern Alaskan sled dogs and may potentially highlight the origins of the unique Alaskan sled dog breed signature.

As previously discussed in Chapter 1, the unique genome structure of dogs created during breed development is an asset of using dogs as a model species for gene mapping. I hope to capitalize on this approach and gather DNA and phenotyping on other working dogs—specifically, the Alaskan Malamutes, Siberian Huskies, and

Samoyeds used for sled dog racing. I am also interested in other working dogs of purebred and mixed breed. Track racing dogs and lure coursing dogs as well as hunting, herding, and agility dogs would also be an interesting addition to analysis options. I note that not all of these working dogs are known for speed or endurance but may be important in localizing genes associated with behavioral attributes such as work ethic and mental stress tolerance, which are major attributes for working dogs. Other avenues that I hope to pursue include the addition of the phenotype of leader ability and correlations of gait efficiency and body structure to my genetic investigation. My research has ingrained in me the importance of solid phenotyping of specific attributes, the range of phenotypic variation, sample numbers, and the utility of using multiple breeds and statistical approaches in the identification and localization of genetic loci.

References

1. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L: **Genetic structure of the purebred domestic dog**. *Science* 2004, **304**(5674):1160-1164.
2. Parker HG, Kukekova AV, Akey DT, Goldstein O, Kirkness EF, Baysac KC, Mosher DS, Aguirre GD, Acland GM, Ostrander EA: **Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds**. *Genome Research* 2007, **17**(11):1562-1571.
3. **Alaska Dog Musher's Association** [<http://www.sleddog.org/2010/03/21/open-north-american-results-day-3-final/>] Accessed Jan 2011.
4. **American Kennel Club- AKC.org** [<http://www.akc.org/>] Accessed Jan 2011.
5. Wendt R: **Alaska Dog Mushing Guide: Facts and Legends**, Fifth edn. Fairbanks: Goldstream Publications; 1999.

6. Vaudrin B (ed.): **Racing Alaskan Sled Dogs**. Anchorage, AK: Alaska Northwest Publishing Company; 1977.
7. Vaughan ND: **With Byrd at the Bottom of the World: The South Pole Expedition of 1928-1930**, 1st edn. Harrisburg: Stackpole Books; 1990.
8. **Sled Dog Central** [<http://www.sleddogcentral.com/>] Accessed Aug 2010.
9. Mosher DS, Quignon P, Bustamante CD, Sutter NB, Mellersh CS, Parker HG, Ostrander EA: **A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs**. *PLoS Genetics* 2007, **3**(5):e79.
10. Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, Soubrier F: **An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels**. *Journal of Clinical Investigation* 1990, **86**(4):1343-1346.
11. Danser AH, Schalekamp MA, Bax WA, van den Brink AM, Saxena PR, Riegger GA, Schunkert H: **Angiotensin-converting enzyme in the human heart. Effect of the deletion/insertion polymorphism**. *Circulation* 1995, **92**(6):1387-1388.
12. Thompson WR, Binder-Macleod SA: **Association of genetic factors with selected measures of physical performance**. *Physical Therapy* 2006, **86**(4):585-591.
13. Bray MS, Hagberg JM, Perusse L, Rankinen T, Roth SM, Wolfarth B, Bouchard C: **The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update**. *Medicine & Science in Sports & Exercise* 2009, **41**(1):35-73.
14. Sharer JD, Shern JF, Van Valkenburgh H, Wallace DC, Kahn RA: **ARL2 and BART enter mitochondria and bind the adenine nucleotide transporter**. *Molecular Biology of the Cell* 2002, **13**(1):71-83.
15. **National Center for Biotechnology Information** [<http://www.ncbi.nlm.nih.gov/>] Accessed April 2011.
16. **University of California, Santa Cruz Genome Browser** [<http://genome.ucsc.edu/>] Accessed April 2011.
17. Varadarajulu J, Lebar M, Krishnamoorthy G, Habelt S, Lu J, Bernard Weinstein I, Li H, Holsboer F, Turck CW, Touma C: **Increased anxiety-related behaviour in Hint1 knockout mice**. *Behavioral Brain Research* 2011, **220**(2):305-311.

18. Barbier E, Wang JB: **Anti-depressant and anxiolytic like behaviors in PKCI/HINT1 knockout mice associated with elevated plasma corticosterone level.** *BMC Neuroscience* 2009, **10**:132.
19. Wilcox B, Walkowicz: **Atlas of Dog Breeds of the World**, vol. 5th Edition, 5th Edition edn. Neptune City: T.F.H. Publications, Inc; 1995.
20. Burniston JG: **Adaptation of the rat cardiac proteome in response to intensity-controlled endurance exercise.** *Proteomics* 2009, **9**(1):106-115.
21. Gray SR, De Vito G, Nimmo MA, Farina D, Ferguson RA: **Skeletal muscle ATP turnover and muscle fiber conduction velocity are elevated at higher muscle temperatures during maximal power output development in humans.** *American Journal of Physiology- Regulatory, Integrative and Comparative Physiology* 2006, **290**(2):R376-382.
22. Ernst J, Kellis M: **Discovery and characterization of chromatin states for systematic annotation of the human genome.** *Nature Biotechnology* 2010, **28**(8):817-825.
23. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al*: **Mapping and analysis of chromatin state dynamics in nine human cell types.** *Nature* 2011, **473**(7345):43-49.