Mamm Genome (2012) 23:178–194
DOI 10.1007/s00335-011-9374-y

ORIGINAL CONTRIBUTION

Breed-specific ancestry studies and genome-wide association analysis highlight an association between the *MYH9* gene and heat tolerance in Alaskan sprint racing sled dogs

Heather J. Huson · Bridgett M. vonHoldt · Maud Rimbault · Alexandra M. Byers · Jonathan A. Runstadler · Heidi G. Parker · Elaine A. Ostrander

Received: 29 August 2011 / Accepted: 20 October 2011 / Published online: 22 November 2011

Abstract Alaskan sled dogs are a genetically distinct population shaped by generations of selective interbreeding with purebred dogs to create a group of high-performance athletes. As a result of selective breeding strategies, sled dogs present a unique opportunity to employ admixture-mapping techniques to investigate how breed composition and trait selection impact genomic structure. We used admixture mapping to investigate genetic ancestry across the genomes of two classes of sled dogs, sprint and long-distance racers, and combined that with genome-wide association studies (GWAS) to identify regions that correlate with performance-enhancing traits. The sled dog genome is enhanced by differential contributions from four non-admixed breeds (Alaskan Malamute, Siberian Husky, German Shorthaired Pointer, and Borzoi). A principal components analysis (PCA) of 115,000 genome-wide SNPs clearly resolved the sprint and distance populations as distinct genetic groups, with longer blocks of linkage disequilibrium (LD) observed in the distance versus sprint dogs (7.5–10 and 2.5–3.75 kb, respectively). Furthermore, we identified eight regions with the genomic signal from either a selective sweep or an association analysis, corroborated by an excess of ancestry when comparing sprint and distance dogs. A comparison of elite and poor-performing sled dogs identified a single region significantly associated with heat tolerance. Within the region we identified seven SNPs within the myosin heavy chain 9 gene (*MYH9*) that were significantly associated with heat tolerance in sprint dogs, two of which correspond to conserved promoter and enhancer regions in the human ortholog.

Electronic supplementary material The online version of this article (doi:10.1007/s00335-011-9374-y) contains supplementary material, which is available to authorized users.

H. J. Huson · M. Rimbault · A. M. Byers · H. G. Parker · E. A. Ostrander (✉)
Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 50 South Drive, Building 50, Room 5351, Bethesda, MD 20892, USA
e-mail: eostrand@mail.nih.gov

H. J. Huson · J. A. Runstadler
Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK 99775, USA

B. M. vonHoldt
Ecology & Evolutionary Biology, University of California Irvine, Irvine, CA 92697, USA

Introduction

The Alaskan sled dog has evolved over the past century from a working dog, originally developed to haul cargo sleds over snow-covered terrain (Collins 1991; Rennick 1987; Vaudrin 1977), to an elite modern-day athlete. Their dominating presence in polar exploration and the boom of the Alaskan Gold Rush gave rise to the "Era of the Sled Dog" from approximately the late 1800 s to the early 1900 s (Wendt 1999). The incorporation of modern transportation methods forced the sled dog into retirement from its necessary role of working dog, transitioning, instead, to a sport-racing dog. Though not recognized by the American Kennel Club (AKC) (AKC 1998) and not developed to meet a physical standard, Alaskan sled dogs are bred for climate-specific athletic performance attributes, which has resulted in a level of genetic distinctiveness comparable to that of AKC-recognized breeds (Huson et al. 2010). Performance selection has given these dogs a common athletic phenotype: a quick and efficient gait, superior pulling strength, and increased endurance. Overall body weight

and coat type, however, can vary depending upon racing style, geographic location, lineage, and cross breeding to purebred lines.

Sled dog racing can be divided into two distinct styles based upon the mileage that teams travel. Long-distance racing covers approximately 1,000 miles over multiple days at moderate racing speeds (13–19 km/h) (e.g., Iditarod and Yukon Quest) (Iditarod 2011; Yukon Quest 2011), while sprint racing is composed of multiple events or classes defined by the number of dogs in the team (4–20), faster racing speeds (29–40 km/h), and shorter distances (~6–38 km). The extreme differences in racing style has led to divergent selection of Alaskan sled dogs for either endurance or speed, resulting in two distinct populations (Fig. 1) (Huson et al. 2010).

As a result of interbreeding practices, the modern sled dog genome is a mosaic of purebred dog ancestry that represents a unique opportunity to document the acquisition of athletic performance traits through both a selection scan and admixture mapping. Admixture mapping has been implemented successfully for genetic variants and disease phenotypes in human populations with mixed ancestry (Buerkle and Lexer 2008; Patterson et al. 2004; Seldin et al. 2011; Winkler et al. 2010). The method scans through a mosaic genome and identifies the ancestry for each chromosomal fragment, provided that parental genomes are defined. The frequency and size of these fragments is influenced by the frequency and direction of interbreeding duration, as well as trait selection. Written pedigrees as well as genetic investigation (Huson et al. 2010) reveal that the Alaskan Malamute, Siberian Husky, Pointer (English and German Shorthaired), Saluki, Borzoi, Irish Setter, Weimaraner, German Shepherd, and Anatolian Shepherd were utilized in generating the Alaskan sled dog (ADMA 2011; Huson et al. 2010). Here, we have used two genome-wide panels consisting of 115,425 and 27,416 single-nucleotide polymorphisms (SNPs) to assess population structure and conduct both admixture mapping and a genome-wide association study (GWAS) to explore the genetics of endurance and heat tolerance in Alaskan sled dogs.

Methods

Sample collection and SNP array genotyping

DNA was extracted from blood samples provided by 150 Alaskan sled dogs, 65 from distance and 85 from sprint racing kennels (see Performance Ratings Section below), and 45 purebred dogs from four AKC-recognized domestic breeds [Alaskan Malamutes (AMAL), $n = 10$; Siberian Huskies (HUSK), $n = 12$; German Shorthaired Pointers (GSHP), $n = 11$; Borzois (BORZ), $n = 12$] (Boyko et al. 2010; Huson et al. 2010). All 45 purebred dogs were unrelated at the grandparent level, as were 19 distance and 27 sprint sled dogs, selected from pedigree analysis. Prior to sample collection all owners provided informed consent,

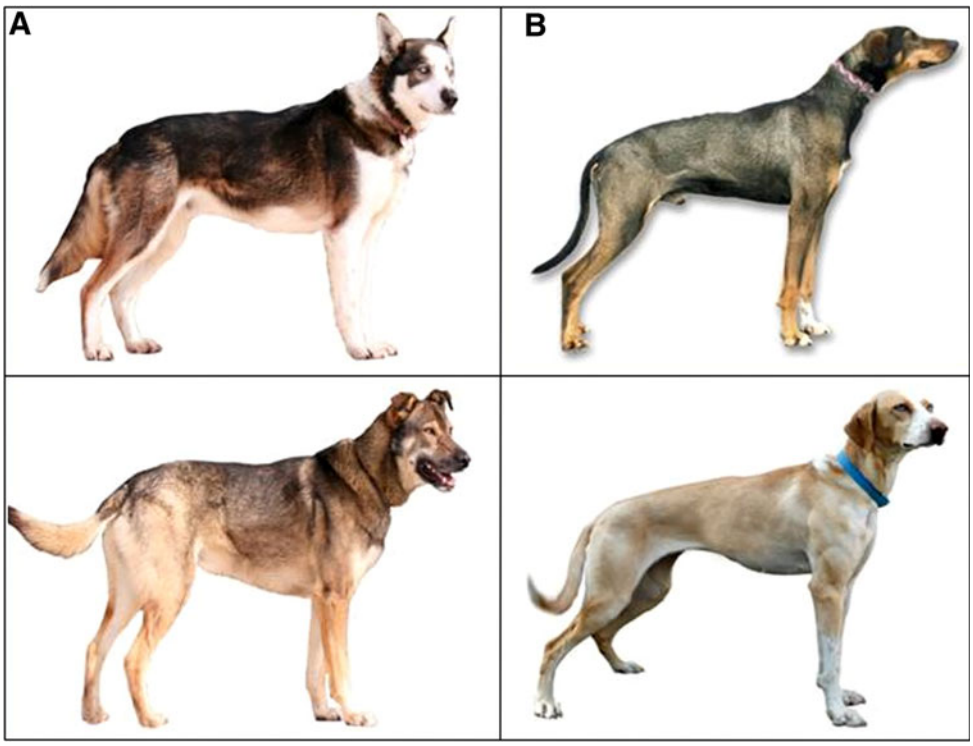

Fig. 1 Alaskan sled dogs are a mixed-breed dog, bred strictly for performance attributes. **a** *Left column* distance racing dogs. **b** *Right column* sprint racing dogs

consistent with NHGRI Animal Care and Use Committee rules. Whole-blood samples were collected from the cephalic vein into 3–5-ml EDTA or ACD tubes. Sled dogs were sampled at their home kennels, while purebred dog samples were obtained through clinics set up at large gatherings, such as conformation competitions, or through their local veterinarian. Samples were stored at 4°C prior to DNA 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, coded, and aliquoted for long-term storage at −80°C. Finally, detailed pedigrees were collected for each sampled individual.

A total of 150 Alaskan sled dogs were genotyped using the Illumina HD Canine SNP array (Illumina, San Diego CA, USA) and a total of 115,425 SNPs were retained after similar quality filtering. The 45 AKC-registered purebred dogs sampled to represent ancestral populations were previously genotyped for 48,716 SNPs (Boyko et al. 2010; VonHoldt et al. 2010) using the Affymetrix v2.0 Canine SNP array (Affymetrix, Santa Clara, CA, USA). For both platforms, SNPs were retained that had a ≥93% genotype call rate, <10% missing genotypes, and >10% minor allele frequency based on data from using Genome Studio (Illumina) and PLINK software (Purcell 2009; Purcell et al. 2007). We identified a set of 27,416 overlapping SNPs between the Illumina and Affymetrix panels to be used for population structure analyses.

Performance ratings

Sled dogs were individually scored for their abilities related to speed, endurance, work ethic, mental stress tolerance, and heat tolerance. Distance dogs ($n = 65$) were sampled from four kennels, all of which finished in the top 15% of competitors for the Yukon Quest or Iditarod race during two consecutive years (2007–2008) of sample collection. Sprint dogs ($n = 85$) were also sampled from four kennels, each of which placed in the top 25% of the International Sled Dog Racing Association points-ranking medal program during the sampling years (2005–2007). All distance kennels maintained similar training regimens with regard to mileage (increasing up to ∼322 km) and speed (13–19 km/h) as it related to fall training through winter racing season (September–March). Sprint kennels also had similar metrics with regard to mileage (increasing up to ∼48 km) and speed (24–40 km/h) during the same time period. The study did not control for individual driver training style. The kennels sampled were located throughout the northern continental United States, including Alaska, as well as northern Canada, with slight variations in weather and terrain. Sampled dogs competed in many of the same races, several of which were held in Alaska. Brand of dry dog food used varied between kennels but was comparable in total protein (∼26–34%) and fat (∼14–20%) content. Each kennel also supplemented diets with either raw meat or meat supplements, particularly during the winter racing season.

Criteria for each athletic attribute were defined and tested by one of the authors (HJH) and reviewed by five professional sled dog drivers. Scorers independently rated a minimum of the same eight sled dogs after a single training run, and scores were reviewed for reliability and repeatability. Distance dogs were scored a single time to obtain their overall performance score for 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 dogs were scored for phenotype during consecutive years (2005–2007). To achieve a single score for each sprint dog that was comparable to those obtained for distance dogs, the last weekly rating for each sprint dog during peak racing season was regarded as their performance score for that year. Consecutive year ratings were obtained for each dog. If a dog's ranking for each attribute (speed, endurance, work ethic, heat tolerance, and mental stress tolerance) did not change over consecutive years, that score was simply used as the dog's overall performance score. For this study, each athletic attribute was viewed independent of the other four. A dog that had different annual scores for any particular athletic attribute was not included for analysis of that trait. In order to obtain suitable numbers, sled dogs were not restricted by age, which ranged from 1 to 6 years at the time of sampling. A disparity in males versus females was observed for sprint versus distance dogs: sprint kennels had a higher percentage of females (60%) and distance kennels had a higher percentage of males (72%). Performance was investigated for sprint and distance dogs separately and no sex disparity was observed in elite versus poorly performing dogs.

Endurance was scored using the average mileage traversed in a race, with dogs ranked 1, 2, or 3 based on their performance. Mileage requirements ranged from 13 to 48 km for sprint dogs and from 1,595 to 1,850 km 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 whether a dog reaches or nears a state of heat exhaustion (inability to reduce body temperature) while running in warm temperatures (approximately −7 to 10°C). The body temperature rise

associated with heat exhaustion causes an increased heart rate, muscle weakness, dizziness or confusion, rapid breathing, nausea, and vomiting. Observational data for the dog's degree of heat exhaustion were substituted as a proxy for the physiological state. Dogs showing no change in their ability to perform were ranked 1. A 2 was given to dogs demonstrating a lower-than-normal performance when running in warm temperatures. Such dogs showed mild signs of heat exhaustion for two or more of the above symptoms. Dogs unable to complete the mileage and demonstrating considerable signs of heat exhaustion (collapse or near collapse) were scored a 3.

Ancestry Informative Marker (AIM) identification

Phase was inferred using the program fastPHASE version 1.4.0 (Scheet and Stephens 2006) across the 27,416-SNP panel for all purebred and sled dogs with a 0.05 masking rate. We specified the number of haplotype clusters (K) to range from two to nine with an interval of one. We selected ancestry informative markers (AIMs) that highly differentiated the reference breeds, selecting one reference breed in comparison to a pool of the other three breeds (e.g., AMAL vs. HUSK/GSHP/BORZ). This allowed us to identify SNPs that were informative for the ancestry of each reference breed. Across all comparisons, the average genome-wide level of differentiation was moderate ($F_{ST} = 0.12$). In order to retain as many SNPs as possible but not compromise the level of differentiation, we included SNPs with an F_{ST} at least 1 SD above the genome-wide mean ($F_{ST} > 0.35$) but also required a genome-wide SNP spacing of $\sim$300 kb. As a result, we then identified a subset of 7,644 AIMs that were diagnostic for the four reference (ancestry) breeds: AMAL, HUSK, GSHP, and BORZ (Cheng et al. 2010; Rosenberg et al. 2010; Tang et al. 2006; Tian et al. 2006). Note that the (English) Pointer, identified in our previous microsatellite work (Huson et al. 2010), was substituted for the GSHP due to the availability of SNP data for that breed only. Both Pointer breeds have been documented as being interbred with Alaskan sled dogs (Parker et al. 2010). However conclusions should be viewed in light of this substitution.

Population structure, linkage disequilibrium, and homozygosity analysis

We conducted a PCA using the *smartpca* function in the EIGENSTRAT package (Price et al. 2006; Shriner 2011) to assess population structure of the unrelated sled dogs (distance, $n = 19$; sprint, $n = 27$) as well as of the entire data set of sled dogs and the four AKC breeds that contributed most to the sled dog genome (AMAL, BORZ, GSHP, and HUSK). In addition, we conducted a PCA with the panel of 7,644 AIM SNPs. This panel was used specifically to test the ability of the AIMs to distinguish individual populations. Using the data from the 115,425 SNPs collected on all sled dogs, we obtained estimates of observed heterozygosity (H_O) per SNP using PLINK (Purcell et al. 2007), and Wright's genetic differentiation (F_{ST}) (Boyko et al. 2010) using the program SCATTER (VonHoldt et al. 2010). We calculated F_{ST} estimates (see AIMs subsection above) for the purebred dogs only using the set of 27,416 overlapping SNPs.

To measure the extent of linkage disequilibrium (LD), we estimated pairwise intermarker genotypic associations (r^2), an estimate of LD using PLINK. We randomly subsampled 19 unrelated sprint dogs to match the sample size of the distance dogs because sample size differences will impact r^2 estimates. Using the panel of 27,416 overlapping SNPs and all unrelated dogs, 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 1,000, as described in Boyko et al. (2010). The distance to LD decay was defined as the distance bin in which the r^2 score dropped below the threshold of 0.5 for each population (Sutter et al. 2004). LD is expected to be more extensive in inbred as opposed to admixed populations (Boyko et al. 2010; Gaut and Long 2003; Gray et al. 2009; Pritchard and Przeworski 2001; Tang et al. 2006). Population distances were also calculated using an r^2 threshold of 0.2, providing a direct comparison to the study of Gray et al. (2009). Additionally, we determined the level of autozygosity within each population by surveying runs of homozygous genotypes (ROH) using the 27,41- SNP panel and PLINK. Homozygous tracks were required to be a minimum of 100 kb in length and to contain at least 25 SNPs, as described by us previously (Boyko et al. 2010).

Selective sweep

We conducted a selective sweep analysis in order to detect genomic regions that differentiated the two performance classes of sled dogs and potentially contained candidates genes linked to endurance and heat tolerance. Four independent criteria were used to 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. Using the genome-wide estimates of H_O, we selected 9,362 SNPs from the lower fifth percentile (distance = 0 H_O; sprint <0.0833 H_O). These SNPs demonstrate a loss of heterozygosity (LOH) defined as the observed heterozygosity being greater than one standard deviation below the genome average (H_O – 1 SD: distance = 0.16; sprint = 0.22). To reduce the number of sites for further investigation, we required that at least one SNP per region

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) as described in VonHoldt et al. (2010). Finally, regions were retained if SNPs were clustered (inter-SNP distance <300 kb), with each SNP in the cluster displaying high levels of LOH. This included 2,145 regions that had two consecutive SNPs <300 kb apart. Sixty regions had both consecutive SNPs and LOH.

Genome-wide association studies (GWAS)

GWAS were run with the data set of 115,425 SNPs in the sled dogs using EMMAX (Kang 2010), which corrects for population stratification and relatedness. To identify SNPs associated with sled dog population differentiation, 27 sprint and 19 distance dogs were compared in a case–control analysis. GWAS were also performed to investigate the performance attributes of endurance and heat tolerance. Age and sex were not considered covariates. All dogs were required to be unrelated through the second generation. Dogs that received scores of 1 were considered elite. Because less than 10% of dogs in this study scored a 3 for either endurance or heat tolerance, the dogs ranked as 2 or 3 were grouped together and considered as poor performers. Significance levels were generated using basic (adaptive) permutation testing in PLINK. SNPs demonstrating genome-wide association in EMMAX (Bonferroni correction equals a P value $\leq 4 \times 10^{-7}$) were required to have a corrected P value $\leq 1 \times 10^{-6}$ in PLINK.

Endurance was tested in sprint (poor, $n = 20$; elite, $n = 21$) and distance dogs (poor, $n = 14$; elite, $n = 19$) separately due to the considerable difference in performance requirements between the two groups, with poor performers (scores of 2 and 3) assigned case status while so-called "elite" performers (score of 1) were controls. Heat tolerance was also tested independently within each sled dog population (sprint: poor, $n = 17$, and elite, $n = 21$; distance: poor, $n = 10$, and elite, $n = 19$). As environmental temperature conditions were comparable for the two groups, an additional GWAS for heat tolerance was conducted by combining the sprint and distance groups, comparing all elite versus all poor performers for this attribute (poor, $n = 27$; elite, $n = 40$).

Modeling ancestry

SABER was utilized for modeling ancestry within the sprint and distance populations. An admixture-mapping approach using this information was taken to identify regions of particular selection within the two sled dog populations (Tang et al. 2006, 2007) SABER 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 while accounting for background LD. We specified a 1.0 cM/Mb recombination rate (Boyko et al. 2010) and used a prior of 10 generations from the initial admixture event ($\tau = 10$) for ancestry block assignments across all 38 autosomes.

SABER generates diploid ancestry block assignments for individual sled dogs. Using the four ancestor populations, ten diploid ancestry states are produced: four states are homozygous for the individual ancestor breeds (AMAL, BORZ, GSHP, and HUSK) and six are heterozygous combinations of the breeds (AMAL/BORZ, AMAL/GSHP, AMAL/HUSK, BORZ/GSHP, BORZ/HUSK, and GSHP/HUSK). 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. To estimate ancestry block frequency within each sled dog group, we used the randomly subsampled 19 unrelated sprint dogs for comparison to the 19 distance dogs. We filtered for ancestry blocks that had at least three contiguous SNPs with the same ancestry assignment in an effort to exclude potentially false ancestry blocks (due to random chance or lack of information). Ancestry blocks were deemed private to a single sled dog population if they had >20% frequency in that population and <5% frequency in the opposing population. Regions showing excess or deficient selection (>1 SD from each ancestral frequency mean) toward a particular ancestor were identified within the distance and sprint sled dogs based upon the highest degree of differential ancestry at consecutive SNPs, defined as the difference between the two populations (Tang et al. 2007). The top 5% of AIMs (382 SNPs) that showed the highest degree of differentiation between the sprint and distance populations was used to identify genomic regions that had undergone the strongest selection. These regions were greater than two standard deviations from the mean ancestry frequency difference (Tang et al. 2007).

Sequencing of the *HINT1* and *MYH9* genes

Two candidate genes were selected for sequencing based on GWAS results. The histidine triad nucleotide binding protein 1 (*HINT1*) gene, identified as a candidate due to a significant association with population variation between the sprint and distance dogs, is located on canine chromosome 11 (CFA11: 22,400,779–22,560,252; CanFam2.0) and consists of four exons that encompass 560 bp. The myosin heavy chain 9 non-muscle type II class A (*MYH9*) gene is located on canine chromosome 10 (CFA10: 31,135,177–31,194,500) and consists of 40 exons, totaling 7,318 bp.

Nineteen distance dogs and 27 sprint dogs, 8 GSHP, and 8 HUSK were sequenced across all exons of *HINT1*. Five amplicons, averaging 550 bp, were necessary to cover the *HINT1* coding region. Forty-three amplicons, averaging 620 bp in length, were sequenced to cover the *MYH9* exons, with an additional 11 amplicons included to cover highly conserved regions flanking the gene. Six elite and six poor performers for the heat tolerance attribute were initially sequenced for all 54 amplicons to identify SNPs. An additional set of 26 poor performers and 15 elite performers were genotyped for 16 *MYH9* SNPs that demonstrated association in the initial 12 dogs. Eight GSHP and six AMAL were also genotyped for the 16 critical SNPs to provide a comparison to the sled dogs.

PCR amplification for both genes was performed in 10-µl volumes containing 10 ng genomic DNA, 1 µl of 10 × TaqGold buffer, 0.05 µl of ABI TaqGold (Applied Biosystems, Carlsbad, CA, USA), 1 µl of 1 mM dNTPs, 0.3 µl of 50 mM $MgCl_2$, 1 µl of both forward and reverse 2 µM primers, and 4.65 µl of 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 45 s, followed by another 20 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 45 s, with a final extension phase of 72°C for 10 min. A small subset of amplicons within each gene required the following PCR protocol for successful amplification: 10-µl total volume containing 10 ng genomic DNA, 5 µl of KOD buffer, 0.2 µl of KOD (EMD Chemicals, Merck, Darmstadt, Germany), 1.6 µl of 2.5 mM dNTPs, and 1.2 µl of both forward and reverse 2 µM primers. The annealing temperature was also adjusted in the touchdown PCR, decreasing by 0.5°C/cycle for the first 20 cycles from 67 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 3730 × 1 capillary electrophoresis unit. Sequence reads were aligned and analyzed using Phred, Phrap, and Consed software (Bhangale et al. 2006; Ewing et al. 1998; Gordon et al. 1998). PolyPhred software was used to identify SNPs (Nickerson et al. 1997). All genetic variations, both SNPs and insertion/deletion polymorphisms, were then analyzed with Haploview 4.2 to assess LD structure, identify haplotypes, and test for association (Barrett et al. 2005).

Results

Population structure, linkage disequilibrium, and homozygosity analysis

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% (Fig. 2a). In a comparison of the domestic breeds and Alaskan sled dogs (Fig. 2b), the first PC (PC1, 16%) separates the Northern breeds (AMAL and HUSK) from the BORZ and GSHP, with both sled dog populations falling between the breed extremities. PC2 (9.8%) separates the BORZ from the GSHP, while PC3 (6.1%) distinguishes the AMAL from the HUSK. PC4 (3.6%) separates all Alaskan sled dogs from all domestic breeds tested, while PC5 (1.9%) separates the sprint from distance sled dogs.

To assess inbreeding patterns associated with the Alaskan sled dog, we estimated the decay of LD and found that both sled dog populations had shorter distances to LD decay ($r^2_{0.5}$; sprint, 2.5–3.75 kb; distance, 7.5–10 kb) than any of the purebred groups (GSHP, 10–15 kb; HUSK, 15–20 kb; AMAL and BORZ, 20–30 kb) (Fig. 3a). For the LD decay threshold of $r^2_{0.2}$, the AMAL, HUSK, BORZ, and distance sled dog populations had longer-range LD (>1 Mb). LD at $r^2_{0.2}$ decayed at approximately 700 kb in GSHP and 80 kb in sprint dogs, which is comparable with previously reported estimates (Gray et al. 2009). 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]) (Fig. 3b) (Boyko et al. 2010). Trends were similar to that of LD decay, with domestic breeds having ROHs that were of longer (>2 Mb) than the sled dogs, indicating a comparatively higher degree of inbreeding in the domestic breeds. However, the distance dogs had a slight inflation of ROHs of large size (∼12 Mb) compared to the sprint dogs, concordant with the previous inbreeding assessments reported for Alaskan sled dogs (Huson et al. 2010).

Selective sweep

We identified 60 genomic regions with a selective sweep signature when comparing the sprint and distance populations using H_O and F_{ST} scans (Supplementary Table 1). Fifty-two (87%) of the regions showed a selective sweep in the distance dogs, while only eight were observed in the sprint dogs. The region of greatest H_O difference (0.833) was on canine chromosome 3 (CFA3) at 83,775,932–83,798,854 bp and was observed in the distance dogs. The region is gene-poor, containing only two annotated genes within 1 Mb of the region boundaries, the most provocative of which is the ADP-ribosylation factor-like 2 binding protein *(ARL2BP)* gene, which is linked to mitochondrial activity in cardiac and skeletal muscle tissues (Sharer et al. 2002). The highest region of heterozygosity difference (0.75) within the sprint dogs was on CFA17 (8,158,751–8,170,123 bp), but there are no obvious candidate genes

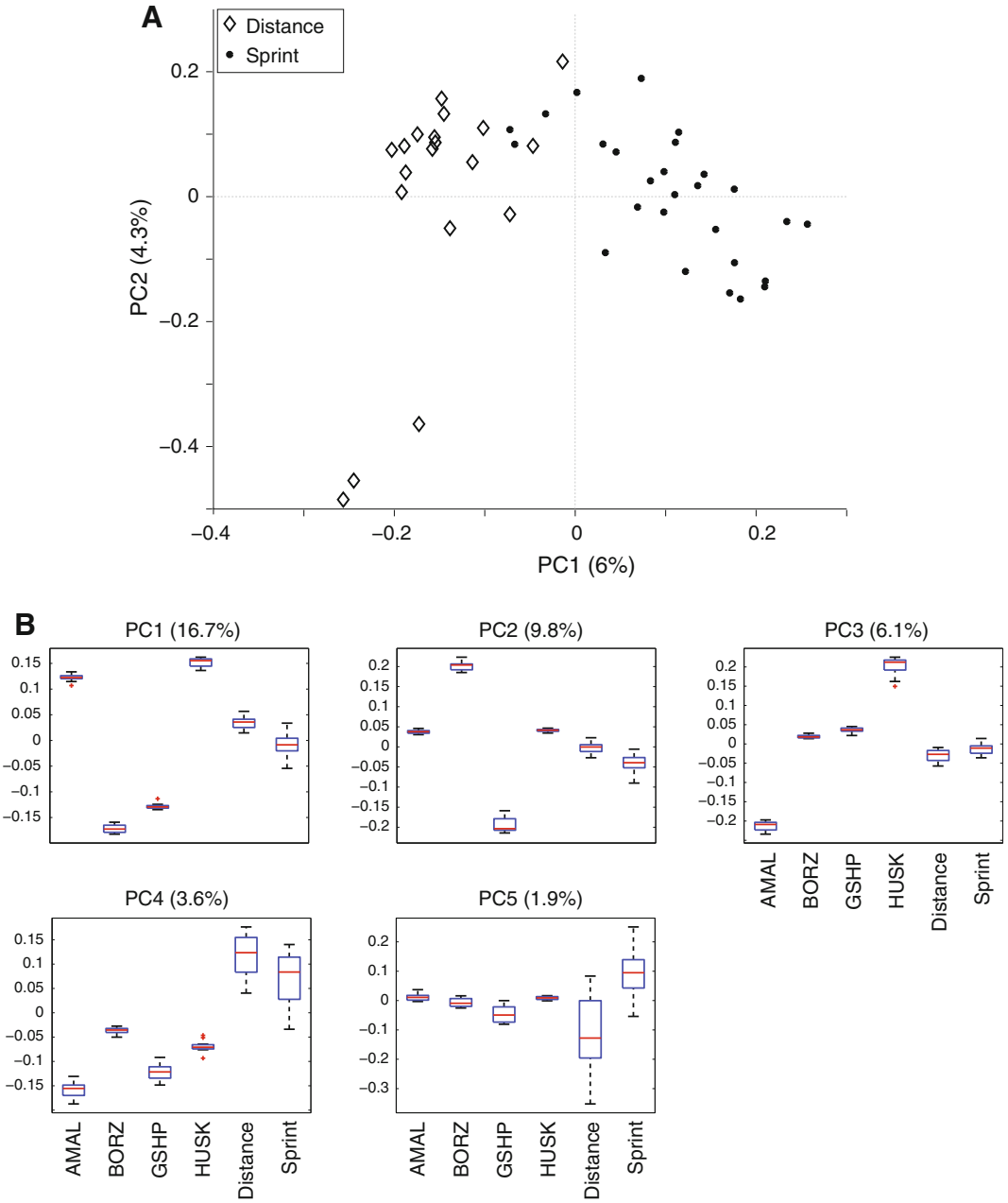

Fig. 2 Principal component analysis plots of Alaskan sled dogs (**a**, **b**) and four ancestry reference breeds (**b**) using a panel of 7,000 highly (FST >0.35) informative SNPs. **a** Alaskan sled dogs from either distance (blue) or sprint (red) racing kennels. **b** Four ancestry reference breeds, including AMAL, BORZ, GSHP, and HUSK as well as Alaskan sled dogs divided into their two populations of distance and sprint

within ± 1 Mb of the region boundaries (Build 2.0, http://genome.ucsc.edu/).

Genome-wide association study (GWAS)

Genome-wide association analyses were performed to identify loci associated with either population differentiation or the performance attributes of endurance or heat tolerance. 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 (Hakonarson and Grant 2011; Parker et al. 2010). Six loci associated with sprint and distance population variation had P values $<4.68 \times 10^{-6}$ (permutated P values $<3 \times 10^{-6}$) (Supplementary Table 2). SNP CFA3.82650187 had the most significant population association, with a P value

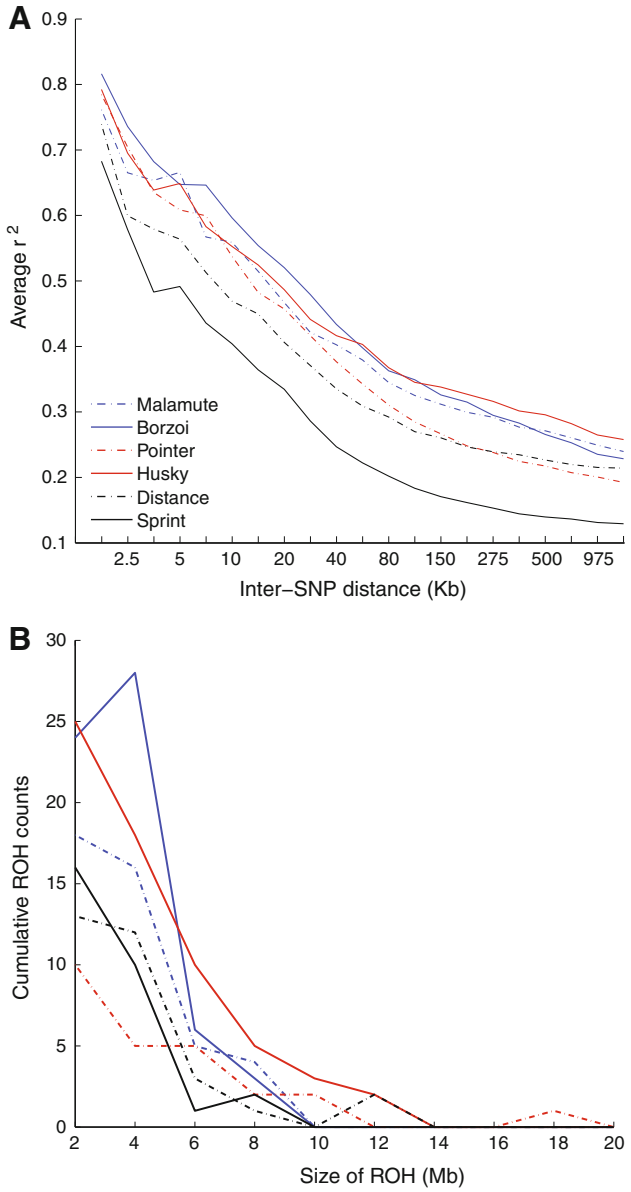

Fig. 3 The estimated decay of linkage disequilibrium and degree of autozygosity among Alaskan sled dogs and their four ancestral component breeds. Alaskan sled dogs are divided into distance and sprint racing styles and are compared with their four ancestral reference populations. **a** The decay of linkage disequilibrium (LD) is estimated from the distance at which the genotypic association, r2, reaches a threshold of 0.5. **b** The degree of autozygosity is determined through the cumulative number of runs of homozygosity (ROH) of various lengths (Mb)

of 1.03×10^{-7}, and is located 1 Mb upstream from the selective sweep region containing *ARL2BP*. The next significantly associated region contained two SNPs in a gene-rich region (25 genes annotated in a ± 1-Mb window) on CFA11 (P values of 1.00×10^{-6}). Of the 25 genes, the histidine triad nucleotide binding protein 1 (*HINT1*) gene, located approximately 70 and 600 kb (http://genome.ucsc.edu/) (UCSC 2011), respectively, from these SNPs, is the most interesting candidate for these studies. *HINT1* was previously associated with anxiety- and stress-coping behaviors in knockout mice (Barbier and Wang 2009; Varadarajulu et al. 2011).

Elite versus poorly performing dogs were assessed for each class of sled dog. While endurance in sprint sled dogs was associated with 15 loci, characterized by SNPs with P values $<1 \times 10^{-6}$, permutation testing proved all sites statistically unstable (P values $>1 \times 10^{-4}$). Performance of the heat tolerance attribute in sprint dogs showed stronger association stability, delineating a region on CFA10 (31,089,847–31,188,654 bp) with four clustered SNPs (P values from 4.53×10^{-6} to 5.57×10^{-7} and permutated P values from 1.20×10^{-5} to 5×10^{-6}) (Fig. 4; Supplementary Table 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 a ±1-Mb window around the critical SNPs, but these either do not have gene or protein function information or they have not demonstrated an association with heat tolerance. The *MYH9* gene makes for an intriguing candidate. It has been associated with muscle efficiency, and differences in protein activity have been observed in an association with variation in muscle temperature (Burniston 2009; Gray et al. 2006; Ingalls et al. 1998).

Ancestry modeling

A genome-wide ancestry profile was generated for the sprint and distance sled dogs to determine regions of ancestry selection based on the four reference breeds (AMAL, HUSK, GSHP, BORZ) (Huson et al. 2010). The genome has an overall mosaic structure in each of the sled dog populations (Fig. 5). However, on average, the distance sled dog genome is composed of 32% AMAL, 26% HUSK, 23% GSHP, and 19% BORZ, whereas the sprint sled dog genome is predominantly GSHP (33%), with 25% AMAL, 22% HUSK, and 20% BORZ (Table 1; Fig. 6a). Notably, GSHP was substantially higher in the sprint dogs, accounting for the largest proportion of ancestry (sprint 33% vs. distance 23%) (Fig. 6a; Table 1). A genome-wide analysis of ancestry block frequencies demonstrated that the most frequent block in distance sled dogs was the AMAL (AMAL/AMAL, 13%; HUSK/AMAL, 13%; GSHP/AMAL, 13%; BORZ/AMAL, 11%), while it was the GSHP in sprint dogs (AMAL/GSHP, 16%; HUSK/GSHP, 13%; GSHP/GSHP, 13%; BORZ/GSHP, 12%) (Fig. 6b; Table 1).

We identified 447 total ancestry blocks in the Alaskan sled dog. A total of 186 unique ancestry blocks were private to either distance ($n = 97$ unique blocks; median length = 1,337 kb) or sprint ($n = 89$ unique blocks; median length = 1,137 kb) dogs (Table 2). The most

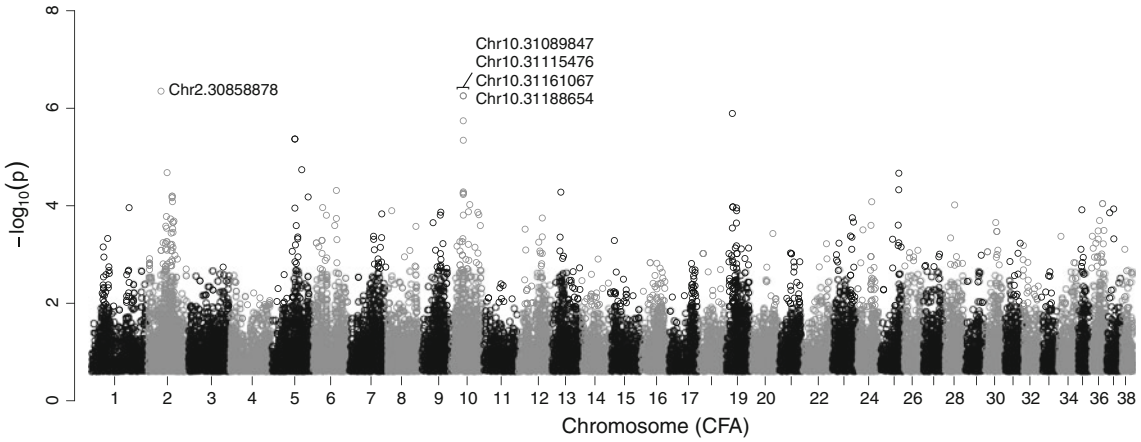

Fig. 4 Genome-wide association results of elite versus poorly performing sprint dogs for the heat-tolerance attribute. Two genomic loci located on CFA 2 and 10 were identified in a comparison of 21 elite and 17 poor-performing sprint dogs with regard to the heat-tolerance attribute. A panel of 115,425 SNPs spanning all autosomes and the X chromosome was tested. The *x* axis denotes SNP positions in increasing genomic order from CFA 1 through 38 and the X chromosome. The *y* axis indicates the −log10 *P* value as determined in an association analysis using the program EMMAX

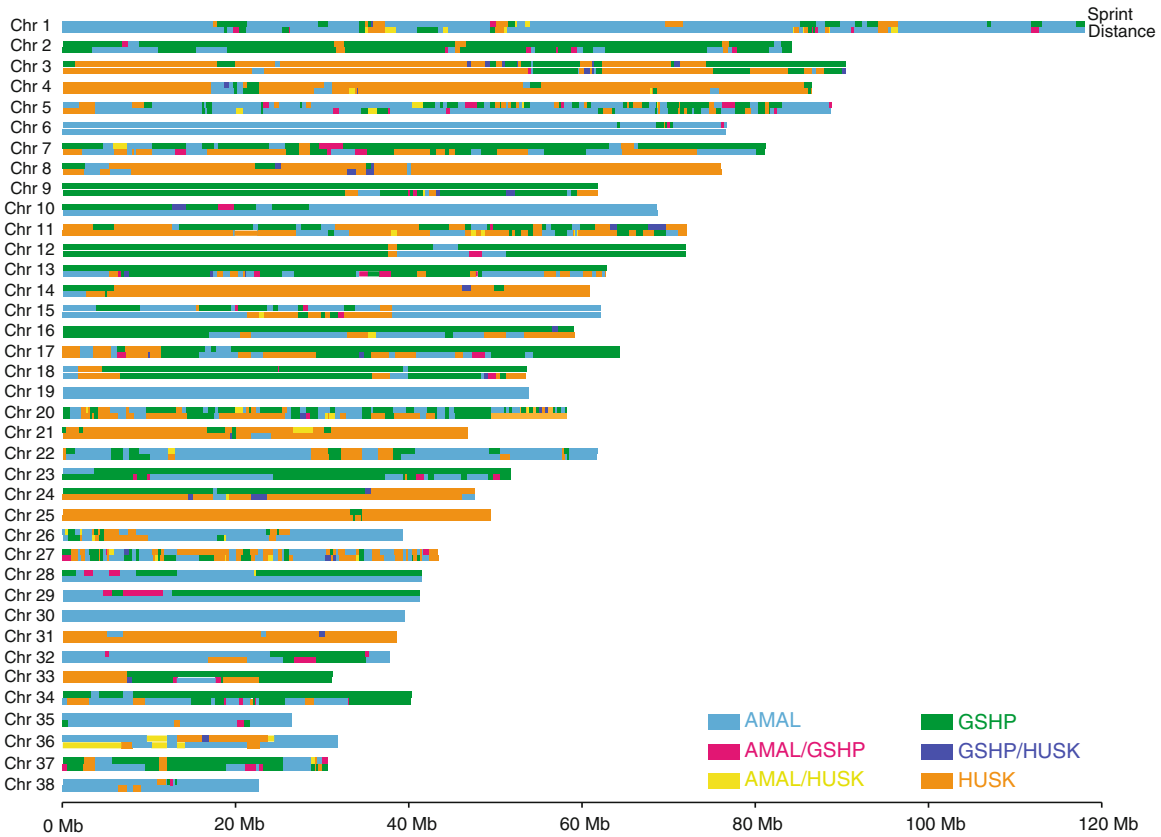

Fig. 5 A comparison of the most prevalent diploid state ancestry blocks across the genomes of sprint and distance sled dogs. Individual chromosomes are indicated on the *y* axis, while the *x* axis denotes genomic position (Mb). The most common diploid ancestry blocks across the genome are visualized using the color scheme with the diploid states (homozygous or heterozygous) as shown in the lower right of the figure

frequent of these blocks in distance dogs was AMAL/GSHP (18%, 80/447) and the longest ancestry block was a homozygous state of AMAL (2,354 kb). Seventeen percent of the blocks private to the sprint dogs were of BORZ/GSHP ancestry, with the longest being of HUSK/HUSK ancestry (1,891 kb).

We further identified 48 regions that showed the substantial ancestry differences between the sprint and

Table 1 The genome-wide frequency (*f*) of individual ancestral populations and their respective diploid ancestry states within the distance and sprint sled dog populations

Breed[b]	AMAL	BORZ	GSHP	HUSK
Distance sled dogs[a]				
AMAL	0.1328	0.1136	0.1279	0.1337
BORZ		0.043	0.0768	0.1074
GSHP			0.0754	0.1096
HUSK				0.0992
Total *f* (distance)	0.3181	0.1884	0.2271	0.2664
Sprint sled dogs[a]				
AMAL	0.071	0.1153	0.1582	0.0966
BORZ		0.0373	0.1194	0.1093
GSHP			0.1275	0.1313
HUSK				0.0535
Total *f* (sprint)	0.2533	0.2059	0.3251	0.2158

[a] A total of 19 distance dogs and 27 sprint dogs, all unrelated at the grandparent generation, were used to generate population frequencies

[b] A matrix of the diploid ancestry states with their respective genome-wide frequencies (*f*)

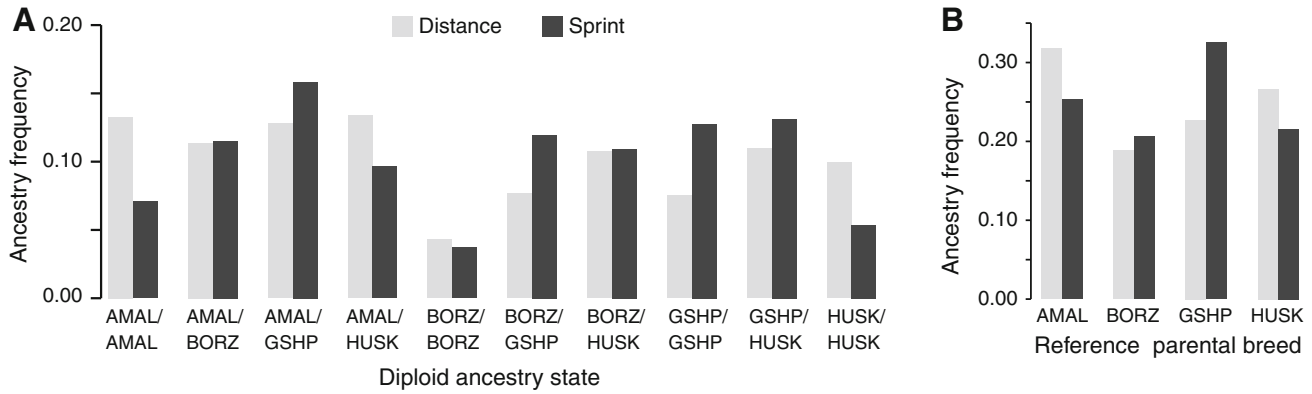

Fig. 6 A comparison of the genome-wide frequency of four ancestral reference breeds within the distance and sprint sled dog populations. **a** The genome-wide proportion of the individual ancestral reference breeds of AMAL, BORZ, GSHP, and HUSK within the distance and sprint populations. **b** The genome-wide proportion of diploid ancestry states within the distance and sprint populations

distance populations (Supplementary Table 3). The minimum ancestral frequency difference in these regions was 0.33, >2 SD from the mean (mean = 0.095; 2 SD = 0.26). The highest ancestral frequency difference was located on CFA11 (18,482,294–23,584,745 bp), a region that also had increased HUSK ancestry (frequency difference = 0.510) in distance dogs. This 5-Mb region contains two fibrillin genes (*FBN1* and *FBN2*) whose protein products are integral to the structure and function of connective tissue, and 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 (March 2006; http://genome.ucsc.edu/). Additionally, *HINT1* is located within this region and corroborates our GWAS results (Barbier and Wang 2009; Varadarajulu et al. 2011). Overall, 19 regions demonstrated a substantial excess of ancestry in sprint dogs, with two regions of excessive BORZ and 17 of excessive GSHP. The remaining 29 regions demonstrated an excess of ancestry in distance dogs and include 15 AMAL, 2 BORZ, 1 GSHP, and 11 HUSK ancestry blocks.

We combined the results from the selective sweep, GWAS, and ancestry analysis to tabulate the regions that have overlapping significant results for the sled dog populations. Here, we attempted to differentiate between random ancestry excess and nonrandom inheritance of variants due to the directional selection for functional phenotypes in the sled dog. Five selective sweep regions overlapped four regions of ancestry selection and were located on CFA3, 10, 16, and 28 (Table 3). CFA3 contains two selective sweeps in distance dogs that also contain a signal of positive selection for HUSK ancestry. This region

Table 2 The overall number, median length, and genome-wide frequency (f) of diploid ancestry blocks found exclusive to either the distance or sprint sled dog populations

	AMAL	BORZ	GSHP	HUSK
Distance sled dogs				
Number of ancestry blocks (total = 447)				
AMAL	45	35	80	72
BORZ		12	35	44
GSHP			40	52
HUSK				32
Median length (kb) of ancestry block				
AMAL	2354	1660	1464	1428
BORZ		1046	1062	1151
GSHP			1740	1083
HUSK				1581
f (ancestry block)				
AMAL	0.101	0.078	0.179	0.161
BORZ		0.027	0.078	0.098
GSHP			0.09	0.116
HUSK				0.071
Sprint sled dogs				
Number of ancestry blocks (total = 392)				
AMAL	17	37	54	58
BORZ		9	65	46
GSHP			39	44
HUSK				23
Median length (kb) of ancestry block				
AMAL	1671	1054	939	967
BORZ		1112	926	1112
GSHP			996	1333
HUSK				1891
f (ancestry block)				
AMAL	0.043	0.094	0.138	0.148
BORZ		0.023	0.166	0.117
GSHP			0.01	0.112
HUSK				0.059

includes the gene solute carrier family 2, member 9 (*SLC2A9*) gene, which is integral to glucose homeostasis as a glucose transporter (March 2006; http://genome.ucsc.edu/). CFA10 also had coinciding selective sweep and GSHP ancestry selection, but in different populations (Table 3). The nearest gene, methionine sulfoxide reductase B3 (*MSRB3*), encodes a protein that performs crucial functions for cell protection against oxidative stress, which may be important for sled dogs that perform under extreme physiological and environmental conditions (Kwak et al. 2009).

Two distinct ancestry patterns occur in the selective sweep on CFA16. There is a large region of positive selection for GSHP ancestry (0.398 frequency difference, Supplementary Table 3) in sprint dogs, and a selective sweep with a 0.25 decrease in AMAL ancestry, coinciding with an increase of 0.25 for HUSK in distance dogs (Fig. 7). Located within this region is the protein tyrosine phosphatase, receptor type, N polypeptide 2 (*PTPRN2*) gene, which functions in insulin binding and β-cell growth regulation within the insulin granule (Suckale and Solimena 2010). CFA28 possessed a selective sweep (29,046,328–29,143,901 bp) with an excess of AMAL (0.312 frequency difference) in distance dogs and a strong frequency difference for GSHP in sprint dogs (0.421). We identified attractin-like 1 (*ATRNL1*) as a candidate gene of interest in this region as it contributes to cognitive functionality, information processing, and distinct morphological characteristics (e.g., dysmorphic facial attributes and toe syndactyly) (Luciano et al. 2011; Stark et al. 2010).

Using both GWAS and ancestry analyses, we further identified two regions, on CFA11 and 32, that significantly differentiated sprint and distance dogs

Table 3 Description of the genetic loci demonstrating the highest degree of interest for population differentiation or performance association within Alaskan sled dogs

Method of identification[a]	Chr	Start (bp)	End (bp)	Block length (bp)	Sled dog population[b]	Ancestry population[c]	GWAS association[d]	Performance candidate genes	No. of genes within region[e]
Selective sweep, SABER	3	71896408	71898732	2,324	Distance	HUSK		*SLC2A9*	11
Selective sweep, SABER	3	72727082	72784438	57,356	Distance	HUSK		*SLC2A9*	14
GWAS	3	82650187					Population	*ARL2BP*	2
Selective sweep	3	83775932	83798854	22,922	Distance			*ARL2BP*	2
Selective sweep, SABER	10	11081762	11121003	39,241	Distance	GSHP		*MSRB3*	15
GWAS, SABER	10	31089847	31188654	98,807		GSHP	Heat Tolerance	*MYH9*	34
SABER	11	18482294	23584745	5,102,451	Distance	HUSK		*FBN1, FBN2, ACSL6, SLC27A6, HINT1*	51
GWAS, SABER	11	22331950	23117401	785,451		GSHP/ HUSK	Population	*HINT1*	25
Selective sweep, SABER	16	23391731	23391985	254	Distance	GSHP		*PTPRN2*	13
Selective sweep, SABER	28	29046328	29143901	97,573	Distance	GSHP		*ATRNL1*	12
GWAS, SABER	32	8774288				GSHP	Population	*HBZ*	11

SNP positions are based on the CanFam2 assembly

[a] Genomic regions of interest were determined by demonstrating an excess of breed ancestry (SABER), selective sweep, or genome-wide association

[b] The sled dog population in which the selective sweep was significant

[c] The reference breed population of excess ancestry

[d] The sled dog population in which the genome-wide association was significant

[e] Total number of human genes annotated within the genomic region of interest as well as 1 Mb upstream and 1 Mb downstream of said region

(P values $< 1 \times 10^{-6}$) (Table 3). The CFA11 locus was highlighted by two SNPs and provided independent confirmation of *HINT1* as a candidate gene (March 2006; http://genome.ucsc.edu/) (Supplementary Table 3). The *MYH9* gene, investigated for its role in heat tolerance, also correlated with positive selection of the GSHP within sprint dogs (frequency difference 0.313). This region was not highlighted in our initial analysis because the frequency of ancestry fell below the 95th percentile threshold (frequency difference ≥ 0.333) (Supplementary Table 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

Direct sequencing of the four *HINT1* exons and their surrounding region produced seven noncoding variants found in sprint and distance dogs. Six of the variants were found in GSHP and four were found within HUSK. None of the variants were found to be associated with the sprint or distance sled dogs.

GWAS identified two SNPs located within the *MYH9* gene with a significant association to heat tolerance, and two additional SNPs located 45 and 20 kb upstream of the 5′ end of the gene that were also significant. Direct sequencing of six elite and six poorly performing sprint

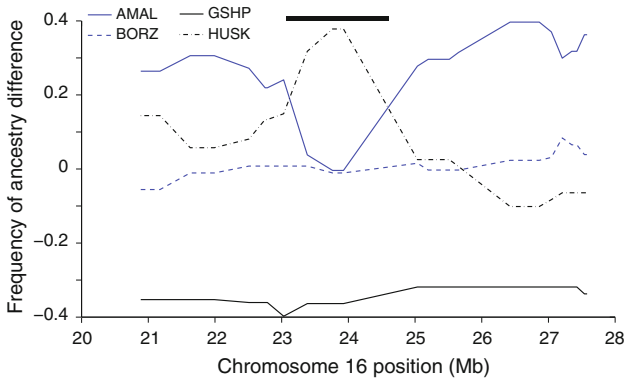

Fig. 7 Chromosome 16 SNP frequency differences for the four ancestral breeds when comparing distance and sprint populations. The difference in frequency scores (*y* axis) between distance and sprint dogs for each ancestral breed was plotted in relation to chromosome 16 SNPs (*x* axis). A more positive 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 highlighted by the black bar denotes an area highlighted as being in the top 5% of genomic regions, demonstrating the greatest degree of ancestry selection between sprint and distance dogs, as well as corresponding to a region of selective sweep within distance dogs

dogs with regard to the heat tolerance attribute through the 40 *MYH9* exons and conserved flanking regions revealed 51 variants. Forty-three variants were within the *MYH9* gene, including 5 SNPs within exons, 43 SNPs within introns, and 2 indels within introns. An additional eight SNPs and two indels were found upstream of the 5′ end of the gene. Synonymous amino acid changes were found in exons 4 (31,155,024 bp), 9 (31,161,766 bp), 18 (31,175,229 bp), 24 (31,181,751 bp), and 29 (31,184,517 bp).

We conducted a preliminary single-marker association analysis of 70 markers (indels and SNPs) that compared six sprint dogs from each of the elite and poorly performing classes for heat tolerance (4 GWAS SNPs + 51 sequencing variants + 8 SNPs + 2 indels upstream of 5′ end of gene +5 synonymous amino acid changes). This analysis revealed 16 SNPs with raw *P* values <0.05 that were associated with poor heat tolerance. An additional set of 26 poor performers and 15 elite performers with regard to heat tolerance were genotyped for 16 *MYH9* SNPs that had demonstrated an association in the initial 12 dogs. Single marker analysis of these 16 SNPs comparing 32 (26 + 6) poor to 21 (15 + 6) elite sprint with regard to heat tolerance yielded 14 SNPs with raw *P* values <0.05 (Table 4). Seven of these SNPs retained permutated *P* values <0.05, with the most significant SNP exhibiting a permutated *P* value of 0.0001 (Table 4). A pairwise comparison of LD among these seven SNPs revealed substantial linkage ($D' > 0.90$) for 65% of the pairwise comparisons, with the remaining 35% demonstrating moderate to strong LD ($D' = 0.6$–0.9).

We also analyzed 16 SNPs from the *MYH9* gene using DNA collected from AMAL and GSHP, two breeds demonstrating excessive ancestry within the sprint dog genome. Three of the SNPs were found to be associated,

Table 4 SNPs within and surrounding the *MYH9* gene on canine chromosome 10 associated with heat tolerance performance in sprint racing Alaskan sled dogs

CanFam2 position	Alleles[a]	Poor HT[b] MAF	Elite HT[c] MAF	Poor HT associated allele	*P*	Permutation *P*
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

HT Heat tolerance, *MAF* minor allele frequency

[a] Major:minor allele

[b] Frequency of the minor allele in 32 poorly performing heat-tolerance sprint dogs (cases)

[c] Frequency of the minor allele in 21 elite performing heat-tolerance sprint dogs (controls)

Table 5 *MYH9* gene SNPs on canine chromosome 10 associated to either the Alaskan Malamute or German Shorthaired Pointer breed

CanFam2 position	Alleles[a]	AMAL[b] MAF	GSHP[c] MAF	AMAL associated allele	*P*	Permutation *P*
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

HT heat tolerance, *MAF* minor allele frequency

[a] Major:minor allele

[b] Frequency of the minor allele in six Alaskan Malamutes (cases)

[c] Frequency of the minor allele in eight German Shorthaired Pointers (controls)

demonstrating both raw and permutated *P* values of <0.05 (Table 5). SNPs CFA10.31115476 and CFA10.31123184 had similar allele frequencies between poorly performing (heat tolerance attribute) sprint dogs ($P = 2.02 \times 10^{-6}$, *G* allele 0.840; $P = 0.0024$, *T* allele 0.553) and AMAL ($P = 4.92 \times 10^{-5}$, *G* allele 0.700; $P = 6.00 \times 10^{-4}$, *T* allele 0.700). SNP CFA10.31156486, for which the *A* allele is associated with the Alaskan Malamute breed, did not show a significant association to heat tolerance after permutation testing. Also, there were no significant differences in the allele frequency regarding poor-performing ($f_A = 0.487$) versus elite ($f_A = 0.310$) sprint dogs with regard to heat tolerance.

Discussion

The Alaskan sled dog is the embodiment of a unique, genetically distinct breed developed solely by selection and breeding for athletic attributes (Huson et al. 2010). They possess a distinct admixed population structure, a consequence of crossing purebred dogs possessing desirable performance traits to what were at the time native Alaskan sled dogs (Huson et al. 2010). The end result is two populations of modern Alaskan sled dogs, optimized for racing short (up to 48 km) or long (~1,609 km) distances. In this study we demonstrated that sprint and distance Alaskan sled dogs are genetically distinct, which corroborates our published findings (Huson et al. 2010) in which microsatellite marker data were used to cluster dogs based on their racing style (Fig. 2a).

We used a set of 7,644 AIM SNPs to model ancestry in sprint and distance sled dog populations with four known reference breeds: Alaskan Malamute, Siberian Husky, German Shorthaired Pointer, and Borzoi. The distance sled dogs had, on average, highest AMAL ancestry (32%) compared to sprint dogs whose highest ancestry was the GSHP (33%). As a result, the most frequent ancestry blocks contained at least one AMAL haplotype in the distance dogs and one GSHP haplotype in sprint dogs (Table 1). This distinct difference in ancestry is likely due to mating strategies that crossed closely related individuals together in order to retain desirable traits. It is likely, therefore, that there are selective advantages for a distance sled dog to have an excess of AMAL ancestry and for sprint dogs to retain GSHP ancestry. Other differences include the fact that distance dogs had a greater number of long (~12 Mb) ROH, a length comparable to those found in purebred Siberian Huskies. Finally, when we compared the ancestry blocks unique to each population, we found that distance dogs have larger private blocks than sprint dogs (Table 2), a result that is concordant with previous microsatellite data and likely reflects the particular breeding strategies used to propagate the population (Huson et al. 2010).

Our ancestry analyses highlight 48 loci that demonstrated a substantial contribution to either the sprint or distance populations (Supplementary Table 3). Investigation of LOH produced 60 regions characterized by selective sweeps, with 87% of those occurring in distance dogs (Supplementary Table 1). While this may be indicative of complex genetic interactions with genes of small effect, we postulate that there are also more attributes under selection within distance dogs; therefore, genomic variation should be, on average, more constrained. Some of these selective sweep regions may signify characteristics that are strictly maintained in distance dogs due to the extreme nature of their racing conditions (e.g., fur length, hair follicle density, hardiness of the toe pads). We utilized a unique approach that combined the ancestry results with selective sweep and GWAS methods to identify a subset of eight regions likely experiencing selective pressure within a sled dog population due to their athletic performance. In all, five areas of selective sweep overlapped four regions of ancestry selection, with potentially interesting candidate genes located at several of the loci (Table 3). CFA3 displayed highly concordant results in the distance dogs. The remaining loci showed a more complex ancestry pattern

where diversity derived from multiple breeds was obviously beneficial. Future research using a denser set of AIMs is required to understand how genes under selection are developed and maintained in breeds whose sole purpose is to perform. A denser set of SNPs is also needed to identify causative variants.

Genome-wide association analyses were used to identify loci associated with either population variation between sprint and distance dogs or the performance attributes of endurance and heat tolerance. Sampling sled dogs from high-performance racing kennels compounded GWAS issues since there were few poor-performing (rank 3) dogs for either endurance or heat tolerance. Therefore, it was necessary to pool dogs scoring 2 and 3 for the GWAS, which decreased our ability to identify the desired loci. However, this created a potential problem with differential relatedness among cases versus controls. To accommodate this problem, we closely matched cases and controls for analysis using EMMAX software, which corrects for both population relatedness and structure. The GWAS results overlapped with three loci that had an excess of ancestry (Table 3), with two of these regions (CFA11 and 32) related to sled dog population differentiation.

The *HINT1* gene on CFA11 allowed us to explore the impact of differential ancestry on sprint versus distance dogs. We originally hypothesized that this gene may account for differences in stress-coping abilities between the two groups. An excess of HUSK ancestry in distance dogs and GSHP in sprint dogs that overlapped the *HINT1* gene supported this idea; however, no association was found with any *HINT1* variants and any population of dogs.

The remaining question of HUSK versus GSHP ancestry, however, is interesting. While the Siberian Husky has been characterized as "stubborn and easily bored" despite its hardy working dog nature, the German Shorthaired Pointer breed is noted for its "ease of training and adaptability" along with its commitment to performing (AKC 1998). 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, with many feeling the cross has an increased desire to perform but may not handle stress as well as the non-Pointer crosses.

The heat tolerance attribute was associated with a cluster of four GWAS SNPs on CFA10, two of which were within the *MYH9* gene and demonstrated genome-wide significance (two highest SNPs had a P value of 5.57×10^{-7}) (Fig. 4; Table 3). Previous research has associated an increase in myosin heavy chain production with increased cardiac output (Burniston 2009). Another study found that a decrease in myosin heavy chain and actin within injured mouse extensor muscles accounts for approximately a 58% reduction in isometric titanic force output (Burniston 2009; Ingalls et al. 1998). Most notably, it has been reported that the percent increase of the myosin heavy chain type II class A within muscle tissue experiencing elevated temperatures (ET = 37.5°C; Normal, N = 34.2°C) correlates with the magnitude of increased power output. It is thought that slight temperature elevations improve muscle fiber power output through an increase in the rate of anaerobic ATP turnover and muscle fiber conduction velocity (Gray et al. 2006). However, the efficiency of muscle contraction actually decreased as temperature rose. Overall, Gray et al. (2006) has concluded that fibers with a high proportion of myosin heavy chain type II class A were the most sensitive to temperature fluctuations (Gray et al. 2006). We found that the *MYH9* gene overlapped with an excess of GSHP and AMAL ancestry in sprint sled dogs, and our fine mapping identified seven SNPs associated with heat tolerance (Table 4). The two most significantly associated (CFA10.31105851, 7.83×10^{-06}, and CFA10.31115476, 2.02×10^{-06}) were 19 and 29 kb upstream from the 5′ end of the *MYH9* gene and in nearly complete LD with the other four SNPs of note in this region ($D' \geq 0.871$). SNP CFA10.31105851 lies within a conserved region of the dog, human, and mouse genomes. In the human genome, chromatin profiling of human skeletal muscle myoblasts shows this region to be an active promoter site. At least seven other human cell types showed signs of having strong enhancers within this region. Likewise, a second SNP, CFA10.31121778, located in an additional conserved region upstream from the 5′ end of the first region, was also found to have a strong enhancer in the analogous region of the human genome (Ernst and Kellis 2010; Ernst et al. 2011). We postulate that variants within promoter and enhancer regulatory sites may be the means by which the canine *MYH9* gene potentially affects heat tolerance performance in sprint sled dogs, although functional studies remain to be done.

SNPs CFA10.31115476 and CFA10.31123184 differentiated significantly between the contributions of AMAL and GSHP (Table 5). Specifically, sprint dogs with poor heat tolerance had allele frequencies similar to those of AMAL. Elite heat-tolerance sprint dogs had a decreased allele frequency at these same two SNPs, with GSHP having the lowest allele frequencies. Previous research associating muscle temperature elevation and power output, combined with our GWAS and fine-mapping results, highlights the *MYH9* gene as an intriguing candidate, potentially affecting the heat tolerance attribute in sprint sled dogs.

Our current study corroborated our previous finding that Alaskan sled dogs are two distinct populations, largely attributed to selective breeding for their divergent racing styles (Huson et al. 2010). A number of candidate genes potentially affecting performance were highlighted by

GWAS and selective sweep analyses within the sled dogs. In addition, we implemented methods from admixture mapping to pinpoint genomic regions that have an excess of a particular reference breed. We found *MYH9* to be associated with heat tolerance performance in sprint dogs, demonstrating the success of researching performance mechanisms within a group of recently admixed dogs. This study provides a foundation for the study of sled dog performance genetics, as well as breed origins. As ever-denser GWAS studies are performed and data sets increase in size, the power to fine map and eventually identify truly causative variants will increase (Ostrander et al. 2009). Finally, our preliminary evidence suggesting a role for the *MYH9* gene in heat tolerance among sprint sled dogs highlights the types of genes and gene families that will likely become the basis of functional studies regarding performance-enhancing genes in the years to come.

Acknowledgments We gratefully acknowledge the Intramural Program of the National Human Genome Research Institute. We thank the many owners and participants who provided DNA samples and information regarding their dogs.

References

ADMA (2011) Alaska Dog Musher's Association, Fairbanks, AK. Available at http://www.sleddog.org

Barbier E, Wang JB (2009) Anti-depressant and anxiolytic like behaviors in PKCI/HINT1 knockout mice associated with elevated plasma corticosterone level. BMC Neurosci 10:132

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265

Bhangale TR, Stephens M, Nickerson DA (2006) Automating resequencing-based detection of insertion-deletion polymorphisms. Nat Genet 38:1457–1462

Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K, Brisbin A, Parker HG, von Holdt BM, Cargill M, Auton A, Reynolds A, Elkahloun AG, Castelhano M, Mosher DS, Sutter NB, Johnson GS, Novembre J, Hubisz MJ, Siepel A, Wayne RK, Bustamante CD, Ostrander EA (2010) A simple genetic architecture underlies morphological variation in dogs. PLoS Biol 8:e1000451

Buerkle CA, Lexer C (2008) Admixture as the basis for genetic mapping. Cell 23:686–694

Burniston JG (2009) Adaptation of the rat cardiac proteome in response to intensity-controlled endurance exercise. Proteomics 9:106–115

Cheng CY, Reich D, Wong TY, Klein R, Klein BE, Patterson N, Tandon A, Li M, Boerwinkle E, Sharrett AR, Kao WH (2010) Admixture mapping scans identify a locus affecting retinal vascular caliber in hypertensive African Americans: the Atherosclerosis Risk In Communities (ARIC) study. PLoS Genet 6:e1000908

Collins MJC (1991) Dog Driver: A Guide for the Serious Musher, 1st edn. Alpine Publications, Crawford

Ernst J, Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat Biotechnol 28:817–825

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473:43–49

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502–1506

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8:195–202

Gray SR, De Vito G, Nimmo MA, Farina D, Ferguson RA (2006) Skeletal muscle ATP turnover and muscle fiber conduction velocity are elevated at higher muscle temperatures during maximal power output development in humans. Am J Physiol Regul Integr Comp Physiol 290:R376–R382

Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne RK (2009) Linkage disequilibrium and demographic history of wild and domestic canids. Genet 181: 1493–1505

Hakonarson H, Grant SF (2011) Planning a genome-wide association study: Points to consider. Ann Med 43:451–460

Huson HJ, Parker HG, Runstadler J, Ostrander EA (2010) A genetic dissection of breed composition and performance enhancement in the Alaskan sled dog. BMC Genet 11:71

Iditarod (2011) The Official Site of the Iditarod, Iditarod Trail Committee, Inc., Anchorage, AK. Available at http://www.iditarod.com/

Ingalls CP, Warren GL, Armstrong RB (1998) Dissociation of force production from MHC and actin contents in muscles injured by eccentric contractions. J Muscle Res Cell Motil 19:215–224

Kang HM (2010) Efficient Mixed_model Association eXpediated (EMMAX). Los Angeles, CA: University of California, Los Angeles

KC A (1998) The Complete Dog Book, Official Publication of The American Kennel Club, 19th edn. Howell Book House, New York

Kwak GH, Kim JR, Kim HY (2009) Expression, subcellular localization, and antioxidant role of mammalian methionine sulfoxide reductases in *Saccharomyces cerevisiae*. BMB Rep 42:113–118

Luciano M, Hansell NK, Lahti J, Davies G, Medland SE, Raikkonen K, Tenesa A, Widen E, McGhee KA, Palotie A, Liewald D, Porteous DJ, Starr JM, Montgomery GW, Martin NG, Eriksson JG, Wright MJ, Deary IJ (2011) Whole genome association scan for genetic polymorphisms influencing information processing speed. Biol Psychol 86:193–202

Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 15: 2745–2751

Ostrander EA, Huson HJ, Ostrander GK (2009) Genetics of atheletic performance. Annu Rev Genomics Hum Genet 10:407–429

Parker HG, Shearin AL, Ostrander EA (2010) Man's best friend becomes biology's best in show: genome analyses in the domestic dog. Annu Rev Genet 44:309–336

Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA et al (2004) Methods for high-density admixture mapping of disease genes. Am J Hum Genet 74:979–1000

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14

Purcell S (2009) PLINK_v1.07, Whole genome association analysis toolset. Available at http://pngu.mgh.harvard.edu/~purcell/plink/download.shtml

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

Rennick P (ed) (1987) Dogs of the North. Alaska Geographic Society, Anchorage

Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M (2010) Genome-wide association studies in diverse populations. Nat Rev Genet 11:356–366

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 78:629–644

Seldin MF, Posaniuc B, Price AL (2011) New approaches to disease mapping in admixed populations. Nat Rev Genet 12:523–528

Sharer JD, Shern JF, Van Valkenburgh H, Wallace DC, Kahn RA (2002) ARL2 and BART enter mitochondria and bind the adenine nucleotide transporter. Mol Biol Cell 13:71–83

Shriner D (2011) Investigating population stratification and admixture using eigenanalysis of dense genotypes. Heredity 107(5): 413–420

Stark Z, Bruno DL, Mountford H, Lockhart PJ, Amor DJ (2010) De novo 325 kb microdeletion in chromosome band 10q25.3 including *ATRNL1* in a boy with cognitive impairment, autism and dysmorphic features. Eur J Med Genet 53:337–339

Suckale J, Solimena M (2010) The insulin secretory granule as a signaling hub. Trends Endocrinol Metab 21:599–609

Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA (2004) Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. Genome Res 14:2388–2396

Tang H, Coram M, Wang P, Zhu X, Risch N (2006) Reconstructing genetic ancestry blocks in admixed individuals. Am J Hum Genet 79:1–12

Tang H, Choudhry S, Mei R, Morgan M, Rodriguez-Cintron W, Burchard EG, Risch NJ (2007) Recent genetic selection in the ancestral admixture of Puerto Ricans. Am J Hum Genet 81: 626–633

Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF (2006) A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. Am J Hum Genet 79:640–649

UCSC (2011) University of California, Santa Cruz Genome Browser, human build March 26; canine version 2.0. Available at http://genome.ucsc.edu/

Varadarajulu J, Lebar M, Krishnamoorthy G, Habelt S, Lu J, Bernard Weinstein I, Li H, Holsboer F, Turck CW, Touma C (2011) Increased anxiety-related behaviour in Hint1 knockout mice. Behav Brain Res 220:305–311

Vaudrin B (ed) (1976) Racing Alaskan Sled Dogs, 1st edn. Alaska Northwest Publishing Company, Anchorage

VonHoldt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A, Reynolds A, Bryc K, Brisbin A, Knowles JC, Mosher DS, Spady TC, Elkahloun A, Geffen E, Pilot M, Jedrzejewski W, Greco C, Randi E, Bannasch D, Wilton A, Shearman J, Musiani M, Cargill M, Jones PG, Qian Z, Huang W, Ding ZL, Zhang YP, Bustamante CD, Ostrander EA, Novembre J, Wayne RK (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature 464:898–902

Wendt R (1999) Alaska Dog Mushing Guide: Facts and Legends, 5th edn. Goldstream Publications, Fairbanks

Winkler CA, Nelson GW, Smith MW (2010) Admixture mapping comes of age. Annu Rev Genomics Hum Genet 11:65–89

Yukon Quest (2011) Yukon Quest Sled Dog Race. Available at http://www.yukonquest.com/