Title

Population Structure and Evidence of Selection in Domestic Dogs and Gray Wolves Based on X Chromosome Single Nucleotide Polymorphisms

Author

Shohfi, Hanna

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Population Structure and Evidence of Selection in Domestic Dogs and Gray Wolves Based on X Chromosome Single Nucleotide Polymorphisms

A thesis submitted in partial satisfaction of the requirements for the degree of Master of Science in Biology

by

Hanna Elisibeth Shohfi

2013

ABSTRACT OF THE THESIS

Population Structure and Evidence of Selection in Domestic Dogs and Gray Wolves Based on X Chromosome Single Nucleotide Polymorphisms

by

Hanna Elisibeth Shohfi

Master of Science in Biology

University of California, Los Angeles, 2013

Professor Robert K. Wayne, Chair

Genomic resources developed for the domestic dog have provided powerful tools for studying canine evolutionary history and dog origins. Although X chromosome data are often excluded from these analyses due to their unique inheritance, comparisons of the X chromosome and the autosomes can illuminate differences in the histories of males and females as well as shed light on the forces of natural selection. Here we use X chromosome single nucleotide polymorphisms (SNPs) to analyze evolutionary relationships among populations of gray wolves worldwide in comparison to domestic dogs, and investigate evidence of selection. The results are concordant with population structure indicated by autosomal data. We additionally conducted a selection scan to identify loci that are putatively under selection.

The thesis of Hanna Elisibeth Shohfi is approved.

Blaire VanValkenberg

Janet S. Sinsheimer

Robert K. Wayne, Committee Chair

University of California, Los Angeles

2013

Table of Contents

INTRODUCTION

Recent advances in genome technologies developed for model species have allowed increasing resolution of the forces that shape the evolution of their genomes. The dog genome sequence has provided an important resource for understanding the evolutionary history of wolf-like canids and the genetic changes that occurred during dog domestication. In a recent study, vonHoldt et al. (2011) used the Affymetrix Canine SNP Genome Mapping Array (version 2) to reveal distinct hierarchical population units within gray wolves based on 48K autosomal loci. Due to the previous exclusion of the X chromosome SNPs, we here provide an independent parallel analysis of these X-linked markers to test conclusions from this study and provide new information on population structure and selection during the evolution of wolf-like canids.

An important feature of the sex chromosomes is that they are present in only a single copy in the heterogametic sex. In mammals, males have only one copy of the X chromosome, so every existing X chromosome has spent two-thirds of its history in females. Because sex chromosomes do not spend equal time in each of the sexes, they will experience different effects from evolutionary processes. The genetic diversity of chromosome X is expected, under equilibrium assumptions, to be three-quarters of that of the autosomes in a population where the two sexes have an identical distribution of offspring numbers. Deviations from this ratio can reveal sex-based differences in mutation, recombination, migration, selection, and changes in population size over time, as predicted by population genetic theory. Another consequence of the smaller population size of the X chromosome is that genetic drift is expected to be faster than for the autosomes. As a result, population differentiation, measured as F_{ST}, should be more pronounced on the X chromosome (Schaffner, 2004).

Recent studies of X-linked variation in human populations have demonstrated accelerated drift during the human dispersal out of Africa resulting in reduced X chromosome to autosomal diversity in non-African populations relative to West Africans, beyond the reduction expected from known historical changes in population size (Gottipati et al., 2011; Keinan et al., 2009). Conversely, Hammer et al. (2008) reported an increased ratio of X-linked to autosomal polymorphism due to background selection, changes in population size, and sex-specific migration in both African and non-African populations. Given the sex-biased dispersal patterns that have been documented in populations of gray wolves (Flagstad et al., 2003; Jędrzejewski et al., 2005; Pilot et al., 2006), it is likely that the ratio of X chromosome to autosomal diversity will deviate from that predicted under neutrality.

Selection is another important evolutionary force that likely influences the genetic diversity of the X chromosome. Because males are hemizygous, any recessive deleterious mutation on the X chromosome would immediately be expressed and, therefore, efficiently removed from the population. Conversely, if a recessive mutation were beneficial, its immediate exposure to selection would be advantageous, and would facilitate adaptive evolution (Charlesworth et al., 1987; Vicoso and Charlesworth, 2006). Genes involved in major biological functions, such as reproduction or cognition, are choice targets for natural selection (Balaresque et al., 2004). The apparent clustering of these genes on sex chromosomes (Saifi and Chandra, 1999; Tao et al., 2003; Thiselton et al., 2002; Wang et al., 2001; Zechner et al., 2001) makes this genomic region interesting with regard to effects of natural selection on genetic variation. Previous genome-wide studies have identified several X chromosome loci associated with traits that may have played an important role during the domestication of dogs and the evolution of their wolf ancestors, such as

traits involved in reproduction, size dimorphism, sexual selection, and cognitive behaviors (Albert et al., 2012; Boyko et al., 2010; Chase et al., 2005; Li et al., 2002; Saetre et al., 2004). However, the extent to which these genes show signals of selection is not known.

Given the smaller effective population size of the X chromosome relative to the autosomes and the complex evolutionary history of wolf-like canids, we can predict deviations from the expected ¾ reduction in genetic variation on the X chromosome relative to autosomes reflecting the demographic forces unique to each population, and well as higher levels of differentiation between geographically isolated populations due to accelerated drift. Finally, we can predict that signatures of selection on the X chromosome will be readily detected because more mutations undergo selection in hemizygous males. To test these predictions, a number of methods are used to characterize the data represented by 412 X-linked SNPs typed in a panel of 163 gray wolves representing their worldwide distribution (Eurasia and North America) in comparison with domestic dogs.

We begin by examining the population structure based on variation in 412 X chromosome SNPs. We then use F_{ST} values and pairwise allele frequency differences to examine population differentiation and explore what the results of these analyses indicate about past demographic patterns. We then scan the X chromosome for genetic structure consistent with the influence of selection. Finally, we discuss several regions identified as being clear outliers from the rest of the chromosome with respect to SNP allele frequency distribution and linkage disequilibrium patterns, which is an indication of loci under positive selection.

MATERIALS AND METHODS

Dataset

This study utilized data derived from the CanMap project (Boyko et al., 2010; vonHoldt et al., 2010) that provided genome-wide SNP data from 912 domestic dogs and 337 wild canids based on genotyping with the Affymetrix Canine version 2 genome-wide SNP mapping array. Quality control filters were applied for the genotyping algorithm (Boyko et al., 2010), from which X chromosome SNP loci were obtained for analysis independent of the 48K autosomal SNP loci from the same data set (vonHoldt et al., 2010; 2011). After exclusion of SNPs from the pseudoautosomal region (PAR; first 6.5Mb of the X chromosome), and loci outside of the PAR that were heterozygous in males (which suggested genotyping errors), a final set of 412 high quality SNPs was obtained.

Genetic diversity

To compare the genetic diversity of the X chromosome SNPs to that of the autosomes, single-marker descriptive statistics (observed and expected heterozygosity) were calculated for the female gray wolves from populations with differing demographic histories (Table 1) as determined by vonHoldt et al. (2011) using PLINK (Purcell et al., 2007) for the 412 SNP data set. The X-to-autosome ratio of genetic diversity was then examined to evaluate if genetic diversity deviates from the predicted effective population size of 0.75 (Gottipati et al., 2011; Hammer et al., 2008; Keinan et al., 2009; Lambert et al., 2010). Additionally, estimates of differentiation (F_{ST}) between the females in populations of gray wolves suggested by principal component analysis (PCA) and STRUCTURE (see below) were calculated using GALAXY (Goecks et al., 2010) with the Weir-Cockerham estimator option (Weir and Cockerham, 1984).

Genetic structure analysis

Given that the X chromosome spends two thirds of its time in females, it is possible that X-linked genomic variation has a different underlying population structure than autosomal variation (Hedrick, 2007; Schaffner, 2004). To investigate this possibility, we analyzed the X chromosome data with STRUCTURE (Pritchard et al., 2000) and PCA. For the initial population genetic structure analysis, all wild canid samples with known sex information from the CanMap SNP data were used, including 148 gray wolves (Europe, n = 71; China, n = 10; Middle East, n = 19; North America, n = 48), 7 Mexican wolves, 8 Great Lake Wolves, 12 red wolves, and 38 coyotes. Additionally, a subset of 50 dogs was selected from the original CanMap SNP data set of 912 dogs from 85 breeds to reduce the impact of the large numbers of dogs relative to wolves. The subset of dogs was selected in a pseudorandom fashion to include a balanced ratio of males and females from each of the 9 ancient breeds (n=18) as well an equal number of male and female samples from 32 randomly selected modern breeds (female: n= 16; male: n=16).

The Bayesian inference program STRUCTURE is a model based clustering method that establishes K populations based on allele frequencies at each locus and then assigns each individual to populations that correspond to their proportional membership in each group. We used STRUCTURE to assess the 412 SNP data set at K=2-10, 10,000 burn-in iterations, 50,000 Markov chain Monte Carlo (MCMC) iterations, with three repetitions of these parameter settings. The alpha and likelihood statistics were verified to reach convergence before the 10,000 burn-in iterations were completed during each repetition for each number of K populations analyzed. Selection of optimal K based on STRUCTURE output was performed with the support of STRUCTURE HARVESTER software (Earl and vonHoldt, 2012). To choose an appropriate K value for presentation, the parameter delta K (Evanno et al. 2005), likelihood values, and

biological information were evaluated (as emphasized in the STRUCTURE manual: Pritchard et al., 2000).

STRUCTURE was used to carry out additional analyses for gray wolves only (Europe, China, Middle East, North America) and North American wild canids (gray wolves, red wolves, and coyotes). Each analysis was performed for males and females separately as well as combined. To assess differentiation between dogs and wolves as well as population subdivision within wolves, the data set was purged of coyotes and wolf populations showing evidence of interspecific admixture in STRUCTURE (vonHoldt et al. 2011) as well as wolves from highly inbred populations from Italy and Spain (Gray et al. 2009). This was done to enable finer resolution of genetic structure between populations of dogs and wolves by avoiding the excessive influence of these extremes in variation on the cluster analysis. The resulting PCA data set included gray wolves (n=125) from Europe, China, Middle East, and North America as well as a subset of domestic dogs of ancient breeds (n= 18; 9 female, 9 male) and modern breeds (n= 39; 24 female, 15 male).

PCA was performed using the R program dapc, part of adegenet-package 1.3-9 (Jombart and Ahmed, 2011) to visualize the dominant relationships in the 412 SNP data set. PCA was performed separately for males, females, and pooled sexes to explore the effect of male haploidy on the patterns of population structure inferred from the X chromosome SNPs. The most contributing SNPs (contributions in the 99^{th} percentile) of PC1 and PC2 were subsequently identified as potential diagnostic SNP loci.

To examine the effects of linkage disequilibrium (LD) on the identification of potential diagnostic SNP loci, a subset of SNPs in approximate linkage equilibrium were identified by linkage disequilibrium (LD)-based SNP pruning in PLINK (Purcell et al., 2007). LD pruning was applied to the 412 SNP data set of all female gray wolves from nonadmixed populations, starting with $r^2 > 0.5$ within a 50 SNP window, and proceeded with progressively lower thresholds to identify the most LD-pruned data set without losing diagnostic SNPs for each LD block. This resulted in a pruned list of 244 SNPs with $r^2 > 0.3$, which were used in subsequent PCA analyses of both males and females in all populations to compare to the same analyses from the 412 SNP data set.

Detecting selection

To identify loci that may have undergone positive selection during the evolution of gray wolves and early dog domestication, we focused on comparing patterns of differentiation between the genetic subdivisions of gray wolves identified in STRUCTURE and PCA (n=125) and a subset of dogs from ancient and modern breeds (n=58), as well as comparing patterns of differentiation between Old World wolves (n=79) and New World wolves (n=46). To explore the differences in selection between males and females, the same population comparisons were conducted independently for each sex as well.

A cross-population composite likelihood ratio (XP-CLR) test (Chen et al. 2010) was performed for the 244 LD pruned SNP data set to scan for highly differentiated regions on the X chromosome between populations as possible targets of selective sweeps. To identify signals of selection, the XP-CLR method combines multi-locus allele frequency differentiation between

two populations with the spatial pattern of allele frequencies along a chromosome as a function of the genetic distance to the advantageous allele.

XP-CLR values were calculated using source code made available by Chen et al. (2010; http://genetics.med.harvard.edu/reich/Reich_Lab/Software.html). The following parameters were used: window size 0.5cm, grid size 1 kb, maximum number of SNPs within a window 50, correlation level from which the SNPs contribution to XP-CLR result was weighed down 0.95. To examine candidate regions for the populations analyzed, segments of the X chromosome where the XP-CLR scores above 99th percentile of the values were identified.

To further assess the SNPs contributing to the differentiation between dogs and wolves as well as between the genetic subdivisions of wolves identified in STRUCTURE and PCA, F_{ST} values were calculated at each of the 412 X chromosome SNPs using GALAXY (Goecks et al. 2010) with the Weir-Cockerham estimator (Weir and Cockerham, 1984). F_{ST} values were calculated for the female, male, and combined data sets, and outlier SNP loci were identified based on ranking in the 99th percentile (Chen et al., 2010).

RESULTS

Genetic diversity

The X chromosome SNP diversity (H_E= 0.09- 0.20) was lower than that of the autosomes (H_E=0.17- 0.26) across all comparable gray wolf populations studied (Table 1). This finding is consistent with predictions of reduced X chromosome diversity due to its presence in a single copy in males. Overall, the lowest observed heterozygosity was found in gray wolf populations from Italy and Spain (0.09 and 0.11 respectively). This is consistent with their similarly low

autosomal diversity (0.17 for both populations) reflecting historic bottlenecks in these populations (Fabbri et al., 2007; Gray et al., 2009; Ramirez et al., 2006). The highest levels of X chromosome SNP variability were found in gray wolf populations from Europe, exclusive of Italy and Spain ($H_E = 0.20$), which is consistent with autosomal data for these large and expanding populations . Gray wolves from China exhibited the highest levels of heterozygosity ($H_E = 0.24$) for X chromosome variability, however autosomal data are not available for comparison.

When the ratio of X-linked to autosomal SNP diversity was examined for populations with comparable estimates, deviations from the ratio predicted under equilibrium assumptions (0.75) were observed. Wolves from Italy and Spain had lower than expected ratios (0.53 and 0.65 respectively), consistent with the signature of reduced diversity resulting from historic bottlenecks. Gray wolves from Yellowstone National Park in North America had a higher X-to-autosome ratio (0.82), although this ratio was calculated from the of autosome diversity of a subset of the population that was used for the X chromosome diversity estimate, making direct comparisons difficult.

As expected, estimates of differentiation between gray wolf populations using X chromosome SNPs (F_{ST}= 0.15- 0.27) were consistently higher than those of the autosomes (F_{ST} 0.05-0.08) in all population comparisons where corresponding data was available (Table 2). The greatest differentiation of X chromosome SNPs was observed between North American and Middle Eastern populations ($F_{ST} = 0.31$) and the least differentiation was between Middle Eastern and European populations ($F_{ST} = 0.15$), which is consistent with continental patterns of population subdivision.

Population Structure

For the initial STRUCTURE analysis, the 263 sample dataset was run for K=2 through 10 and a maximum delta K value was observed at K=4, corresponding to clusters including North American gray wolves and Mexican wolves, Eurasian gray wolves (excluding Chinese populations), coyotes, and the domestic dog (Figure 1A). Red wolves and Great Lakes wolves show signals of admixture with coyotes, which is consistent with the autosomal data. Gray wolves from China exhibit mixed signals from each of the other groupings indicating a history of admixture. This signal is consistent with the high levels of heterozygosity observed in female gray wolves from China previously referenced.

When analyzed separately (Figures 1B and 1C), females and males of the 263 sample data set both exhibit a maximum delta K value at K=3. In females at K=3, coyotes and North American gray wolves cluster together distinct from European and Middle Eastern gray wolves. However, in males at K=3, North American gray wolves show a substantial proportion of shared ancestry with domestic dogs not indicated by female X chromosome SNPs or autosomal SNP analyses. At K=4, this difference between males and females disappears and reflects the same pattern of population structure as indicated when males and females are analyzed together at K=4.

To further resolve the population structure within gray wolves, STRUCTURE was run separately on the data set of 148 gray wolves for K=2 through 8 (Figure 2). The first partitions (K=2) correspond to North American wolves and European/Middle Eastern wolves, with wolves from China showing signals from both groups. At K=3, the Middle Eastern wolves show a signal of admixture with Europe, North American wolves remain distinct, and Chinese wolves show signals from all three groups. At K=4, a partition forms identifying Italian wolves, and at K=5,

the Chinese wolves signal becomes more distinct, partitioning gray wolves into Chinese wolves, Eastern European wolves, Italian wolves, Middle Eastern wolves, and North American wolves. At K=6, the maximum delta K value, no further resolution to the partitions at K=5 is produced. When male and female gray wolves are analyzed separately (Figure 3), the results produce the same partitions when analyzed together at K=2; however, at K=3 the distinct profile of the Italian wolves appears in males but not in females until K=5.

Principal component analysis of 183 individuals of gray wolves and dogs for 412 SNPs was found to clearly discriminate domestic dogs and populations of gray wolves on the first two PC axes (Figure 4). The first principal component (PC1; 57.8% of variation) represents a wild versus domestic canid axis, whereas the second principal component (PC2; 26.2% of variation) separates New World (North America) and Old World (Europe and Middle East) gray wolves with wolves from China as the intermediate between the two. PCA of the LD pruned data set (244 SNPs) produced the same pattern of clustering as the 412 SNP data set, with PC1 (67.5% of variation) discriminating dogs from wolves, and PC2 (16.2% of variation) separating populations of gray wolves (Figure 5).

Analysis of the PCA loading plot of the 412 SNP set for diagnostic SNPs contributing the most to the differentiation between dogs and wolves on PC1 identified the highest loading SNP located at CFAX.42400795 (Figure 6A). The same analysis for diagnostic loci contributing to differences between North American and Old World gray wolves on PC2 identified two adjacent SNPs located at CFAX.86813164 and CFAX.87234117 (Figure 6B). Analysis of the loading plots for the LD pruned data set identified the same diagnostic SNP contributing to the variation between dogs and wolves (CFAX.42400795; Figure 6C) as well as a SNP located in the same

region as the SNPs contributing to the variation between North American and Old World gray wolves (CFAX.88575097; Figure 6D).

When the sexes were analyzed separately for the 244 SNP set, the female comparison (Figure 7A) matched the PCA of the combined data in the number and relative position of clusters. The male comparison (Figure 7B) identified fewer clusters explaining the variation between the populations. This is expected, because for every X chromosome locus, females can be assigned to three potential genotype combinations compared to the two possible genotypes in hemizygous males. Across all loci, males will have fewer combinations of possible genotypes than females, resulting in fewer clusters describing the variation in males compared to females. While there were fewer clusters in the male PCA, PC1 continued to represent a wild versus domestic canid axis and PC2 primarily differentiated Old World from North American gray wolves. Male gray wolves from China were found in both Old World and New World clusters, while the female wolves from China clustered in between.

Analysis of the PCA loading plots for each sex identifies the same diagnostic SNPs as when the sexes are pooled (Figure 8A-8D), however the ranking of the highest loading SNPs contributing to the variation between male Old World and North American gray wolves varies. In females and the pooled analyses, the SNP located at or near CFAX.88575097 has the highest loading on PC2 whereas in males, the SNP located at CFAX.107746728 has the highest loading and CFAX.88575097 has the second highest.

Detecting Selection

The XP-CLR statistic was used to identify candidate regions for selective sweeps along the X chromosome in the male, female, and combined data sets for the dog-wolf comparison and Old World and New World wolf comparison. In the comparison of male and female gray wolves and dogs, the X chromosome region with the strongest XP-CLR scores (in the 99th percentile) occurs between CFAX.42326906- 42616906, with additional strong signals occurring at CFAX.13846906-14141906, CFAX.66601906-66671906, and CFAX.110471906-110716906 (Figure 9). The strongest signal at CFAX.42326906- 42616906 is resolved to separate regions when males and females are analyzed separately, the female signal matching that of the combined data set (Figure 10A) and the male signal occurring 2Mb away at CFAX.40071906-40311906 (Figure 10B). In the male only comparison, the strongest signal occurs at CFAX.13846906-14141906. While, this region is identified among the top scores in the female only comparisons, the signal is not as strong as in males. Additional high XC-PCLR score regions that were identified only in the female comparison occur between CFAX.8436906-8486906 and CFAX.66641906-66671906. Both males and females separately have a 99th percentile score between CFAX.110471906-110716906.

In the comparison between Old World and New World gray wolves, the X chromosome region with the strongest XP-CLR scores is located at CFAX.88406906-88701906, with additional signals at CFAX.33161906-33551906, CFAX.38001906-38241906, and CFAX.104591906-104676906 (Figure 11). When analyzed separately, the CFAX.88406906-88701906 region scores are strong in both males and females (Figures 12A and 12B). The female only analysis indicates the regions at CFAX.38001906-38241906, CFAX.104591906-104676906, and CFAX.106866906-106986906 are in the top 99th percentile of scores, but these

regions are not top scoring in the male analyses. Additionally there are signals in males at CFAX.74026906-74136906 and CFAX.95816906-951991906 that are not ranked in the 99th percentile of scores for females.

Analysis of SNPs with F_{ST} values in the 99th percentile found several chromosomal regions that overlap with the diagnostic loci identified by high PCA loadings and XP-CLR scores. Five regions were identified consistently in at least one of the same sex comparisons across all methods (CFAX.13846906-14141906, CFAX.42326906-42621906, CFAX.88406906-88701906, CFAX.104591906-104676906, CFAX.110637383-110689568), and three were identified in the combined sex comparisons across all methods (CFAX.13846906-14141906, CFAX.42326906-42621906, CFAX.88406906-88701906).

Two X chromosome regions were identified as outliers in all sex comparisons in all statistical methods employed. SNPs the region of CFAX.13846906-14141906 were consistently identified in comparisons between male and female dogs and wolves, and CFAX.86813164-88876389 were consistently identified in comparisons between male and female North American gray wolves and Old World wolves (excluding China). Additionally the SNPs located in the region of CFAX.42326906-42621906 were consistently identified as outliers in all comparisons between dogs and wolves, except in the male only XP-CLR analysis.

Table 3 lists regions that were found by a combination of PCA, XP-CLR and F_{ST} statistics and their associated genes or traits. All regions listed were investigated using the UCSC genome browser. Because the annotation of the X chromosome sequence for dogs is minimal, predictions for genes in the outlier regions were found based on sequence alignment in known protein-

coding and non-protein-coding genes in other mammals (Kent et al., 2002). Three of the regions detected have also been identified in two independent studies investigating the genetic basis of morphology in dogs (Boyko et al., 2010; Chase et al., 2005).

DISCUSSION

Compared to the autosomes, measurements of the genetic diversity of X chromosome SNPs in wolf-like canids were consistently lower across all populations and measurements of population differentiation were consistently higher, as predicted, due to the lower effective population size of the X chromosome. The genetic variability and differentiation at the population level are consistent with autosomal patterns of species-level and geographic subdivisions (vonHoldt et al., 2011). Specifically, gray wolf populations from China, Eastern Europe and Spain, Italy, Middle East, and North America were identified as genetically distinct units. Additionally, X chromosome SNP analysis confirms substantial coyote ancestry for Red wolves as well their distinct but admixed evolutionary history with Great Lakes wolves as inferred by autosomal SNPs. Furthermore, Great Lakes wolves appear to be genetically distinct but derived from Western gray wolves. While the suggested patterns of population structure might be overinflated because it was inferred using the original 412 X chromosome SNP data set that was not LD pruned, the pattern is consistent with the genetic structure results of genome-wide autosomal SNP data (48K).

Deviations from the X-to-autosome ratio predicted under neutrality were also observed suggesting there are additional evolutionary forces affecting these populations. X-linked diversity will generally be less than 75% of autosomal diversity in populations that have

undergone recent bottlenecks (Fay and Wu, 1999). This is consistent with findings reported here in populations of gray wolves from Italy and Spain, as well as populations with a smaller effective number of mating females as demonstrated in human populations dispersing out of Africa (Keinan et al., 2009). X-to-autosome ratios of effective population sizes that are greater than 75% could be an indication of systematic difference between the sexes in the variance in reproductive success, such as the historical excess of breeding females over the number of breeding males as a result of polygyny in human populations (Hammer et al., 2008). In addition to these parameters being likely to vary between populations, differences in population sizes, composition, and diversity estimates between studies make it difficult to directly compare genetic distance measurements and reach definitive conclusions about the relative importance of various forces on the population genetics of the X chromosome.

A number of statistical methods have been developed to infer selection from SNP data in population surveys. The XP-CLR statistic appears to currently be best suited to deal with the complexity of adaptive processes, uncertainty about demographic history of a population, older signals of selection (XP-EHH signals are expected to break down after several hundred generations), as well as selection on standing variation (Chen et al., 2010).

Several candidate regions were identified through a combination of statistical methods used to infer patterns of selection. All genes associated with the outlier regions (Table 3) are predictions based on estimates of gene conservation with other organisms. Some of the gene regions identified by more than one statistic are of biological interest. For example, two of the regions identified in this study were also found to be associated with morphological variation in a genome-wide association study (Boyko et al. 2010). The authors of the study identified a

significant genomic association for limb and tail length in an LD block with high F_{ST} (0.658) located at CFAX.86813164- 87299370. We have also identified this this region as an outlier in all three statistics for analyses with varying degrees of overlap. Both F_{ST} and PCA statistics detect outlier SNPs within the region at CFAX.86813164. Although XP-CLR was run on the LD pruned data set in which this SNP region was removed, the selection scan detected a strong signal at a region immediately adjacent to CFAX.86813164, located at CFAX.88406906-88701906. The other common region located at CFAX.104591906-104676906 is associated with body size and skull shape. Both SNP regions associated with body size were identified in comparisons of Old World and North American gray wolves suggesting that these regions may contain genes contributing to the evolution of these trait variations between these two groups. Further analysis of habitat groupings of New World gray wolves may allow greater resolution of the candidate SNPs differentiating the New World from Old World wolves.

While these outliers provide evidence for signatures of natural selection, the gene regions need to be investigated further to be able to rule out the effects of additional forces in shaping the evolution of wolf-like canids. Fine-mapping of the X chromosome will also be required in order to confirm single candidate genes for all regions of interest and explore their role in canine evolution.

SUPPORTING MATERIAL

Figure Legends

Figure 1. STRUCTURE clustering analysis of domestic and wild canids using the 412 X SNP data set with (A) males and females, (B) females only, and (C) males only.

Figure 2. STRUCTURE clustering analysis of gray wolves using the 412 SNP data set with males and females combined at K=2-6.

Figure 3. STRUCTURE clustering analysis of gray wolves using the 412 SNP data set of (A) females and (B) males at K=3-5.

Figure 4. (A) PCA of 412 SNP data set for 183 individuals: dogs (n=48) and gray wolves (n=125), with males and females combined.

Figure 5. (A) PCA of the LD-pruned 244 SNP data set for 183 individuals: dogs (n=48) and gray wolves (n=125), with males and females combined.

Figure 6. The PCA loading plots identifying the SNPs in the 99^{th} percentile of loading scores for the each principal component. PC1 of the 412 SNP data set (A) represents the SNPs contributing to the dog versus wolf axis and PC2 (B) represents the SNPs contributing to the OW gray wolf versus NW gray wolf axis. (C) and (D) represent PC1 and PC2 for the LD-pruned 244 SNP data set respectively.

Figure 7. PCA of 244 SNPs of (A) females and (B) males.

Figure 8. The loading plots represent female (A) PC1- dog/wolf axis and (B) PC2-OW/NW wolf axis and male (C) PC1- dog/wolf axis (D) and PC2-OW/NW wolf axis. The horizontal line represents the SNPs in the 99%tile of loading scores.

Figure 9. Plot of XP-CLR scores of the dog versus wolf comparison with males and females combined.

Figure 10. Plot of XP-CLR scores of the dog versus wolf comparison for females only (A) and males only (B).

Figure 11. Plot of XP-CLR scores of the Old World Wolf versus New World Wolf comparison with males and females combined.

Figure 12. Plot of XP-CLR scores of the Old World Wolf versus New World Wolf comparison for females only (A) and males only (B).

Figure 1.

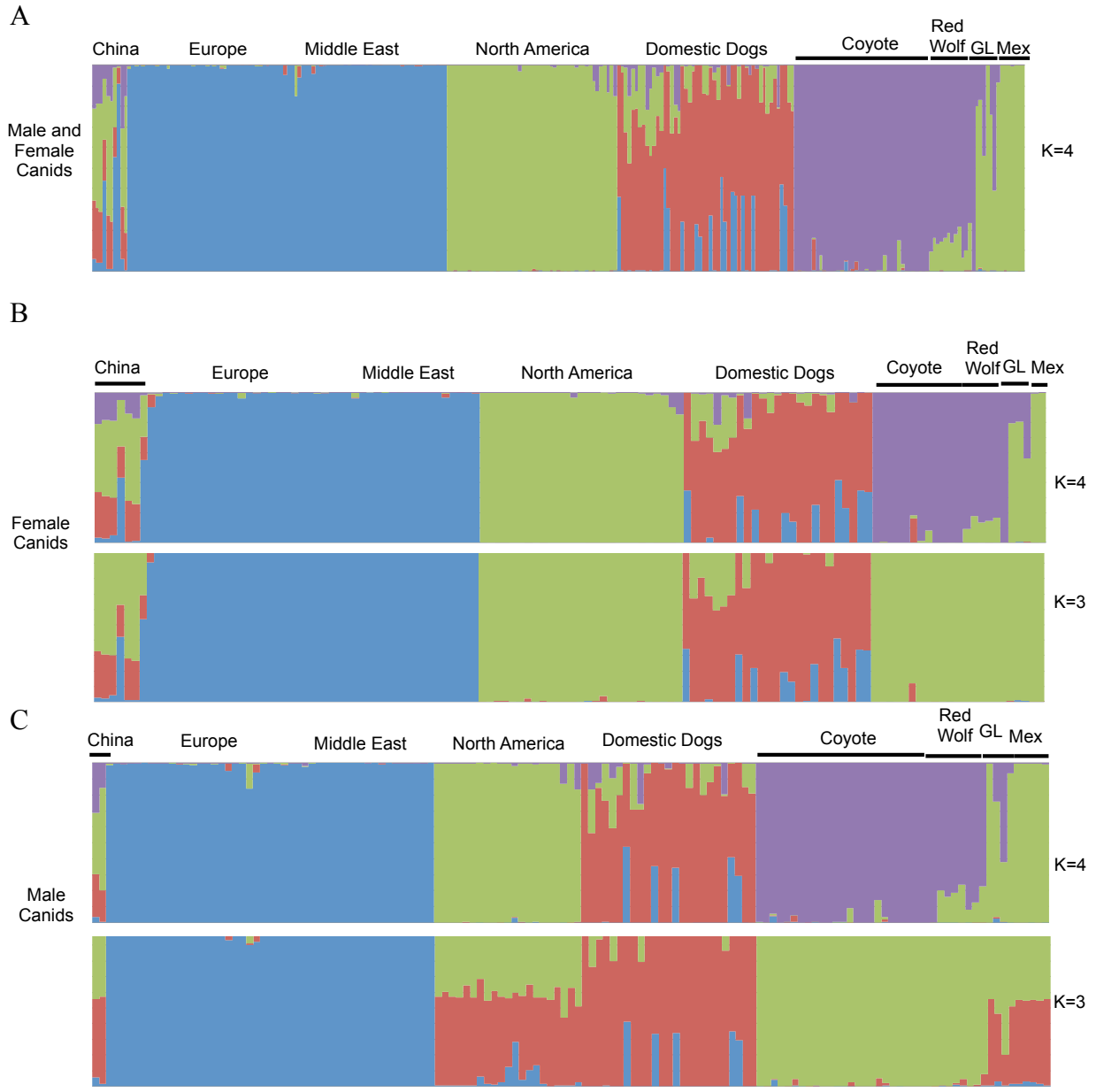
A
China
Europe
Middle East
North America
Domestic Dogs
Coyote
Red Wolf
GL
Mex
Male and Female Canids
K=4
B
China
Europe
Middle East
North America
Domestic Dogs
Coyote
Red Wolf
GL
Mex
Female Canids
K=4
K=3
C
China
Europe
Middle East
North America
Domestic Dogs
Coyote
Red Wolf
GL
Mex
Male Canids
K=4
K=3

Figure 2.

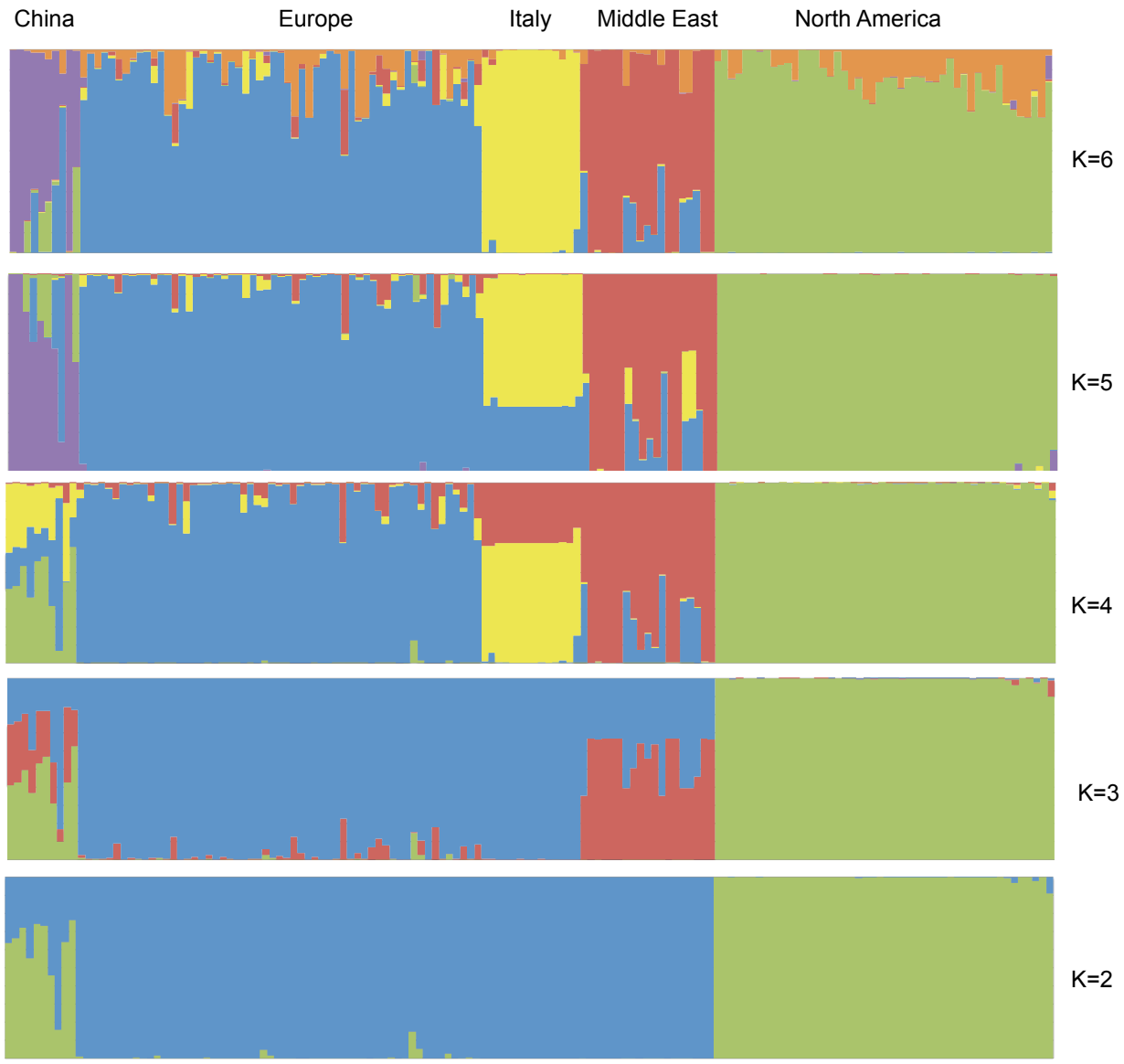

Figure 3.

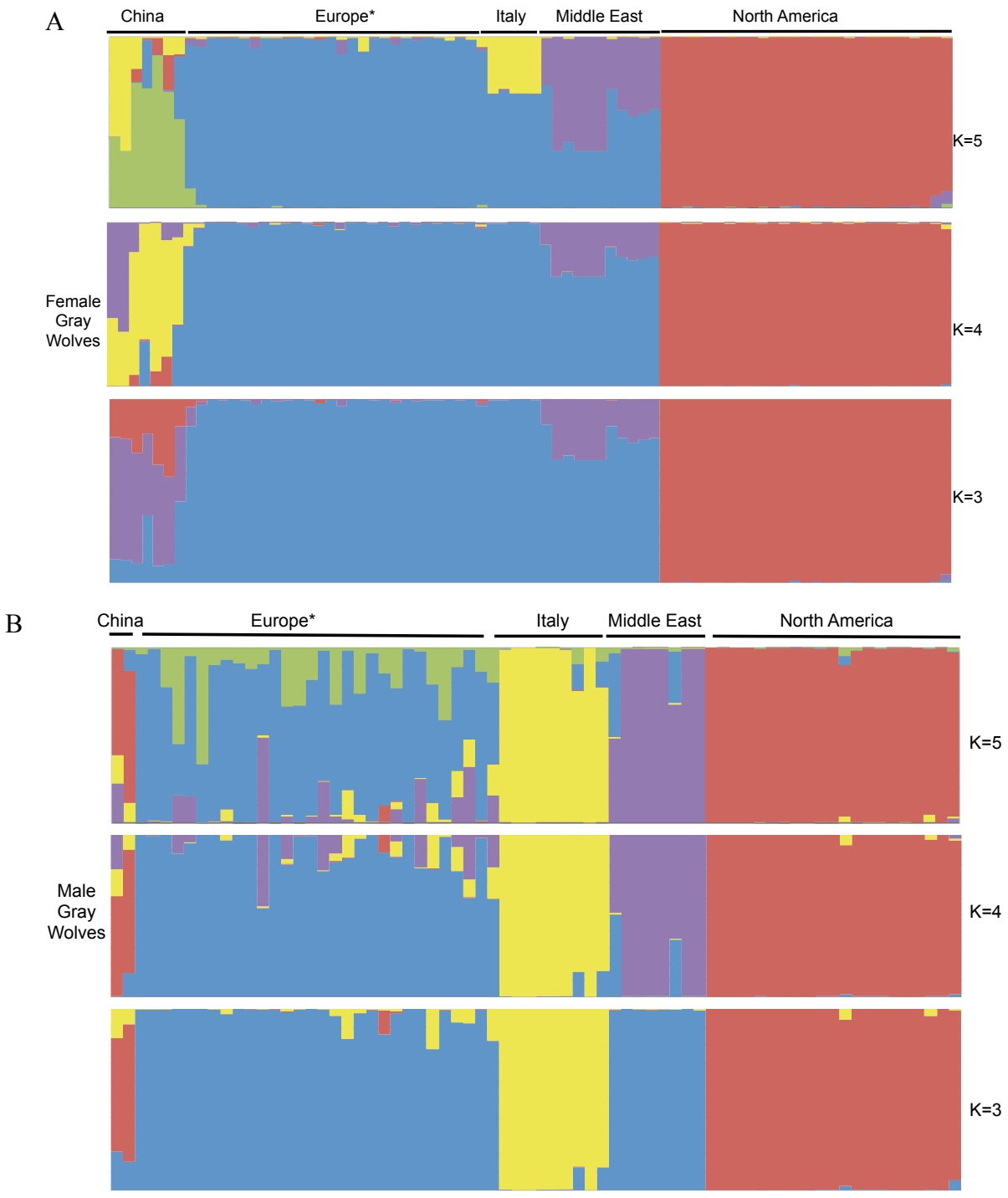
A
China
Europe*
Italy
Middle East
North America
K=5
Female Gray Wolves
K=4
K=3
B
China
Europe*
Italy
Middle East
North America
K=5
Male Gray Wolves
K=4
K=3

Figure 4.

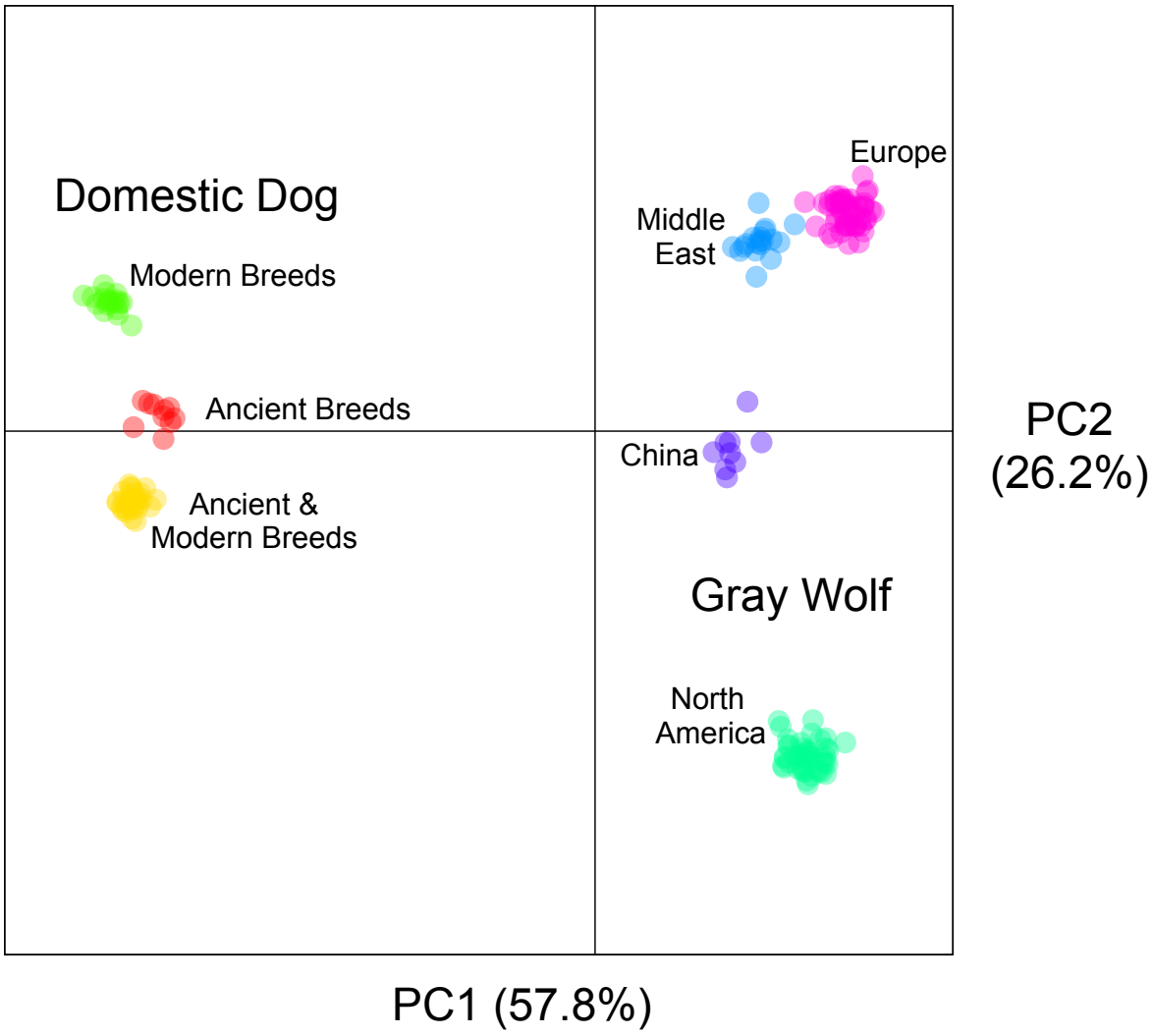
Domestic Dog
Modern Breeds
Ancient Breeds
Ancient &
Modern Breeds
Europe
Middle
East
China
Gray Wolf
North
America
PC2
(26.2%)
PC1 (57.8%)

Figure 5.

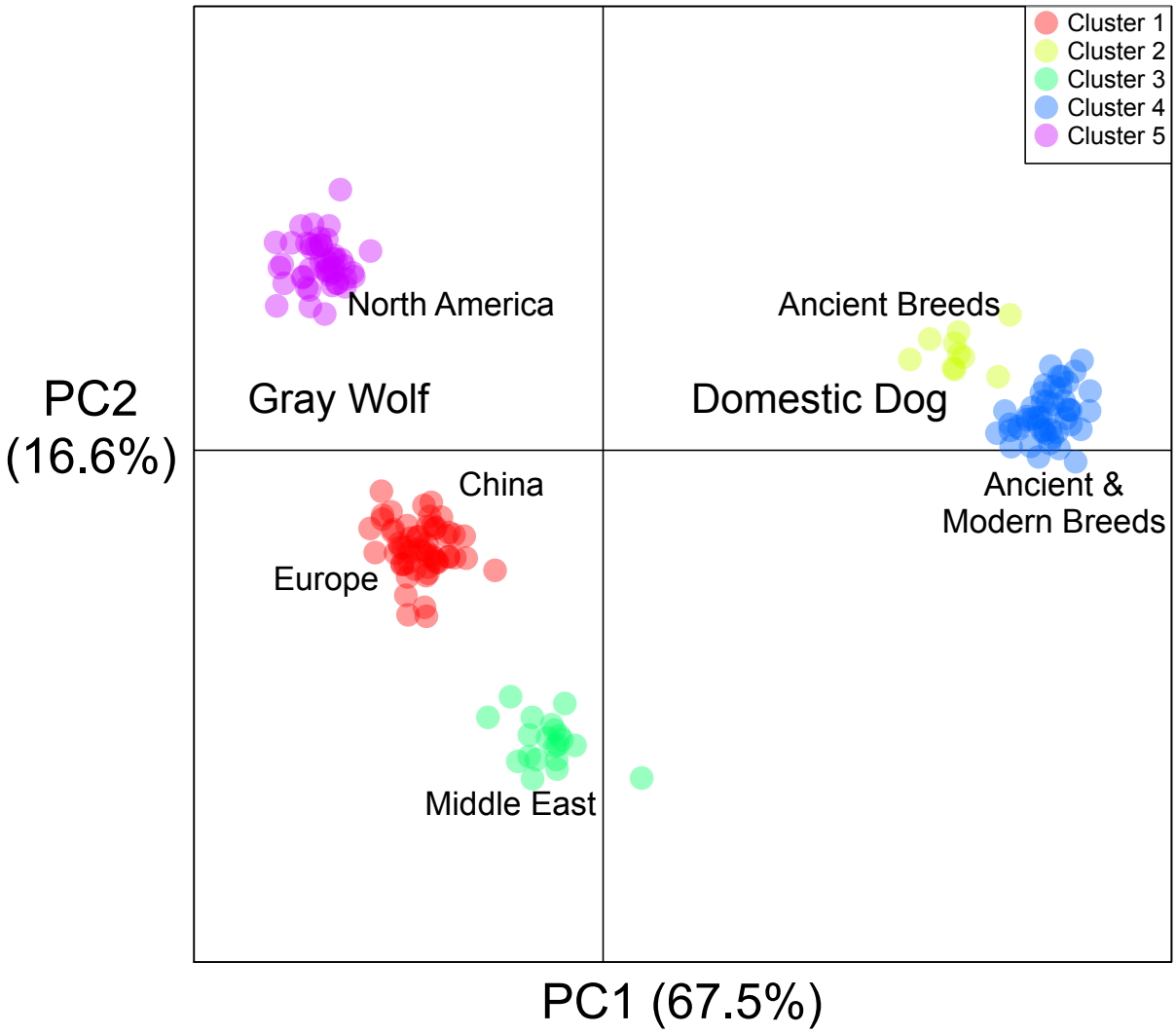
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
North America
Gray Wolf
Ancient Breeds
Domestic Dog
China
Europe
Ancient &
Modern Breeds
Middle East
PC2
(16.6%)
PC1 (67.5%)

Figure 6.

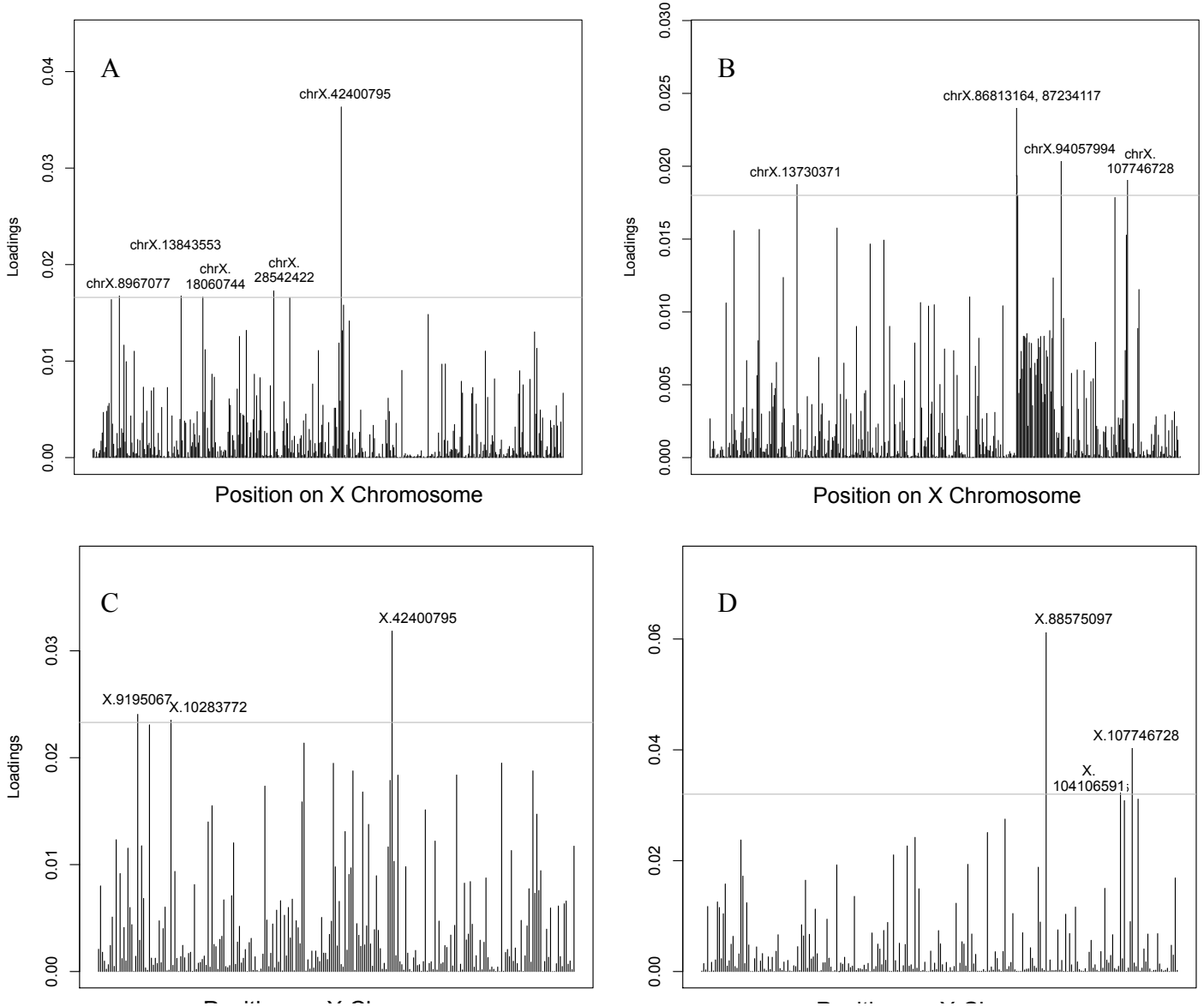
A
chrX.42400795
chrX.13843553
chrX.
18060744
chrX.
28542422
chrX.8967077
Loadings
0.00
0.01
0.02
0.03
0.04
Position on X Chromosome
B
chrX.86813164, 87234117
chrX.94057994
chrX.
107746728
chrX.13730371
Loadings
0.000
0.005
0.010
0.015
0.020
0.025
0.030
Position on X Chromosome
C
X.42400795
X.9195067
X.10283772
Loadings
0.00
0.01
0.02
0.03
D
X.88575097
X.107746728
X.
104106591
0.00
0.02
0.04
0.06

Figure 7.

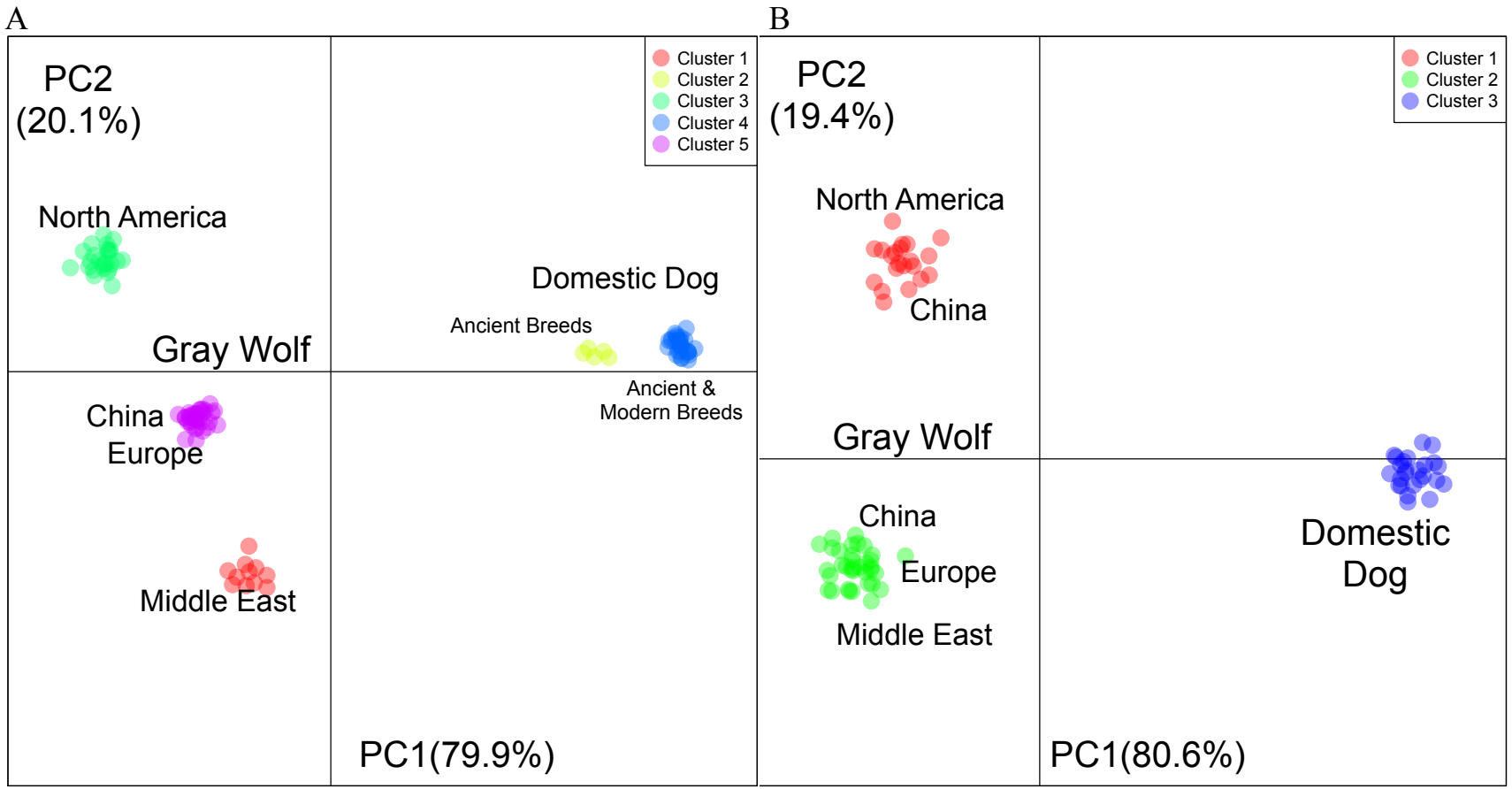
A
PC2
(20.1%)
North America
Gray Wolf
China
Europe
Middle East
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Domestic Dog
Ancient Breeds
Ancient &
Modern Breeds
PC1(79.9%)
B
PC2
(19.4%)
North America
China
Gray Wolf
China
Europe
Middle East
Cluster 1
Cluster 2
Cluster 3
Domestic
Dog
PC1(80.6%)

Figure 8.

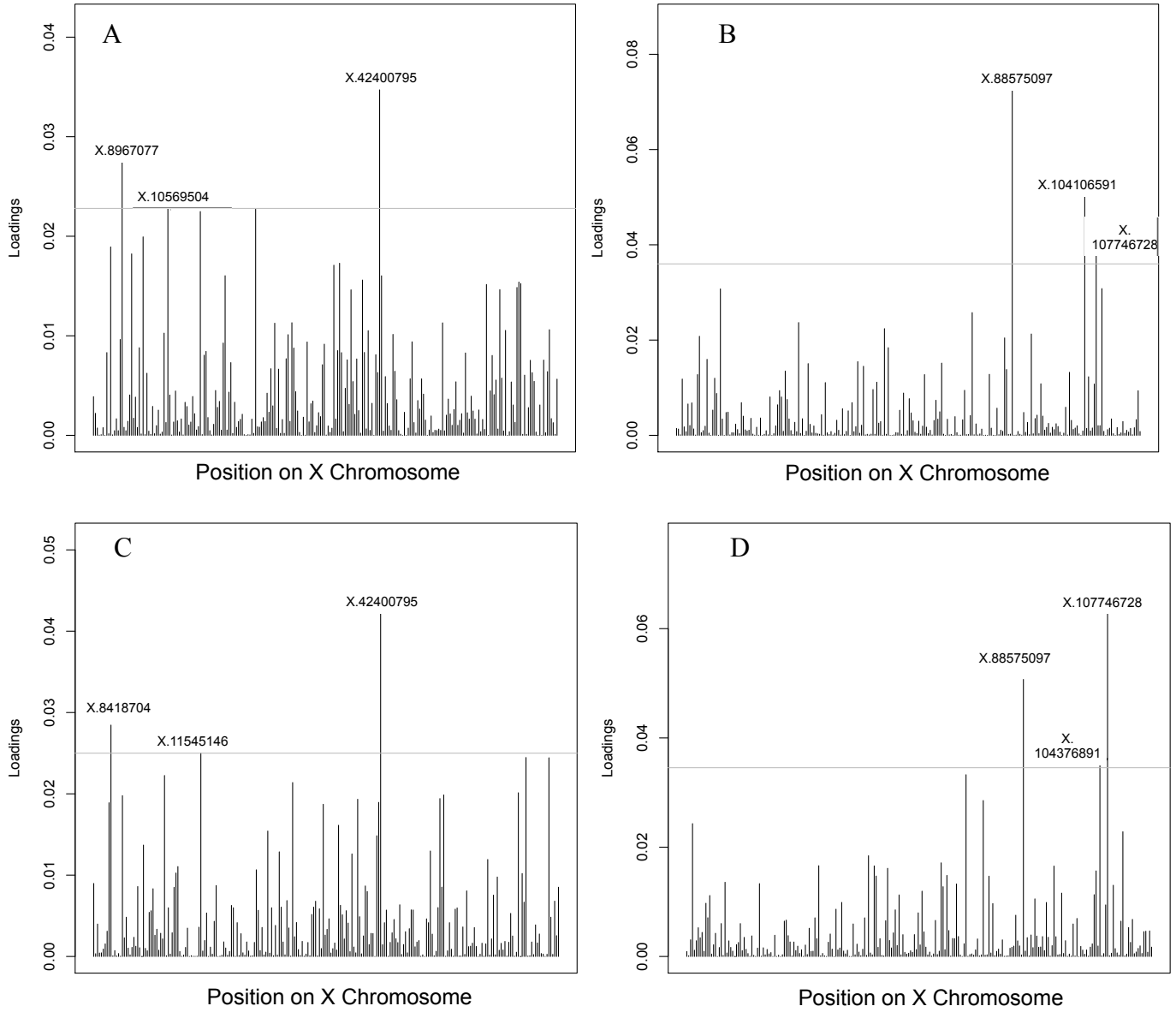
A
X.42400795
X.8967077
X.10569504
Loadings
0.00
0.01
0.02
0.03
0.04
Position on X Chromosome
B
X.88575097
X.104106591
X.
107746728
Loadings
0.00
0.02
0.04
0.06
0.08
Position on X Chromosome
C
X.42400795
X.8418704
X.11545146
Loadings
0.00
0.01
0.02
0.03
0.04
0.05
Position on X Chromosome
D
X.107746728
X.88575097
X.
104376891
Loadings
0.00
0.02
0.04
0.06
Position on X Chromosome

Figure 9.

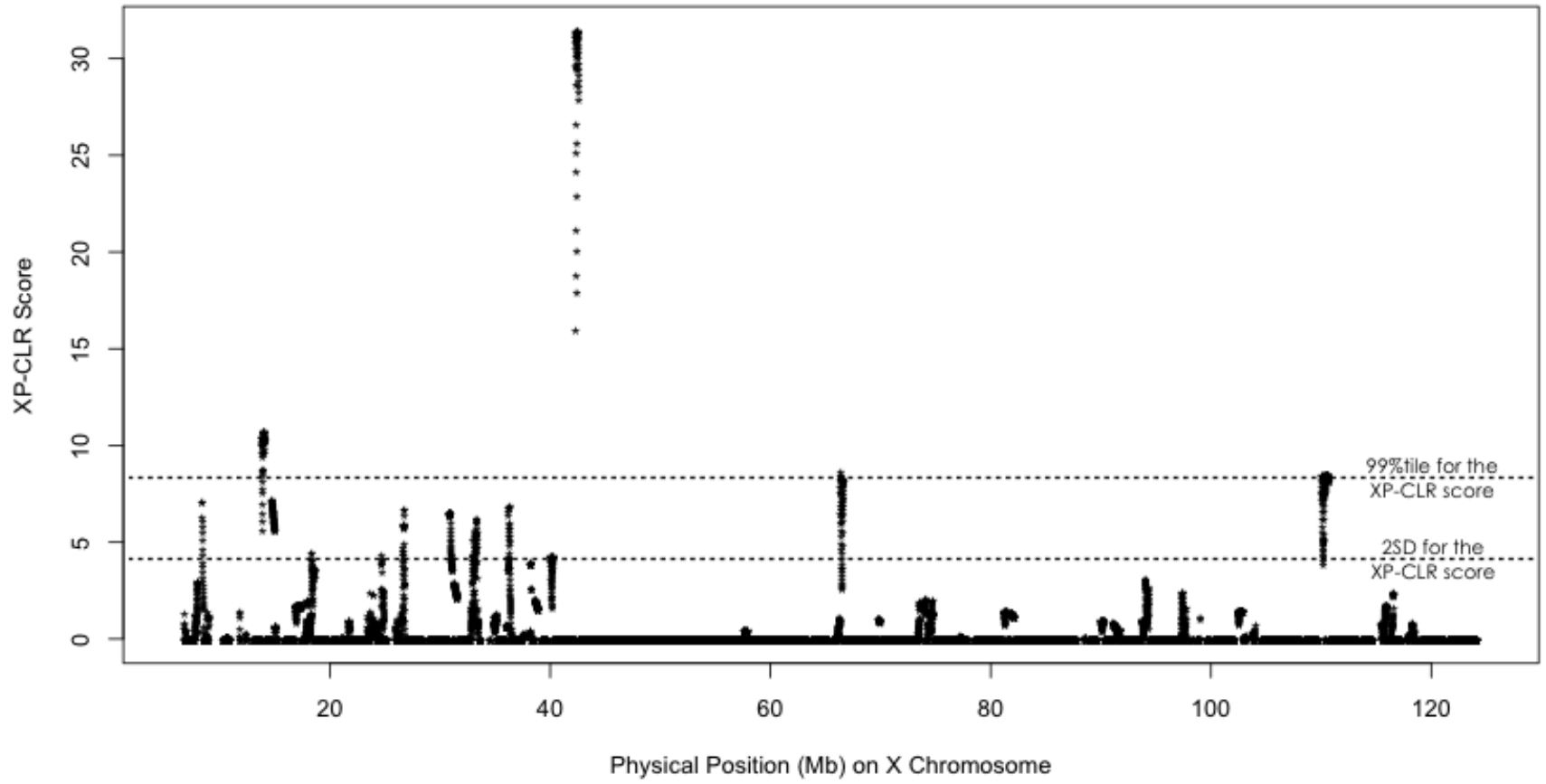
XP-CLR Score
0
5
10
15
20
25
30
99%tile for the XP-CLR score
2SD for the XP-CLR score
20
40
60
80
100
120
Physical Position (Mb) on X Chromosome

Figure 10.

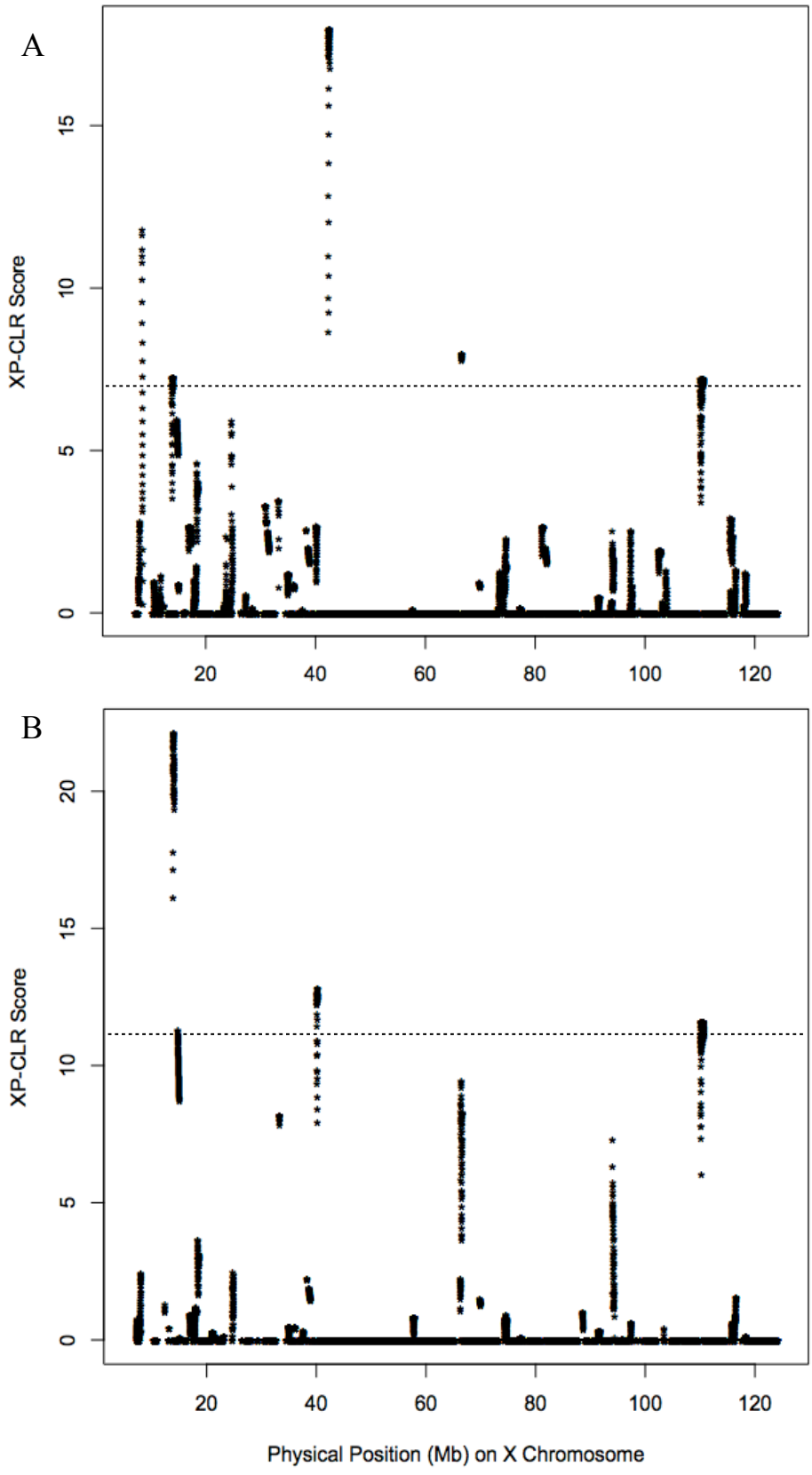
A
XP-CLR Score
0
5
10
15
20
40
60
80
100
120
B
XP-CLR Score
0
5
10
15
20
20
40
60
80
100
120
Physical Position (Mb) on X Chromosome

Figure 11.

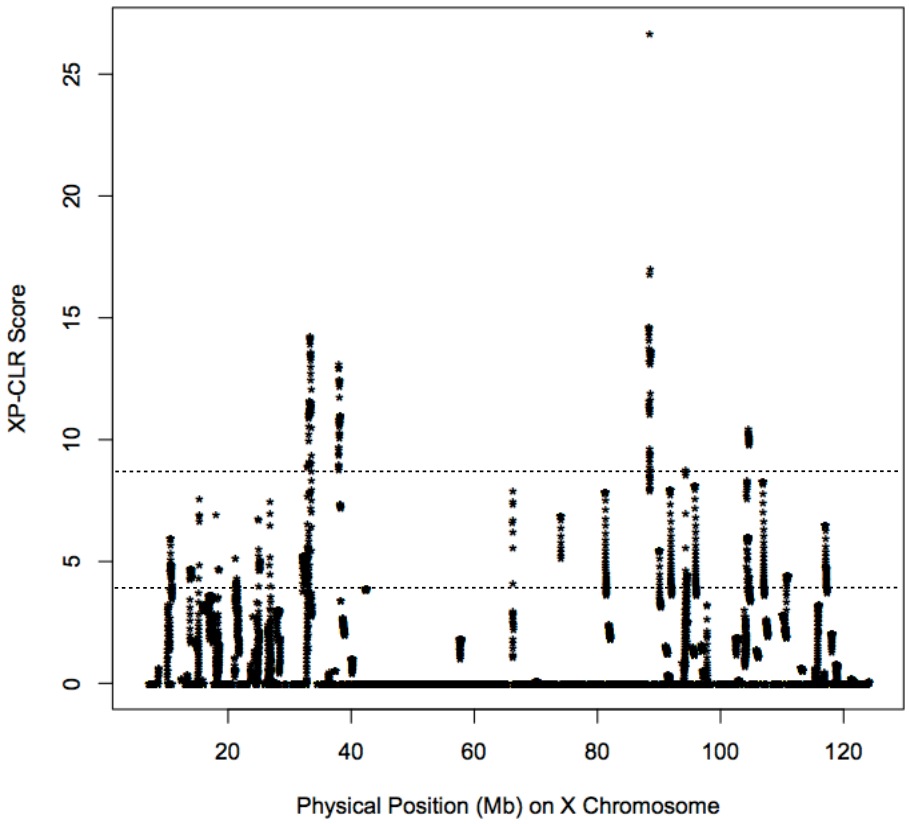
XP-CLR Score
0
5
10
15
20
25
20
40
60
80
100
120
Physical Position (Mb) on X Chromosome

Figure 12.

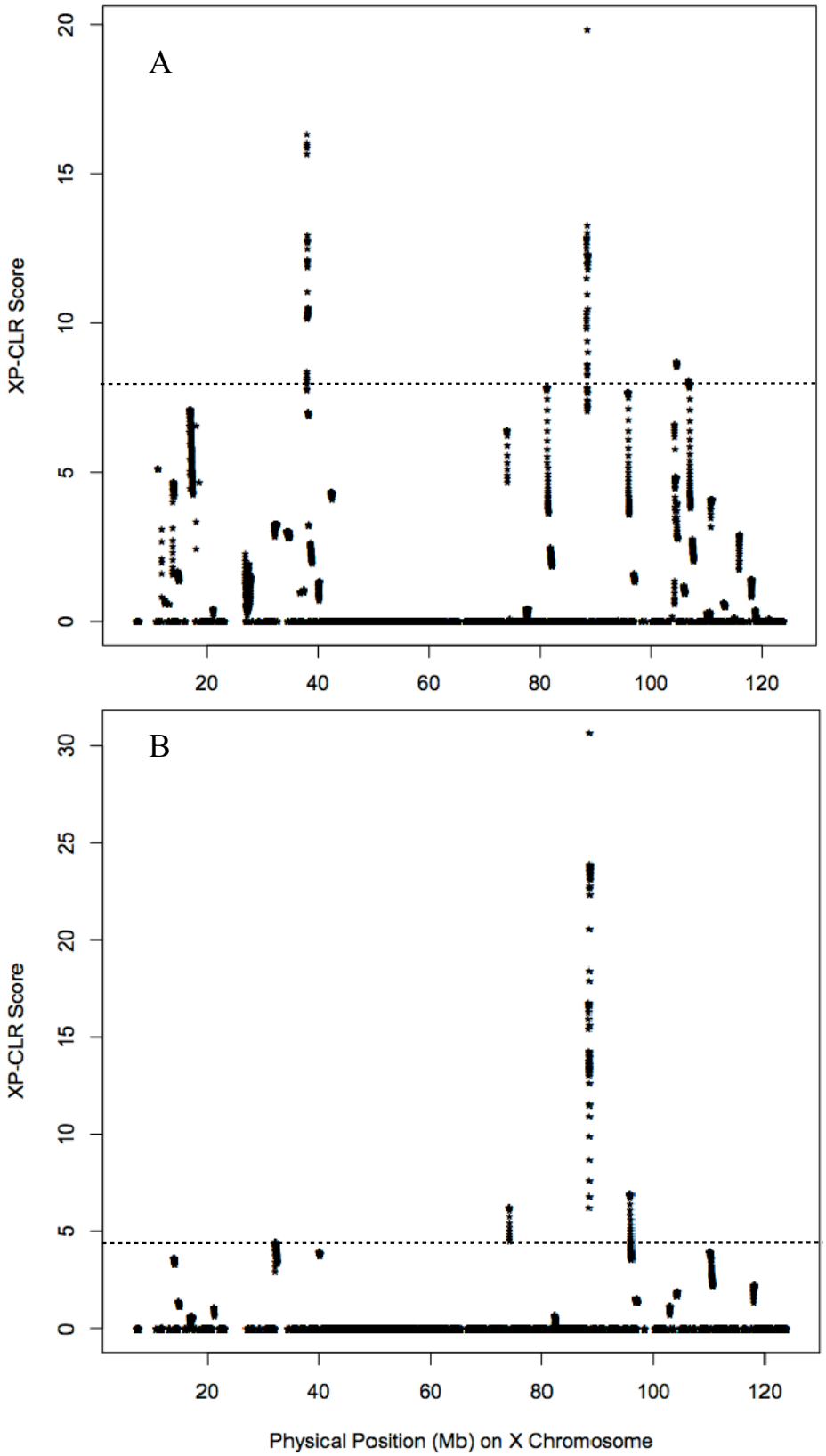
A
B
XP-CLR Score
XP-CLR Score
0
5
10
15
20
25
30
20
40
60
80
100
120
Physical Position (Mb) on X Chromosome

Table 1. Comparison of average observed (H_O) and expected (H_E)) heterozygosity for 412 X chromosome SNPs and 48K autosomal SNPs (vonHoldt et al., 2011) in the populations examined in this study when available (n, sample size).

Population	X Chromosome H_O (H_E)	n	Autosomal H_O (H_E)	n	X/A Ratio (H_E)
Europe*	0.18 (0.20)	24	0.24 (0.26)	57	0.77
Italy	0.10 (0.09)	5	0.15 (0.17)	20	0.53
Spain	0.09 (0.11)	5	0.18 (0.17)	10	0.65
Middle East	0.13 (0.15)	11	—	—	—
China	0.21 (0.24)	8	—	—	—
North America	0.14 (0.18)	27	YNP=0.22 (0.22) Canada=0.22 (0.24)	18 13	0.82 0.75

* Excludes Italian and Spanish wolves

Table 2. F_{ST} for the 412 X chromosome SNP data set in female Old World and North American gray wolves above the diagonal, and 48K autosomal SNP data set for Old World gray wolves below the diagonal (vonHoldt et al. 2011).

Group	China (F)	Europe (F)	Middle East (F)	North America (F)
China (F)		0.20	0.27	0.22
Europe (F)	0.05		0.15	0.22
Middle East (F)	0.06	0.08		0.31
North America (F)				

Table 3. Summary of outlier SNP loci identified by multiple statistics with names of genes overlapping the region and associated traits (if known).

Dog versus Wolf comparison

XP-CLR				FST				PCA				Genes of Interest	Trait Associations
X Chromosome Region	M	F	M/F	Marker	M	F	M/F	Marker	M	F	M/F		
8436906- 8486906	-	+	-	8418704	-	+	-	8418704	+	-	-	Frmpd4	
13846906-14141906	+	+	+	13843553	+	+	+	13843553	+	+	+	RAI1	
40071906-40311906	+	-	-	40077301, 40119780	+	+	+	-	-	-	-	KRBOX4 ZNF674, CHST7, COMMD6, SLC9A7, SLC9A6	
42326906-42621906	-	+	+	42400795	+	+	+	42400795	+	+	+	PPP1R3F, POLDIP2, USP27X, USP51, USP22	
66601906-66671906	-	+	+	66375156	+	+	+	-	-	-	-	CHM locus	Chase et al. (2005); interaction with IGF1 regulates size sexual dimorphism
110086906-110146906	+	+	+	110084946, 110104000	+	+	+	-	-	-	-	CD40LG, ARHGEF6	

North American versus Old World Wolf comparison

XP-CLR				FST				PCA				Genes of Interest	Trait Associations
X Chromosome Region	M	F	M/F	Marker	M	F	M/F	Marker	M	F	M/F		
88406906-88701906	+	+	+	86813164-88876389	+	+	+	86813164, 88575097	+	+	+	NHP2L1, CHRDL1, CDC42, DCX, PAK3, TADA3, ALG13, TPM4, TRPC5, ZCCHC16, LHFPL1, AMOT	Boyko et al. (2010); associated with limb and tail length
104591906-104676906	-	+	+	104106591	+	-	-	104106591	+	+	+	FSIP2, RBMX2	Boyko et al. (2010); associated with body size and skull shape
110637383-110689568	-	+	-	110637383, 110689568	+	+	-	107746728	+	+	+		

REFERENCES

Albert, F.W., Somel, M., Carneiro, M., Aximu-Petri, A., Halbwax, M., Thalmann, O., Blanco-Aguiar, J.A., Plyusnina, I.Z., Trut, L., Villafuerte, R., et al. (2012). A Comparison of Brain Gene Expression Levels in Domesticated and Wild Animals. PLoS Genet. *8*.

Balaresque, P., Toupance, B., Quintana-Murcilluis, Crouau-Roy, B., and Heyer, E. (2004). Sex-specific selection on the human X chromosome? Genet. Res. *83*, 169–176.

Boyko, A.R., Quignon, P., Li, L., Schoenebeck, J.J., Degenhardt, J.D., Lohmueller, K.E., Zhao, K., Brisbin, A., Parker, H.G., vonHoldt, B.M., et al. (2010). A Simple Genetic Architecture Underlies Morphological Variation in Dogs. PLoS Biol. *8*.

Charlesworth, B., Coyne, J.A., and Barton, N.H. (1987). The Relative Rates of Evolution of Sex Chromosomes and Autosomes. Am. Nat. *130*, 113–146.

Chase, K., Carrier, D.R., Adler, F.R., Ostrander, E.A., and Lark, K.G. (2005). Interaction between the X chromosome and an autosome regulates size sexual dimorphism in Portuguese Water Dogs. Genome Res. *15*, 1820–1824.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Res. *20*, 393–402.

Earl, D.A., and vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. *4*, 359–361.

Fabbri, E., Miquel, C., Lucchini, V., Santini, A., Caniglia, R., Duchamp, C., Weber, J.-M., Lequette, B., Marucco, F., Boitani, L., et al. (2007). From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (Canis lupus) population. Mol. Ecol. *16*, 1661–1671.

Fay, J.C., and Wu, C.I. (1999). A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol. Biol. Evol. *16*, 1003–1005.

Flagstad, Ø., Walker, C.W., Vilà, C., Sundqvist, A.-K., Fernholm, B., Hufthammer, A.K., Wiig, Ø., Koyola, I., and Ellegren, H. (2003). Two centuries of the Scandinavian wolf population: patterns of genetic variability and migration during an era of dramatic decline. Mol. Ecol. *12*, 869–880.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.

Gottipati, S., Arbiza, L., Siepel, A., Clark, A.G., and Keinan, A. (2011). Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. Nat. Genet. *43*, 741–743.

Gray, M.M., Granka, J.M., Bustamante, C.D., Sutter, N.B., Boyko, A.R., Zhu, L., Ostrander, E.A., and Wayne, R.K. (2009). Linkage Disequilibrium and Demographic History of Wild and Domestic Canids. Genetics *181*, 1493–1505.

Hammer, M.F., Mendez, F.L., Cox, M.P., Woerner, A.E., and Wall, J.D. (2008). Sex-Biased Evolutionary Forces Shape Genomic Patterns of Human Diversity. PLoS Genet. *4*.

Hedrick, P.W. (2007). Sex: Differences in Mutation, Recombination, Selection, Gene Flow, and Genetic Drift. Evolution *61*, 2750–2771.

Jędrzejewski, W., Branicki, W., Veit, C., MeĐugorac, I., Pilot, M., Bunevich, A.N., Jędrzejewska, B., Schmidt, K., Theuerkauf, J., Okarma, H., et al. (2005). Genetic diversity and relatedness within packs in an intensely hunted population of wolvesCanis lupus. Acta Theriol. (Warsz.) *50*, 3–22.

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics.

Keinan, A., Mullikin, J.C., Patterson, N., and Reich, D. (2009). Accelerated genetic drift on chromosome X during the human dispersal out of Africa. Nat. Genet. *41*, 66–70.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, and D. (2002). The Human Genome Browser at UCSC. Genome Res. *12*, 996–1006.

Lambert, C.A., Connelly, C.F., Madeoy, J., Qiu, R., Olson, M.V., and Akey, J.M. (2010). Highly Punctuated Patterns of Population Structure on the X Chromosome and Implications for African Evolutionary History. Am. J. Hum. Genet. *86*, 34–44.

Li, W.-H., Yi, S., and Makova, K. (2002). Male-driven evolution. Curr. Opin. Genet. Dev. *12*, 650–656.

Pilot, M., Jedrzejewski, W., Branicki, W., Sidorovich, V.E., Jedrzejewska, B., Stachura, K., and Funk, S.M. (2006). Ecological factors influence population genetic structure of European grey wolves. Mol. Ecol. *15*, 4533–4553.

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

Ramirez, O., Altet, L., Enseñat, C., Vilà, C., Sanchez, A., and Ruiz, A. (2006). Genetic assessment of the Iberian wolf Canis lupus signatus captive breeding program. Conserv. Genet. *7*, 861–878.

Saetre, P., Lindberg, J., Leonard, J.A., Olsson, K., Pettersson, U., Ellegren, H., Bergström, T.F., Vilà, C., and Jazin, E. (2004). From wild wolf to domestic dog: gene expression changes in the brain. Brain Res. Mol. Brain Res. *126*, 198–206.

Saifi, G.M., and Chandra, H.S. (1999). An apparent excess of sex- and reproduction-related genes on the human X chromosome. Proc. R. Soc. B Biol. Sci. *266*, 203–209.

Schaffner, S.F. (2004). The X chromosome in population genetics. Nat. Rev. Genet. *5*, 43–51.

Tao, Y., Chen, S., Hartl, D.L., and Laurie, C.C. (2003). Genetic dissection of hybrid incompatibilities between Drosophila simulans and D. mauritiana. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. Genetics *164*, 1383–1397.

Thiselton, D.L., McDowall, J., Brandau, O., Ramser, J., d' Esposito, F., Bhattacharya, S.S., Ross, M.T., Hardcastle, A.J., and Meindl, A. (2002). An Integrated, Functionally Annotated Gene Map of the DXS8026–ELK1 Interval on Human Xp11.3–Xp11.23: Potential Hotspot for Neurogenetic Disorders. Genomics *79*, 560–572.

Vicoso, B., and Charlesworth, B. (2006). Evolution on the X chromosome: unusual patterns and processes. Nat. Rev. Genet. *7*, 645–653.

vonHoldt, B.M., Pollinger, J.P., Lohmueller, K.E., Han, E., Parker, H.G., Quignon, P., Degenhardt, J.D., Boyko, A.R., Earl, D.A., Auton, A., et al. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature *464*, 898–902.

vonHoldt, B.M., Pollinger, J.P., Earl, D.A., Knowles, J.C., Boyko, A.R., Parker, H., Geffen, E., Pilot, M., Jedrzejewski, W., Jedrzejewska, B., et al. (2011). A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Res. *21*, 1294–1305.

Wang, P.J., McCarrey, J.R., Yang, F., and Page, D.C. (2001). An abundance of X-linked genes expressed in spermatogonia. Nat. Genet. *27*, 422–426.

Weir, B.S., and Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution *38*, 1358–1370.

Zechner, U., Wilda, M., Kehrer-Sawatzki, H., Vogel, W., Fundele, R., and Hameister, H. (2001). A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? Trends Genet. *17*, 697–701.