

1 **Linkage Disequilibrium and Demographic History of Wild and Domestic Canids**

2

3

4 Melissa M. Gray^{*}, Julie M. Granka[†], Carlos D. Bustamante[†], Nathan B. Sutter[‡], Adam R.

5 Boyko[†], Lan Zhu[§], Elaine A. Ostrander^{**}, and Robert K. Wayne^{*}.

6

7 ^{*}Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA

8 90095.

9 [†]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY

10 14853.

11 [‡]Department of Clinical Sciences, College of Veterinary Medicine, VMC C3-179, Cornell

12 University, Ithaca, NY 14853.

13 [§]Department of Statistics, Oklahoma State University, 301C MSCS Bldg, Stillwater, OK 74078.

14 ^{**}Cancer Genetics Branch, National Human Genome Research Institute, National Institute of

15 Health, Bethesda, MD 20892.

16

17 Running Header: LD in Wild and Domestic Canids

18

19 Key Words: Linkage disequilibrium, gray wolf, domestic dog, domestication, coyote,
20 demographic history

21

22 Correspondence to:

23 Melissa M. Gray

24 621 Charles E. Young Dr. So.

25 Los Angeles, CA 90095

26 Phone: 310-825-5014

27 Fax: 310-206-3987

28 email:mgray9@ucla.edu

29

30

31

ABSTRACT

32 Assessing the extent of linkage disequilibrium (LD) in natural populations of a non-model
33 species has been difficult due to the lack of available genomic markers. However, with advances
34 in genotyping and genome sequencing, genomic characterization of natural populations has
35 become feasible. Using sequence data and SNP genotypes, we measured linkage disequilibrium
36 (LD) and modeled the demographic history of wild canid populations and domestic dog breeds.
37 In 11 gray wolf and one coyote population we find that the extent of LD as measured by the
38 distance at which r^2 equals 0.2, extends <10kb in outbred populations to over 1.7Mb in
39 populations that have experienced significant founder events and bottlenecks. This large range in
40 the extent of LD parallels that observed in 18 dog breeds where the r^2 value varies from ~20kb to
41 >5Mb. Furthermore, in modeling demographic history under a composite likelihood framework,
42 we find that two of five wild canid populations exhibit evidence of a historical population
43 contraction. Five domestic dog breeds display evidence for a minor population contraction
44 during domestication, and a more severe contraction during breed formation. Only a 5%
45 reduction in nucleotide diversity was observed as a result of domestication, whereas the loss of
46 nucleotide diversity with breed formation averaged 35%.

47

INTRODUCTION

48 Recombination, recurrent mutation, selection, admixture, and mate choice are all factors
49 that can affect the extent of linkage disequilibrium (LD) within a species (DEONIER *et al.* 2005;
50 GAUT and LONG 2003; MUELLER 2004). The extent of LD is one of several factors that affects
51 how readily phenotypic traits in natural populations can be mapped using whole genome
52 association studies. Extensive LD allows associations to be detected more readily using a small
53 number of distantly placed but informative markers, whereas low LD necessitates fine-scale
54 mapping (KOHN *et al.* 2006; SLATE 2005; STEINER *et al.* 2007). Several studies have measured
55 the extent of LD in plant, invertebrate and domestic vertebrate species (CUTTER *et al.* 2006;
56 FARNIR *et al.* 2000; HADDRILL *et al.* 2005; HARMEGNIES *et al.* 2006; INGVARSSON 2005; MCRAE
57 *et al.* 2002; REMINGTON *et al.* 2001); however, little is known about the extent of LD in wild
58 populations of non-human vertebrates.

59 Among populations of the same species that share similar rates of recombination and
60 mutation and where selection is weak, a critical variable for determining the extent of LD is
61 demographic history. In general, populations which have remained large for a substantial period
62 of time or have rapidly expanded, demonstrate lower levels of LD than those that are small or
63 have experienced recent population bottlenecks (GAUT and LONG 2003; MUELLER 2004;
64 PRITCHARD and PRZEWORSKI 2001; REICH *et al.* 2001). Therefore, the extent of LD can be used
65 as a tool to infer demographic history. However, with the exception of a few model species and
66 humans, the extent of LD has not been used to explore population history, primarily because
67 large numbers of markers need to be typed in a substantial number of samples. To date, there
68 have been only a few wild vertebrate species for which the extent of LD has been carefully
69 documented: the collared flycatcher (*Ficedula albicollis*) (BACKSTROM *et al.* 2006), red deer

70 (*Cervus elaphus*) (SLATE and PEMBERTON 2007), and wild mice (*Mus musculus domesticus*)
71 (LAURIE *et al.* 2007). Additionally, LD was modeled in a wild Soay Sheep population using
72 simulations that included parameters of population history (MCRAE *et al.* 2005).

73 The domestic dog (*Canis familiaris*) is emerging as an important model for understanding
74 the genetic basis of morphology, behavior, and disease in mammals (OSTRANDER and WAYNE
75 2005; PARKER and OSTRANDER 2005; SUTTER and OSTRANDER 2004; WAYNE and OSTRANDER
76 2007). In 2005, a 7.8x whole-genome shotgun sequence and assembly of the Boxer was
77 completed (LINDBLAD-TOH *et al.* 2005). In addition, a 1.5x survey sequence of the Standard
78 Poodle became publically available in 2003 (KIRKNESS *et al.* 2003). These two resources,
79 together with 100,000 random sequence reads from nine other dogs of unrelated breeds and
80 20,000 sequence reads from each of four gray wolves (*Canis lupus*) and one coyote (*Canis*
81 *latrans*) (LINDBLAD-TOH *et al.* 2005), provide extensive resources for identifying markers for
82 large scale genetic analysis of wild canid species.

83 In this paper, we utilize dog-derived single nucleotide polymorphisms (SNPs) as well as
84 extensive resequencing to obtain estimates of LD in wild and domestic canids. Sutter *et al.*
85 (2004) first characterized the extent of LD in five dog breeds across five 1Mb regions, which
86 was followed by Linbladh-toh *et al.* (2005), who examined the extent of LD within 10 dog
87 breeds across a 15 Mb region. We expand the number of domestic dog breeds that the extent of
88 LD is estimated (12 new breeds), although our primary goal is to determine the range that LD
89 extends in a large panel of wild canid populations, including gray wolves and coyotes, and
90 compare these estimates to those from domestic dog breeds. Furthermore, we explore the
91 relationship of LD and demographic history by comparing estimates of LD to known population
92 histories. We also modeled population histories using the site frequency spectra (SFS) of each

93 population based on origination from an ancestral wolf population followed by specific
94 demographic scenarios. These predicted SFS are then compared to those observed.

95 METHODS

96 Blood, tissue, or buccal swab samples were collected from 908 individuals: 18 dog
97 breeds, $n=546$ (unrelated at the grandparent level); 14 gray wolf populations, $n=344$; and one
98 coyote population, $n=18$ (Table 1). To determine the rate of successful amplification of dog
99 derived molecular markers in distant relatives of the domestic dog, an additional 93 samples were
100 typed from golden jackal (*Canis aureus*), bat-eared fox (*Otocyon megalotis*), gray fox (*Urocyon*
101 *cinareoargenteus*) and Channel Island fox (*Urocyon littoralis*). For samples with low DNA
102 concentrations, whole genome amplification was performed according to manufacturer
103 guidelines (Qiagen REPLI-g kit; QIAGEN INC., VALENCIA, CA).

104 Gray wolf populations sampled varied in demographic history, and include individuals
105 from large outbred populations and from smaller inbred or recently bottlenecked populations
106 (Table S1). Furthermore, the populations chosen here have been the focus of previous genetic
107 research (LEHMAN *et al.* 1992; LEHMAN and WAYNE 1991; LEONARD *et al.* 2005; MUSIANI *et al.*
108 2007; RAMIREZ *et al.* 2006; ROY *et al.* 1994; ROY *et al.* 1996; VILÀ *et al.* 1999a) such that
109 demographic and population genetic conclusions from LD patterns can be independently
110 verified.

111 The domestic dog breeds included in this study also vary in relatedness and demographic
112 history; thus, they provide a test of the use of LD to assess population demography across a
113 variety of timescales and population sizes. American Kennel Club (AKC) registration statistics
114 (AKC website: http://www.akc.org/reg/dogreg_stats.cfm) were used as a proxy for effective
115 population size and recent demographic history. Kendall's tau and a Mantel's test were

116 performed to determine the significance of the association between LD and the log of the total
117 number of registered individuals.

118 All 1001 samples were genotyped on an ABI 3730 (APPLIED BIOSYSTEMS INC., FOSTER
119 CITY, CA) for 106 SNP loci (Table 1; Figure S1) using a custom set of primers designed for the
120 SNPlex genotyping system (APPLIED BIOSYSTEMS INC., FOSTER CITY, CA). The 106 SNPs
121 genotyped were chosen as a representative subset of the 200 SNPs described previously in Sutter
122 et al. (2004), which were ascertained by direct resequencing of 5 loci on 5 chromosomes each
123 spanning a non-contiguous 5Mb region. Genemapper 4.0 was used to make genotype calls for
124 each of the SNPs (APPLIED BIOSYSTEMS INC., FOSTER CITY, CA).

125 To determine the effect of ascertainment bias on LD estimates from genotype data and to
126 model demographic history (see below), sequencing of 18 amplicons spaced across a non-
127 contiguous 5Mb region of chromosome 1 (similar to Sutter et al. 2004) was performed on 188
128 individuals (a subset of genotyped individuals): 5 breeds of dog, $n=97$ (same as Sutter et al.); 4
129 gray wolf populations, $n=73$; a coyote population, $n=17$; and one golden jackal (Table S2 &
130 Figure 1). Eleven of the 18 amplicons were reported previously in Sutter et al. (2004). The 11
131 amplicons were chosen to minimize the amount of sequencing but still measure low to medium
132 range LD (i.e. 10-100kb). We designed and sequenced an additional seven amplicons, spaced at
133 50kb intervals from the central region on chromosome 1, to enhance this latter goal (Figure 1 &
134 Table S2). Sequences were run on an ABI 3730 and polymorphisms were identified and viewed
135 using Phred/Phrap/Consed/Polyphred (EWING and GREEN 1998; EWING *et al.* 1998; GORDON *et*
136 *al.* 1998; NICKERSON *et al.* 1997). All data will be made available on upon request.

137 **Analysis**

138 GENEPOP (RAYMOND and ROUSSET 1995) was used to calculate levels of heterozygosity,
139 and Hardy-Weinberg equilibrium. Estimates of nucleotide diversity were calculated from
140 sequence data representing a sampling of one chromosome per individual across 2000 iterations.
141 This was done to account for inbreeding within breeds and potential sampling of closely related
142 individuals in wild populations. To determine if and to what degree there was a loss of diversity
143 during the domestication of dogs, we sampled one chromosome from each of five breeds and
144 four wild canid populations and calculated loss of diversity as $1 - (\pi \text{ in dogs}) / (\pi \text{ in wolves})$.
145 Additionally, the loss of nucleotide diversity at breed formation ($1 - (\pi \text{ in breed}) / (\pi \text{ in dogs})$) was
146 calculated by sampling one chromosome from each breed of dog and five chromosomes within
147 one breed. Haplotypes were inferred across each population using the software program PHASE
148 (STEPHENS and DONNELLY 2003; STEPHENS *et al.* 2001). The percent of haplotypes within and
149 among breeds/populations were calculated as in Sutter *et al.* (2004). The software program
150 Haploview (BARRETT *et al.* 2005) was used to calculate D' and r^2 , which were plotted by
151 matching allele frequencies between pairs of SNPs with an allele frequency difference of <10%
152 (EBERLE *et al.* 2006). Median values for each distance category were calculated and a
153 logarithmic curve was fitted to the data. An r^2 of 0.2 was arbitrarily chosen as the value for
154 which the extent of LD was compared between populations and species.

155 **Ascertainment Bias:**

156 The discovery panel for genotyped loci contained 5 breeds of dog: one breed from each
157 of 5 distinct phylogenetic groupings (PARKER *et al.* 2004; PARKER *et al.* 2007; SUTTER *et al.*
158 2004). Genotyping these loci in populations outside the discovery panel will likely result in
159 ascertainment bias because SNPs at high frequency in one population/breed may be rare or
160 absent in others. To assess the level of ascertainment bias in breeds genotyped outside of the

161 discovery panel, simulations were carried out replicating the ascertainment study design (Figure
162 S2). Using the original sequence data (SUTTER *et al.* 2004) we performed the following: (1) one
163 breed of dog was randomly chosen as the focus breed while individuals from the 4 other breeds
164 were designated as the ascertainment panel; (2) one individual was randomly selected from the
165 ascertainment panel; (3) along each amplicon, another individual was randomly selected from
166 the same ascertainment panel and compared with the individual from step 2. If any marker was
167 segregating, it was flagged as a SNP and was “genotyped” in the focus breed. We did this to
168 increase the diversity of sequence comparisons and account for the possible unobserved
169 sequences due to recombination among amplicons. Also, we wanted to reduce the bias caused by
170 using a limited set of starting points; (4) repeat steps 2 and 3 for 2000 bootstraps. (5) calculate
171 the extent of LD for the focus breed and compare it to the extent of LD obtained from the
172 observed sequence data.

173 **Population Structure:**

174 Principal component analysis (PCA) using the program EIGENSTRAT (PRICE *et al.* 2006)
175 was performed on genotyped loci to determine if the 106 SNP dataset identified population
176 substructure. Twstats (PATTERSON *et al.* 2006) in EIGENSTRAT, was used to determine the
177 number of significant principal components. One plus the number of significant principal
178 components is considered “a rule of thumb” that identifies the number of groups determined by
179 the PCA.

180 **Demographic Modeling:**

181 To explore the demographic history of wild canid populations and domestic dog breeds,
182 we used the program PRFREQ (BOYKO *et al.* 2008; WILLIAMSON *et al.* 2005) to estimate
183 demographic parameters under a composite likelihood framework. The program utilizes the

184 Poisson Random Field (PRF) approach (SAWYER and HARTL 1992), which predicts the
185 distribution of allele frequencies across sites based on single-locus diffusion theory.
186 Demographic parameters are then estimated by maximum likelihood using the site frequency
187 spectra (SFS). Assumptions of the program are the Wright-Fisher model of mutation and
188 independence among sites. However, the majority of loci utilized in this study were in linkage
189 disequilibrium. Therefore, the likelihoods should be interpreted as a “composite-likelihood”
190 function, an approximation of the true likelihood (CAICEDO *et al.* 2007), as the assumption of
191 independence between sites is violated. Because of the violation of linkage equilibrium,
192 simulations of linked sites within amplicons were performed to verify our p-values (see
193 supplemental text). Two likelihood functions were used to make inferences. The first is based on
194 the number of SNPs in each frequency class (denoted "Poisson"), and the second is based on the
195 proportion of SNPs in each class (denoted "multinomial"). In our experience, the Poisson
196 likelihood function is much more powerful for inference of bottlenecks, since it takes into
197 account the degree of reduction in diversity as well as the skew in allele frequency distribution.
198 The multinomial likelihood function captures only the latter, but has the advantage of not
199 requiring *a priori* assumptions regarding the mutation rate (or equivalently, the effective
200 population size prior to bottleneck). Multinomial calculations were found to be qualitatively
201 similar to Poisson calculations (see supplement Tables S5-S6). Across models, the significance
202 of incorporating additional demographic parameters was assessed by using the likelihood ratio
203 test ($2\log[L(\text{model 1})/L(\text{model 2})]$).

204 **Domestication Model:**

205 To model the initial domestication event in dogs (Figure 2a), sequence data from Sutter et
206 al. (2004) derived from five breeds of dog were combined with genotype data from this study,

207 producing a combined dataset containing 22 dog breeds. Only those loci that were common to
208 both datasets and were not segregating in the ancestral state (determined by comparison to the
209 golden jackal) were retained for further analysis (82 SNPs). In order to generate an unfolded site
210 frequency spectrum representative of the ancestral domesticated dog population used to estimate
211 demographic parameters in PRFREQ, we randomly selected one chromosome from each breed for
212 2000 iterations and constructed a SFS averaged across iterations. However, as genotypes were
213 missing for individuals for many of the SNPs, sample sizes varied per SNP. In order to create a
214 valid SFS, we selected a new sample size of 14, slightly greater than half of the largest sample
215 size of 22. Although the choice of 14 is somewhat arbitrary, this approach reflects a trade off
216 between having a larger number of entries in the SFS and excluding more SNPs with low sample
217 size. SNPs with sample size less than 14 were excluded, leaving a total of 76 SNPs; SNPs with
218 sample sizes greater than 14 were “projected” to a sample size of 14. This involved using the
219 hypergeometric distribution to calculate the probability of the latter falling into each frequency
220 class of a SFS with sample size 14, and summing over all SNPs in each frequency class to create
221 the final SFS (CLARK *et al.* 2005). This projection makes no assumptions regarding missing data,
222 as each SNP is projected to a sample size smaller than its original sample size.

223 To control for the effect of ascertainment bias on our observed SFS, we chose to also
224 create a corrected site frequency spectrum as outlined in Nielsen *et al.* (2004). Corrections were
225 made under the basic model, assuming all SNPs were ascertained at a discovery panel depth of
226 five (for the five initial dog breeds in the ascertainment panel). We then found the maximum
227 likelihood of the true probabilities of each entry in the site frequency spectrum given our
228 observed values of entries.

229 PRFREQ was then used to estimate demographic parameters based on two site frequency
230 spectra (corrected for ascertainment bias and uncorrected). A constant population size of 21,591
231 in gray wolf (N_{e_WOLF}) was assumed over time, which was estimated from $\theta = 4N\mu$, using a
232 mutation rate (μ) of 1×10^{-8} per generation (LINDBLAD-TOH *et al.* 2005) and Watterson's θ
233 calculated from pairwise estimates between all wolves. This value was re-evaluated leaving out
234 populations found later to exhibit some evidence of a past contraction event (see results) but was
235 minimally different (22,600). The initiation of the domestication of domestic dogs was assumed
236 to be 15,000 years before present (τ) (OLSEN 1985; SAVOLAINEN *et al.* 2002).

237 We estimated the following parameters: length of the domestication event (τ_B), the
238 bottleneck population size (ω_B), and the domesticated dog population size after the bottleneck
239 (ω) (Figure 2a & Table 2). The model that best fit the observed data was determined by
240 comparing the likelihoods of the nested models. The Constant Size model estimated the
241 likelihood of a constant population size. The Contraction at Fixed Time model estimated the size
242 of a population after a contraction (ω) while keeping the contraction time constant (τ). The
243 Contraction at Unknown Time model estimated both ω and τ . The Bottleneck of Fixed Size and
244 Unknown Size model estimated populations that have undergone two population size changes.
245 The Bottleneck of Fixed Size model was fixed for a bottleneck with a 10-fold population
246 reduction at domestication of length τ_B followed by a second population size change of severity
247 ω whereas Bottleneck of Unknown Size modeled a breed bottleneck estimating τ_B , ω_B , and ω .

248 **Breed Formation Model:**

249 To further model the formation of breeds (Figure 2b), we used the complete sequence
250 data from Sutter *et al.* (2004) from five chromosomes in five dog breeds. However, as a result of
251 breeding programs, dog breeds are highly inbred, and an individual's chromosomes are more

252 similar than expected under random mating. Since this could potentially affect our demographic
253 inferences, we attempted to reduce the effects of inbreeding within breeds by sampling one
254 chromosome per individual per breed 2000 times, creating a site frequency spectrum for each
255 iteration, and averaging across iterations. Because the data were generated by sequencing and
256 sample sizes of SNPs were consistent across loci, corrections for sample size and ascertainment
257 bias were not necessary as with modeling the domestication event.

258 Likelihoods between several nested models were compared to determine which model
259 best fit the observed data (Table 2). The Constant Size, Contraction at Fixed Time, and
260 Contraction at Unknown Time models are breed contraction models identical to domestication
261 models of the same name. For the Contraction at Fixed Time model, a population contraction
262 time was set at 100 generations for all breeds, corresponding to roughly 300 years ago. While
263 this may not be appropriate for all breeds, it provided a general estimate allowing for better
264 comparison of contraction severities between them. Bottleneck of Fixed Length and Bottleneck
265 of Unknown Length, both model a population that has undergone a contraction followed by an
266 expansion, where Bottleneck of Fixed Length modeled a fixed bottleneck of short length (τ_B) and
267 Bottleneck of Unknown Length estimated this parameter.

268 **Wild Canid Model:**

269 To model the formation of wild canid populations (Figure 2c) , we used chromosome 1
270 sequence data generated in this study across 4 gray wolf populations and one coyote population.
271 To account for potential inbreeding in wild canid populations, and to ensure that unknown
272 closely-related individuals did not bias our estimates, we chose to perform the same chromosome
273 sampling done for dog breeds. As in the breed formation models, corrections for sample size and
274 ascertainment bias were not necessary.

297 nucleotide diversity suggests a minimal loss of diversity as a result of the domestication event
298 (0.05) whereas the average loss of diversity due to breed formation was much larger (0.35).

299 Using data from five chromosomes (1, 2, 3, 34, and 37), 105 of 106 loci were
300 successfully genotyped in 18 dog breeds, and a variety of wild canid species (Table S4). In the
301 most distantly related species, approximately 93% of loci in golden jackals, and 80% of loci in
302 bat-eared fox, gray fox, and island fox were successfully amplified. Observed heterozygosity
303 values ranged from 0.24 in golden jackals to 0.31, 0.29, and 0.33 in the bat-eared, gray, and
304 island fox, respectively.

305 **Population Structure:**

306 Population structure within and between wild canids and among domestic dog breeds was
307 explored through principal component analyses (PCA) of the genotype data using the program
308 EIGENSTRAT (Figures 3a-d). Fourteen principal components (PCs) were found to be significant
309 for the seven canid species analyzed (Figure 3a). Domestic dog, gray wolf, coyote, golden jackal,
310 and foxes were found to separate along principal component axes 1 and 2. Red wolf was found to
311 overlap both coyote and gray wolf but more so with the latter species.

312 PCA of 13 gray wolf populations revealed 11 significant principal components (Figure
313 3b). The most distinct pattern was observed along the first axis of variation separating Old World
314 and New World wolf populations. Minimal overlap between the two groups was evident. PCA
315 was then performed separately on Old and New World populations (Figures 3c & d). Seven Old
316 World gray wolf populations were found to have seven significant principal components. The
317 first axis of variation visibly separated the majority of the populations particularly from Swedish
318 gray wolves. Six New World gray wolf populations were found to have six significant principal
319 components. Isle Royale, Minnesota, and Northern Quebec define distinct clusters with Alaska,

320 Canada, and Yellowstone forming a fourth cluster, but showing considerable overlap with each
321 other.

322 PCA of 18 domestic dog breeds exhibited 15 significant principal components. There was
323 considerable overlap between breeds. However, Akita displayed virtually no overlap with any
324 other domestic dog breed along the first axis of variation (data not shown). Additionally,
325 Pekingese exhibits some separation along the first axis. Along the second axis of variation,
326 Mastiffs and to a lesser extent, Portuguese Water Dogs exhibited separation from the main
327 cluster of breeds.

328 **Ascertainment Bias:**

329 Ascertainment bias typically produces a pattern characterized by a decrease in low
330 frequency alleles and an increase in higher frequency alleles (CLARK *et al.* 2005; ROSENBLUM
331 and NOVEMBRE 2007). A shift in allele frequencies was observed between the sequence and
332 genotype data; however, there was no discernable pattern (Figure S3). Simulations showed that
333 the degree to which the difference in allele frequency affects estimates of LD varies for each
334 breed (Figure S4). Ascertainment bias in Labrador Retriever was observed to have minimal
335 effects; however, in Golden Retrievers the effect was large. Despite the variance in estimates, the
336 rank order of breeds based on estimates of LD remained the same.

337 Site frequency spectra generated from sequence and genotype data for gray wolf and
338 coyote populations were similar to the domestic dog in showing no distinguishable patterns
339 (Figure S3). When LD estimates from sequence data were compared with LD estimates from
340 genotype data, sequence data generally gave lower estimates of LD (Figure 4). The exception to
341 this pattern was found in the Spanish gray wolf, which had an increase in LD measured from

342 sequenced data compared with genotyped data. Despite the shift in LD estimates from sequence
343 to genotype data, the rank order of LD estimates of each population remained the same.

344 Given that ascertainment bias is present within the genotype dataset, we proceeded with
345 caution by focusing on general trends (e.g., strong association between the extent of LD and
346 demographic history). Additionally, we relied on estimates from re-sequencing to make unbiased
347 estimates and comparisons of genetic diversity and extent of LD between domestic and wild
348 canids.

349 **Linkage Disequilibrium:**

350 The extent of LD estimated from genotyped data in gray wolf populations ranged from
351 <10kb in Alaskan gray wolves to >5Mb in gray wolves from Isle Royale ($r^2_{0.2}$; Table 5). The
352 extent of LD was consistent with the known demographic history of each population. Large
353 outbreeding populations such as Alaska, Minnesota, Canada, Yellowstone, and Northern Quebec
354 exhibited such low levels of LD that the decay curves did not extend to an r^2 value of 0.2 (Figure
355 4 and Table 5). Therefore, we take a conservative approach and consider these populations to
356 generally have LD levels lower than 10kb. However, small/bottlenecked populations such as Isle
357 Royale, Spanish, Italian, and Swedish gray wolves exhibited high levels of LD ($r^2_{0.2}>500\text{kb}$).
358 Lastly, coyotes exhibited levels of LD below an r^2 value of 0.2 (Figure 4) consistent with their
359 large population size in southern California (FEDRIANI *et al.* 2001; VILÀ *et al.* 1999a).

360 Estimates of LD from genotyped data in dog breeds ranged from 20kb to >5Mb ($r^2_{0.2}$;
361 Table 5). The extent of LD was found to be significantly correlated to the log of registered
362 individuals for both Kendal's tau rank correlation (p-value = 0.02) and Mantel's test (p-value =
363 0.0001). However, three breeds had sample numbers below the minimum cutoff ($n<17$) used by
364 Sutter *et al.* (2004), introducing potentially greater bias into measures of LD. When these breeds

365 were excluded from the analysis, the correlation statistics for the remaining 14 breeds were still
366 significant (Kendal's tau, p-value = 0.003; Mantel's test, p-value = 0.0002). Thus, the level of
367 LD within dog breeds was found to be well correlated with 2006 registration numbers (Figure 5).

368 Values of LD based on sequence data were highly correlated with those based on SNP
369 genotypes (Kendal's tau, p-value = 0.02; Mantel's test, p-value = 0.0001). Sequence data
370 comparisons of the extent of LD between species demonstrated that gray wolves and coyotes
371 have less LD (<10kb to 1.7Mb) than the domestic dog (785kb to >5Mb; Table 5). The extent of
372 LD seen in the Spanish gray wolf population was much higher than any other sequenced gray
373 wolf population (1.7Mb). We explored the possibility of relatedness among the samples by
374 eliminating individuals with high levels of allele sharing based on 11 microsatellite loci
375 (VONHOLDT *et al.* 2008), and confirmed that high levels of LD are still present in a sample set of
376 reduced allele sharing ($r^2_{0.2} \geq 1.5\text{Mb}$).

377 **Domestication Modeling:**

378 Parameter estimates were scaled in terms of the estimated gray wolf effective population
379 size (i.e., $\omega = N_{e\text{DOG}}/N_{e\text{WOLF}}$). The Contraction at Fixed Time model, with a single contraction
380 event fixed at 15,000 years ago, significantly explained the data better than the null model of
381 constant population size (Table 6). This applied for both the ascertainment bias corrected (p-
382 value = 2.27×10^{-6}) and uncorrected data sets (p-value = 4.01×10^{-8}). The corrected Poisson
383 calculations suggest this contraction was followed by a population expansion (Bottleneck of
384 Fixed Size, Table 6), although the improvement in the model fit is slight (p-value = 0.033) and
385 unlikely to be significant after correcting for linkage in the dataset. Therefore, we focus on the
386 Contraction at Fixed Time model findings. The estimate of ω for the Poisson calculation of the
387 Contraction at Fixed Time model was 0.23 for the uncorrected data, indicating the dog ancestral

388 population size was 0.23 times the size of the wolf ancestral population, and 0.25 for the
389 corrected data. Therefore, results suggest a single minor contraction event was associated with
390 domestication of the dog.

391 **Breed Formation Modeling:**

392 Demographic parameters based on the discussed model of breed formation were
393 estimated for each of 5 breeds from the averaged sampled site frequency spectra. Across all
394 breeds and calculations, no models had a higher likelihood than the Contraction at Fixed Time
395 model (Table 7), which indicates a contraction without a subsequent increase in population size.
396 Under the Contraction at Fixed Time model, Bernese Mountain Dog and Pekingese were
397 observed to have the largest bottleneck with a current effective population size approximately
398 0.0055 and 0.0056 that of the ancestral dog effective population. Labrador Retriever, Golden
399 Retriever, and Akita exhibited a weaker reduction in population size with values of 0.0095,
400 0.011, and 0.012, respectively.

401 Although not significantly better than the Contraction at Fixed Time model, both ω and τ
402 were optimized under the Contraction at Unknown Time model allowing examination of the
403 timing of breed contractions. Under this model, Pekingese was observed to have a severe
404 reduction in population size ~65 generations ago ($\omega=0.0035$), while the Akita and Golden
405 Retriever were observed to have similar contraction times at ~92 generations ($\omega=0.0113$ &
406 0.0100). It is important to note that when estimating both ω and τ , timing estimates may not be
407 entirely realistic, as there is a tradeoff between having a recent τ and a severe population decline
408 and having a more distant τ but less severe decline. This trade-off is exemplified by the Bernese
409 Mountain Dog which was founded by a small number of individuals and maintained as a small
410 population to the present day. Therefore, the breed founding prediction of 755 generations ago is

411 likely overestimated. Regardless, the bottleneck at breed formation is orders of magnitude more
412 severe, and more recent, than an ancient domestication event and more likely to impact
413 differences in LD among breeds (see below).

414 **Wild Canid Modeling:**

415 As with the inference of breed formation, we used Poisson calculations to determine the
416 presence and severity of a bottleneck within gray wolf and coyote populations. Only the
417 Contraction at Unknown Time model of the Spanish and Israeli gray wolf populations was found
418 to be significantly different from the null model (Table 7). The Spanish gray wolf was observed
419 to have undergone a contraction in population size ($\omega=0.028$) about 226 generations ago, or
420 slightly less than 700 years ago and the Israeli wolf population was observed to have undergone
421 a more mild population decline ($\omega=0.25$) over 10,000 generations, or 30,000 years, ago. Again,
422 these estimates may not be entirely accurate as they may represent the tradeoff between a recent
423 large population decline and a more ancient and mild population decline. Lastly, no significant
424 evidence was found to support a change in population size in Alaskan and Yellowstone gray
425 wolf or coyote populations.

426 DISCUSSION

427 The extent of LD and its relationship to demographic history has been well documented
428 in domesticated and model organisms (ARDLIE *et al.* 2002; DUNNING *et al.* 2000; LAURIE *et al.*
429 2007; PRITCHARD and PRZEWORSKI 2001). However, little research has been done to explore the
430 extent of LD in wild populations, particularly vertebrate species. As mentioned previously, only
431 a few studies to date have measured the extent of LD in naturally occurring vertebrate
432 populations. Utilizing SNP markers developed in the domestic dog and extensive resequencing,

433 we explored the extent of LD and modeled demographic history in several populations of wild
434 canids. Additionally, we calculated the same measures in the domestic dog for comparison.

435 Five domestic dog breeds, four gray wolf, and one coyote population were sequenced for
436 11,279 bps on chromosome 1. Levels of LD in domestic dogs were consistent with previous
437 studies (LINDBLAD-TOH *et al.* 2005; SUTTER *et al.* 2004) and in general, we found that gray wolf
438 and coyote populations exhibited lower levels of LD (<10kb to 1.4Mb) than domestic dog breeds
439 (785kb to >5Mb; Table 5). Barley (CALDWELL *et al.* 2006), soybean (HYTEN *et al.* 2007), sheep
440 (MCRAE *et al.* 2002; MCRAE *et al.* 2005), and house mice (LAURIE *et al.* 2007) display a
441 consistent pattern of reduced levels of LD in wild populations compared to their domesticates.
442 This is expected since domestication likely results in a bottleneck event. However, across wild
443 populations, demographic history can still be observed to strongly influence levels of LD. For
444 example, the Spanish wolf population had LD levels higher than some domestic dog breeds
445 ($r^2_{0.2}=1.7\text{Mb}$). In the past century, gray wolves from Spain were hunted to near extinction, but
446 have steadily risen in numbers since the enactment of hunting restrictions (RAMIREZ *et al.* 2006).
447 In contrast, Labrador Retrievers exhibited levels of LD similar to wild gray wolf populations
448 ($r^2_{0.2}=785\text{kb}$) as they are the most popular breed in the U.S. today with about 150,000 new
449 registrations per year (www.AKC.org). Lastly, coyotes were found to display the lowest levels of
450 LD ($r^2_{0.2}<10\text{kb}$) relative to all domestic dog breeds and gray wolf populations. Consistent with
451 low levels of LD, coyote population sizes are reportedly an order of magnitude greater than the
452 gray wolf (VILÀ *et al.* 1999a).

453 As seen with the sequence data, LD levels from SNP genotype data were found to also
454 correspond well with the known demographic history for the 11 gray wolf populations. For
455 example, the Isle Royale gray wolf population is a small population of wolves that inhabit an

456 island in Lake Superior off the coast of Minnesota. The population was founded by a single
457 breeding pair of wolves in 1950 (PETERSON *et al.* 1998). Previous genetic research found
458 population heterozygosity levels half that observed in the mainland progenitor population
459 (Wayne *et al.* 1991). The extent of LD in the Isle Royale population ($r^2_{0.2} > 5\text{Mb}$) is consistent
460 with that expected in small and/or severely bottlenecked populations (GAUT and LONG 2003;
461 MUELLER 2004; PRITCHARD and PRZEWORSKI 2001). Other populations that are known to have a
462 history of population contraction or small population size had high levels of LD including
463 Spanish, Swedish, and Italian wolves (for supporting demographic and genetic research see:
464 FABBRI *et al.* 2007; LEHMAN *et al.* 1992; RAMIREZ *et al.* 2006; VILÀ *et al.* 1999a; WAYNE *et al.*
465 1992). At the other end of the spectrum, populations of Alaskan, Canadian, and Northern Quebec
466 gray wolves have been large and of constant size for a long time, and exhibit low levels of LD
467 (MUSIANI *et al.* 2007; WECKWORTH *et al.* 2005). Supporting this finding, genetic studies (ROY *et*
468 *al.* 1994; VILÀ *et al.* 1999b; WAYNE *et al.* 1992) of Alaskan and Northern Canadian gray wolf
469 populations found high variability and reduced population differentiation suggesting a large
470 population size and higher levels of gene flow than among European wolf populations which
471 were more structured. Similarly, LD estimates in dog breeds from SNP genotype data
472 corroborate findings from sequence data as exemplified by a significant correlation to popularity
473 of the breed based on registration numbers (Figure 5). Thus, the extent of LD measured from the
474 SNP genotype data also support the correlation between LD and demographic history in wild and
475 domestic populations.

476 **Demographic modeling:**

477 Previous studies based on mtDNA analysis (SAVOLAINEN *et al.* 2002; VILÀ *et al.* 1997)
478 have indicated that four to six matriline of gray wolf were involved in the founding of the

479 domestic dog. In contrast, analysis of major histocompatibility (MHC) loci suggested several
480 hundred founders or extensive backcrossing with wild canids is needed to explain present day
481 diversity in domestic dogs (VILÀ *et al.* 2005). Linbladh-Toh *et al.* (2005) found evidence for two
482 major bottlenecks in modern dog breeds, the first occurring as a result of domestication from
483 wolves, supported by short range LD estimates, and the second occurring as a result of breed
484 formation, supported by long range LD. Linbladh-Toh *et al.* (2005) simulated the demographic
485 history of domestic dogs over a coarse grid of demographic parameter values, and compared the
486 observed and simulated rates of pair-wise polymorphism across ten 15 Mb regions. They then
487 selected the domestication parameters for which the simulations resulted in polymorphism values
488 that were the closest to observed values. Although they do find evidence for two major
489 bottlenecks, they do not use a rigorous likelihood framework, and thus are not able to perform
490 any hypothesis testing or formal model selection. In our work, we search a denser grid of
491 domestication parameter values and examine the site frequency spectrum of dogs rather than
492 pair-wise polymorphism. For domestication events with parameters searched over this grid, we
493 calculate the likelihood of the observed domesticated dog site frequency spectrum. In this
494 likelihood framework, we are able to perform nested likelihood ratio tests to test the null
495 hypothesis of constant population size and make meaningful comparisons between models.

496 From our demographic modeling, we found evidence for a modest population contraction
497 approximately 15,000 years ago (5,000 generations ago) and a severe contraction at breed
498 formation. The contraction due to breed formation was found to be an order of magnitude greater
499 than the domestication contraction based on analysis of the site frequency spectra. From
500 nucleotide diversity estimates, only a 5% reduction in diversity was observed as a result of
501 domestication whereas an average loss of nucleotide diversity of 35% was observed due to breed

502 formation. This severe breed formation contraction was expected, as continued inbreeding within
503 a given breed may act to maintain a small effective population size even if the census population
504 size has actually increased since breed formation. The absence of a strong signal for a
505 contraction at domestication may reflect continued interbreeding between early dogs and wolves
506 or multiple domestication events (RANDI and LUCCHINI 2002; TSUDA *et al.* 1997; VILÀ *et al.*
507 1997). Indeed, high levels of diversity observed in domestic dogs may have been maintained
508 through a modest population bottleneck, backcrossing with wolf populations, and rapid
509 population expansion (VILÀ *et al.* 2005; WAYNE and OSTRANDER 2007).

510 Lastly, demographic modeling of the site frequency spectra of wild canid populations and
511 dog breeds were found to be concordant with estimates of LD and known population history. In
512 wild canid populations, a significant population decline was observed for the Spanish gray wolf
513 and to a lesser extent the Israeli gray wolf population, which was expected from known historical
514 data. Furthermore, neither coyote nor Alaskan and Yellowstone gray wolf populations showed
515 significant evidence of a population size change. In modeling the demographic history of
516 domestic dog breeds, Pekingese and Bernese Mountain Dog exhibited the greatest population
517 contraction and more modest contractions were observed in Golden Retriever and Labrador
518 Retriever. The strong concordance observed in this study between the extent of LD, demographic
519 modeling and known demographic history support the use of LD to infer population history not
520 only in model organisms but also in wild populations.

521 **Population Structure:**

522 Eighty percent or greater of SNPs that were discovered in dogs successfully amplified in
523 the most distantly related species (gray and island fox) and polymorphism levels ranged from
524 25% to 40%. Genetic isolation and/or admixture revealed in the PCA was consistent with

525 previous studies (LEHMAN *et al.* 1992; LEONARD *et al.* 2005; ROY *et al.* 1994; VILÀ *et al.* 1999a;
526 WAYNE *et al.* 1992; WAYNE *et al.* 1991). Within gray wolves, PCA identified strong geographic
527 differentiation between Old and New World populations as well as between populations within
528 each continent. Similar relationships have been observed in mtDNA studies of gray wolves (ROY
529 *et al.* 1994; VILÀ *et al.* 1999a; WAYNE *et al.* 1992). Patterns in the PCA plots were also
530 consistent with previous phylogenetic studies (LEONARD *et al.* 2005; LINDBLAD-TOH *et al.* 2005;
531 ROY *et al.* 1996; VILÀ *et al.* 1999a). For example, PC one supports the fundamental genetic
532 distance between wild canids and domestic dogs. PC two distinguishes wild canids from each
533 other with coyotes and golden jackals positioned nearest to the gray wolves and red wolves
534 overlapping coyotes and gray wolves. The overlap of red wolves with both species is consistent
535 with extensive hybridization in the past (WAYNE and JENKS 1991). The high degree of SNP
536 amplification success between species suggests that dog-derived SNP markers may be useful in
537 mapping phenotypic traits in wild canid species such as wolves and coyotes. To support this
538 conclusion, Kukekova *et al.* (2007) used dog derived microsatellite markers to develop a genetic
539 map for the silver fox and Sacks and Louie (2008) and Seddon *et al.* (2005) sequenced SNP loci
540 from the dog genome to develop new SNPs for genetic studies in gray wolf, coyote, red fox, and
541 gray fox.

542 **Conclusions:**

543 The extent of LD in natural vertebrate populations has been difficult to assess in the past
544 because large scale genomic surveys were only possible in model species. However, with the
545 availability of high throughput genotyping and information from genome sequencing projects, a
546 new era has emerged in the genetic characterization of natural populations. Utilizing these
547 resources, we have estimated LD in 11 natural populations of gray wolf, one population of

548 coyote and 18 dog breeds. Additionally, because a causal relationship exists between LD and
549 population history, we have made inferences about the demographic and evolutionary processes
550 in wild and domestic canids. Our results suggest that a relatively minor population contraction
551 was associated with domestication in dogs and that genetic variation was preserved in the rapid
552 expansion that followed. However, this variation is now partitioned in dog breeds which
553 generally have high and variable amounts of LD. The high level of LD in some wolf populations
554 further suggests the possibility of trait mapping in natural populations. For example, in North
555 America, approximately half of wolves are dark colored (ANDERSON *et al.* in review; MUSIANI
556 *et al.* 2007), and given the recent identification of coat color mutants in dogs associated with
557 black color (CALDWELL *et al.* 2006), similar mutants may now be identified through association
558 studies in wild wolves. Finally, we demonstrate how simulation models in general can be used to
559 make inferences about population demography and show that predictions generally fit with
560 observed levels of LD and known population history. Consequently, our approach may have
561 wide applicability to other species with extensive genomic resources and to their close relatives.

ACKNOWLEDGEMENTS

562
563
564
565
566
567
568
569
570
571

We would like to thank the following individuals for their helpful comments and discussion: three anonymous reviewers, Matthew Stephens, John Novembre, Olaf Thalmann, Klaus Koepfli, Pascal Quignon, Bridgett VonHoldt, and John Pollinger. We would also like to thank Dan Stahler, Seth Riley, Eli Geffen, Kevin Chase, Gordon Lark, and countless dog owners and breeders for sample contribution. For analytical assistance, we thank Katarzyna Bryc, Badri Padhukasahasram, and Ryan Hernandez. This study was supported by National Institute of Health training grant 5 T32 HG002536 (MMG), National Science Foundation grants 0516310 (CDB), 0733033 (RKW), National Institute of Health grant 5 U01 HL084706-02 (ARB) and by the Intramural program of the National Human Genome Research Institute.

572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592

LITERATURE CITED

ANDERSON, T. M., B. M. VONHOLDT, S. I. CANDILLE, M. MUSIANI, C. GRECO *et al.*, in review
Molecular and Evolutionary History of Melanism in North American Gray Wolves.

ARDLIE, K. G., L. KRUGLYAK and M. SEIESTAD, 2002 Patterns of Linkage Disequilibrium in the
Human Genome. *Nature Reviews Genetics* **3**: 299-309.

BACKSTROM, N., A. OVARNSTROM, L. GUSTAFSSON and H. ELLEGREN, 2006 Levels of linkage
disequilibrium in a wild bird population. *Biology Letters* **2**: 435-438.

BARRETT, J. C., B. FRY, J. MALLER and M. J. DALY, 2005 Haploview: analysis and visualization
of LD and haplotype maps. *Bioinformatics* **21**: 263-265.

BOYKO, A. R., S. H. WILLIAMSON, A. R. INDAP, J. D. DEGENHARDT, R. D. HERNANDEZ *et al.*,
2008 Assessing the evolutionary impact of amino acid mutations in the human genome.
PLoS Genetics **4**: e1000083.

CAICEDO, A. L., S. H. WILLIAMSON, R. D. HERNANDEZ, A. BOYKO, A. FLEDEL-ALON *et al.*, 2007
Genome-wide patterns of nucleotide polymorphism in domesticated rice. *Plos Genetics*
3: 1745-1756.

CALDWELL, K. S., J. RUSSELL, P. LANGRIDGE and W. POWELL, 2006 Extreme population-
dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum*
vulgare. *Genetics* **172**: 557-567.

CLARK, A. G., M. J. HUBISZ, C. D. BUSTAMANTE, S. H. WILLIAMSON and R. NIELSEN, 2005
Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*
15: 1496-1502.

593 CUTTER, A. D., S. E. BAIRD and D. CHARLESWORTH, 2006 High nucleotide polymorphism and
594 rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*.
595 *Genetics* **174**: 901-913.

596 DEONIER, R. C., S. TAVARE and M. S. WATERMAN, 2005 *Computation Genome Analysis: An*
597 *Introduction*. Springer Science, New York, New York.

598 DUNNING, A. M., F. DUROCHER, C. S. HEALEY, M. D. TEARE, S. E. MCBRIDE *et al.*, 2000 The
599 extent of linkage disequilibrium in four populations with distinct demographic histories.
600 *American Journal of Human Genetics* **67**: 1544-1554.

601 EBERLE, M. A., M. J. RIEDER, L. KRUGLYAK and D. A. NICKERSON, 2006 Allele frequency
602 matching between SNPs reveals an excess of linkage disequilibrium in genic regions of
603 the human genome. *Plos Genetics* **2**: 1319-1327.

604 EWING, B., and P. GREEN, 1998 Basecalling of automated sequencer traces using phred. II. Error
605 probabilities. *Genome Research* **8**: 186-194.

606 EWING, B., L. HILLIER, M. WENDL and P. GREEN, 1998 Basecalling of automated sequencer
607 traces using phred. I. Accuracy assessment. *Genome Research* **8**: 175-185.

608 FABRI, E., C. MIQUEL, V. LUCCHINI, A. SANTINI, R. CANIGLIA *et al.*, 2007 From the Apennines
609 to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*)
610 population. *Molecular Ecology* **16**: 1661-1671.

611 FARNIR, F., W. COPPIETERS, J. J. ARRANZ, P. BERZI, N. CAMBISANO *et al.*, 2000 Extensive
612 genome-wide linkage disequilibrium in cattle. *Genome Research* **10**: 220-227.

613 FEDRIANI, J. M., T. K. FULLER and R. M. SAUVAJOT, 2001 Does availability of anthropogenic
614 food enhance densities of omnivorous mammals? An example with coyotes in southern
615 California. *Ecography* **24**.

616 GAUT, B. S., and A. D. LONG, 2003 The lowdown on linkage disequilibrium. *Plant Cell* **15**:
617 1502-1506.

618 GORDON, D., C. ABAJIAN and P. GREEN, 1998 Consed: a graphical tool for sequence finishing.
619 *Genome Research* **8**: 195-202.

620 HADDRILL, P. R., K. R. THORNTON, B. CHARLESWORTH and P. ANDOLFATTO, 2005 Multilocus
621 patterns of nucleotide variability and the demographic and selection history of *Drosophila*
622 *melanogaster* populations. *Genome Research* **15**: 790-799.

623 HARMEGNIES, N., F. FARNIR, F. DAVIN, N. BUYS, M. GEORGES *et al.*, 2006 Measuring the extent
624 of linkage disequilibrium in commercial pig populations. *Animal Genetics* **37**: 225-231.

625 HYTEN, D. L., I. Y. CHOI, Q. J. SONG, R. C. SHOEMAKER, R. L. NELSON *et al.*, 2007 Highly
626 variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics*
627 **175**: 1937-1944.

628 INGVARSSON, P. K., 2005 Nucleotide polymorphism and linkage disequilibrium within and among
629 natural populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics* **169**:
630 945-953.

631 KIRKNESS, E. F., V. BAFNA, A. L. HALPERN, S. LEVY, K. REMINGTON *et al.*, 2003 The dog
632 genome: Survey sequencing and comparative analysis. **301**: 1898-1903.

633 KOHN, M. H., W. J. MURPHY, E. A. OSTRANDER and R. K. WAYNE, 2006 Genomics and
634 conservation genetics. *Trends in Ecology & Evolution* **21**: 629-637.

635 KUKKOVA, A. V., L. N. TRUT, I. N. OSKINA, J. L. JOHNSON, S. V. TEMNYKH *et al.*, 2007 A
636 meiotic linkage map of the silver fox, aligned and compared to the canine genome. **17**:
637 387-399.

638 LAURIE, C. C., D. A. NICKERSON, A. D. ANDERSON, B. S. WEIR, R. J. LIVINGSTON *et al.*, 2007
639 Linkage Disequilibrium in Wild Mice. *Plos Genetics* **3**: e144.

640 LEHMAN, N., P. CLARKSON, L. D. MECH, T. J. MEIER and R. K. WAYNE, 1992 A Study of Genetic
641 Relationships Withing and Among Wolf Packs Using DNA Filgerprinting and
642 mitochondrial DNA. *Behavioral Ecology and SocioBiology* **30**: 83-94.

643 LEHMAN, N., and R. K. WAYNE, 1991 Analysis of Coyote Mitochondrial-DNA Genotype
644 Frequencies - Estimation of the Effective Number of Alleles. *Genetics* **128**: 405-416.

645 LEONARD, J. A., C. VILÀ and R. K. WAYNE, 2005 Legacy lost: genetic variability and population
646 size of extirpated US grey wolves (*Canis lupus*). *Molecular Ecology* **14**: 9-17.

647 LINDBLAD-TOH, K., C. M. WADE, T. S. MIKKELSEN, E. K. KARLSSON, D. B. JAFFE *et al.*, 2005
648 Genome sequence, comparative analysis and haplotype structure of the domestic dog.
649 *Nature* **438**: 803-819.

650 MCRAE, A. F., J. C. MCEWAN, K. G. DODDS, T. WILSON, A. M. CRAWFORD *et al.*, 2002 Linkage
651 disequilibrium in domestic sheep. *Genetics* **160**: 1113-1122.

652 MCRAE, A. F., J. M. PEMBERTON and P. M. VISSCHER, 2005 Modeling linkage disequilibrium in
653 natural populations: The example of the Soay sheep population of St. Kilda, Scotland.
654 *Genetics* **171**: 251-258.

655 MUELLER, J. C., 2004 Linkage disequilibrium for different scales and applications. *Briefings in*
656 *Bioinformatics* **5**: 355-364.

657 MUSIANI, M., J. A. LEONARD, H. D. CLUFF, C. C. GATES, S. MARIANI *et al.*, 2007 Differentiation
658 of tundra and boreal coniferous forest wolves: genetics, coat color and foraging ecology.
659 *Molecular Ecology* **16**: 4149-4170.

660 NICKERSON, D. A., V. O. TOBE and S. L. TAYLOR, 1997 Polyphred: automating the detection and
661 genotyping of single nucleotide substitutions using fluorescence-based resequencing.
662 *Nucleic Acids Research* **25**: 2745-2751.

663 NIELSEN, R., M. J. HUBISZ and A. G. CLARK, 2004 Reconstituting the Frequency Spectrum of
664 Ascertained Single-Nucleotide Polymorphism Data. *Genetics* **168**: 2373-2382.

665 OLSEN, S. J., 1985 *Origins of the Domestic Dog*. University of Arizona Press, Tuscon.

666 OSTRANDER, E. A., and R. K. WAYNE, 2005 The canine genome. *Genome Research* **15**: 1706-
667 1716.

668 PARKER, H. G., L. V. KIM, N. B. SUTTER, S. CARLSON, T. D. LORENTZEN *et al.*, 2004 Genetic
669 structure of the purebred domestic dog. *Science* **304**: 1160-1164.

670 PARKER, H. G., A. V. KUKEKOVA, D. T. AKEY, O. GOLDSTEIN, E. F. KIRKNESS *et al.*, 2007 Breed
671 relationships facilitate fine-mapping studies: A 7.8-kb deletion cosegregates with Collie
672 eye anomaly across multiple dog breeds. **17**: 1562-1571.

673 PARKER, H. G., and E. A. OSTRANDER, 2005 Canine genomics and genetics: Running with the
674 pack. *Plos Genetics* **1**: 507-513.

675 PATTERSON, N. J., A. L. PRICE and D. REICH, 2006 Population Structure and Eigenanalysis. *Plos*
676 *Genetics* **2**: e190.

677 PETERSON, R. O., N. J. THOMAS, J. M. THURBER, J. A. VUCETICH and T. A. WAITE, 1998
678 Population limitation and the wolves of Isle Royale. *Journal of Mammalogy* **79**: 828-841.

679 PRICE, A. L., N. J. PATTERSON, R. M. PLENGE, M. E. WEINBLATT, N. A. SHADICK *et al.*, 2006
680 Principal components analysis corrects for stratification in genome-wide association
681 studies. *Nature Genetics* **38**: 904-909.

682 PRITCHARD, J. K., and M. PRZEWORSKI, 2001 Linkage disequilibrium in humans: Models and
683 data. *American Journal of Human Genetics* **69**: 1-14.

684 RAMIREZ, O., L. ALTET, C. ENSENAT, C. VILA, A. SANCHEZ *et al.*, 2006 Genetic assessment of
685 the Iberian wolf *Canis lupus signatus* captive breeding program. *Conservation Genetics*
686 **7**: 861-878.

687 RANDI, E., and V. LUCCHINI, 2002 Detecting rare introgression of domestic dog genes into wild
688 wolf (*Canis lupus*) populations by Bayesian admixture analyses of microsatellite
689 variation. *Conservation Genetics* **3**: 31-45.

690 RAYMOND, M., and F. ROUSSET, 1995 Genepop (Version-3.4) - Population-Genetics Software for
691 Exact Tests and Ecumenicism. *Journal of Heredity* **86**: 248-249.

692 REICH, D. E., M. CARGILL, S. BOLK, J. IRELAND, P. C. SABETI *et al.*, 2001 Linkage disequilibrium
693 in the human genome. *Nature* **411**: 199-204.

694 REMINGTON, D. L., J. M. THORNSBERRY, Y. MATSUOKA, L. M. WILSON, S. R. WHITT *et al.*, 2001
695 Structure of linkage disequilibrium and phenotypic associations in the maize genome.
696 *Proceedings of the National Academy of Sciences of the United States of America* **98**:
697 11479-11484.

698 ROSENBLUM, E. B., and J. NOVEMBRE, 2007 Ascertainment Bias in Spatially Structured
699 Populations: A Case Study in the Eastern Fence Lizard. *Journal of Heredity* **98**: 331-336.

700 ROY, M. S., E. GEFFEN, D. SMITH, E. A. OSTRANDER and R. K. WAYNE, 1994 Patterns of
701 Differentiation and Hybridization in North-American Wolflike Canids, Revealed by
702 Analysis of Microsatellite Loci. *Molecular Biology and Evolution* **11**: 553-570.

703 ROY, M. S., E. GEFFEN, D. SMITH and R. K. WAYNE, 1996 Molecular genetics of pre-1940 red
704 wolves. *Conservation Biology* **10**: 1413-1424.

705 SACKS, B. N., and S. LOUIE, 2008 Using the dog genome to find single nucleotide
706 polymorphisms in red foxes and other distantly related members of the Canidae
707 *Molecular Ecology Resources* **8**: 35-49.

708 SAVOLAINEN, P., Y. P. ZHANG, J. LUO, J. LUNDEBERG and T. LEITNER, 2002 Genetic evidence for
709 an East Asian origin of domestic dogs. *Science* **298**: 1610-1613.

710 SAWYER, S. A., and D. L. HARTL, 1992 Population-Genetics of Polymorphism and Divergence.
711 *Genetics* **132**: 1161-1176.

712 SEDDON, J. M., H. G. PARKER, E. A. OSTRANDER and H. ELLEGREN, 2005 SNPs in ecological and
713 conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology* **14**:
714 503-511.

715 SLATE, J., 2005 Quantitative trait locus mapping in natural populations: progress, caveats and
716 future directions. *Molecular Ecology* **14**: 363-379.

717 SLATE, J., and J. M. PEMBERTON, 2007 Admixture and Patterns of Linkage Disequilibrium in a
718 Free-living Vertebrate Population. *European Society For Evolutionary Biology* **20**: 1415-
719 1427.

720 STEINER, C. C., J. N. WEBER and H. E. HOEKSTRA, 2007 Adaptive Variation in Beach Mice
721 Produced by Two Interacting Pigmentation Genes. *Plos Biology* **5**: e219.

722 STEPHENS, M., and P. DONNELLY, 2003 A comparison of Bayesian methods for haplotype
723 reconstruction from population genotype data. *American Journal of Human Genetics* **73**:
724 1162-1169.

725 STEPHENS, M., N. J. SMITH and P. DONNELLY, 2001 A new statistical method for haplotype
726 reconstruction from population data. *American Journal of Human Genetics* **68**: 978-989.

727 SUTTER, N. B., M. A. EBERLE, H. G. PARKER, B. J. PULLAR, E. F. KIRKNESS *et al.*, 2004
728 Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome*
729 *Research* **14**: 2388-2396.

730 SUTTER, N. B., and E. A. OSTRANDER, 2004 Dog star rising: The canine genetic system. *Nature*
731 *Reviews Genetics* **5**: 900-910.

732 TSUDA, K., Y. KIKKAWA, H. YONEKAWA and Y. TANABE, 1997 Extensive interbreeding occurred
733 among multiple matriarchal ancestors during the domestication of dogs: Evidence from
734 inter- and intraspecies polymorphisms in the D-loop region of mitochondrial DNA
735 between dogs and wolves. *Genes & Genetic Systems* **72**: 229-238.

736 VILÀ, C., I. R. AMORIM, J. A. LEONARD, D. POSADA, J. CASTROVIEJO *et al.*, 1999a Mitochondrial
737 DNA phylogeography and population history of the grey wolf *Canis lupus*. *Molecular*
738 *Ecology* **8**: 2089-2103.

739 VILÀ, C., J. E. MALDONADO and R. K. WAYNE, 1999b Phylogenetic relationships, evolution, and
740 genetic diversity of the domestic dog. *Journal of Heredity* **90**: 71-77.

741 VILÀ, C., P. SAVOLAINEN, J. E. MALDONADO, I. R. AMORIM, J. E. RICE *et al.*, 1997 Multiple and
742 ancient origins of the domestic dog. *Science* **276**: 1687-1689.

743 VILÀ, W., J. SEDDON and H. ELLEGREN, 2005 Genes of domestic mammals augmented by
744 backcrossing with wild ancestors. *Trends in Genetics* **21**: 214-218.

745 VONHOLDT, B. M., D. R. STAHLER, D. W. SMITH, D. A. EARL, J. P. POLLINGER *et al.*, 2008 The
746 Genealogy and Genetic Viability of Reintroduced Yellowstone Grey Wolves. *Molecular*
747 *Ecology* **17**: 252-274.

748 WAYNE, R. K., and S. M. JENKS, 1991 Mitochondrial DNA analysis implying extensive
749 hybridization of the endangered red wolf *Canis rufus*. *Nature* **351**: 565-568.

- 750 WAYNE, R. K., N. LEHMAN, M. W. ALLARD and R. L. HONEYCUTT, 1992 Mitochondrial-DNA
751 Variability of the Gray Wolf - Genetic Consequences of Population Decline and Habitat
752 Fragmentation. *Conservation Biology* **6**: 559-569.
- 753 WAYNE, R. K., N. LEHMAN, D. GIRMAN, P. J. P. GOGAN, D. A. GILBERT *et al.*, 1991 Conservation
754 Genetics of the Endangered Isle Royale Gray Wolf. *Conservation Biology* **5**: 41-51.
- 755 WAYNE, R. K., and E. A. OSTRANDER, 2007 Lessons Learned from the Dog Genome. *Trends in*
756 *Genetics* **23**: 557-567.
- 757 WECKWORTH, B. V., S. TALBOT, G. K. SAGE, D. K. PERSON and J. COOK, 2005 A signal for
758 independent coastal and continental histories among North American wolves. *Molecular*
759 *Ecology* **14**: 917-931.
- 760 WILLIAMSON, S. H., R. HERNANDEZ, A. FLEDEL-ANON, L. ZHU, R. NIELSEN *et al.*, 2005
761 Simultaneous inference of selection and population growth from patterns of variation in
762 the human genome. *Proceedings of the National Academy of Sciences of the United*
763 *States of America* **102**: 7882-7887.

Table 1 Summary of samples genotyped and sequenced.

Genotyping (106 SNP Loci)			Sequencing (18 Amplicons)**		
# of Populations	# Individuals	Species	# of Populations	# Individuals	Species
14	344	Gray Wolf	4	73	Gray Wolf
1	18	Coyote	1	17	Coyote
18	546	Domestic Dog Breeds	5	97	Domestic Dog Breeds*
1	18	Golden Jackal	1	1	Golden Jackal
1	20	Bat-eared Fox			
1	29	Gray Fox			
1	26	Channel Island Fox			
	1001			188	

**DNA sequencing of amplicons involved a subset of samples that were SNP genotyped.

* Same samples set used in Sutter et al. (2004).

Table 2. Parameters and demographic models tested.

	Parameters				Demographic Models		
					Domestic Dog		Wild Canids
	τ	ω	τ_B	ω_B	Domestication	Breed Formation	Wild Population History
Nested Likelihood Models	τ	ω	τ_B	ω_B	Domestication	Breed Formation	Wild Population History
Constant Population Size	-	-	-	-	X	X	X
Contraction at Fixed Time	fixed	vary	-	-	X	X	-
Contraction at Unknown time	vary	vary	-	-	X	X	X
Bottleneck of Fixed Size	fixed	vary	vary	fixed	X	-	-
Bottleneck of Unknown Size	fixed	vary	vary	vary	X	-	-
Bottleneck of Fixed Length	vary	vary	fixed	vary	-	X	-
Bottleneck of Unknown Length	vary	vary	vary	vary	-	X	X

- not applicable or not tested

Table 3. Proportion of shared SNPs and haplotypes.

	Dog Breeds	Wolf populations	Coyote	uniquely polymorphic SNPs
Dog Breeds	0.15 / 0.85	0.82	0.52	0.12
Wolf populations	0.17	0.27 / 0.90	0.61	0.15
Coyote	0.14	0.19	N/A	0.23

Lower left matrix displays the proportion of SNPs shared

Upper right matrix displays the proportion of haplotypes shared

N/A - not applicable for comparison

Table 4. Nucleotide diversity.

Species	Comparison	π
Dog Breeds	among	0.0026
	within	0.0006
Wolf populations	among	0.0012
	within	0.0009
Coyote		0.001

Table 5. Estimated extent of linkage disequilibrium measured as the distance at which r^2 decays to 0.2.

Breed	Domestic Dog			Wild Canid		
	Genotype Data	Sequence Data	2006 AKC Registration	Populations	Genotype Data	Sequence Data
Labrador Retriever	20kb	785kb	123,760	Alaska	<10kb	<10kb
Saint Bernard	83kb		3,828	Canada	<10kb	
Pomeranian	200kb		18,218	Yellowstone	<10kb	<10kb
Poodle	250kb		29,939	Minnesota	<10kb	
Standard Poodle	318kb			Israel (older)	<10kb	
Toy Poodle	415kb			Northern Quebec	<10kb	
Dachshund	427kb		36,033	Israel (2006)	564kb	241kb
Whippet	500kb		1,836	Spain	906kb	1.7Mb
Pembroke Welsh Corgi	639kb		10,250	Sweden	1Mb	
Old English Sheepdog	672kb		1,403	Italy	3.2Mb	
Miniature Poodle	672kb			Isle Royale	*	
Scottish Terrier	716kb		3,545			

Doberman Pinscher	826kb		11,546	Coyote	<10kb	<10kb
Portuguese Water Dog	846kb		1,454			
Golden Retriever		1.4Mb	42,962			
Akita	1.3Mb	1.4kb	2,457			
Border Terrier	1.3Mb		873			
Briard	2Mb		284			
Irish Wolfhound	2.2Mb		1,016			
Basset Hound	2.4Mb		7,844			
Pekingese	4.2Mb	*	3,155			
Mastiff	*		6,913			
Bernese Mountain Dog		*	3,714			

* indicates the decay curve did not reach an r^2 of 0.2.

Table 6. Poisson calculated parameter estimates under nested likelihood models of domestication from analysis of site frequency spectra in PRFREQ of ascertainment bias corrected and uncorrected genotype data.

Models	Contraction at Fixed Time			Contraction at Unknown time			Bottleneck of Fixed Size					Bottleneck of Unknown Size				
	df = 1			df = 2			df = 2					df = 3				
	$\tau = \text{fixed}$	$\omega = \text{vary}$	p-value	$\tau = \text{vary}$	$\omega = \text{vary}$	p-value	$\tau = \text{fixed}$	$\omega = \text{vary}$	$\tau_B = \text{vary}$	$\omega_B = 0.1$	p-value	$\tau = \text{fixed}$	$\omega = \text{vary}$	$\tau_B = \text{vary}$	$\omega_B = \text{vary}$	p-value
Uncorrected	5000	0.23	4.01×10^{-8}	2245	0.12	0.51	5000	64.5	1835	0.1	0.51	5000	0.11	3023	0.9	0.8
Corrected	5000	0.25	2.27×10^{-5}	16193	0.38	0.11	5000	2.5	2591	0.1	0.03	5000	1.8	2159	0.08	0.1

τ time of contraction event in number of generations

τ_B length of time the contraction occurred in number of generations

ω effective population size following the contraction scaled in terms of the ancestral wolf effective population size

ω_B effective population size during the contraction scaled in terms of the ancestral wolf effective population size.

p-values calculated from likelihood ratio test, where a p-value < 0.05 indicates the alternative model fits the data better than the previous model

null model ($\tau = \infty$ and $\omega = 1$)

Table 7. Poisson calculated parameter estimates under nested likelihood models of breed formation and wild canid population history.

Breed / Wild Canid Population	Contraction at Fixed Time			Contraction at Unknown Time		
	df = 1			df = 2		
	$\tau = \text{fixed}$	$\omega = \text{vary}$	p-value	$\tau = \text{vary}$	$\omega = \text{vary}$	p-value
Akita	100	0.012	6.74×10^{-8}	92	0.011	0.96
Bernese Mountain Dog	100	0.006	0	756	0.039	0.77
Golden Retriever	100	0.011	4.72×10^{-10}	92	0.01	0.99
Labrador Retriever	100	0.01	1.03×10^{-10}	367	0.035	0.83
Pekingese	100	0.006	0	65	0.004	0.87
Alaska Wolf	-	-	-	97	0.053	0.61
Israel Wolf	-	-	-	10796	0.25	0.002
Spain Wolf	-	-	-	227	0.028	0.014
Yellowstone Wolf	-	-	-	16	0.045	0.99
Coyote	-	-	-	20512	1.63	0.26

τ time of contraction event in number of generations

ω effective population size following the contraction scaled in terms of the ancestral wolf

effective population size

p-values calculated from likelihood ratio test, where a p-value < 0.05 indicates the alternative

model fits the data better than the previous model

null model ($\tau = \infty$ and $\omega = 1$)

Figure 1. Schematic representation of sequencing on chromosome 1. Red text indicates new regions sequenced and black text represents regions resequenced from Sutter et al (2005).

Figure 2. Demographic model of (a) dog domestication, (b) dog breed formation from past to present, and (c) wolf population size change. N_{eWOLF} , N_{eB} , N_{eDOG} , N_{eBreed} , N_{eWOLFP} , and N_{eWOLFA} are the effective population sizes of the gray wolf (constant over time), of the population during the bottleneck, the effective population size of dogs, the effective population size of breeds, the effective population size of gray wolves prior to a population size change, and the effective population size of wolves after a population size change, respectively. τ is the time of the domestication event from the present, and τ_B is the bottleneck duration.

Figure 3. Principal component analysis of SNP genotype data within and across species: (a) all species; (b) Gray wolves; (c) Old World Wolves; (d) New World Wolves; and (e) Poodles.

Figure 4. Decay plots of LD (r^2) estimates from sequence data, and genotype data in: (a,b,c) Alaskan gray wolf; (d,e,f) Israeli gray wolf; (g,h,i) Spanish gray wolf; (j,k,l) Yellowstone gray wolf; and (m,n,o) Coyote. The 1st column of data represents decay plots from chromosome 1 sequence data. The 2nd column of data represent data from SNP genotypes on all 5 chromosome, and the 3rd column of data represents decay plots from SNP genotypes just on chromosome 1. Gray diamonds are pairwise r^2 . Black squares are average r^2 for each distance group for which a logarithmic trend line was fitted to the data (black line).

Figure 5. Correlation between the extent of LD at $r^2=0.2$ and the log of the number of AKC registered individuals across 14 breeds of dog

Figure 1

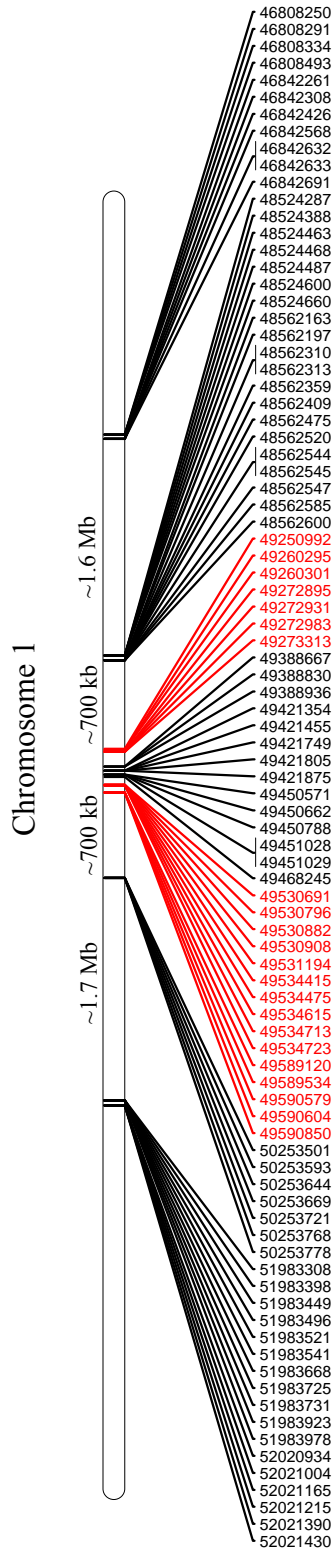


Figure 2.

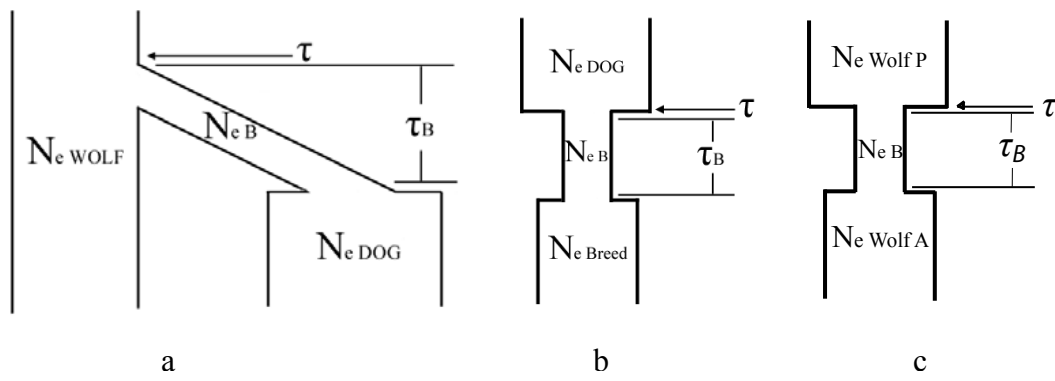


Figure 3

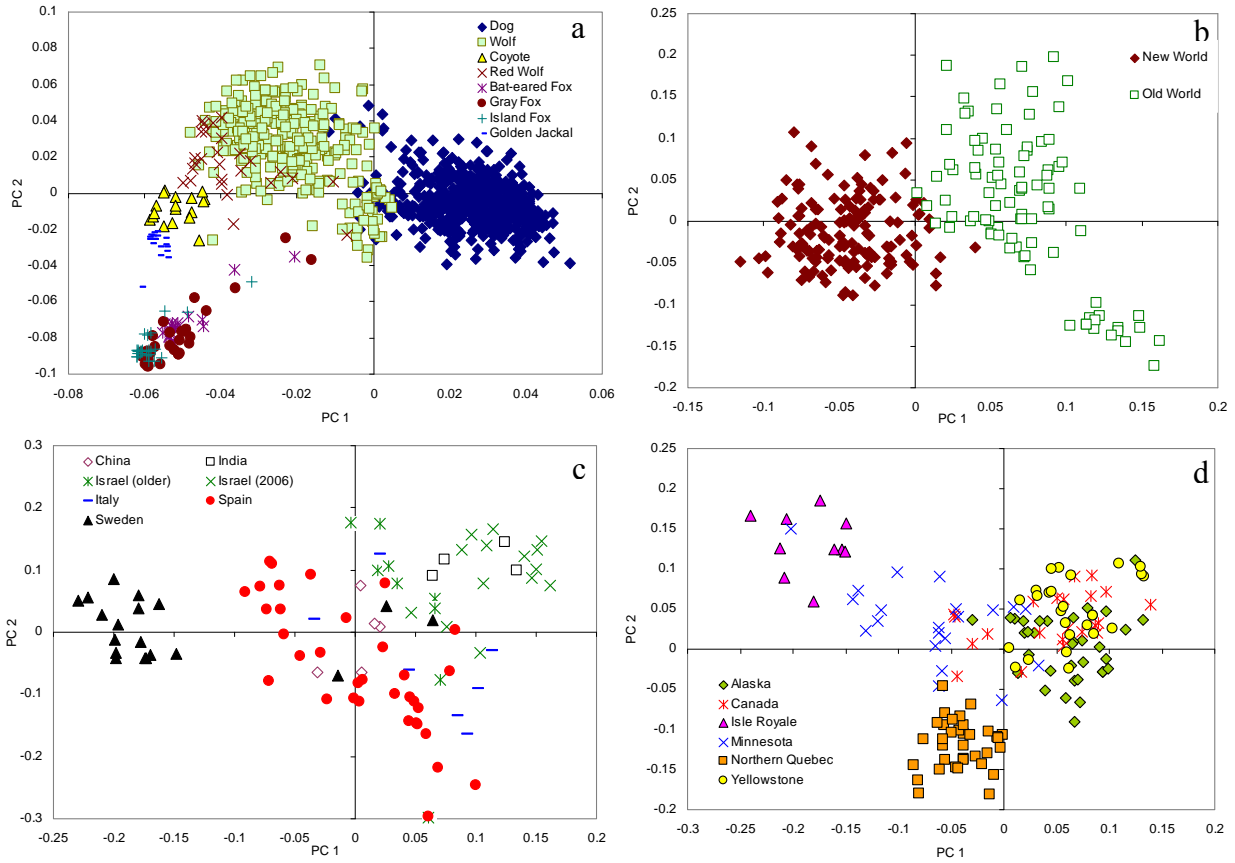


Figure 4

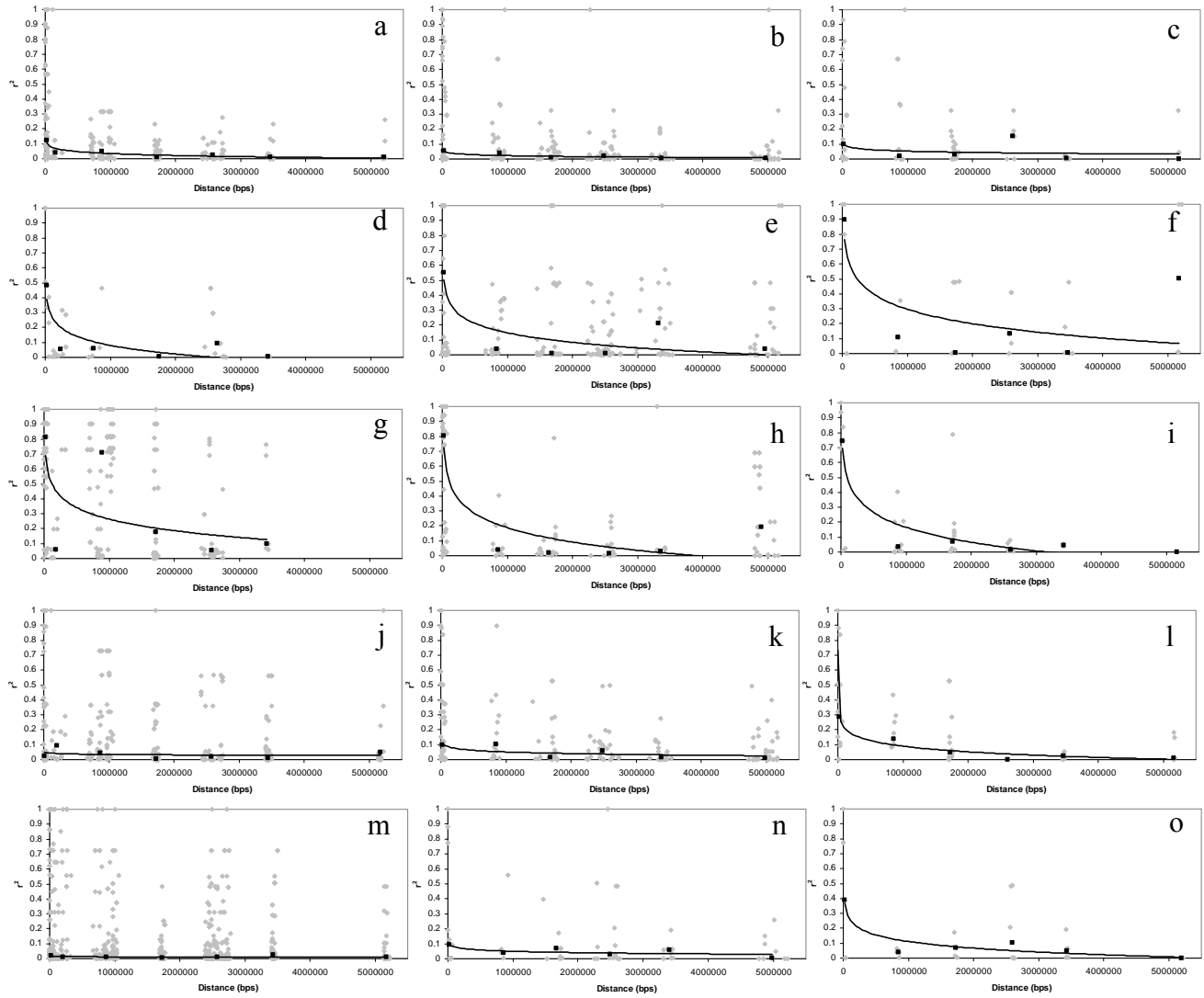


Figure 5

