

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

2

3

4   Melissa M. Gray<sup>\*</sup>, Julie M. Granka<sup>†</sup>, Carlos D. Bustamante<sup>†</sup>, Nathan B. Sutter<sup>‡</sup>, Adam R.  
5   Boyko<sup>†</sup>, Lan Zhu<sup>§</sup>, Elaine A. Ostrander<sup>\*\*</sup>, and Robert K. Wayne<sup>\*</sup>.

6

7   <sup>\*</sup>Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA  
8   90095.

9   <sup>†</sup>Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY  
10   14853.

11   <sup>‡</sup>Department of Clinical Sciences, College of Veterinary Medicine, VMC C3-179, Cornell  
12   University, Ithaca, NY 14853.

13   <sup>§</sup>Department of Statistics, Oklahoma State University, 301C MSCS Bldg, Stillwater, OK 74078.

14   <sup>\*\*</sup>Cancer Genetics Branch, National Human Genome Research Institute, National Institute of  
15   Health, Bethesda, MD 20892.

16

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

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

47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69

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

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

70 PEMBERTON 2007), and wild mice (*Mus musculus domesticus*) (LAURIE *et al.* 2007).  
71 Additionally, LD was modeled in a wild Soay Sheep population using simulations that included  
72 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, followed  
86 by Linbladh-toh *et al.* (2005), who examined the extent of LD within 10 dog breeds across a 15  
87 Mb region. We expand the number of domestic dog breeds for which the extent of LD is  
88 estimated (12 new breeds), although our primary goal is to determine the range that LD extends  
89 in a large panel of wild canid populations, including gray wolves and coyotes, and compare these  
90 estimates to those from domestic dog breeds. Furthermore, we explore the relationship of LD and  
91 demographic history by comparing estimates of LD to known population histories. We also  
92 model population histories using the site frequency spectra (SFS) of each population based on

93 origination from an ancestral wolf population followed by specific demographic scenarios. These  
94 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](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 association between LD and the log of the number of  
117 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 SNP locus (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 S2). 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 S2 &  
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 domestic dog  
144 breeds and four wild canid populations and calculated loss of diversity as  $1 - (\pi \text{ in dogs}) / (\pi \text{ in}$   
145  $\text{wolves})$ . Additionally, the loss of nucleotide diversity at breed formation ( $1 - (\pi \text{ in breed}) / (\pi \text{ in}$   
146  $\text{dogs})$ ) was calculated by sampling one chromosome from each individual of each breed of dog  
147 and one chromosome from each individual across all breeds. Haplotypes were inferred across  
148 each population using the software program PHASE (STEPHENS and DONNELLY 2003; STEPHENS  
149 *et al.* 2001). The percent of haplotypes within and among breeds/populations were calculated as  
150 in Sutter *et al.* (2004). The software program Haploview (BARRETT *et al.* 2005) was used to  
151 calculate  $r^2$  and  $D'$ , which were plotted by matching allele frequencies between pairs of SNPs  
152 with an allele frequency difference of <10% (EBERLE *et al.* 2006). Median values for each  
153 distance category were calculated and a logarithmic curve was fitted to the data. An  $r^2$  of 0.2 was  
154 arbitrarily chosen as the value for which the extent of LD was compared between populations  
155 and species.

#### 156 **Ascertainment Bias:**

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



161 absent in others. Thus, the level of ascertainment bias present in this study was assessed by  
162 performing simulations that replicated the ascertainment scheme (Figure S3). This scheme  
163 utilized the original sequence data (SUTTER *et al.* 2004) in the following procedure: (1) one breed  
164 of dog was randomly chosen as the focus breed and individuals from the 4 other breeds were  
165 designated as the ascertainment panel; (2) For each amplicon, two individuals were randomly  
166 selected from the ascertainment panel and compared. If any marker was segregating, it was  
167 flagged as a SNP and “genotyped” in the focus breed. This sampling was repeated for 2000  
168 bootstraps; (3) the extent of LD was calculated for the focus breed and compared to the extent of  
169 LD based on the observed sequence data. This sampling increased the diversity of sequence  
170 comparisons used and accounted for possible unobserved sequences due to recombination among  
171 amplicons. Also, we wanted to reduce the bias caused by using a limited set of starting points.

#### 172 **Population Structure:**

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

#### 178 **Demographic Modeling:**

179 To explore the demographic history of wild canid populations and domestic dog breeds,  
180 we used the program PRFREQ (BOYKO *et al.* 2008; WILLIAMSON *et al.* 2005) to estimate  
181 demographic parameters under a composite likelihood framework. The program utilizes the  
182 Poisson Random Field (PRF) approach (SAWYER and HARTL 1992), which predicts the  
183 distribution of allele frequencies across sites based on single-locus diffusion theory.

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

#### 200 **Domestication Model:**

201 To model the initial domestication event in dogs (Figure 1a), sequence data from Sutter et  
202 al. (2004) derived from five breeds of dog were combined with genotype data from this study,  
203 producing a combined dataset containing 22 dog breeds. Loci common to both datasets and not  
204 segregating in the outgroup (golden jackal) were retained for further analysis (82 SNPs). To  
205 generate an unfolded site frequency spectrum representative of the ancestral domesticated dog  
206 population to estimate demographic parameters, we randomly selected one chromosome from

207 each breed for 2000 iterations and constructed a SFS averaged across iterations. SNPs with  
208 sample sizes less than 14 (due to missing data) were excluded, leaving a total of 76 SNPs. SNPs  
209 with sample sizes greater than 14 were “projected” to a sample size of 14. This involved using  
210 the hypergeometric distribution to calculate the probability of the latter falling into each  
211 frequency class of a SFS with sample size 14, and summing over all SNPs in each frequency  
212 class to create the final SFS (CLARK *et al.* 2005).

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

219 PRFREQ was then used to estimate demographic parameters based on two site frequency  
220 spectra (corrected for ascertainment bias and uncorrected). We estimated the following  
221 parameters: length of the domestication event ( $\tau_B$ ), the bottleneck population size ( $\omega_B$ ), and the  
222 domesticated dog population size after the bottleneck ( $\omega$ ) (Figure 1a & Table 2). The model that  
223 best fit the observed data was determined by comparing the likelihoods of the nested models. A  
224 constant population size of 21,591 in gray wolf ( $N_{e\_WOLF}$ ) was assumed over time, estimated from  
225 Watterson’s  $\theta$  ( $\theta = 4N_e\mu$ ) calculated from pairwise estimates between all wolves and using a  
226 mutation rate ( $\mu$ ) of  $1 \times 10^{-8}$  per generation (LINDBLAD-TOH *et al.* 2005). This value was re-  
227 evaluated excluding populations found later to exhibit some evidence of a past contraction event  
228 (see results) but was minimally different (22,600). The initiation of the domestication of

229 domestic dogs ( $\tau$ ) was assumed to be 15,000 years before present (OLSEN 1985; SAVOLAINEN *et*  
230 *al.* 2002).

### 231 **Breed Formation Model:**

232 To further model the formation of breeds (Figure 1b), we used the complete sequence  
233 data from Sutter *et al* (2004) from five chromosomes in five dog breeds. However, as a result of  
234 breeding programs, dog breeds are highly inbred, and an individual's chromosomes are more  
235 similar than expected under random mating. Since this could potentially affect our demographic  
236 inferences, we attempted to reduce the effects of inbreeding within breeds by sampling one  
237 chromosome per individual per breed creating a site frequency spectrum for each of 2000  
238 iterations, and averaging across iterations. Because the data were generated by sequencing and  
239 sample sizes of SNPs were consistent across loci, corrections for sample size and ascertainment  
240 bias were not necessary as with modeling the domestication event. Likelihoods between several  
241 nested models were compared to determine which model best fit the observed data (Figure 1b &  
242 Table 2).

### 243 **Wild Canid Model:**

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

250 Nested models identical to those examined for breed formation were used to model the  
251 wild canid populations (Figure 1c & Table 2). However, models fixing the time of contraction

252 were not tested, as we did not know *a priori* whether we should expect to see a population  
253 contraction or expansion nor the timing of such an event. For those populations with evidence for  
254 a population decline, we tested models varying the contraction event.

## 255 RESULTS

### 256 **Genetic Diversity:**

257 To study the effects of demographic history on LD in wild canids and domestic dogs, we  
258 sequenced 11,279 base pairs spanning ~5.2Mb on dog chromosome 1 (Figure S2 & Table S2). A  
259 total of 92 SNP loci (Table S2 & S3) were identified of which 54 were polymorphic across four  
260 gray wolf populations, 48 were polymorphic in one coyote population, and 43 were polymorphic  
261 across five breeds of dog.

262 Eighteen percent of SNPs on average were shared within and between canid species with  
263 gray wolf populations exhibiting the highest sharing (27%; Table S4). Interestingly, 24 loci  
264 (26%) were observed to have the derived allele fixed in domestic dogs but polymorphic in gray  
265 wolves which likely reflects the bottlenecks associated with domestication or breed formation.  
266 The average proportion of shared haplotypes within and between species was 74%. Wolf  
267 populations had the highest average percentage of haplotype sharing (90%; Table S4). Average  
268 nucleotide diversity among dog breeds was significantly different from that within dog breeds (t-  
269 test p-value <0.001; Table S5) and among wolf populations (p-value <0.001). Furthermore,  
270 examination of the ratio of nucleotide diversity suggests a minimal loss of diversity as a result of  
271 the domestication event (0.05) whereas the average loss of diversity due to breed formation was  
272 much larger (0.35).

273 Using data from five chromosomes (1, 2, 3, 34, and 37), 105 of 106 loci were  
274 successfully genotyped in 18 dog breeds, and a variety of wild canid species (Table S6). In the

275 most distantly related species, approximately 93% of loci in golden jackals, and 80% of loci in  
276 bat-eared fox, gray fox, and island fox were successfully amplified. Observed heterozygosity  
277 values ranged from 0.24 in golden jackals to 0.31, 0.29, and 0.33 in the bat-eared, gray, and  
278 island fox, respectively.

### 279 **Population Structure:**

280 Population structure across a variety of wild canid species and domestic dog breeds was  
281 explored through PCA of the genotype data (Figures 2a-d). Fourteen principal components (PCs)  
282 were found to be significant for the seven canid species analyzed (Figure 2a). Domestic dog,  
283 gray wolf, coyote, golden jackal, and foxes were found to separate along principal component  
284 axes 1 and 2. Red wolf was found to overlap both coyote and gray wolf but more so with the  
285 latter species.

286 PCA of 13 gray wolf populations revealed 11 significant principal components (Figure  
287 2b). The most distinct pattern was observed along the first axis of variation separating Old World  
288 and New World wolf populations. Minimal overlap between the two groups was evident. PCA  
289 was then performed separately on Old and New World populations (Figures 2c & d). Seven Old  
290 World gray wolf populations were found to have seven significant principal components. The  
291 first axis of variation visibly separated the majority of the populations with a distinct separation  
292 of Swedish gray wolves. Six New World gray wolf populations were found to have six  
293 significant principal components. Isle Royale, Minnesota, and Northern Quebec define distinct  
294 clusters with Alaska, Canada, and Yellowstone forming a fourth cluster.

295 PCA of 18 domestic dog breeds exhibited 15 significant principal components (data not  
296 shown). There was considerable overlap between breeds. However, Akita displayed virtually no  
297 overlap with any other domestic dog breed while Pekingese exhibited slight separation along the

298 first axis of variation. Along the second axis of variation, Mastiffs and to a lesser extent,  
299 Portuguese Water Dogs exhibited separation from the main cluster of breeds.

### 300 **Ascertainment Bias:**

301 Ascertainment bias typically produces a pattern characterized by a decrease in low  
302 frequency alleles and an increase in higher frequency alleles (CLARK *et al.* 2005; ROSENBLUM  
303 and NOVEMBRE 2007). A shift in allele frequencies was observed between the sequence and  
304 genotype data; however, there was no discernable pattern (Figure S4). Simulations showed that  
305 the degree to which the difference in allele frequency affects estimates of LD varies for each  
306 breed (Figure S5). Ascertainment bias in Labrador Retriever was observed to have minimal  
307 effects; however, in Golden Retrievers the effect was large. Despite the variance in estimates, the  
308 rank order of breeds based on estimates of LD remained the same.

309 Site frequency spectra generated from sequence and genotype data for gray wolf and  
310 coyote populations were similar to the domestic dog in showing no distinguishable patterns  
311 (Figure S4). When LD estimates from sequence data were compared with LD estimates from  
312 genotype data, sequence data generally gave lower estimates of LD (Figure 3). Spanish gray wolf  
313 was the exception with an increase in LD measured from genotyped data to sequenced data.  
314 Despite the shift in LD estimates from sequence to genotype data, the rank order of LD estimates  
315 of each population remained the same.

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

321 **Linkage Disequilibrium:**

322           The extent of LD estimated from genotyped data in gray wolf populations ranged from  
323 <10kb in Alaskan gray wolves to >5Mb in gray wolves from Isle Royale ( $r^2_{0.2}$ ; Table 3; see  
324 Figure S6 for  $D'$  estimates). The extent of LD was consistent with the known demographic  
325 history of each population. Large outbreeding populations such as Alaska, Minnesota, Canada,  
326 Yellowstone, and Northern Quebec exhibited such low levels of LD that the decay curves did not  
327 extend to an  $r^2$  value of 0.2 (Figure 3 and Table 3). Therefore, we take a conservative approach  
328 and consider these populations to generally have LD levels lower than 10kb. However,  
329 small/bottlenecked populations such as Isle Royale, Spanish, Italian, and Swedish gray wolves  
330 exhibited high levels of LD ( $r^2_{0.2}>500\text{kb}$ ). Lastly, coyotes exhibited levels of LD below an  $r^2$   
331 value of 0.2 (Figure 3) consistent with their large population size in southern California  
332 (FEDRIANI *et al.* 2001; VILÀ *et al.* 1999a).

333           Estimates of LD from genotyped data in dog breeds ranged from 20kb to >5Mb ( $r^2_{0.2}$ ;  
334 Table 3). The extent of LD was found to be significantly correlated to the log of registered  
335 individuals for both Kendal's tau rank correlation (p-value = 0.02) and Mantel's test (p-value =  
336 0.0001). However, three breeds had sample numbers below the minimum cutoff ( $n<17$ ) used by  
337 Sutter *et al.* (2004), introducing potentially greater bias into our measures of LD. When these  
338 breeds were excluded from the analysis, the correlation statistics for the remaining 14 breeds  
339 were still significant (Kendal's tau, p-value = 0.003; Mantel's test, p-value = 0.0002). Thus, the  
340 level of LD within dog breeds was found to be well correlated with 2006 registration numbers  
341 (Figure S7).

342           Values of LD based on sequence data were highly correlated with those based on SNP  
343 genotypes (Kendal's tau, p-value = 0.02; Mantel's test, p-value = 0.0001). The extent of LD



344 between species based on sequence data demonstrated that gray wolves and coyotes have less  
345 LD (<10kb to 1.7Mb) than the domestic dog (785kb to >5Mb; Table 3). The extent of LD seen in  
346 the Spanish gray wolf population was much higher than any other sequenced gray wolf  
347 population (1.7Mb). We explored the possibility of relatedness among the samples by  
348 eliminating individuals with high levels of allele sharing based on 11 microsatellite loci  
349 (VONHOLDT *et al.* 2008), and confirmed that high levels of LD are still present in a sample set of  
350 reduced allele sharing ( $r^2_{0.2} \geq 1.5\text{Mb}$ ).

### 351 **Domestication Modeling:**

352 Parameter estimates were scaled in terms of the estimated gray wolf effective population  
353 size (i.e.,  $\omega = N_{e\text{DOG}}/N_{e\text{WOLF}}$ ) and multinomial calculations were found to be qualitatively similar  
354 to Poisson calculations (see supplement Tables S7-S8 & Figures S8-S10). The Contraction at  
355 Fixed Time model, with a single contraction event fixed at 15,000 years ago, explained the data  
356 significantly better than the null model of constant population size (Table 4). This applied for  
357 both the ascertainment bias corrected (p-value =  $2.27 \times 10^{-6}$ ) and uncorrected data sets (p-value =  
358  $4.01 \times 10^{-8}$ ). The corrected Poisson calculations suggest that this contraction was followed by a  
359 population expansion (Bottleneck of Fixed Size, Table 4), although the improvement in the  
360 model fit is slight (p-value = 0.033) and unlikely to be significant after correcting for linkage in  
361 the dataset. Therefore, we focus on the Contraction at Fixed Time model findings. The estimate  
362 of  $\omega$  for the Poisson calculation of the Contraction at Fixed Time model was 0.23 for the  
363 uncorrected data, indicating the dog ancestral population size was 0.23 times the size of the wolf  
364 ancestral population, and 0.25 for the corrected data. Therefore, results suggest a single minor  
365 contraction event was associated with domestication of the dog.

### 366 **Breed Formation Modeling:**

367 Demographic parameters based on the discussed model of breed formation were  
368 estimated for each of 5 breeds from the averaged sampled site frequency spectra. Across all  
369 breeds and calculations, no models had a higher likelihood than the Contraction at Fixed Time  
370 model (Table 5), which indicates a contraction without a subsequent increase in population size.  
371 Under the Contraction at Fixed Time model, Bernese Mountain Dog and Pekingese were  
372 observed to have the largest bottleneck with a current effective population size approximately  
373 0.0055 and 0.0056 that of the ancestral dog effective population (assumed to be equal to the gray  
374 wolf effective population size). Labrador Retriever, Golden Retriever, and Akita exhibited a  
375 weaker reduction in population size with values of 0.0095, 0.011, and 0.012, respectively.

376 Although not significantly better than the Contraction at Fixed Time model, both  $\omega$  and  $\tau$   
377 were optimized under the Contraction at Unknown Time model allowing examination of the  
378 timing of breed contractions. Under this model, Pekingese was observed to have a severe  
379 reduction in population size ~65 generations ago ( $\omega=0.0035$ ), while the Akita and Golden  
380 Retriever were observed to have similar contraction times at ~92 generations ( $\omega=0.0113$  &  
381 0.0100). It is important to note that when estimating both  $\omega$  and  $\tau$ , timing estimates may not be  
382 entirely realistic, as there is a tradeoff between having a recent  $\tau$  and a severe population decline  
383 and having a more distant  $\tau$  but less severe decline. For example, the founding prediction of 755  
384 generations ago for Bernese Mountain Dog, originating from a small number of individuals and  
385 maintained as a small population, is likely an overestimate. Regardless, the bottleneck at breed  
386 formation across breeds is orders of magnitude more severe, and more recent, than an ancient  
387 domestication event and more likely to impact differences in LD among breeds (see below).

388 **Wild Canid Modeling:**



412 (785kb to >5Mb; Table 3). Barley (CALDWELL *et al.* 2006), soybean (HYTEN *et al.* 2007), sheep  
413 (MCRAE *et al.* 2002; MCRAE *et al.* 2005), and house mice (LAURIE *et al.* 2007) display a  
414 consistent pattern of reduced levels of LD in wild populations compared to their domesticates.  
415 This is expected since domestication likely results in a bottleneck event. However, across wild  
416 populations, demographic history can still be observed to strongly influence levels of LD. For  
417 example, the Spanish wolf population had LD levels higher than some domestic dog breeds  
418 ( $r^2_{0.2}=1.7\text{Mb}$ ). In the past century, gray wolves from Spain were hunted to near extinction, but  
419 have steadily risen in numbers since the enactment of hunting restrictions (RAMIREZ *et al.* 2006).  
420 In contrast, Labrador Retrievers exhibited levels of LD similar to wild gray wolf populations  
421 ( $r^2_{0.2}=785\text{kb}$ ) as they are the most popular breed in the U.S. today with about 150,000 new  
422 registrations per year (www.AKC.org). Lastly, coyotes were found to display the lowest levels of  
423 LD ( $r^2_{0.2}<10\text{kb}$ ) relative to all domestic dog breeds and gray wolf populations. Consistent with  
424 low levels of LD, coyote population sizes are reportedly an order of magnitude greater than the  
425 gray wolf (VILÀ *et al.* 1999a).

426         As seen with the sequence data, LD levels from SNP genotype data were found to  
427 correspond with known demographic history for the 11 gray wolf populations. For example, the  
428 Isle Royale gray wolf population is a small population of wolves that inhabit an island in Lake  
429 Superior off the coast of Minnesota. The population was founded by a single breeding pair in  
430 1950 (PETERSON *et al.* 1998). Previous genetic research found population heterozygosity levels  
431 half that observed in the mainland progenitor (Wayne *et al.* 1991). The extent of LD in the Isle  
432 Royale population ( $r^2_{0.2}>5\text{Mb}$ ) is consistent with that expected in small and/or severely  
433 bottlenecked populations (GAUT and LONG 2003; MUELLER 2004; PRITCHARD and PRZEWORSKI  
434 2001). Other populations that are known to have undergone a contraction or have a history of

435 small population size had high levels of LD (Spanish, Swedish, and Italian gray wolves; (for  
436 supporting demographic and genetic research see: FABBRI *et al.* 2007; LEHMAN *et al.* 1992;  
437 RAMIREZ *et al.* 2006; VILÀ *et al.* 1999a; WAYNE *et al.* 1992). At the other end of the spectrum,  
438 populations of Alaskan, Canadian, and Northern Quebec gray wolves have been large and of  
439 constant size for a long time, and exhibit low levels of LD (MUSIANI *et al.* 2007; WECKWORTH *et*  
440 *al.* 2005). Supporting this finding, genetic studies (ROY *et al.* 1994; VILÀ *et al.* 1999b; WAYNE *et*  
441 *al.* 1992) of Alaskan and Northern Canadian gray wolf populations found high variability and  
442 reduced population differentiation suggesting a large population size and higher levels of gene  
443 flow than among European wolf populations which were more structured. Similarly, LD  
444 estimates in dog breeds from SNP genotype data corroborate findings from sequence data as  
445 exemplified by a significant correlation to popularity of the breed based on registration numbers  
446 (Figure S7). Thus, the extent of LD measured from the SNP genotype data also support the  
447 correlation between LD and demographic history in wild and domestic populations.

#### 448 **Demographic modeling:**

449 Previous studies based on mtDNA analysis (SAVOLAINEN *et al.* 2002; VILÀ *et al.* 1997)  
450 have indicated that four to six matriline of gray wolf were involved in the founding of the  
451 domestic dog. In contrast, analysis of major histocompatibility (MHC) loci suggested several  
452 hundred founders or extensive backcrossing with wild canids is needed to explain present day  
453 diversity in domestic dogs (VILÀ *et al.* 2005). Linbladh-Toh *et al.* (2005) found evidence for two  
454 major bottlenecks in modern dog breeds, the first occurring as a result of domestication from  
455 wolves, supported by short range LD estimates, and the second occurring as a result of breed  
456 formation, supported by long range LD. Linbladh-Toh *et al.* (2005) simulated the demographic  
457 history of domestic dogs over a coarse grid of demographic parameter values, and compared the

458 observed and simulated rates of pairwise polymorphism across ten 15 Mb regions. They then  
459 selected the domestication parameters for which the simulations resulted in polymorphism values  
460 that were the closest to observed values. Although they did find evidence for two major  
461 bottlenecks, they did not use a rigorous likelihood framework, and thus are not able to perform  
462 any hypothesis testing or formal model selection. For our analysis, we searched a denser grid of  
463 domestication parameter values and examined the site frequency spectrum of dogs rather than  
464 pairwise polymorphism. For domestication events with parameters that were searched over this  
465 grid, we calculated the likelihood of the observed domesticated dog site frequency spectrum. In  
466 this likelihood framework, we were able to perform nested likelihood ratio tests to test the null  
467 hypothesis of constant population size and make meaningful comparisons between models.

468         From our demographic modeling, we found evidence for a modest population contraction  
469 approximately 15,000 years ago (5,000 generations ago) and severe contractions at breed  
470 formation. The contractions due to breed formation were found to be an order of magnitude  
471 greater than the domestication contraction based on analysis of the site frequency spectra. From  
472 nucleotide diversity estimates, only a 5% reduction in diversity was observed as a result of  
473 domestication whereas an average of 35% of nucleotide diversity was lost due to breed  
474 formation. This severe contraction at breed formation was expected, as continued inbreeding  
475 within a given breed may act to maintain a small effective population size even if the census  
476 population size has actually increased since breed formation. The absence of a strong signal for a  
477 contraction at domestication may reflect continued interbreeding between early dogs and wolves  
478 or multiple domestication events (RANDI and LUCCHINI 2002; TSUDA *et al.* 1997; VILÀ *et al.*  
479 1997). Indeed, high levels of diversity observed in domestic dogs may have been maintained

480 through a modest population bottleneck, backcrossing with wolf populations, and rapid  
481 population expansion (VILÀ *et al.* 2005; WAYNE and OSTRANDER 2007).

482         Lastly, demographic modeling of the site frequency spectra of wild canid populations and  
483 dog breeds were found to be concordant with estimates of LD and known population history. In  
484 wild canid populations, a significant population decline was observed for the Spanish gray wolf  
485 and to a lesser extent the Israeli gray wolf population, which was expected from known historical  
486 data. Furthermore, neither coyote nor Alaskan and Yellowstone gray wolf populations showed  
487 significant evidence of a population size change. In modeling the demographic history of  
488 domestic dog breeds, Pekingese and Bernese Mountain Dog exhibited the greatest population  
489 contraction and more modest contractions were observed in Golden Retriever and Labrador  
490 Retriever. The strong concordance observed in this study between the extent of LD, demographic  
491 modeling, and known demographic history supports the use of LD to infer population history not  
492 only in model organisms but also in wild populations.

#### 493 **Population Structure:**

494         Eighty percent or greater of SNPs that were discovered in dogs successfully amplified in  
495 the most distantly related species and polymorphism levels ranged from 25% to 40%. Genetic  
496 isolation and/or admixture revealed in the PCA was consistent with previous studies (LEHMAN *et*  
497 *al.* 1992; LEONARD *et al.* 2005; ROY *et al.* 1994; VILÀ *et al.* 1999a; WAYNE *et al.* 1992; WAYNE  
498 *et al.* 1991). Within gray wolves, PCA identified strong geographic differentiation between Old  
499 and New World populations as well as populations within each continent. Similar relationships  
500 have been observed in mtDNA studies of gray wolves (ROY *et al.* 1994; VILÀ *et al.* 1999a;  
501 WAYNE *et al.* 1992). Patterns in the PCA plots were consistent with previous phylogenetic  
502 studies (LEONARD *et al.* 2005; LINDBLAD-TOH *et al.* 2005; ROY *et al.* 1996; VILÀ *et al.* 1999a).

503 For example, PC one supports the fundamental genetic distance between wild canids and  
504 domestic dogs. PC two distinguishes wild canids from each other with coyotes and golden  
505 jackals positioned nearest to the gray wolves and red wolves overlapping coyotes and gray  
506 wolves. The overlap of red wolves with both species is consistent with extensive hybridization in  
507 the past (WAYNE and JENKS 1991). The high degree of SNP amplification success between  
508 species suggests that dog-derived SNP markers may be useful in mapping phenotypic traits in  
509 wild canid species. To support this conclusion, Kukekova et al. (2007) used dog derived  
510 microsatellite markers to develop a genetic map for the silver fox and Sacks and Louie (2008)  
511 and Seddon et al.(2005) sequenced SNP loci from the dog genome to develop new SNPs for  
512 genetic studies in gray wolf, coyote, red fox, and gray fox.

### 513 **Conclusions:**

514 The extent of LD in natural vertebrate populations has been difficult to assess in the past  
515 because large scale genomic surveys were only possible in model species. However, with the  
516 availability of high throughput genotyping and information from genome sequencing projects, a  
517 new era has emerged in the genetic characterization of natural populations. Utilizing these  
518 resources, we have estimated LD in 11 natural populations of gray wolf, one population of  
519 coyote and 18 dog breeds. Additionally, because a causal relationship exists between LD and  
520 population history, we have made inferences about the demographic and evolutionary processes  
521 in wild and domestic canids. Our results suggest that a relatively minor population contraction  
522 was associated with domestication in dogs and that genetic variation was preserved in the rapid  
523 expansion that followed. However, this variation is now partitioned in dog breeds which  
524 generally have high and variable amounts of LD. The high level of LD in some wolf populations  
525 further suggests the possibility of trait mapping in natural populations. For example, in North



526 America, approximately half of wolves are dark colored (ANDERSON *et al.* in press; MUSIANI *et*  
527 *al.* 2007), and given the recent identification of coat color mutants in dogs associated with black  
528 color (CALDWELL *et al.* 2006), similar mutants may now be identified through association studies  
529 in wild wolves. Finally, we demonstrate how simulation models in general can be used to make  
530 inferences about population demography and show that predictions generally fit with observed  
531 levels of LD and known population history. Consequently, our approach may have wide  
532 applicability to other species with extensive genomic resources and their close relatives.  
533

## ACKNOWLEDGEMENTS

533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543

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 (NBS).

- 544 ANDERSON, T. M., B. M. VONHOLDT, S. I. CANDILLE, M. MUSIANI, C. GRECO *et al.*, in press  
 545 Molecular and Evolutionary History of Melanism in North American Gray Wolves.  
 546 Science.
- 547 ARDLIE, K. G., L. KRUGLYAK and M. SEIESTAD, 2002 Patterns of Linkage Disequilibrium in the  
 548 Human Genome. *Nature Reviews Genetics* **3**: 299-309.
- 549 BACKSTROM, N., A. OVARNSTROM, L. GUSTAFSSON and H. ELLEGREN, 2006 Levels of linkage  
 550 disequilibrium in a wild bird population. *Biology Letters* **2**: 435-438.
- 551 BARRETT, J. C., B. FRY, J. MALLER and M. J. DALY, 2005 Haploview: analysis and visualization  
 552 of LD and haplotype maps. *Bioinformatics* **21**: 263-265.
- 553 BOYKO, A. R., S. H. WILLIAMSON, A. R. INDAP, J. D. DEGENHARDT, R. D. HERNANDEZ *et al.*,  
 554 2008 Assessing the evolutionary impact of amino acid mutations in the human genome.  
 555 *PLoS Genetics* **4**: e1000083.
- 556 CAICEDO, A. L., S. H. WILLIAMSON, R. D. HERNANDEZ, A. BOYKO, A. FLEDEL-ALON *et al.*, 2007  
 557 Genome-wide patterns of nucleotide polymorphism in domesticated rice. *Plos Genetics*  
 558 **3**: 1745-1756.
- 559 CALDWELL, K. S., J. RUSSELL, P. LANGRIDGE and W. POWELL, 2006 Extreme population-  
 560 dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum*  
 561 *vulgare*. *Genetics* **172**: 557-567.
- 562 CLARK, A. G., M. J. HUBISZ, C. D. BUSTAMANTE, S. H. WILLIAMSON and R. NIELSEN, 2005  
 563 Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*  
 564 **15**: 1496-1502.
- 565 CUTTER, A. D., S. E. BAIRD and D. CHARLESWORTH, 2006 High nucleotide polymorphism and  
 566 rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*.  
 567 *Genetics* **174**: 901-913.
- 568 DEONIER, R. C., S. TAVARE and M. S. WATERMAN, 2005 *Computation Genome Analysis: An*  
 569 *Introduction*. Springer Science, New York, New York.
- 570 DUNNING, A. M., F. DUROCHER, C. S. HEALEY, M. D. TEARE, S. E. MCBRIDE *et al.*, 2000 The  
 571 extent of linkage disequilibrium in four populations with distinct demographic histories.  
 572 *American Journal of Human Genetics* **67**: 1544-1554.
- 573 EBERLE, M. A., M. J. RIEDER, L. KRUGLYAK and D. A. NICKERSON, 2006 Allele frequency  
 574 matching between SNPs reveals an excess of linkage disequilibrium in genic regions of  
 575 the human genome. *Plos Genetics* **2**: 1319-1327.
- 576 EWING, B., and P. GREEN, 1998 Basecalling of automated sequencer traces using phred. II. Error  
 577 probabilities. *Genome Research* **8**: 186-194.
- 578 EWING, B., L. HILLIER, M. WENDL and P. GREEN, 1998 Basecalling of automated sequencer  
 579 traces using phred. I. Accuracy assessment. *Genome Research* **8**: 175-185.
- 580 FABBRI, E., C. MIQUEL, V. LUCCHINI, A. SANTINI, R. CANIGLIA *et al.*, 2007 From the Apennines  
 581 to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*)  
 582 population. *Molecular Ecology* **16**: 1661-1671.
- 583 FARNIR, F., W. COPPIETERS, J. J. ARRANZ, P. BERZI, N. CAMBISANO *et al.*, 2000 Extensive  
 584 genome-wide linkage disequilibrium in cattle. *Genome Research* **10**: 220-227.
- 585 FEDRIANI, J. M., T. K. FULLER and R. M. SAUVAJOT, 2001 Does availability of anthropogenic  
 586 food enhance densities of omnivorous mammals? An example with coyotes in southern  
 587 California. *Ecography* **24**.

588 GAUT, B. S., and A. D. LONG, 2003 The lowdown on linkage disequilibrium. *Plant Cell* **15**:  
589 1502-1506.

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

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

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

597 HYTEN, D. L., I. Y. CHOI, Q. J. SONG, R. C. SHOEMAKER, R. L. NELSON *et al.*, 2007 Highly  
598 variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics*  
599 **175**: 1937-1944.

600 INGVARSSON, P. K., 2005 Nucleotide polymorphism and linkage disequilibrium within and among  
601 natural populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics* **169**:  
602 945-953.

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

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

607 KUKKOVA, A. V., L. N. TRUT, I. N. OSKINA, J. L. JOHNSON, S. V. TEMNYKH *et al.*, 2007 A  
608 meiotic linkage map of the silver fox, aligned and compared to the canine genome. **17**:  
609 387-399.

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

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

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

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

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

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

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

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

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

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

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

637 OLSEN, S. J., 1985 *Origins of the Domestic Dog*. University of Arizona Press, Tucson.

638 OSTRANDER, E. A., and R. K. WAYNE, 2005 The canine genome. *Genome Research* **15**: 1706-  
639 1716.

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

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

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

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

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

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

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

656 RAMIREZ, O., L. ALTET, C. ENSENAT, C. VILA, A. SANCHEZ *et al.*, 2006 Genetic assessment of  
657 the Iberian wolf *Canis lupus signatus* captive breeding program. *Conservation Genetics*  
658 **7**: 861-878.

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

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

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

666 REMINGTON, D. L., J. M. THORNSBERRY, Y. MATSUOKA, L. M. WILSON, S. R. WHITT *et al.*, 2001  
667 Structure of linkage disequilibrium and phenotypic associations in the maize genome.  
668 *Proceedings of the National Academy of Sciences of the United States of America* **98**:  
669 11479-11484.

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

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

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

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

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

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

684 SEDDON, J. M., H. G. PARKER, E. A. OSTRANDER and H. ELLEGREN, 2005 SNPs in ecological and  
685 conservation studies: a test in the Scandinavian wolf population. Molecular Ecology **14**:  
686 503-511.

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

689 SLATE, J., and J. M. PEMBERTON, 2007 Admixture and Patterns of Linkage Disequilibrium in a  
690 Free-living Vertebrate Population. European Society For Evolutionary Biology **20**: 1415-  
691 1427.

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

694 STEPHENS, M., and P. DONNELLY, 2003 A comparison of Bayesian methods for haplotype  
695 reconstruction from population genotype data. American Journal of Human Genetics **73**:  
696 1162-1169.

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

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

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

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

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

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

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

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

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

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

722 WAYNE, R. K., N. LEHMAN, M. W. ALLARD and R. L. HONEYCUTT, 1992 Mitochondrial-DNA  
723 Variability of the Gray Wolf - Genetic Consequences of Population Decline and Habitat  
724 Fragmentation. *Conservation Biology* **6**: 559-569.

725 WAYNE, R. K., N. LEHMAN, D. GIRMAN, P. J. P. GOGAN, D. A. GILBERT *et al.*, 1991 Conservation  
726 Genetics of the Endangered Isle Royale Gray Wolf. *Conservation Biology* **5**: 41-51.

727 WAYNE, R. K., and E. A. OSTRANDER, 2007 Lessons Learned from the Dog Genome. *Trends in*  
728 *Genetics* **23**: 557-567.

729 WECKWORTH, B. V., S. TALBOT, G. K. SAGE, D. K. PERSON and J. COOK, 2005 A signal for  
730 independent coastal and continental histories among North American wolves. *Molecular*  
731 *Ecology* **14**: 917-931.

732 WILLIAMSON, S. H., R. HERNANDEZ, A. FLEDEL-ANON, L. ZHU, R. NIELSEN *et al.*, 2005  
733 Simultaneous inference of selection and population growth from patterns of variation in  
734 the human genome. *Proceedings of the National Academy of Sciences of the United*  
735 *States of America* **102**: 7882-7887.

736  
737

**Table 1** 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	

\*\* 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
					Breed		Wild Population
Nested Likelihood Models	$\tau$	$\omega$	$\tau_B$	$\omega_B$	Domestication	Formation	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.

$\tau$ , time of contraction event in number of generations.

$\tau_B$ , length of time the contraction occurred in number of generations.

$\omega$ , effective population size following the contraction scaled in terms of the ancestral wolf/dog effective population size.

$\omega_B$ , effective population size during the contraction scaled in terms of the ancestral wolf/dog effective population size.

**Table 3.** 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	
				Northern		
Toy Poodle	415kb			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	672kb		1,403	Italy	3.2Mb	

Sheepdog					
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.

Models	Contraction at Fixed Time			Contraction at Unknown time			Bottleneck of Fixed Size					
	df = 1			df = 2			df = 2					
	$\tau = \text{fixed}$	$\omega = \text{vary}$	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	$\tau = \text{fi}$
Uncorrected	5000	0.23	$4.01 \times 10^{-8}$	2245	0.12	0.51	5000	64.5	1835	0.1	0.51	500
Corrected	5000	0.25	$2.27 \times 10^{-5}$	16193	0.38	0.11	5000	2.5	2591	0.1	0.03	500

See table 2 for description of parameters.

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 5.** 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	$6.74 \times 10^{-8}$	92	0.011	0.96
Bernese Mountain Dog	100	0.006	0	756	0.039	0.77
Golden Retriever	100	0.011	$4.72 \times 10^{-10}$	92	0.01	0.99
Labrador Retriever	100	0.01	$1.03 \times 10^{-10}$	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

See table 2 for description of parameters.

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. Demographic model of (a) dog domestication, (b) dog breed formation, 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.  $\tau$  is the time of the domestication event from the present, and  $\tau_B$  is the bottleneck duration.

Figure 2. 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.

Figure 3. 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 1<sup>st</sup> column of data represents decay plots from chromosome 1 sequence data. The 2<sup>nd</sup> column of data represent data from SNP genotypes on all 5 chromosome, and the 3<sup>rd</sup> 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).







